11-2016

Semiparametric single index panel data models with interactive fixed effects: Theory and practice

Guohua FENG

Bin PENG

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Thomas Tao YANG

Semiparametric Single-Index Panel Data Models with Interactive Fixed Effects: Theory and Practice

GUOHUA FENG[§], BIN PENG[†], LIANGJUN SU[⋆] AND THOMAS TAO YANG[*]

[§]University of North Texas, [†]University of Technology Sydney, [⋆]Singapore Management University and [*]Australian National University

Abstract

In this paper, we propose a single-index panel data model with unobserved multiple interactive fixed effects. This model has the advantages of being flexible and of being able to allow for common shocks and their heterogeneous impacts on cross sections, thus making it suitable for the investigation of many economic issues. We derive asymptotic theories for both the case where the link function is integrable and the case where the link function is non-integrable. Our Monte Carlo simulations show that our methodology works well for large N and T cases. In our empirical application, we illustrate our model by analyzing the returns to scale of large commercial banks in the U.S. Our empirical results suggest that the vast majority of U.S. large banks exhibit increasing returns to scale.

Keywords: Asymptotic theory; Nonlinear panel data model; Interactive fixed effects; Orthogonal series expansion method

JEL classification: C13, C14, C23, C51

- Guohua Feng, University of North Texas, U.S.A. Email: Guohua.Feng@unt.edu.
- Bin Peng, University of Technology Sydney, Australia. Email: Bin.Peng@uts.edu.au.
- Liangjun Su, Singapore Management University, Singapore. Email: ljsu@smu.edu.sg.
- Thomas Tao Yang, Australian National University, Australia. Email: Tao.Yang@anu.edu.au.

1 Introduction

Models with interactive fixed effects have drawn considerable attention in the last decade or so. Two well-known models are studied respectively by Pesaran (2006) and Bai (2009), where the interactive fixed effects (also widely known as factor structure) are used to model unobserved common shocks or time-varying heterogeneity existing in micro- and macro-economic data. Building on these two excellent works, different types of generalization have been proposed. For example, Bai et al. (2009) and Kapetanios et al. (2011) respectively allow time-varying factors to be non-stationary and establish the corresponding asymptotic results; Su and Jin (2012) extend Pesaran (2006) to a non-parametric setting and provide non-parametric versions of common correlated effects mean group (CCEMG) and common correlated effects pooled (CCEP) estimators by a sieve estimation technique; Li et al. (2016), by using a least absolute shrinkage and selection operator, extend Bai (2009) to allow for structural breaks.

The purpose of this study is to contribute to this literature by extending Bai (2009) to a semi-parametric single-index setting. Precisely, our model is specified as follows:

$$y_{it} = g_o(x_{it}'\theta_o) + \gamma_{o,i}' f_{o,t} + \varepsilon_{it} \quad \text{with} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1.1}$$

where the regressor x_{it} is a $d \times 1$ vector, both the factors $f_{o,t}$ and the factor loadings $\gamma_{o,i}$ are $m \times 1$ unknown vectors, and g_o is the so-called link function and is unknown. In addition, we assume both d and m are known and finite. The subscript "$_o$" indicates true values or true functions throughout the paper. For notational simplicity, let $F_o = (f_{o,1}, \ldots, f_{o,T})'$, $Y_i = (y_{i1}, \ldots, y_{iT})'$, $X_i = (x_{i1}, \ldots, x_{iT})'$ and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, where $i = 1, \ldots, N$. Then (1.1) can be expressed in matrix notation as

$$Y_i = \phi_i[\theta_o, g_o] + F_o\gamma_{o,i} + \varepsilon_i \quad \text{with} \quad i = 1, \ldots, N, \tag{1.2}$$

where $\phi_i[\theta, g] = (g(x_{i1}'\theta), \ldots, g(x_{iT}'\theta))'$ for any given θ (belonging to $\mathbb{R}^d$) and g (defined on $\mathbb{R}$).

The single-index model with interactive fixed effects can be widely used in many sub-fields of economics such as production economics, economic growth, international trade, etc. Consider the issue to be examined in the application of this study – the issue of economies of scale of commercial banks in the U.S. This issue has received a considerable amount of attention in the past three decades, because during this period the U.S. banking industry has undergone an unprecedented transformation – one marked by a substantial decline in the number of commercial banks and savings institutions and a growing concentration of industry assets among large financial institutions (Jones and Critchfield (2005)).

Most studies investigating this issue use a fully-parametric translog cost (distance or profit) function without interactive fixed effects to represent the production technology of banks. Compared with this commonly-used fully-parametric translog functional form, the single-index panel data model has two advantages. First, the link function (i.e., $g_o(\cdot)$) is more flexible than the commonly-used translog functional form. Specifically, the translog function is merely a quadratic specification in log-space, and thus limits the variety of shapes the cost function is permitted to take. In contrast, the single-index model does not assume that $g_o(\cdot)$ is known and hence it is more flexible and less restrictive. Second, the single-index panel data model allows for common shocks, which have increased in the past three decades due to the following reasons: (1) the industry has become increasingly dominated by large banks that share similar business models and offer similar ranges of products, which has in turn increased the exposure of those banks to common shocks; and (2) as the process of banking and financial integration has progressed, bank linkages have also risen, which has also increased the exposure of large banks to common shocks (see Houston and Stiroh (2006); Brasili and Vulpes (2008)). To give another of example, the single-index model can also be used for investigating determinants of economic growth. Compared with the commonly-used fully-parametric growth regression models, the single-index model has two advantages: (1) the link function is more flexible; and (2) it allows common shocks to affect all countries, such as the recent global financial crisis. In sum, these examples show that the flexibility of the link function and the presence of the interactive fixed effects make the single-index panel data model very useful for investigating many important economic issues.

There are two dominant approaches to studying single-index models. The first approach involves using nonparametric kernels to implement estimation (e.g., Ichimura (1993), Hardle et al. (1993), Carroll et al. (1997) and so on). An excellent review on this approach can be found in Xia (2006). The second approach involves using the sieve method which provides good and computable approximations to an unknown function (see Chen (2007) for an excellent review). For example, Yu and Ruppert (2002) employ penalized spline estimation to investigate partially linear single-index models. Dong et al. (2015) and Dong et al. (2016) use Hermite polynomials and Hermite functions to investigate single-index models in panel data and non-stationary time series data frameworks respectively. In this paper, we adopt the latter approach.

In order to derive asymptotic properties for single-index models: (1) one can restrict both the space of parameter θ_o and the space of regressor x_{it} to compact sets (cf., Assumptions 5.2 and 5.3 of Ichimura (1993)); or (2) one can bound the link function on the whole real axis (cf., condition C2 of Xia (2006)). However, these settings instantly rule out the linear model considered in Bai (2009), i.e.,

$$y_{it} = g_o(x_{it}'\theta_o) + \gamma_{o,i}' f_{o,t} + \varepsilon_{it} \text{ with } g(w) = \alpha_o w \text{ and } w \text{ being defined on } \mathbb{R}. \tag{1.3}$$

Similar concerns have also been raised by Hansen (2015) recently, where he points out that in nonparametric sieve regression a very limited number of works have been done to cover the cases with unbounded regressors.

Thus, an intriguing question is how to ensure our proposed model (i.e., (1.1)) nests such parametric models as those studied by Bai (2009). To tackle this problem, one method is to introduce a weighted sup norm (e.g., Chen (2007), Su and Jin (2012) and Lee and Robinson (2016)) without specifying too many details of the basis functions. One can also follow Chen and Christensen (2015) to truncate the unbounded support by a compact set depending on the sample size. Alternatively, Dong et al. (2015) and Dong et al. (2016) use Hermite functions and Hermite polynomials as basis functions respectively. In what follows, we adopt the latter approach. It is easy to see that with this approach, (1.3) is nested as a special case of (1.1), due to the nature of function space $L^2(\mathbb{R}, \exp(-w^2))$.

The structure of this paper is as follows: Section 2 presents the basic settings for model (1.1). Section 3 discusses a simple case with an integrable link function. Section 4 further discusses a more involved case with a non-integrable link function. Section 5 provides some guidance on choosing the truncation parameter and the number of factors. A Monte Carlo study and an empirical application are provided in Sections 6 and 7 respectively; Section 8 concludes. All the proofs are provided in the Appendix A and the supplementary file of this paper.

Throughout this paper, we will use the following notations: $\lambda_{min}(W)$ and $\lambda_{max}(W)$ denote the minimum and maximum eigenvalues of a square matrix W, respectively; I_q denotes the identity matrix with dimension q; $M_W = I_T - P_W$ denotes the projection matrix generated by matrix W, where $P_W = W(W'W)^{-1}W'$, and W is a $T \times q$ matrix with rank q; for matrices W_1 and W_2, $W_1 \otimes W_2$ defines the Kronecker product; $\to_P$ and $\to_D$ stand for convergence in probability and convergence in distribution, respectively; $\|\cdot\|$ denotes the Euclidean norm; $\|\cdot\|_{\text{sp}}$ denotes the spectral norm; $\lfloor a \rfloor$ means the largest integer not exceeding a; ρ_1, $O(1)$ and A always denote constants and may be different at each appearance.

2 Basic Settings

As is well-known, single-index models and factor models usually require some restrictions for the purpose of identification (cf., Ichimura (1993) and Bai (2009)). Following this tradition, we begin by introducing identification restrictions needed for (1.1).

For the parameter vector $\theta_o = (\theta_{o,1}, \dots, \theta_{o,d})'$, we follow Xia (2006) and Dong et al. (2016) and define it as follows:

$$\theta_o \in \Theta \text{ with } \Theta \text{ being a compact subset of } \mathbb{R}^d,\ \|\theta_o\| = 1 \text{ and } \theta_{o,1} > 0. \tag{2.1}$$

Throughout this paper, let $\hat{F} = (\hat{f}_1, \dots, \hat{f}_T)'$ and $\hat{\Gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_N)'$ denote the estimates of the factors and factor loadings respectively. For the purpose of identification, we impose the following restrictions:

$$\begin{aligned}
&\hat{F} \in D_F, \text{ where } D_F = \left\{ \underset{T\times m}{F} = (f_1, \dots, f_T)' :\ F'F/T = I_m \right\};\\
&\hat{\Gamma} \in D_\Gamma, \text{ where } D_\Gamma = \left\{ \underset{N\times m}{\Gamma} = (\gamma_1, \dots, \gamma_N)' :\ \Gamma'\Gamma \text{ being diagonal} \right\}.
\end{aligned} \tag{2.2}$$

Note (2.2) has nothing to do with the data generating process of model (1.1), and is only used for identification purpose. For the true factors $F_o = (f_{o,1}, \dots, f_{o,T})'$ and factor loadings $\Gamma_o = (\gamma_{o,1}, \dots, \gamma_{o,N})'$, we require them to satisfy:

$$\begin{aligned}
&f_{o,t} \text{ is identically distributed across } t,\ \frac{F_o'F_o}{T} \to_P \Sigma_F > 0, \text{ and } E\|f_{o,1}\|^4 < \infty;\\
&\gamma_{o,i} \text{ is identically distributed across } i,\ \frac{\Gamma_o'\Gamma_o}{N} \to_P \Sigma_\Gamma > 0, \text{ and } E\|\gamma_{o,1}\|^4 < \infty.
\end{aligned} \tag{2.3}$$

The requirement of identical distributions in (2.3) is only for notational simplicity.

For the link function g_o, we will consider two cases in Sections 3 and 4 respectively. In the first case in Section 3, we assume that the link function is integrable, that is

$$g_o \in \Im_1 \text{ with } \Im_1 \text{ being a compact subset of } L^2(\mathbb{R}). \tag{2.4}$$

However, this would rule out the linear model (i.e., (1.3)) studied in Bai (2009). To deal with this issue, we will consider another case in Section 4, where the link function is not integrable, that is

$$g_o \in \Im_2 \text{ with } \Im_2 \text{ being a compact subset of } L^2\left(\mathbb{R}, \exp(-w^2)\right). \tag{2.5}$$

It is easy to see (2.5) nests model (1.3) due to the nature of function space $L^2(\mathbb{R}, \exp(-w^2))$.

3 The Case with Integrable Link Function

In this section, we investigate the first case of the single-index panel data model in (1.1), where the link function $g_o(w)$ is integrable (i.e., (2.4) holds).

3.1 Estimation

In order to recover the link function, we use the physicists' Hermite polynomial system in this study. Specifically, let $\{H_n(w), n = 0, 1, 2, \ldots\}$ be the physicists' Hermite polynomial system orthogonal with respect to $\exp(-w^2)$. The orthogonality of the system reads

$$\int H_n(w)H_m(w)\exp(-w^2)dw = \sqrt{\pi}2^n n!\delta_{nm}, \tag{3.1}$$

where δ_{nm} is the Kronecker delta.

By (3.1), $\forall g \in L^2(\mathbb{R})$ can be expanded into an infinite series as follows:

$$g(w) = \sum_{n=0}^{\infty} c_n \mathbb{H}_n(w) = g_k(w) + \delta_k(w), \tag{3.2}$$

where

$$\begin{aligned}
&\mathbb{H}_n(w) = \frac{1}{\sqrt[4]{\pi}\sqrt{2^n n!}} H_n(w)\exp(-w^2/2) \quad \text{with} \quad n = 0, 1, 2, \ldots, \\
&g_k(w) = \sum_{n=0}^{k-1} c_n \mathbb{H}_n(w) = C'\mathcal{H}(w), \quad c_n = \int g(w)\mathbb{H}_n(w)dw, \quad \delta_k(w) = \sum_{n=k}^{\infty} c_n \mathbb{H}_n(w), \\
&\mathcal{H}(w) = (\mathbb{H}_0(w), \ldots, \mathbb{H}_{k-1}(w))', \quad C = (c_0, \ldots, c_{k-1})'.
\end{aligned}$$

Throughout this paper, k denotes the truncation parameter, so that $g_k(w)$ is the partial sum of the infinite series, which converges to $g(w)$ under certain conditions (mathematical derivations are omitted at this stage for conciseness). Correspondingly, the true link function in this case can be written as

$$g_o(w) = \sum_{n=0}^{\infty} c_{o,n} \mathbb{H}_n(w) = g_{o,k}(w) + \delta_{o,k}(w), \tag{3.3}$$

where $g_{o,k}(w) = C_o'\mathcal{H}(w)$, $\delta_{o,k}(w) = \sum_{n=k}^{\infty} c_{o,n}\mathbb{H}_n(w)$ and $C_o = (c_{o,0}, \ldots, c_{o,k-1})'$.

Given that our interest lies in both θ_o and g_o, we define a norm $\|\cdot\|_w$ for the 2-fold Cartesian product space formed by $\mathbb{R}^d$ and $L^2(\mathbb{R})$:

$$\|(\theta, g)\|_w = \left\{\|\theta\|^2 + \|g\|_{L^2}^2\right\}^{1/2}, \tag{3.4}$$

where $\theta \in \mathbb{R}^d$, $g \in L^2(\mathbb{R})$, and $\|g\|_{L^2} = \left\{\int g^2(w)dw\right\}^{1/2} = \{\sum_{n=0}^{\infty} c_n^2\}^{1/2}$.

Remark 3.1. $\|\cdot\|_w$ *satisfies the definition of a norm and is consistent with the notations used in Newey and Powell (2003, p. 1569). In this section, when we discuss the set* $\Theta \times \Im_1$ *that* (θ_o, g_o) *belongs to, compactness is always imposed with respect to* $\|\cdot\|_w$.

Multiplying (1.2) by $M_{F_o} = I_T - F_o(F_o'F_o)^{-1}F_o'$ and then using (3.3), we obtain

$$M_{F_o}Y_i = M_{F_o}\phi_i[\theta_o, g_{o,k}] + M_{F_o}\phi_i[\theta_o, \delta_{o,k}] + M_{F_o}\varepsilon_i \quad \text{with} \quad i = 1, \ldots, N, \tag{3.5}$$

where $\phi_i[\theta, g]$ with $i = 1, \ldots, N$ are defined in (1.2).

According to (3.5), the objective function is intuitively defined as

$$S_{NT}(\theta, C, F) = \frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - \phi_i[\theta, g_k]\right)' M_F \left(Y_i - \phi_i[\theta, g_k]\right), \tag{3.6}$$

where $(\theta, C, F) \in \Lambda$, and Λ is defined as

$$\Lambda = \{(\theta, C, F) : \theta \in \Theta, \quad g_k(w) = C'\mathcal{H}(w) \in \Im_{1k}, \quad \text{(2.2) being satisfied}\} \tag{3.7}$$

with $\Im_{1k} = \Im_1 \cap \text{span}\{\mathbb{H}_0(w), \mathbb{H}_1(w), \ldots, \mathbb{H}_{k-1}(w)\}$ and $\Im_1$ being defined in (2.4). Then the estimator is obtained as follows:

$$(\hat{\theta}, \hat{C}, \hat{F}) = \underset{(\theta, C, F)\in\Lambda}{\text{argmin}}\ S_{NT}(\theta, C, F). \tag{3.8}$$

Note that simple algebra shows that $\frac{\partial S_{NT}(\hat{\theta},\hat{C},\hat{F})}{\partial\theta} = 0$ and $\frac{\partial S_{NT}(\hat{\theta},\hat{C},\hat{F})}{\partial C} = 0$ are respectively equivalent to $\frac{\partial S^*_{NT}(\hat{\theta},\hat{C})}{\partial\theta} = 0$ and $\frac{\partial S^*_{NT}(\hat{\theta},\hat{C})}{\partial C} = 0$, where

$$S^*_{NT}(\theta, C) = \frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - \phi_i[\theta, g_k]\right)' M_{\hat{F}} \left(Y_i - \phi_i[\theta, g_k]\right).$$

Further note that the objective function given in (3.6) can be written as

$$(NT)\cdot S_{NT}(\theta, C, F) = \text{tr}(W'M_F W) = \text{tr}(W'W) - \text{tr}(F'WW'F)/T, \tag{3.9}$$

where $W = (W_1, \ldots, W_N)$ is a $T \times N$ matrix and $W_i = Y_i - \phi_i[\theta, g_k]$ for $i = 1, \ldots, N$. (3.9) implies that minimizing $\text{tr}(W'M_F W)$ with respect to F is equivalent to maximizing $\text{tr}(F'WW'F)$. Therefore, the estimate of F is equal to the first m eigenvectors (multiplied by $\sqrt{T}$ due to the restriction $F'F/T = I_T$) associated with the first m largest eigenvalues of the matrix $WW' = \sum_{i=1}^{N} W_i W_i'$.

Therefore, (3.8) can be decomposed into the following two expressions:

$$(\hat{\theta}, \hat{C}) = \underset{\theta, C}{\text{argmin}}\ \frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - \phi_i[\theta, g_k]\right)' M_{\hat{F}} \left(Y_i - \phi_i[\theta, g_k]\right),$$
$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - \phi_i[\hat{\theta}, \hat{g}_k]\right)\left(Y_i - \phi_i[\hat{\theta}, \hat{g}_k]\right)'\right]\hat{F} = \hat{F}V_{NT}, \tag{3.10}$$

where $\hat{g}_k(w) = \hat{C}'\mathcal{H}(w)$, and V_{NT} is a diagonal matrix with the diagonal being the m largest eigenvalues of the following matrix

$$\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - \phi_i[\hat{\theta},\hat{g}_k]\right)\left(Y_i - \phi_i[\hat{\theta},\hat{g}_k]\right)'$$

arranged in descending order. Finally, $\hat{\Gamma}$ is expressed as a function of $(\hat{\theta},\hat{C},\hat{F})$ such that

$$\hat{\Gamma}' = (\hat{\gamma}_1,\ldots,\hat{\gamma}_N) = \frac{1}{T}\left[\hat{F}'\left(Y_1 - \phi_1[\hat{\theta},\hat{g}_k]\right),\ldots,\hat{F}'\left(Y_N - \phi_N[\hat{\theta},\hat{g}_k]\right)\right].$$

3.2 Consistency

To show the consistency of the estimator in (3.8), we make the following assumptions.

Assumption 1:

1. $\{(y_{it},\ x_{it}), 1 \le i \le N, 1 \le t \le T\}$ *are observable. Both* m *and* d *are known and finite.*

2. *The distribution of* $\{x_{i1},\ldots,x_{iT};\varepsilon_{i1},\ldots,\varepsilon_{iT}\}$ *is identical across* i*. Denote* $x_t = (x_{1t},\ldots,x_{Nt})'$ *and* $\xi_t = (\varepsilon_{1t},\ldots,\varepsilon_{Nt})'$*. Let* $\{(x_t,\xi_t), t \ge 1\}$ *be strictly stationary and* α*-mixing. Let* $\alpha_{ij}(|t-s|)$ *denote the* α*-mixing coefficient between* $(x_{it}',\varepsilon_{it})$ *and* $(x_{js}',\varepsilon_{js})$*, such that for a* $\nu_1 > 0$, $\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\alpha_{ij}(|t-s|))^{\frac{\nu_1}{4+\nu_1}} = O(NT)$*. For the same* ν_1, $E[\|x_{11}\| + \|\varepsilon_{11}\|]^{4+\nu_1} \le A < \infty$.

3. *For the error terms, assume further* $E[\varepsilon_{11}] = 0$, $E[\varepsilon_{11}^2] = \sigma_\varepsilon^2$ *and* $\{\varepsilon_{it}, 1 \le i \le N, 1 \le t \le T\}$ *is independent of* $\{x_{it}, 1 \le i \le N, 1 \le t \le T\}$, Γ_o *and* F_o*. Moreover, assume*
   (a) $\|\varepsilon\|_{\mathrm{sp}} = O_P(max\{\sqrt{N},\sqrt{T}\})$*, where* $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_N)'$ *and* ε_i *is defined in (1.2);*
   (b) $\sum_{t\neq s}\sum_{i\neq j}|E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{jt}\varepsilon_{js}]| = O(NT)$;
   (c) $\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{is}\varepsilon_{js}]| = O(NT)$.

4. *Let (2.1)-(2.3) hold, i.e.,*
   (a) $\theta_o \in \Theta$ *with* Θ *being a compact subset of* $\mathbb{R}^d$, $\|\theta_o\| = 1$ *and* $\theta_{o,1} > 0$;
   (b) *Assume* $\hat{F} \in D_F = \{F : F'F/T = I_m\}$ *and* $\hat{\Gamma} \in D_\Gamma = \{\Gamma : \Gamma'\Gamma$ *being diagonal*$\}$;
   (c) $f_{o,t}$ *is identically distributed across* t, $\frac{F_o'F_o}{T} \to_P \Sigma_F > 0$*, and* $E\|f_{o,1}\|^4 \le A < \infty$; $\gamma_{o,i}$ *is identically distributed across* i, $\frac{\Gamma_o'\Gamma_o}{N} \to_P \Sigma_\Gamma > 0$*, and* $E\|\gamma_{o,1}\|^4 \le A < \infty$.

Assumption 1.1 is standard in the literature. For notational simplicity, we assume that the distribution of $\{x_{i1}, \dots, x_{iT}; \varepsilon_{i1}, \dots, \varepsilon_{iT}\}$ is identical across i in Assumption 1.2. One can remove this condition to allow for non-identical distributions; however, this would result in notational clutter when deriving asymptotic results. Imposing strict stationarity on (x_t, ξ_t) is the same as Assumption A4 of Chen et al. (2012*a*) and Assumption A.2 of Chen et al. (2012*b*). Relevant discussions about various mixing conditions can be found in Bradley (2005), Fan and Yao (2003) and Gao (2007). Assumption 1.3 and the mixing conditions of Assumption 1.2 together are equivalent to Assumptions B and C of Bai (2009), Assumption iii of Li et al. (2016) and Assumption A.1.v of Lu and Su (2016). Assumption 1.4 serves the purpose of identification, and is discussed above in Section 2.

Assumption 2:

1. *(i) There exists a positive integer $r > 2$ such that $w^{r-s} g_o^{(s)}(w) \in L^2(\mathbb{R})$ for $s = 0, 1, \dots, r$, where $g_o^{(s)}$ defines the s^{th} derivative of g_o. (ii) Assume $\sup_{(\theta,w)\in\Theta\times\mathbb{R}} f_\theta(w) \leq A < \infty$, where $f_\theta(w)$ defines the density function of $w = x_{11}'\theta$.*

2. *Assume that (2.4) holds, i.e., $g_o \in \Im_1$ with $\Im_1$ being a compact subset of $L^2(\mathbb{R})$. Further assume $L(\theta, g)$ has a unique minimum on $\Theta \times \Im_1$ at (θ_o, g_o), where*

$$L(\theta, g) = E[\Delta g(x_{11}'\theta)]^2 - E[\Delta g(x_{11}'\theta) f_{o,1}'] \Sigma_F^{-1} E[f_{o,1} \Delta g(x_{11}'\theta)]$$

and $\Delta g(x_{11}'\theta) = g(x_{11}'\theta) - g_o(x_{11}'\theta_o)$.

Assumption 2.1 is fairly standard in the literature and ensures that the approximation of the unknown function $g_o(w)$ by an orthogonal expansion has a fast rate of convergence (cf., Condition C2 of Xia (2006) and Assumption B of Dong et al. (2016)). Assumption 2.1 can be further simplified, if one adopts the norm provided in Assumption 3 of Newey (1997) for the function space. Assumption 2.2 is for the purpose of identification, and is in the same spirit as Assumption 1 of Newey and Powell (2003).

Consider the least squares projection of $\Delta g(x_{it}'\theta) = g\left(x_{it}'\theta\right) - g_o\left(x_{it}'\theta_o\right)$ on $f_{o,t}$:

$$\Delta g(x_{it}'\theta) = f_{o,t}'\beta^* + u_{it},$$

where $\beta^* = \Sigma_F^{-1} E\left[f_{o,1} \Delta g(x_{11}'\theta)\right]$ and $u_{it} := u_{it}\left(\theta, g\right) = \Delta g(x_{it}'\theta) - f_{o,t}'\Sigma_F^{-1} E\left[f_{o,1}\Delta g(x_{11}'\theta)\right]$. If the link function is linear, then we have $g\left(w\right) = g_o\left(w\right) = w$, which immediately gives

$$\Delta g(x_{it}'\theta) = x_{it}'\left(\theta - \theta_o\right) \text{ and } u_{it} = \left\{x_{it} - E\left[x_{11} f_{o,1}'\right] \Sigma_F^{-1} f_{o,t}\right\}'\left(\theta - \theta_o\right).$$

For this special case, $L(\theta, g)$ can be written as

$$L(\theta) = (\theta - \theta_o)' E\left\{\left[x_{11} - E\left[x_{11} f'_{o,1}\right]\Sigma_F^{-1} f_{o,1}\right]\left[x_{11} - E\left[x_{11} f'_{o,1}\right]\Sigma_F^{-1} f_{o,1}\right]'\right\}(\theta - \theta_o),$$

which is uniquely minimized at θ_o provided that

$$E\left\{\left[x_{11} - E\left[x_{11} f'_{o,1}\right]\Sigma_F^{-1} f_{o,1}\right]\left[x_{11} - E\left[x_{11} f'_{o,1}\right]\Sigma_F^{-1} f_{o,1}\right]'\right\}$$

is full rank, a typical condition for the consistency of regression coefficients in linear panel data models with interactive fixed effects.

With the above assumptions, the consistency of the estimator (3.8) can be established as follows.

Theorem 3.1. *Under Assumptions 1 and 2, as* $(N, T) \to (\infty, \infty)$,

1. $\|P_{\hat{F}} - P_{F_o}\| \to_P 0$, *where* $P_{\hat{F}}$ *and* P_{F_o} *are the idempotent matrices generated by* $\hat{F}$ *and* F_o *respectively;*
2. $\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w \to_P 0$, *where* $\hat{g}_k(w) = \hat{C}'\mathcal{H}(w)$ *and* $\mathcal{H}(w)$ *is defined in (3.2).*

It is well understood that F_o is identifiable up to a non-singular matrix in the literature, so we establish the consistency for the idempotent matrix $P_{\hat{F}}$ rather than $\hat{F}$ itself. Note that in the second result of Theorem 3.1 the consistency is established with respect to the norm defined in (3.4).

Using Theorem 3.1, we obtain the following rates of convergence.

Lemma 3.1. *Let* $\eta_{NT} = \frac{1}{\min\{\sqrt{N}, \sqrt{T}\}}$ *and* $Q^{-1} = V_{NT}(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}$. *Under Assumptions 1-2, as* $(N, T) \to (\infty, \infty)$,

1. $V_{NT} \to_P V$, *where* V_{NT} *is defined in (3.10), and* V *is an* $m \times m$ *diagonal matrix consisting of the eigenvalues of* $\Sigma_F\Sigma_\Gamma$, *and* Σ_F *and* Σ_Γ *are defined in (2.3);*
2. $\left\|\frac{1}{\sqrt{T}}(\hat{F}Q^{-1} - F_o)\right\| = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P(\eta_{NT})$;
3. $\left\|\frac{1}{T}F_o'(\hat{F} - F_oQ)\right\| = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P(\eta_{NT}^2)$;
4. $\|P_{\hat{F}} - P_{F_o}\|^2 = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P(\eta_{NT}^2)$.

The results of Lemma 3.1 are analogous to Proposition A.1 and Lemma A.7 of Bai (2009). However, due to the semiparametric setting, estimates of both the parameter of interest and link function have impacts on the rate of convergence for the estimates of unknown factors.

3.3 Asymptotic Normality of $\hat{\theta}$

In this subsection, we establish asymptotic normality of $\hat{\theta}$.

Assumption 3:

Let ϵ be a sufficiently small positive number, $\Omega(\epsilon) = \{(\theta, g) :\ \|(\theta, g) - (\theta_o, g_o)\|_{L^2} \leq \epsilon\}$.

1. *Assume that*

   (a) $$\sup_{(\theta_1,g_1),(\theta_2,g_2)\in\Omega(\epsilon)} \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} G_{it}(\theta_1, g_1, \theta_2, g_2) - E[G_{11}(\theta_1, g_1, \theta_2, g_2)] \right\| = o_P(1),$$
   where $G_{it}(\theta_1, g_1, \theta_2, g_2) = g_1(x_{it}'\theta_1) g_2^{(2)}(x_{it}'\theta_2) x_{it} x_{it}'$;

   (b) $$\max_{1\leq i\leq N} \sup_{(\theta,g)\in\Omega(\epsilon)} \left\| \frac{1}{T}\sum_{t=1}^{T} x_{it} g^{(1)}(x_{it}'\theta) f_{o,t}' - E\left[x_{11} g^{(1)}(x_{11}'\theta) f_{o,1}'\right] \right\| = o_P(1).$$

2. *Assume that*

   (a) $0 < \rho_1 \leq \inf_{\{\theta:\ \|\theta-\theta_o\|\leq\epsilon\}} \lambda_{min}(\Sigma_0(\theta)) \leq \sup_{\{\theta:\ \|\theta-\theta_o\|\leq\epsilon\}} \lambda_{max}(\Sigma_0(\theta)) \leq A$ *uniformly in k, where* $\Sigma_0(\theta) = E[\mathcal{H}(x_{11}'\theta)\mathcal{H}(x_{11}'\theta)']$;

   (b) $\sup_{\{\theta:\ \|\theta-\theta_o\|\leq\epsilon\}} \lambda_{max}(\Sigma_1(\theta)) \leq A$, *where* $\Sigma_1(\theta) = E[\dot{\mathcal{H}}(x_{11}'\theta)\dot{\mathcal{H}}(x_{11}'\theta)']$ *uniformly in k, where* $\dot{\mathcal{H}}(w) = (\mathbb{H}_0^{(1)}(w), \ldots, \mathbb{H}_{k-1}^{(1)}(w))'$.

3. *Let* $\max_{i,j,t_1,t_2,t_3,t_4} E\left[\|x_{it_1}\|^2\|x_{jt_2}\|^2\|f_{o,t_1}\|^2\|f_{o,t_2}\|^2\|f_{o,t_3}\|^2\|f_{o,t_4}\|^2\right] \leq A < \infty$, *where* $1 \leq i, j \leq N$ *and* $1 \leq t_1, t_2, t_3, t_4 \leq T$.

4. *Assume that* $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \psi_{1i}[\theta_o, g_o^{(1)}]' M_{F_o} \varepsilon_i \ \to_D\ N(0, \tilde{\Sigma})$, *where, for* $i = 1, \ldots, N$, $\psi_{1i}[\theta_o, g_o] = (g_o(x_{i1}'\theta_o)x_{i1}, \ldots, g_o(x_{iT}'\theta_o)x_{iT})'$.

5. *For the same r defined in Assumption 2.1, let* $\frac{NT}{k^r} \to 0$, *as* $(N, T, k) \to (\infty, \infty, \infty)$.

Assumption 3.1 is the same as Assumption 2 of Yu and Ruppert (2002) and Assumptions iii-iv of Li et al. (2016), and requires uniform convergence in a small neighbourhood of (θ_o, g_o). We can further decompose Assumption 3.1 by using Lemma A2 of Newey and Powell (2003) and prove the uniform convergence by following a procedure similar to those given for (1) of Lemma A2. However, this would entail a lengthy derivation. Assumption 3.2 is not needed, if the norm of Assumption 3 of Newey (1997) is adopted. Assumption 3.3 can be removed, if one is willing to bound higher moments of x_{it} and $f_{o,t}$. Assumption 3.4 is in the same spirit as Assumption E of Bai (2009). Assumption 3.5 ensures the truncation residual goes to 0 sufficiently fast, so it can be smoothed out when establishing the asymptotic normality.

Using the above assumptions, the asymptotic normality can be obtained as follows.

Theorem 3.2. *Under Assumptions 1-3, as* $(N, T) \to (\infty, \infty)$,

$$\sqrt{NT}\left(\hat{\theta} - \theta_o + V_*^{-1}\Pi_{NT1} + V_*^{-1}\Pi_{NT2}\right) \to_D N(0, \Sigma^*),$$

where $\Sigma_* = V_*^{-1}\tilde{\Sigma}V_*^{-1}$, $\tilde{\Sigma}$ *and* $\psi_{1i}[\theta_o, g_o]$ *with* $i = 1, \ldots, N$ *are defined in Assumption 3.4, and*

$$V_* = V_1 - V_2\Sigma_F^{-1}V_2', \quad V_1 = E\left[\left(g_o^{(1)}(x_{11}'\theta_o)\right)^2 x_{11}x_{11}'\right], \quad V_2 = E\left[g_o^{(1)}(x_{11}'\theta_o)x_{11}f_{o,1}'\right],$$

$$\Pi_{NT1} = \frac{1}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta_o, \hat{g}_k^{(1)}]'M_{\hat{F}}(\phi_i[\theta_o, g_o] - \phi_i[\theta_o, \hat{g}_k]), \quad \hat{g}_k(w) = \hat{C}'\mathcal{H}(w),$$

$$\Pi_{NT2} = \frac{1}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta_o, \hat{g}_k^{(1)}]'M_{\hat{F}}F_o\gamma_{o,i}.$$

Remark 3.2.

1. Π_{NT1} *is similar to condition (2.6) of Theorem 2 in Chen et al. (2003), where they point out that the verification of this type of condition is in some cases difficult, and is itself the subject of a long paper Newey (1994). Although* Π_{NT1} *is exactly 0 uniformly in* N *and* T *when* $\hat{g}_k = g_o$, *we cannot further decompose* Π_{NT1} *without imposing further restrictions (e.g., "Linearization" condition of Assumption 5.1 of Newey (1994)).*

2. Π_{NT2} *is equivalent to the biased terms in Bai (2009). Due to the semi-parametric nature of the single-index model, it is not helpful to further decompose* Π_{NT2} *as* $\Pi_{NT2} = \frac{1}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta_o, \hat{g}_k^{(1)}]'M_{\hat{F}}(F_o - \hat{F}Q^{-1})\gamma_i$, *where* Q *is given in Lemma 3.1. This is because further decomposing* $F_o - \hat{F}Q^{-1}$ *will introduce the residual* $O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w)$ *into the system. For parametric models, decomposing* $F_o - \hat{F}Q^{-1}$ *will only introduce* $\hat{\theta} - \theta_o$ *to the system. After further rearranging and imposing conditions like* $N/T \to A$, *asymptotic normality can be established as in Bai (2009). However, for our semi-parametric model, it is not the case any more.*

3. *If* $g_o(w) = \alpha w$ *with* α *being a constant is known, we can obtain* $g_o^{(1)} = \alpha$. *Then* Π_{NT1} *will disappear from the system, and we will be able to further decompose* Π_{NT2} *as Bai (2009). In this case, we will get exactly the same components (i.e.,* B, C *and* D_0 *on page 1247 of Bai (2009)) for our asymptotic distribution.*

4. *While theoretically it remains unknown if the biases (i.e., Π_{NT1} and Π_{NT2}) can be removed, in practice one can always follow the discussion in Li et al. (2013, p. 558) and use bootstrapping techniques to make statistical inferences on θ_o.*

3.4 Rate of Convergence of $\hat{g}_k$

In this section we derive the rate of convergence of $\|\hat{g}_k - g_o\|_{L^2}$. For this purpose, we make the following assumptions.

Assumption 4:

1. *Let* $\Upsilon(\theta) = E[\mathcal{H}(x_{11}'\theta)\mathcal{H}(x_{11}'\theta)'] - E[\mathcal{H}(x_{11}'\theta)f_{o,1}']\Sigma_F^{-1}E[f_{o,1}\mathcal{H}(x_{11}'\theta)']$. *Assume that the minimum eigenvalue of* Υ *is bounded away from 0 uniformly in* k *and a small neighborhood of* θ_o *(i.e.,* $\inf_{\{\theta:\ \|\theta-\theta_o\|\leq\epsilon\}}\lambda_{min}(\Upsilon(\theta)) \geq \rho_1 > 0$ *uniformly in* k*.)*
2. Assume that $\frac{k^2\ln(NT)}{NT} \to 0$ and $\frac{k\ln(T)}{T} \to 0$ as $(N,T,k)\to(\infty,\infty,\infty)$.

Assumption 4.2 is standard in the literature (e.g., Assumption A5 of Chen et al. (2012*b*)). We now show that Assumption 4.1 is reasonable. Suppose that the elements of x_{11} follow a normal distribution with variance $\frac{1}{2}$, and are independent of each other. Then it is easy to show that $x_{11}'\theta$ follows a normal distribution with variance $\frac{1}{2}\|\theta\|^2$. With the identification restriction $\|\theta\|^2 = 1$, we obtain $f_\theta(w) = \frac{1}{\sqrt{\pi}}\exp(-w^2)$, where $f_\theta(w)$ denotes the pdf of $x_{11}'\theta$. Then it is easy to see that $E[\mathcal{H}(x_{11}'\theta)\mathcal{H}(x_{11}'\theta)']$ reduces to I_k. In the special case where $f_{o,1}$ is independent of x_{11} and has mean 0, we immediately obtain that $\Upsilon(\theta) = I_k$. In this case, it is straightforward to show that Assumption 4 holds.

Theorem 3.3. *Under Assumptions 1-4, as* $(N,T)\to(\infty,\infty)$,

$$\|\hat{g}_k - g_o\|_{L^2} = O_P\left(\sqrt{k}\eta_{NT}\right) + O_P(k^{-r/2}), \tag{3.11}$$

where η_{NT} *is denoted in Lemma 3.1.*

Due to the presence of the factor structure, the leading term $\sqrt{k}\eta_{NT}$ is much slower than $\sqrt{k(NT)^{-1}}$, which is a result commonly found in traditional nonparametric panel data models without interactive fixed effects (e.g., Chen et al. (2012*b*) and Dong et al. (2015)). The term $O_P(k^{-r/2})$ represents the rate of convergence of truncation residual and is quite standard in the literature (e.g., Newey (1997)). Since we consider g_o on the Hilbert space $L^2(\mathbb{R})$ in this section, $x_{11}'\theta_o$ can be defined on the whole real axis. Thus, we will not establish uniform convergence for $\hat{g}_k$. Moreover, one can follow Su et al. (2015) to develop a hypothesis testing procedure for the function form of g_o. Considering the development of such a procedure can form a different paper (cf., Bai (2009) and Su et al. (2015)), we leave it for future research.

4 Beyond Integrability

Following the spirit of Dong and Gao (2014), in this section we consider the second case of the single-index panel data model in (1.1), where the link function is not integrable (i.e., (2.5)). As discussed above, this case nests the model (i.e., (1.3)) studied by Bai (2009) as a special case.

Accordingly, a new norm for the 2-fold Cartesian product space formed by $\mathbb{R}^d$ and $L^2\left(\mathbb{R}, \exp(-w^2)\right)$ is defined as follows:

$$\|(\theta, g)\|_{\widetilde{w}} = \left\{\|\theta\|^2 + \|g\|^2_{\widetilde{L}^2}\right\}^{1/2}, \tag{4.1}$$

where $\theta \in \mathbb{R}^d$, $g \in L^2(\mathbb{R}, \exp(-w^2))$, and $\|g\|_{\widetilde{L}^2} = \left\{\int g^2(w)\exp(-w^2)dw\right\}^{1/2}$. Remark 3.1 applies to the new norm $\|\cdot\|_{\widetilde{w}}$.

By (3.1), $\forall g \in L^2(\mathbb{R}, \exp(-w^2))$ can be expanded into an infinite series as follows:

$$g(w) = \sum_{n=0}^{\infty} c_n h_n(w) = g_k(w) + \delta_k(w), \tag{4.2}$$

where

$$\begin{aligned}
&h_n(w) = \frac{1}{\sqrt[4]{\pi}\sqrt{2^n n!}} H_n(w) \quad \text{with} \quad n = 0, 1, 2, \ldots, \\
&g_k(w) = \sum_{n=0}^{k-1} c_n h_n(w) = C'\mathsf{H}(w), \quad c_n = \int g(w) h_n(w) dw, \quad \delta_k(w) = \sum_{n=k}^{\infty} c_n h_n(w), \\
&\mathsf{H}(w) = (h_0(w), \ldots, h_{k-1}(w))', \quad C = (c_0, \ldots, c_{k-1})'.
\end{aligned}$$

Correspondingly, the true link function in this case can be written as

$$g_o(w) = \sum_{n=0}^{\infty} c_{o,n} h_n(w) = g_{o,k}(w) + \delta_{o,k}(w), \tag{4.3}$$

where $g_{o,k}(w) = C_o'\mathsf{H}(w)$, $\delta_{o,k}(w) = \sum_{n=k}^{\infty} c_{o,n} h_n(w)$ and $C_o = (c_{o,0}, \ldots, c_{o,k-1})'$.

The objective function is then rewritten as follows:

$$\widetilde{S}_{NT}(\theta, C, F) = \frac{1}{NT} \sum_{i=1}^{N} \left(Y_i - \phi_i[\theta, g_k]\right)' M_F \left(Y_i - \phi_i[\theta, g_k]\right), \tag{4.4}$$

where $(\theta, C, F) \in \widetilde{\Lambda}$, and $\widetilde{\Lambda}$ is defined as

$$\widetilde{\Lambda} = \{(\theta, C, F) : \theta \in \Theta, \quad g_k(w) = C'\mathsf{H}(w) \in \Im_{2k}, \quad \text{(2.2) being satisfied}\}$$

with $\Im_{2k} = \Im_2 \cap \text{span}\{h_0(w), h_1(w), \ldots, h_{k-1}(w)\}$ and $\Im_2$ being defined in (2.5). For (4.4), the estimate $(\hat{\theta}, \hat{C}, \hat{F})$ is obtained in exactly the same way as in (3.8) and (3.10).

Remark 4.1. *The only difference between the two objective functions (3.6) and (4.4) lies in the basis functions used to recover the link function. In practice, choosing the correct basis functions requires prior knowledge of the economic model under study. A detailed discussion and explanation can be found in Chen (2007).*

To derive consistency with respect to the norm $\|\cdot\|_{\widetilde{w}}$, we make the following assumptions.

Assumption 2*:

1. (a) $\sup_{\{(\theta,w):\ \Theta\times\mathbb{R}\}} \exp(w^2)\, f_\theta(w) \leq A$, *where* $f_\theta(w)$ *defines the probability density function (pdf) of* $w = x_{11}'\theta$ *for* $\forall\theta \in \Theta$.
   (b) $\sup_{\{(\theta,g):\ \Theta\times\Im_2\}} E\left[\|x_{11}\|^2 \left\{g^{(1)}(x_{11}'\theta)\right\}^2\right] \leq A < \infty$.
   (c) *For the decomposition (4.3), assume that* $\sum_{n=k}^{\infty} c_{o,n}^2 = O(k^{-r})$ *with* $r \geq 2$ *being a positive constant.*

2. *Assume (2.5) holds, i.e.,* $g_o \in \Im_2$ *with* $\Im_2$ *being a compact subset of* $L^2(\mathbb{R}, \exp(-w^2))$. *Assume that* $L(\theta, g)$ *has a unique minimum on* $\Theta \times \Im_2$ *at* (θ_o, g_o), *where*

$$L(\theta, g) = E[\Delta g(x_{11}'\theta)]^2 - E[\Delta g(x_{11}'\theta) f_{o,1}']\Sigma_F^{-1} E[f_{o,1}\Delta g(x_{11}'\theta)]$$

   and $\Delta g(x_{11}'\theta) = g(x_{11}'\theta) - g_o(x_{11}'\theta_o)$. *For* $\forall(\theta, g) \in \Theta \times \Im_2$, *assume further that* $L_{NT}(\theta, g) \to_P L(\theta, g)$, *where*

$$L_{NT}(\theta, g) = \frac{1}{NT}\sum_{i=1}^{N} (\phi_i[\theta, g] - \phi_i[\theta_o, g_o])' M_{F_o} (\phi_i[\theta, g] - \phi_i[\theta_o, g_o]).$$

In Assumption 2*.1.a, due to the fact that the link function is not integrable and unbounded, we need to impose a stronger assumption on the density of $x_{11}'\theta$. However, it rules out some heavy-tailed distributions for $x_{11}'\theta$.

For Assumption 2*.1.b, if we adopt Assumption 5.3.1 of Ichimura (1993) (i.e., x_{11} belongs to a compact set), it can be further simplified as follows:

- If x_{11} belongs to a compact set, we can write

$$E\left[\|x_{11}\|^2\{g^{(1)}(x_{11}'\theta)\}^2\right] \leq O(1)E\left|g^{(1)}(x_{11}'\theta)\right|^2 = O(1)\int\left|g^{(1)}(w)\right|^2 f_\theta(w)dw$$
$$\leq O(1)\int\left|g^{(1)}(w)\right|^2 \exp(-w^2)\cdot\exp(w^2)f_\theta(w)dw \leq O(1)\int\left|g^{(1)}(w)\right|^2\exp(-w^2)dw,$$

  where $f_\theta(w)$ is the same as that defined in Assumption 2*.1.a. In this case, Assumption 2*.1.b reduces to $g^{(1)}(w) \in L^2(\mathbb{R}, \exp(-w^2))$ for $\forall g \in \Im_2$.

Assumption 2*.1.c is in the same spirit as Assumption 2.1, and is used to ensure the truncation residual $\delta_{o,k}(x_{11}'\theta_o)$ converges to zero at a rate of $k^{-\frac{r}{2}}$ in probability one. A detailed explanation is given as follows.

- By (4.3), write

$$E\left[\delta_{o,k}^2(x_{11}'\theta_o)\right] = \int\left\{\sum_{n=k}^{\infty}c_{o,n}h_n(w)\right\}^2\exp(-w^2)\cdot\exp(w^2)f_{\theta_o}(w)dw$$
$$\leq O(1)\int\left\{\sum_{n=k}^{\infty}c_{o,n}h_n(w)\right\}^2\exp(-w^2)dw = O(1)\sum_{n=k}^{\infty}c_{o,n}^2 = O(k^{-r}), \tag{4.5}$$

  where the first inequality and the third equality follow from Assumptions 2*.1.a and 2*.1.c respectively.

We now show Assumption 2*.2 is reasonable.

- If $\sup_{(\theta,g)\in\Theta\times\Im_2} E|g(x_{11}'\theta)|^{4+\nu_1} < \infty$ for the same ν_1 defined in Assumption 1, we can show, by applying the same procedure as for (9) of Lemma A2, that for $\forall(\theta, g) \in \Theta\times\Im_2$, $L_{NT}(\theta, g) = L(\theta, g) + O_P\left(\frac{1}{\sqrt{NT}}\right)$. Thus, $\sup_{(\theta,g)\in\Theta\times\Im_2} E|g(x_{11}'\theta)|^{4+\nu_1} < \infty$ is sufficient for Assumption 2*.2.

Based on the above setting, the consistency can be stated as follows.

Theorem 4.1. *Under Assumptions 1 and 2*, as $(N, T) \to (\infty, \infty)$,*

1. $\|P_{\hat{F}} - P_{F_o}\| \to_P 0$, *where $P_{\hat{F}}$ and P_{F_o} are the idempotent matrices generated by $\hat{F}$ and F_o respectively;*

2. $\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_{\tilde{w}} \to_P 0$, *where $\hat{g}_k(w) = \hat{C}'\mathsf{H}(w)$ and $\mathsf{H}(w)$ is defined in (4.2).*

All the discussions following Theorem 3.1 also apply to Theorem 4.1, and thus are omitted here. Using Theorem 4.1, it is straightforward to establish the rates of convergence as follows.

Lemma 4.1. *Let η_{NT} and Q^{-1} be those defined in Lemma 3.1. Under Assumptions 1 and 2*, as $(N,T) \to (\infty, \infty)$,*

1. *$V_{NT} \to_P V$, and V is an $m \times m$ diagonal matrix consisting of the eigenvalues of $\Sigma_F \Sigma_\Gamma$;*
2. *$\left\| \frac{1}{\sqrt{T}} (\hat{F} Q^{-1} - F_o) \right\| = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_{\widetilde{w}}) + O_P(\eta_{NT})$;*
3. *$\left\| \frac{1}{T} F_o' (\hat{F} - F_o Q) \right\| = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_{\widetilde{w}}) + O_P(\eta_{NT}^2)$;*
4. *$\|P_{\hat{F}} - P_{F_o}\|^2 = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_{\widetilde{w}}) + O_P(\eta_{NT}^2)$.*

Note that more restrictions are needed in order to prove asymptotic normality for $\hat{\theta}$ and establish the rate of convergence for $\hat{g}_k$. However, these additional restrictions will rule out more potential functions for $g_o(\cdot)$. These proofs are very similar to those for Theorems 3.2 and 3.3, and thus are omitted.

5 Determination of k and m

In this section, we discuss how to choose the truncation parameter k and the number of factors m in this section. Assuming the link function is smooth enough, $k = \lfloor (NT)^{1/5} \rfloor$ always satisfies the assumptions in Sections 3 and 4. Although $\lfloor (NT)^{1/5} \rfloor$ may not be the optimal k, all the asymptotic results derived above remain valid. Therefore, in the numerical studies in Section 6 and the empirical application in Section 7, we always use $k = \lfloor (NT)^{1/5} \rfloor$ (see Su and Jin (2012) and Dong et al. (2016) for similar arguments and settings).

With the choice of the truncation parameter, we choose the number of factors by minimizing the following criterion function:

$$CI(m) = \ln\left(\frac{1}{NT} \sum_{i=1}^{N} \left(Y_i - \phi_i[\hat{\theta}, \hat{g}_k] - \hat{F}\hat{\gamma}_i \right)' \left(Y_i - \phi_i[\hat{\theta}, \hat{g}_k] - \hat{F}\hat{\gamma}_i \right) \right) + m \cdot \frac{N+T}{NT} \cdot \ln\left(\frac{NT}{N+T} \right). \quad (5.1)$$

The basic idea behind (5.1) is exactly the same as that in Bai (2003) and thus is not discussed further here. We investigate the performance of the criterion function in (5.1) via the Monte Carlo simulations in the following section. Our simulation results suggest that overall the criterion function works well for choosing the number of factors.

In general, how to simultaneously choose the optimal truncation parameter and the number of factors remains an open issue. For the case of cross-sectional or time series data, previous studies have examined the choices of optimal bandwidth and truncation parameter (e.g., Gao et al. (2002), Li and Racine (2010) and Li et al. (2013)) for nonparametric and varying-coefficient models. For the case of panel data, even the choice of optimal bandwidth (or truncation parameter) alone remains unresolved (see, for example, Sun et al. (2009), Su and Jin (2012) and Chen et al. (2012*b*)); let alone simultaneous choice of optimal truncation parameter and the number of factors. For single-index panel data models with interactive fixed effects, simultaneous choice of k and m is even more daunting, and thus we will leave it for future work.

6 Monte Carlo Study

In this section, we perform Monte Carlo simulations to investigate the finite sample properties of our models and estimators. The data generating process (DGP) is as follows:

$$y_{it} = g_o(x_{it}'\theta_o) + \gamma_{o,i}' f_{o,t} + \varepsilon_{it}.$$

Let $d = 2$ and $m = 3$. The factor loadings $\gamma_{o,i}$'s are generated as $\gamma_{o,i} \sim$ i.i.d. $N(i_m, I_m)$ for $i = 1, \ldots, N$, where i_m denotes an $m \times 1$ one factor and I_m denotes an $m \times m$ identity matrix. The factors $f_{o,t}$'s are generated as $f_{o,t} \sim$ i.i.d. $N(0_m, I_m)$ for $t = 1, \ldots, T$, where 0_m denotes an $m \times 1$ zero factor. The error terms (denoted by $\xi_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$) are generated using

$$\xi_t = 0.5\xi_{t-1} + v_t,$$

where $v_t \sim$ i.i.d. $N(0_N, \Sigma_v)$ for $t = 1, \ldots, T$ and the $(i,j)^{th}$ element of Σ_v is $0.5^{|i-j|}$ for $1 \le i, j \le N$. For regressors, let

$$x_{it} = 1 + \sum_{j=1}^{m} |\gamma_{o,i,j} \cdot f_{o,t,j}| + \delta_{it},$$

where $\delta_{it} \sim$ i.i.d. $N(0_d, I_d)$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and $\gamma_{o,i,j}$ and $f_{o,t,j}$ denote the j^{th} elements of $\gamma_{o,i}$ and $f_{o,t}$ respectively.

For the link function, we consider two cases, where the first case is integrable, while the second case is not integrable. Specifically, these two cases are given as follows:

1. Case 1: $g_o(w) = (1 + w^2) \exp(-w^2)$ such that $g_o \in L^2(\mathbb{R})$;
2. Case 2: $g_o(w) = \exp((0.6 + w)^2/8)$ such that $g_o \in L^2(\mathbb{R}, \exp(-w^2))$.

Case 1 is estimated using the method outlined in Section 3, while Case 2 is estimated using the method described in Section 4.

For simplicity, we refer to the single-index panel data model with interactive fixed effects as the "SI_Factor model". When estimating the "SI_Factor model", the truncation parameter k and the number of factors m are always chosen as described in Section 5. For comparison purposes, we also use two other models to estimate θ_o. The first model is the linear panel data model with interactive fixed effect proposed Bai (2009). For simplicity, we refer to this model as the "L_Factor model". The second model is the traditional panel data models with fixed effects. We refer to this latter model as the "F_Panel model" below.

For each generated data set and each of the three models mentioned above, we calculate the following bias and squared error ("se" for short): $bias = \hat{\theta}_j - \theta_{o,j}$ and $se = (\hat{\theta}_j - \theta_{o,j})^2$ for $j = 1, \ldots, d$, where $\hat{\theta}_j$ and $\theta_{o,j}$ denote the j^{th} elements of $\hat{\theta}$ and θ_o respectively. We repeat the above procedure 1000 times and calculate the mean of these biases (Bias) and the root of the mean of these squared errors (RMSE). Results for Cases 1 and 2 are summarized in Tables 1 and 2 respectively.

Two findings emerge from Tables 1 and 2. First, the biases and RMSEs associated with the SI_Factor model are much smaller than those associated with the other two models. For example, when $N = T = 80$, the biases for $\hat{\theta}_1$ and $\hat{\theta}_2$ associated with the SI_Factor model are 0.001 and 0.004 respectively, whereas those associated with the L_Factor model are -0.8839 and 0.8404; and those associated with the F_Panel model are -0.9001 and 0.8266. This suggests that compared with those obtained from the other two models, parameter estimates obtained from the SI_Factor are more accurate. Second, the biases associated with the SI_Factor model become virtually zero when both N and T are greater than or equal to 160. In contrast, those associated with the other two models are well above zero even when $N = T = 200$.

It is of interest to examine how the accuracy of the estimated link function changes as N and T increase. For this purpose, we plot in Figure 1 (Figure 2) the 95% upper and lower bounds of $\hat{g}_k(w)$ for Case 1 (Case 2) on a selected interval based on 1,000 replications. As can be seen, the bounds become tighter and tighter as N and T increase, implying the estimates of the link function become more and more accurate as N and T increase.

It is also of interest to examine how the accuracy of the estimates of the number of factors, chosen using (5.1), changes as sample size increases. To this end, we calculate the percentage of replications where the number of factors is underestimated (i.e., $\hat{m} < 3$), overestimated (i.e., $\hat{m} > 3$), or accurately estimated (i.e., $\hat{m} = 3$). The results are reported in Table 3. This table shows that as N and T increase, the percentage of replications, where the number of factors are accurately estimated, quickly approaches one. This suggests that the procedure for selecting the number of factors given in (5.1) work very well when sample size is large.

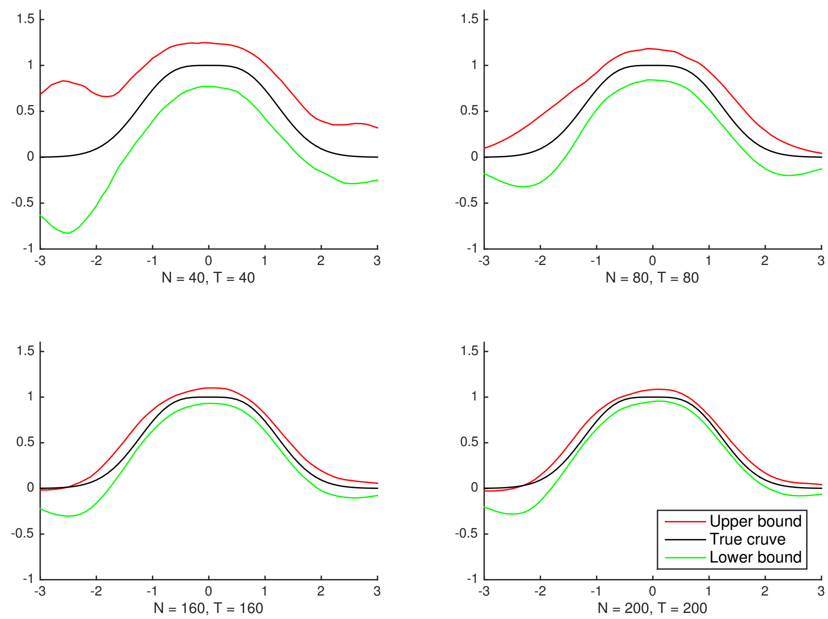

Figure 1: Estimated Link Function for Case 1 Where $g_o(w) = (1 + w^2)\exp(-w^2)$

7 An Application to U.S. Banking Data

In this section, we provide an application of the single-index panel data with interactive fixed effects to the analysis of returns to scale of large commercial banks in the U.S. Over the past three decades, the increasing dominance of large banks in the U.S. banking industry, caused by fundamental regulatory changes and technological and financial innovations, has stimulated considerable research into returns to scale at large banks in the U.S. Major regulatory changes include the removal of restrictions on interstate banking and branching and the elimination of restrictions against combinations of banks, securities firms, and

Table 1: Biases and RMSEs for Case 1 Where $g_o(w) = (1 + w^2)\exp(-w^2)$

		Bias								RMSE							
		$\hat{\theta}_1$				$\hat{\theta}_2$				$\hat{\theta}_1$				$\hat{\theta}_2$			
	$N \setminus T$	40	80	160	200	40	80	160	200	40	80	160	200	40	80	160	200
SI_Factor	40	-0.0012	-0.0001	0.0002	0.0004	0.0044	0.0003	0.0004	0.0007	0.0403	0.0138	0.0096	0.0085	0.0745	0.0185	0.0128	0.0113
	80	-0.0003	0.0001	0.0000	0.0001	0.0002	0.0004	0.0001	0.0002	0.0152	0.0097	0.0067	0.0060	0.0202	0.0130	0.0089	0.0080
	160	0.0004	-0.0002	0.0000	0.0000	0.0007	-0.0001	0.0000	0.0000	0.0105	0.0070	0.0046	0.0040	0.0140	0.0093	0.0061	0.0054
	200	0.0002	0.0003	0.0002	-0.0001	0.0005	0.0005	0.0003	-0.0001	0.0092	0.0059	0.0040	0.0036	0.0123	0.0078	0.0054	0.0048
L_Factor	40	-0.8867	-0.8850	-0.8840	-0.8852	0.8354	0.8382	0.8407	0.8396	0.8877	0.8855	0.8843	0.8855	0.8364	0.8386	0.8410	0.8398
	80	-0.8852	-0.8839	-0.8850	-0.8844	0.8380	0.8404	0.8404	0.8414	0.8858	0.8841	0.8851	0.8845	0.8385	0.8406	0.8405	0.8415
	160	-0.8857	-0.8854	-0.8855	-0.8854	0.8388	0.8408	0.8406	0.8411	0.8861	0.8856	0.8856	0.8855	0.8391	0.8409	0.8407	0.8411
	200	-0.8843	-0.8855	-0.8849	-0.8851	0.8392	0.8399	0.8412	0.8414	0.8846	0.8857	0.8849	0.8852	0.8394	0.8400	0.8413	0.8414
F_Panel	40	-0.9025	-0.8995	-0.9014	-0.9012	0.8254	0.8279	0.8270	0.8287	0.9049	0.9008	0.9020	0.9017	0.8282	0.8293	0.8278	0.8293
	80	-0.8996	-0.9001	-0.9010	-0.9021	0.8257	0.8266	0.8258	0.8260	0.9016	0.9011	0.9015	0.9025	0.8277	0.8275	0.8263	0.8264
	160	-0.9006	-0.9016	-0.9026	-0.9010	0.8279	0.8260	0.8246	0.8261	0.9020	0.9023	0.9030	0.9013	0.8295	0.8269	0.8250	0.8265
	200	-0.9024	-0.9015	-0.9026	-0.9015	0.8243	0.8261	0.8244	0.8267	0.9039	0.9022	0.9030	0.9018	0.8260	0.8269	0.8248	0.8270

Table 2: Biases and RMSEs for Case 2 Where $g_o(w) = \exp((0.6 + w)^2/8)$

		Bias								RMSE							
		$\hat{\theta}_1$				$\hat{\theta}_2$				$\hat{\theta}_1$				$\hat{\theta}_2$			
	$N \setminus T$	40	80	160	200	40	80	160	200	40	80	160	200	40	80	160	200
SI_Factor	40	0.0004	-0.0001	0.0000	0.0000	0.0006	-0.0001	0.0000	0.0000	0.0050	0.0029	0.0019	0.0016	0.0067	0.0039	0.0025	0.0021
	80	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0000	0.0001	0.0032	0.0019	0.0012	0.0021	0.0042	0.0025	0.0016	0.0030
	160	0.0001	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0019	0.0011	0.0008	0.0006	0.0026	0.0014	0.0010	0.0009
	200	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0001	0.0000	0.0017	0.0010	0.0032	0.0006	0.0022	0.0014	0.0052	0.0008
L_Factor	40	-0.0970	-0.0794	-0.0697	-0.0693	0.2733	0.2671	0.2623	0.2598	0.1253	0.0980	0.0813	0.0799	0.2831	0.2730	0.2666	0.2640
	80	-0.0838	-0.0704	-0.0622	-0.0578	0.2735	0.2640	0.2561	0.2533	0.1038	0.0822	0.0706	0.0652	0.2797	0.2679	0.2588	0.2555
	160	-0.0714	-0.0622	-0.0558	-0.0534	0.2650	0.2569	0.2520	0.2496	0.0891	0.0705	0.0617	0.0587	0.2695	0.2596	0.2537	0.2511
	200	-0.0702	-0.0607	-0.0540	-0.0536	0.2632	0.2558	0.2510	0.2503	0.0854	0.0689	0.0594	0.0584	0.2679	0.2584	0.2525	0.2517
F_Panel	40	-0.0508	-0.0460	-0.0487	-0.0500	0.2361	0.2375	0.2375	0.2392	0.1172	0.0979	0.0817	0.0760	0.2571	0.2502	0.2460	0.2457
	80	-0.0501	-0.0530	-0.0496	-0.0474	0.2372	0.2404	0.2373	0.2356	0.1192	0.0813	0.0702	0.0664	0.2588	0.2478	0.2420	0.2399
	160	-0.0409	-0.0481	-0.0505	-0.0468	0.2318	0.2348	0.2347	0.2355	0.1496	0.0742	0.0663	0.0648	0.2563	0.2412	0.2382	0.2382
	200	-0.0451	-0.0495	-0.0506	-0.0503	0.2315	0.2356	0.2358	0.2382	0.0905	0.0734	0.0628	0.0617	0.2445	0.2417	0.2387	0.2406

Table 3: Choice of the Number of Factors

		Percentage of replications where $m < 3$					Percentage of replications where $m = 3$					Percentage of replications where $m > 3$				
	$N \setminus T$	20	40	80	160	200	20	40	80	160	200	20	40	80	160	200
Case 1	20	0.2770	0.1250	0.0600	0.0300	0.0450	0.6530	0.8570	0.9400	0.9700	0.9550	0.0700	0.0180	0.0000	0.0000	0.0000
	40	0.1410	0.0120	0.0000	0.0000	0.0000	0.8390	0.9870	1.0000	1.0000	1.0000	0.0200	0.0010	0.0000	0.0000	0.0000
	80	0.0750	0.0020	0.0000	0.0000	0.0000	0.9230	0.9980	1.0000	1.0000	1.0000	0.0020	0.0000	0.0000	0.0000	0.0000
	160	0.0560	0.0000	0.0000	0.0000	0.0000	0.9440	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	200	0.0420	0.0000	0.0000	0.0000	0.0000	0.9580	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Case 2	20	0.2850	0.1300	0.0630	0.0350	0.0460	0.6620	0.8590	0.9370	0.9650	0.9540	0.0530	0.0110	0.0000	0.0000	0.0000
	40	0.1450	0.0150	0.0000	0.0000	0.0000	0.8390	0.9840	1.0000	0.9990	0.9970	0.0160	0.0010	0.0000	0.0010	0.0030
	80	0.0780	0.0020	0.0000	0.0000	0.0000	0.9210	0.9980	1.0000	1.0000	0.9990	0.0010	0.0000	0.0000	0.0000	0.0010
	160	0.0530	0.0000	0.0000	0.0000	0.0000	0.9470	0.9990	0.9990	0.9990	0.9990	0.0000	0.0010	0.0010	0.0010	0.0010
	200	0.0420	0.0000	0.0000	0.0010	0.0000	0.9560	0.9980	0.9980	0.9960	1.0000	0.0020	0.0020	0.0020	0.0030	0.0000

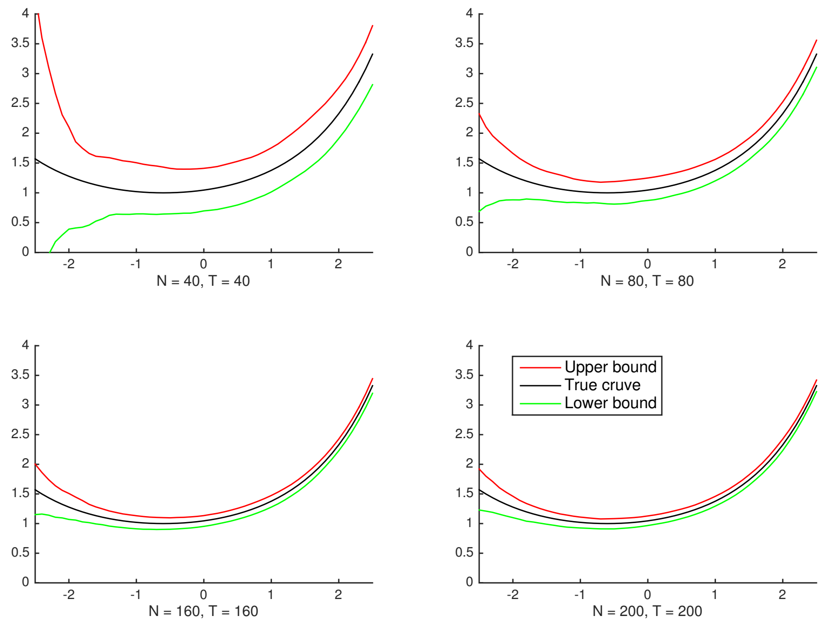

Figure 2: Estimated Link Function for Case 2 Where $g_o(w) = \exp((0.6 + w)^2/8)$

insurance companies, while technological and financial innovations include, but are not limited to, information processing and telecommunication technologies, the securitization and sale of bank loans, and the development of derivatives markets. One of the most important consequences of these changes is the increasing concentration of industry assets among large banks. According to Jones and Critchfield (2005), the asset share of large banks (those with assets in excess of \$1 billion) increased from 76 percent in 1984 to 86 percent in 2003. In the meantime, the average size of those banks increased from \$4.97 billion to \$15.50 billion (in 2002 dollars). This has raised concern that some banks might be too large to operate efficiently, stimulating a substantial body of research into returns to scale at large banks in the U.S. For excellent reviews, see Berger et al. (1999).

Compared with the conventional fully-parametric translog cost function, the model (1.1) is more suitable for modeling the production technology of U.S. large banks for the following two reasons. First, the single-index setting (i.e., $g_o(x_{it}'\theta_o)$) is more flexible than the commonly-used translog linear form, which limits the variety of shapes the cost function is permitted to take. Therefore, use of a semiparametric single-index model reduces the risk of obtaining misleading results. Second, our model (1.1) allows for common shocks, which have increased in the U.S. banking industry in the past few decades. As discussed above, the U.S. banking industry experienced massive deregulation in the 1980s and 1990s. As a result, banks in the U.S. have become increasingly exposed to common shocks. There are two reasons for this. (i) The industry has become increasingly dominated by large banks that share similar business models and offer similar ranges of products. The similarity in business model and product portfolio has in turn increased the exposure of those banks to common shocks. (ii) As the process of banking and financial integration has progressed, bank linkages have also risen, which has also increased the exposure of large banks to common shocks (cf., Houston and Stiroh (2006) and Brasili and Vulpes (2008)). As is well-known in the panel data literature, failure to account for common shocks that affect all cross-sectional units results in correlated residuals. Furthermore, if these unobserved common shocks are correlated with other regressors, the resulting parameter estimates will be biased and inconsistent.

The data used in this application are obtained from the Reports of Income and Condition (Call Reports) published by the Federal Reserve Bank of Chicago. The sample covers the period 1986-2005. We examine only continuously operating large banks with assets of at least \$1 billion (in 1986 dollars) to avoid the impact of entry and exit and to focus on the performance of a core of healthy, surviving institutions. This gives a total of 466 banks over 20 years (i.e., 80 quarters, so $N = 466$ and $T = 80$). To select the relevant variables, we follow the commonly-accepted intermediation approach (Sealey and Lindley (1977)). In this paper, three output quantities and three input prices are identified. The three outputs are consumer loans, y_1; non-consumer loans, y_2, is composed of industrial and commercial loans and real estate loans; and securities, y_3, includes all non-loan financial assets, i.e., all financial and physical assets minus the sum of consumer loans, non-consumer loans, securities, and equity. All outputs are deflated by the GDP deflator to the base year 1986. The three input prices include: the wage rate for labor, w_1; the interest rate for borrowed funds, w_2; and the price of physical capital, w_3. The wage rate equals total salaries and benefits divided by the number of full-time employees. The price of capital equals expenses on premises and equipment divided by premises and fixed assets. The price of deposits and purchased funds equals total interest expense divided by total deposits and purchased funds. Total cost is thus the sum of these three input costs. This specification of outputs and input prices is the same as or similar to most of the previous studies in this literature (e.g., Stiroh (2000) and Berger and Mester (2003)).

7.1 The Single-Index Cost Function with Interactive Fixed Effects

The variables defined above suggest the following mapping:

$$\left(\frac{w_1}{w_3},\ \frac{w_2}{w_3},\ y_1,\ y_2,\ y_3\right) \rightarrow\ \frac{CT}{w_3},$$

where CT represents total costs, and all the other variables have been defined as above. Note that we divide CT, w_1, and w_2 by w_3 to maintain linear homogeneity with respect to input prices. Then our single-index cost function with interactive fixed effects can be written as:

$$\begin{aligned} \ln\frac{CT_{it}}{w_{3,it}} &= C\left(\frac{w_{1,it}}{w_{3,it}},\ \frac{w_{2,it}}{w_{3,it}},\ y_{1,it},\ y_{2,it},\ y_{3,it}\right) + \gamma_{o,i}' f_{o,t} + \varepsilon_{it} \\ &= g_o(x_{it}'\theta_o) + \gamma_{o,i}' f_{o,t} + \varepsilon_{it}, \end{aligned} \tag{7.1}$$

where $x_{it} = \left(\ln\frac{w_{1,it}}{w_{3,it}},\ \ln\frac{w_{2,it}}{w_{3,it}},\ \ln y_{1,it},\ \ln y_{2,it},\ \ln y_{3,it},\ 1\right)'$; (2.5) is imposed on g_o to facilitate comparison of empirical results between this model and the two linear models given in (7.3) and (7.4) below; $C(\cdot)$ represents the normalized cost function; $f_{o,t}$ represents a vector of unobserved common shocks that simultaneously affect all banks in the sample; $\gamma_{o,i}$ represents the responses of large banks to the common shocks albeit with different degrees; and ε_{it} is a random error. Given the estimation of (7.1), it is possible to compute returns to scale as follows:

$$RTS = \left[\sum_{j=1}^{3}\frac{\partial}{\partial\ln y_j}C\left(\frac{w_1}{w_3},\ \frac{w_2}{w_3},\ y_1,\ y_2,\ y_3\right)\right]^{-1}, \tag{7.2}$$

where $\frac{\partial}{\partial\ln y_j}C\left(\frac{w_1}{w_3},\ \frac{w_2}{w_3},\ y_1,\ y_2,\ y_3\right)$ is the cost elasticity of output j with $j = 1,\ 2,\ 3$.

For comparison purposes, we also consider two other parametric models: (1) a fully-parametric translog cost function with fixed effects (referred to as "F_Panel"), which is commonly used in studies investigating returns to scale at U.S. banks; and (2) a fully parametric translog cost function with interactive fixed effects (referred to "L_Factor")[2]. Specifically, these two models are written as

$$\frac{CT_{it}}{w_{3,it}} = x_{it}'\beta_o + \alpha_i + \varepsilon_{it}, \tag{7.3}$$

$$\frac{CT_{it}}{w_{3,it}} = x_{it}'\beta_o + \gamma_{o,i}' f_{o,t} + \varepsilon_{it}, \tag{7.4}$$

where, in order to implement within transformation for (7.3), we exclude constant 1 from x_{it} when estimating (7.3).

[2] In the literature, one usually adds quadratic terms and interaction terms to form a translog function. However, it is easy to see that the translog function is just a special case of (7.1). To place the three models in (7.1), (7.3) and (7.4) on equal footing, we use the same set of regressors for all these models.

7.2 Empirical Results

We estimate the single-index cost function with interactive fixed effects in (7.1), the fully parametric translog cost function with fixed effects in (7.3), and the fully parametric translog cost function with interactive fixed effects in (7.4) for the U.S. large banks.[3] With regard to the number of factors for the first model (i.e., (7.1)), we choose it by using (5.1). Specifically, with $N = 466$, $T = 80$, and $k = \lfloor (NT)^{1/5} \rfloor = 8$, the number of factors is chosen as $\widehat{m}_1 = 4$. Similarly, the number of factors is chosen as $\widehat{m}_2 = 4$ for the model in (7.4). Parameter estimates and the associated standard errors for the three models are reported in Table 4.

Moreover, to test the linearity of g_o, we implement the hypothesis test outlined in Su et al. (2015). A detailed description of the procedure can be found therein. For our case, we reject the null hypothesis that g_o is linear (i.e., $\Pr(g_o(x_{it}'\theta_o) = x_{it}'\theta_o) = 1$) at the 1% significant level. This suggests that the single-index model outperforms the other two models. The nonlinearity of g_o can also be clearly seen from Figure 3, which plots the estimated cost function (the solid line) together its 95% confidence intervals (the dashed lines) on $[1.6,\ 2.3]$.

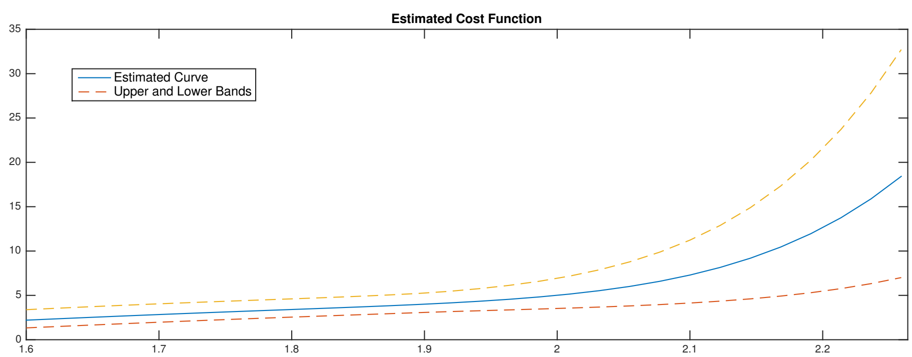

Figure 3: Estimated Cost Function

To further compare the performance of these three models, we follow Dong et al. (2015) and Dong et al. (2016) to compute mean squared error (MSE) for each of the three models using $\text{MSE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{y}_{it} - y_{it})^2$, where $\hat{y}_{it}$ is the estimated normalized total costs

[3]Following the literature of production econometrics, standard deviations and confidence intervals in this subsection are constructed using the wild bootstrap method with 300 replications. Here we use (7.1) as an example to illustrate how this bootstrap method works. We first implement (3.10) with $k = \lfloor (NT)^{1/5} \rfloor$ and select $\hat{m}$. Given $\hat{m}$, we then calculate $\hat{y}_{it} = \hat{g}_k(x_{it}'\hat{\theta}) + \hat{\gamma}_i'\hat{f}_t$ and $\hat{\varepsilon}_{it} = \hat{y}_{it} - y_{it}$ based on the estimates. For each replication we generate new dependent variables using $y_{it}^* = \hat{y}_{it} + \hat{\varepsilon}_{it}e_{it}$, where e_{it} is generated from a standard normal distribution and is independent over i and t. After that, we implement (3.10) based on $\{y_{it}^*, x_{it}, 1 \le i \le N, 1 \le t \le T\}$ using the same $\hat{m}_1$ for each replication.

for bank i in period t and y_{it} is the corresponding actual normalized total costs. The MSE values for the three models are reported in Table 4. As can be seen, the single-index model has the lowest MSE, confirming that the single-index cost function with interactive fixed effects outperforms the other two models. This is not surprising given the fact that the single-index cost function with interactive fixed effects nests the other two models as special cases.

Table 4: MSEs and Estimates of Coefficients

	SI_Factor		L_Factor		F_Panel	
MSE	0.0024		0.0047		0.0105	
Variables	Est	std	Est	std	Est	std
$\ln \frac{w_1}{w_3}$	0.0636	0.0503	0.4216	0.0154	0.3728	0.0044
$\ln \frac{w_2}{w_3}$	0.0042	0.0039	0.0190	0.0007	0.0177	0.0017
$\ln y_1$	0.0636	0.0534	0.3883	0.0024	0.3560	0.0045
$\ln y_2$	0.0156	0.0133	0.0938	0.0015	0.0997	0.0027
$\ln y_3$	0.0897	0.0721	0.4786	0.0033	0.4213	0.0049
constant	0.9805	0.0157	0.5120	0.0344		

In what follows, we focus on empirical results from the single-index cost function with interactive fixed effects. Table 5 presents the annual average returns to scale (RTS) estimates for each year, obtained by averaging over all sampled banks in that year. Two findings emerge from this table. First, the annual average RTS is greater than unity for all sample years, ranging from 1.1229 to 1.1607, indicating that on average the large banks show increasing returns to scale during the sample period. This finding is consistent with Wheelock and Wilson (2012), who, using a non-parametric local-linear estimator to estimate the cost relationship for commercial banks in the U.S. over the period 1984-2006, find that U.S. banks operated under increasing returns to scale. This finding partially explains why bank mergers and acquisitions in 1990's and early 2000's occurred at an unprecedented rate[4], because mergers and acquisitions allow banks to exploit economies of scale.

Second, the annual average RTS shows a general downward trend, falling gradually from 1.1607 in 1986 to 1.1238 in 2005. A possible reason for the decline in RTS is that U.S. large banks grew rapidly during the sample period, which allowed those banks to exploit economies of scale and at the same time lowered their returns to scale. For instance, the

[4]For example, from 1990 through 1998, the number of banks dropped from 12,347 to 8,774 resulting in a 28.9% decline. During this same period, there were 4,625 unassisted mergers with only 569 failures. Thus the major contributor to the 28.9% decline in the number of banks was attributable to the merger activity within the industry.

average size of the large banks in our sample increased from $1.22 billion in 1986 to $7.63 billion in 2005 (in 1986 dollars), which, for a given production technology, would certainly lead to lower returns to scale. However, it should be noted here that there is another factor that might have increased RTS as the banks grew bigger – new technologies. Specifically, as banks grow bigger, they are more likely to afford new technologies. The adoption of new technologies increases the banks' optimal scales over time, which results in higher RTS for given bundles of inputs. This explains why the annual average RTS does not decline rapidly during the sample period, despite the rapid growth of the large banks.

Table 5: Annual Average Returns To Scale

Year	RTS	95 % Confidence Intervals
1986	1.1607	(1.1415, 1.1781)
1987	1.1579	(1.1406, 1.1731)
1988	1.1598	(1.1423, 1.1748)
1989	1.1596	(1.1422, 1.1747)
1990	1.1589	(1.1421, 1.1735)
1991	1.1581	(1.1420, 1.1719)
1992	1.1531	(1.1385, 1.1669)
1993	1.1492	(1.1345, 1.1644)
1994	1.1466	(1.1320, 1.1628)
1995	1.1469	(1.1326, 1.1624)
1996	1.1458	(1.1315, 1.1614)
1997	1.1435	(1.1294, 1.1593)
1998	1.1406	(1.1267, 1.1567)
1999	1.1381	(1.1244, 1.1544)
2000	1.1366	(1.1231, 1.1525)
2001	1.1354	(1.1221, 1.1510)
2002	1.1296	(1.1152, 1.1455)
2003	1.1253	(1.1105, 1.1411)
2004	1.1229	(1.1085, 1.1384)
2005	1.1238	(1.1093, 1.1390)
Average	1.1446	(1.1295, 1.1601)

Having examined annual average RTS estimates, we now turn to RTS estimates at individual bank level. We calculate the percentage of banks facing increasing, constant, or decreasing returns to scale for each year. Specifically, we count the number of cases where the 95% confidence intervals are strictly greater than 1.0 (indicating increasing returns to scale or IRS for short), strictly less than 1.0 (indicating decreasing returns to scale or DRS for short), or contain 1.0 (indicating constant returns to scale or CRS for short).

The results are reported in Table 6. Two findings in this table are noteworthy. First, on average the vast majority (97.85%) of the banks face increasing returns to scale, a tiny percentage (1.33%) face decreasing returns to scale, and an even tinier percentage (0.82%) face constant returns to scale. Second, the percentage of banks facing increasing returns to scale shows a general downward trend, while the percentage of banks facing constant (decreasing) returns to scale shows a general upward trend. Specifically, the percentage of banks facing increasing returns to scale decreases markedly from 99.35% in 1986 to 95.49% in 2005; the percentage of banks facing constant returns to scale increases noticeably from 0.43% in 1986 to 2.15% in 2005; and the percentage of banks facing decreasing returns to scale steadily from 0.22% in 1986 to 2.36% in 2005. The results presented here are consistent with our previous discussion, that is, due to their rapid growth, more and more banks have reached or passed their optimal scales, leaving more and more banks operating under constant or decreasing returns to scale.

Table 6: Returns To Scale at Individual Bank Level

Year	IRS	CRS	DRS
1986	99.35%	0.43%	0.22%
1987	99.14%	0.43%	0.43%
1988	99.14%	0.43%	0.43%
1989	99.14%	0.64%	0.22%
1990	98.93%	0.86%	0.21%
1991	99.36%	0.42%	0.22%
1992	99.14%	0.43%	0.43%
1993	98.71%	0.64%	0.65%
1994	98.71%	0.64%	0.65%
1995	98.93%	0.43%	0.64%
1996	98.50%	0.64%	0.86%
1997	98.07%	1.29%	0.64%
1998	97.64%	1.29%	1.07%
1999	97.42%	1.72%	0.86%
2000	96.78%	2.58%	0.64%
2001	96.57%	2.58%	0.85%
2002	96.35%	2.58%	1.07%
2003	95.06%	3.22%	1.72%
2004	94.64%	3.22%	2.14%
2005	95.49%	2.15%	2.36%
Average	97.85%	1.33%	0.82%

8 Conclusion

This paper proposes a new model that extends Bai (2009) to a semi-parametric single-index setting. The new model is flexible due to the use of a link function; at the same time, it allows for common shocks or time-varying heterogeneity because of the presence of the interactive fixed effects. We derive asymptotic theories for both the case where the link function is integrable and the case where the link function is not. We also investigate the finite sample properties of our single-index model via Monte Carlo experiments in large N and T cases.

Finally, we show how our model and methodology can be used by analyzing the returns to scale of large commercial banks in the U.S. over the period 1986-2005. Specifically, we estimate a single-index cost function with interactive fixed effects. We then compare this cost function with two other parametric cost functions: (1) a fully-parametric translog cost function with fixed effects, and (2) a fully-parametric translog cost function with interactive fixed effects. Our results show that the former cost function outperforms the latter two. Our empirical results from the single-index cost function with interactive fixed effects show that the vast majority of U.S. large banks face increasing returns to scale, and that the percentage of U.S. large banks that face increasing returns to scale has declined over time.

Appendix A: Proofs of the Main Results

Note that proofs of Lemma 3.1 and Theorems 3.1, 3.2 and 3.3 are provided in the main text of the paper, while those of Lemma 4.1 and Theorem 4.1 are provided in the supplementary file of the paper. The proofs of all preliminary results (e.g., Lemmas A1-A3) are also provided in the supplementary file of this paper. In what follows, when no misunderstanding arises, we may suppress indexes i and t for notational simplicity. In addition, ρ_1, $O(1)$ and A always denote constants and may be different at each appearance in the following development.

Lemma A1. *Assume that $g(w)$ is differentiable on $\mathbb{R}$ and $w^{r-j}g^{(j)}(w) \in L^2(\mathbb{R})$ for $j = 0, 1, \ldots, r$ and $r \geq 1$. For the expansion*

$$g(w) = \sum_{n=0}^{\infty} c_n \mathbb{H}_n(w) = g_k(w) + \delta_k(w),$$

where

$$\mathbb{H}_n(w) = \frac{1}{\sqrt[4]{\pi}\sqrt{2^n n!}} H_n(w) \exp(-w^2/2) \quad \text{with} \quad n = 0, 1, 2, \ldots,$$

$$g_k(w) = \sum_{n=0}^{k-1} c_n \mathbb{H}_n(w) = C'\mathcal{H}(w), \quad c_n = \int g(w)\mathbb{H}_n(w)dw, \quad \delta_k(w) = \sum_{n=k}^{\infty} c_n \mathbb{H}_n(w),$$
$$\mathcal{H}(w) = (\mathbb{H}_0(w), \ldots, \mathbb{H}_{k-1}(w))', \quad C = (c_0, \ldots, c_{k-1})',$$

the following results hold:
(1) $\int w^2 \mathbb{H}_n^2(w)dw = n + 1/2$; (2) $\max_w |\delta_k(w)| = O(1)k^{-\frac{r}{2}+\frac{5}{12}}$; (3) $\int \delta_k^2(w)dw = O(1)k^{-r}$;
(4) $\int \|\mathcal{H}(w)\|dw = O(1)k^{11/12}$; (5) $\int \|\mathcal{H}(w)\|^2 dw = k$; (6) $\|\mathcal{H}(w)\|^2 = O(1)k$ uniformly on $\mathbb{R}$;
(7) $\int |\delta_k(w)|dw = O(1)k^{-\frac{r}{2}+\frac{11}{12}}$; (8) $\int |\mathbb{H}_n(w)|dw = O(1)n^{5/12}$; (9) $\int w^2\|\mathcal{H}(w)\|^2 dx = O(1)k^2$;
(10) $\max_w |\delta_k^{(1)}(w)| = O(1)k^{-\frac{r}{2}+\frac{11}{12}}$.

Lemma A1 is a part of Lemma C.1 of the supplementary file of Dong et al. (2016), wherein a detailed proof can be found. ■

Lemma A2. *Let $\phi_i[\theta, g] = (g(x_{i1}'\theta), \ldots, g(x_{iT}'\theta))'$ with $i = 1, \ldots, N$. We consider the following limits on 3-fold Cartesian product space formed by $\Theta \times \Im_1 \times D_F$. Under Assumptions 1 and 2, as $(N,T) \to (\infty, \infty)$,*

1. $\sup_{(\theta,g)} |L_{NT}(\theta, g) - L(\theta, g)| = o_P(1)$, where $L(\theta, g)$ is defined in Assumption 2 and

$$L_{NT}(\theta, g) = \frac{1}{NT}\sum_{i=1}^{N} (\phi_i[\theta, g] - \phi_i[\theta_o, g_o])' M_{F_o} (\phi_i[\theta, g] - \phi_i[\theta_o, g_o]);$$

2. $\sup_F \left\| \frac{1}{NT} \sum_{i=1}^N \varepsilon_i' P_F \varepsilon_i \right\| = o_P(1)$;
3. $\sup_F \left\| \frac{1}{NT} \sum_{i=1}^N \gamma_{o,i}' F_o' M_F \varepsilon_i \right\| = o_P(1)$;
4. $\sup_{(\theta,g,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \phi_i[\theta, g]' M_F \varepsilon_i \right\| = o_P(1)$;
5. $\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \phi_i[\theta, \delta_{o,k}]' M_F \varepsilon_i \right\| = o_P(1)$;
6. $\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \phi_i[\theta, \delta_{o,k}]' M_F \phi_i[\theta, \delta_{o,k}] \right\| = o_P(1)$;
7. $\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \phi_i[\theta, \delta_{o,k}]' M_F F_o \gamma_{o,i} \right\| = o_P(1)$;
8. $\sup_{(\theta_1,\theta_2,g,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \phi_i[\theta_1, \delta_{o,k}]' M_F \phi_i[\theta_2, g] \right\| = o_P(1)$;
9. $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it}^2 = \sigma_\varepsilon^2 + O_P\left(\frac{1}{\sqrt{NT}}\right)$;
10. $\sup_{(\theta,g)} \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left(g^{(1)}(x_{it}'\theta)\right)^2 x_{it}x_{it}' - E\left[\left(g^{(1)}(x_{11}'\theta)\right)^2 x_{11}x_{11}'\right] \right\| = o_P(1)$;

11. $\sup_{(\theta,g)} \left\| \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it} g^{(2)}(x_{it}'\theta)x_{it}x_{it}' \right\| = o_P(1)$;

12. $\sup_F \left| \frac{1}{NT}\sum_{i=1}^N (\phi_i[\theta_o, g_o] - \phi_i[\theta, g])' M_F F_o \gamma_{o,i} \right| = O_P(\|(\theta_o, g_o) - (\theta, g)\|_w)$ for $\forall(\theta, g)$ in a sufficient small neighbourhood of (θ_o, g_o).

    In addition, assume that Assumption 4 holds for the following results. Then

13. $\sup_{(\theta,F)} \left\| \frac{1}{NT}\sum_{i=1}^N \mathcal{H}_i(\theta)' M_F \varepsilon_i \right\| = O_P(\sqrt{k}\eta_{NT})$, where $\mathcal{H}_i(\theta) = (\mathcal{H}(x_{i1}'\theta), \ldots, \mathcal{H}(x_{iT}'\theta))'$ for $i = 1, \ldots, N$;

14. $\sup_\theta \left\| \frac{1}{NT}\sum_{i=1}^N \mathcal{H}_i(\theta)' M_{F_o} \mathcal{H}_i(\theta) - \Upsilon(\theta) \right\| = o_P(1)$, where $\mathcal{H}_i(\theta) = (\mathcal{H}(x_{i1}'\theta), \ldots, \mathcal{H}(x_{iT}'\theta))'$ for $i = 1, \ldots, N$.

The proof of Lemma A2 is provided in the supplementary file of this paper.

Lemma A3. *Let $\phi_i[\theta, g] = (g(x_{i1}'\theta), \ldots, g(x_{iT}'\theta))'$ and $\psi_{1i}[\theta, g] = (g(x_{i1}'\theta)x_{i1}, \ldots, g(x_{iT}'\theta)x_{iT})'$ with $i = 1, \ldots, N$. We consider the following limits on 3-fold Cartesian product space formed by $\Theta \times \Im_1 \times D_F$. Under Assumptions 1-3,*

1. $\left\| \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \left(g^{(1)}(x_{it}'\theta_o) - g_o^{(1)}(x_{it}'\theta_o) \right) x_{it}\varepsilon_{it} \right\| = o_P\left(\frac{1}{\sqrt{NT}}\right)$, as $\|g^{(1)} - g_o^{(1)}\|_{L^2} = o(1)$;

2. $\frac{1}{NT}\sum_{i=1}^N \left(\psi_{1i}[\theta_o, g^{(1)}] - \psi_{1i}[\theta_o, g_o^{(1)}] \right)' P_{F_o}\varepsilon_i = o_P\left(\frac{1}{\sqrt{NT}}\right)$, as $\|g^{(1)} - g_o^{(1)}\|_{L^2} = o(1)$;

3. $\left\| \frac{1}{NT}\sum_{i=1}^N \psi_{1i}[\theta_o, g_o^{(1)}]' P_T \varepsilon_i \right\| = o_P\left(\frac{1}{\sqrt{NT}}\right)$, as $\|P_T\| = o(1)$.

The proof of Lemma A3 is provided in the supplementary file of this paper.

Proof of Theorem 3.1:

(1). Expanding $S_{NT}(\theta, C, F) - S_{NT}(\theta_o, C_o, F_o)$, we have

$$
\begin{aligned}
& S_{NT}(\theta, C, F) - S_{NT}(\theta_o, C_o, F_o) \\
= & \frac{1}{NT}\sum_{i=1}^N \left\{ (Y_i - \phi_i[\theta, g_k])' M_F (Y_i - \phi_i[\theta, g_k]) - (Y_i - \phi_i[\theta_o, g_{o,k}])' M_{F_o} (Y_i - \phi_i[\theta_o, g_{o,k}]) \right\} \\
= & \frac{1}{NT}\sum_{i=1}^N \left\{ (Y_i - \phi_i[\theta, g_k])' M_F (Y_i - \phi_i[\theta, g_k]) - (\phi_i[\theta_o, \delta_{o,k}] + \varepsilon_i)' M_{F_o} (\phi_i[\theta_o, \delta_{o,k}] + \varepsilon_i) \right\} \\
= & \mathbb{S}_{NT}(\theta, C, F) \\
& + \frac{1}{NT}\sum_{i=1}^N \sum_{s=1}^T \varepsilon_i' (M_F - M_{F_o}) \varepsilon_i + \frac{2}{NT}\sum_{i=1}^N \gamma_{o,i}' F_o' M_F \varepsilon_i + \frac{2}{NT}\sum_{i=1}^N (\phi_i[\theta_o, g_{o,k}] - \phi_i[\theta, g_k])' M_F \varepsilon_i
\end{aligned}
$$

$$+\frac{2}{NT}\sum_{i=1}^{N}\phi_i[\theta_o,\delta_{o,k}]'\left(M_F-M_{F_o}\right)\varepsilon_i+\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\phi_i[\theta_o,\delta_{o,k}]'\left(M_F-M_{F_o}\right)\phi_i[\theta_o,\delta_{o,k}]$$
$$+\frac{2}{NT}\sum_{i=1}^{N}(\phi_i[\theta_o,g_{o,k}]-\phi_i[\theta,g_k])'M_F\phi_i[\theta_o,\delta_{o,k}]+\frac{2}{NT}\sum_{i=1}^{N}\gamma_{o,i}'F_o'M_F\phi_i[\theta_o,\delta_{o,k}],$$

where g_k is defined in (3.2), $g_{o,k}$ and $\delta_{o,k}$ are defined in (3.3), and

$$\mathbb{S}_{NT}(\theta,C,F)=\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\theta_o,g_{o,k}]-\phi_i[\theta,g_k]+F_o\gamma_i\right)'M_F\left(\phi_i[\theta_o,g_{o,k}]-\phi_i[\theta,g_k]+F_o\gamma_i\right).$$

By (2)-(9) of Lemma A2, we immediately obtain

$$S_{NT}(\theta,C,F)-S_{NT}(\theta_o,C_o,F_o)=\mathbb{S}_{NT}(\theta,C,F)+o_P(1). \tag{9.1}$$

Let $b_i=\phi_i[\theta,g_k]-\phi_i[\theta_o,g_{o,k}]$ and $b=(b_1',\ldots,b_N')'$. As in the proof of Lemma A2, we can readily argue that $(NT)^{-1}b'b=O(1)$ uniformly in (θ,C). Let $\eta=\operatorname{vec}(M_FF_o)$, $A_1=I_N\otimes M_F$, $A_2=(\Gamma_o'\Gamma_o)\otimes I_T$, and $A_3=(\gamma_{o,1}\otimes I_T,\ldots,\gamma_{o,N}\otimes I_T)$. Then

$$\begin{aligned}\mathbb{S}_{NT}(\theta,C,F)&=\frac{1}{NT}\sum_{i=1}^{N}b_i'M_Fb_i+\frac{1}{NT}\sum_{i=1}^{N}\gamma_{o,i}'F_o'M_FF_o\gamma_{o,i}-\frac{2}{NT}\sum_{i=1}^{N}b_i'M_FF_o\gamma_{o,i}\\&=\frac{1}{NT}\sum_{i=1}^{N}b_i'M_Fb_i+\frac{1}{NT}\operatorname{tr}\left(M_FF_o\Gamma_o'\Gamma_oF_o'M_F\right)-\frac{2}{NT}\sum_{i=1}^{N}\operatorname{tr}\left(M_FF_o\gamma_{o,i}b_i'\right)\\&=\frac{1}{NT}b'A_1b+\frac{1}{NT}\eta'A_2\eta-\frac{2}{NT}\eta'\sum_{i=1}^{N}\left(\gamma_{o,i}\otimes I_T\right)b_i\\&=\frac{1}{NT}b'A_1b+\frac{1}{NT}\eta'A_2\eta-\frac{2}{NT}b'A_3'\eta\\&=\frac{1}{NT}b'A_1b+\frac{1}{NT}\vartheta'A_2\vartheta-\frac{1}{NT}b'A_3'A_2^{-1}A_3b,\end{aligned} \tag{9.2}$$

where $\vartheta=\eta-A_2^{-1}A_3b$, and the third equality follows from the fact that

$$\operatorname{tr}(B_1B_2B_3)=\operatorname{vec}(B_1)'\left(B_2\otimes I\right)\operatorname{vec}\left(B_3'\right)\ \text{ and }\operatorname{tr}(B_1B_2B_3B_4)=\operatorname{vec}(B_1)'\left(B_2\otimes B_4'\right)\operatorname{vec}\left(B_3'\right)$$

for any conformable matrices B_1,B_2, B_3, B_4 and an identity matrix I (see, e.g., Bernstein (2009, p. 253)). We now show that the last term in (9.2) is $o_P(1)$ uniformly in b. Note that $(NT)^{-1}b'A_3'A_2^{-1}A_3b\le\lambda_{\max}\left(A_3'A_2^{-1}A_3\right)(NT)^{-1}b'b=o_P(1)$ for any $(NT)^{-1}b'b=O(1)$ provided $\lambda_{\max}\left(A_3'A_2^{-1}A_3\right)=o_P(1)$. Also, note that $\lambda_{\max}\left(A_3'A_2^{-1}A_3\right)\le\left[\lambda_{\min}\left(A_2/N\right)\right]^{-1}\lambda_{\max}\left(N^{-1}A_3'A_3\right)=\underline{c}_\Gamma^{-1}\lambda_{\max}\left(N^{-1}A_3'A_3\right)$ where $\underline{c}_\Gamma^{-1}\equiv\left[\lambda_{\min}\left(\Gamma_o'\Gamma_o/N\right)\right]^{-1}=O_P(1)$. Define the following upper block-triangular matrix

$$C_1=\left(\begin{array}{cccc}\gamma_{o,1}'\gamma_{o,1}I_T & \gamma_{o,1}'\gamma_{o,2}I_T & \cdots & \gamma_{o,1}'\gamma_{o,N}I_T\\ 0 & \gamma_{o,2}'\gamma_{o,2}I_T & \cdots & \gamma_{o,2}'\gamma_{o,N}I_T\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \gamma_{o,N}'\gamma_{o,N}I_T\end{array}\right).$$

Noting that the $NT\times NT$ matrix $A_3'A_3$ has a typical $T\times T$ block submatrix $T^{-1}\gamma_{o,i}'\gamma_{o,j}I_T$, we have $A_3'A_3=C_1+C_1'-C_d$ where C_d =diag$(\gamma_{o,1}'\gamma_{o,1}I_T,\ldots,\gamma_{o,N}'\gamma_{o,N}I_T)$. By the fact that the eigenvalues of a block upper/lower triangular matrix are the combined eigenvalues of its diagonal block matrices, Weyl's inequality, and Assumption 1.4, we have

$$\begin{aligned}\lambda_{\max}\left(N^{-1}A_3'A_3\right)&\leq N^{-1}\left\{2\lambda_{\max}\left(C_1\right)-\lambda_{\min}(C_d)\right\}\leq 2N^{-1}\max_{1\leq i\leq N}\|\gamma_{o,i}\|^2\\&=N^{-1}o_P(N^{1/2})=o_P(1),\end{aligned}\tag{9.3}$$

where the first equality follows from the fact that $\max_{1\leq i\leq N}\|\gamma_{o,i}\|^2=o_P(N^{1/2})$ under Assumption 1.4 by the Markov inequality. It follows that $\lambda_{\max}\left(A_3'A_2^{-1}A_3\right)=o_P\left(1\right)$ and $\frac{1}{NT}b'A_3'A_2^{-1}A_3b=o_P\left(1\right)$ uniformly in b. This, in conjunction with (9.1)-(9.2) and the fact that $S_{NT}(\hat{\theta},\hat{C},\hat{F})-S_{NT}(\theta_o,C_o,F_o)\leq 0$, implies that

$$\left(NT\right)^{-1}\hat{b}'\hat{A}_1\hat{b}=\left(NT\right)^{-1}\sum_{i=1}^N\hat{b}_i'M_{\hat{F}}\hat{b}_i=o_P\left(1\right),\tag{9.4}$$

where $\hat{A}_1=I_N\otimes M_{\hat{F}}$, $\hat{b}=(\hat{b}_1',\ldots,\hat{b}_N')'$, and $\hat{b}_i=\phi_i[\hat{\theta},\hat{g}_k]-\phi_i[\theta_o,g_{o,k}]$.

By (9.1), (9.2), (9.4), and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}0&\geq S_{NT}(\hat{\theta},\hat{C},\hat{F})-S_{NT}(\theta_o,C_o,F_o)\\&=\frac{1}{NT}\hat{b}'\hat{A}_1\hat{b}+\frac{1}{NT}\mathrm{tr}\left[\left(F_o'M_{\hat{F}}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right]-\frac{2}{NT}\sum_{i=1}^N\hat{b}_i'M_{\hat{F}}F_o\gamma_{o,i}+o_P\left(1\right)\\&\geq\frac{1}{N}\hat{b}'\hat{A}_1\hat{b}+\frac{1}{NT}\mathrm{tr}\left[\left(F_o'M_{\hat{F}}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right]\\&\quad-2\left\{\frac{1}{N}\hat{b}'\hat{A}_1\hat{b}\right\}^{1/2}\left\{\frac{1}{NT}\mathrm{tr}\left[\left(F_o'M_{\hat{F}}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right]\right\}^{1/2}+o_P\left(1\right)\\&=o_P\left(1\right)+\frac{1}{NT}\mathrm{tr}\left[\left(F_o'M_{\hat{F}}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right]-2o_P\left(1\right)\left\{\frac{1}{NT}\mathrm{tr}\left[\left(F_o'M_{\hat{F}}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right]\right\}^{1/2}.\end{aligned}$$

It follows that $\frac{1}{NT}\mathrm{tr}\left[\left(F_o'M_{\hat{F}}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right]=o_P\left(1\right)$. As in Bai (2009, p.1265), this further implies that $\frac{1}{T}\mathrm{tr}\left(F_o'M_{\hat{F}}F_o\right)=o_P\left(1\right)$, $\frac{1}{T}\hat{F}'F_o$ is invertible, and $\left\|P_{\hat{F}}-P_{F_o}\right\|=o_P\left(1\right)$.

(2). By (9.4) and noting that

$$\left|(NT)^{-1}\sum_{i=1}^{N}\hat{b}_i'\left(M_{F_o}-M_{\hat{F}}\right)\hat{b}_i\right|\leq\left\|P_{\hat{F}}-P_{F_o}\right\|(NT)^{-1}\sum_{i=1}^{N}\hat{b}_i'\hat{b}_i=o_P(1),$$

we have

$$o_P(1)=(NT)^{-1}\sum_{i=1}^{N}\hat{b}_i'M_{\hat{F}}\hat{b}_i=(NT)^{-1}\sum_{i=1}^{N}\hat{b}_i'M_{F_o}\hat{b}_i-o_P(1).$$

Then $(NT)^{-1}\sum_{i=1}^{N}\hat{b}_i'M_{F_o}\hat{b}_i=o_P(1)$. Noting $\hat{b}_i=\phi_i[\hat{\theta},\hat{g}_k]-\phi_i[\theta_o,g_{o,k}]=\left(\phi_i[\hat{\theta},\hat{g}_k]-\phi_i[\theta_o,g_o]\right)+\phi_i[\theta_o,\delta_{o,k}]$, we have

$$\begin{aligned}
o_P(1)&=(NT)^{-1}\sum_{i=1}^{N}\hat{b}_i'M_{F_o}\hat{b}_i\\
&=(NT)^{-1}\sum_{i=1}^{N}\left(\phi_i[\hat{\theta},\hat{g}_k]-\phi_i[\theta_o,g_o]\right)'M_{F_o}\left(\phi_i[\hat{\theta},\hat{g}_k]-\phi_i[\theta_o,g_o]\right)\\
&\quad+(NT)^{-1}\sum_{i=1}^{N}\left(\phi_i[\hat{\theta},\hat{g}_k]-\phi_i[\theta_o,g_o]\right)'M_{F_o}\phi_i[\theta_o,\delta_{o,k}]\\
&\quad+(NT)^{-1}\sum_{i=1}^{N}\left(\phi_i[\theta_o,\delta_{o,k}]\right)'M_{F_o}\left(\phi_i[\theta_o,\delta_{o,k}]\right)\\
&=(NT)^{-1}\sum_{i=1}^{N}\left(\phi_i[\hat{\theta},\hat{g}_k]-\phi_i[\theta_o,g_o]\right)'M_{F_o}\left(\phi_i[\hat{\theta},\hat{g}_k]-\phi_i[\theta_o,g_o]\right)+o_P(1)\\
&=L(\hat{\theta},\hat{g}_k)+o_P(1)
\end{aligned}$$

where the third equality follows from (6) and (8) of Lemma A2; and the fourth equality follows from (1) of Lemma A2. Consequently, $L(\hat{\theta},\hat{g}_k)=o_P(1)$.

Note that we have shown $L(\theta,g)$ is continuous with respect to $\|\cdot\|_w$ on $\Theta\times\Im_1$ in the proof of (1) of Lemma A2. This implies that, for $\forall(\theta,C)$ satisfying $\|(\theta,C)-(\theta_o,C_o)\|\geq\epsilon>0$, we have $L(\theta,g_k)$ does not converge to zero for $g_k(\cdot)=C'\mathcal{H}(\cdot)$. In other words, if $\|(\hat{\theta},\hat{g}_k)-(\theta_o,g_o)\|_w\not\to_P 0$, we have $L(\hat{\theta},\hat{g}_k)\neq o_P(1)$. See Bai (2009, p. 1265) and Newey and Powell (2003, p. 1576) for a similar argument. Therefore, we have proved $\|(\hat{\theta},\hat{g}_k)-(\theta_o,g_o)\|_w\to_P 0$.

The proof is now complete. ■

Proof of Lemma 3.1:

(1). We now consider V_{NT} and write

$$\hat{F}V_{NT}=\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i-\phi_i[\hat{\theta},\hat{g}_k]\right)\left(Y_i-\phi_i[\hat{\theta},\hat{g}_k]\right)'\right]\hat{F}$$

$$
\begin{aligned}
&= \left[\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\theta_o, g_o] + F_o\gamma_{o,i} + \varepsilon_i - \phi_i[\hat{\theta}, \hat{g}_k]\right)\left(\phi_i[\theta_o, g_o] + F_o\gamma_{o,i} + \varepsilon_i - \phi_i[\hat{\theta}, \hat{g}_k]\right)'\right]\hat{F} \\
&= \left[\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta}, \hat{g}_k]\right)\left(\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta}, \hat{g}_k]\right)'\right]\hat{F} \\
&\quad + \left[\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta}, \hat{g}_k]\right)(F_o\gamma_{o,i})'\right]\hat{F} + \left[\frac{1}{NT}\sum_{i=1}^{N}(F_o\gamma_i)\left(\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta}, \hat{g}_k]\right)'\right]\hat{F} \\
&\quad + \left[\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta}, \hat{g}_k]\right)\varepsilon_i'\right]\hat{F} + \left[\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\left(\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta}, \hat{g}_k]\right)'\right]\hat{F} \\
&\quad + \left[\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'\right]\hat{F} + \left[\frac{1}{NT}\sum_{i=1}^{N}F_o\gamma_{o,i}\varepsilon_i'\right]\hat{F} + \left[\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\gamma_{o,i}'F_o'\right]\hat{F} \\
&\quad + \left[\frac{1}{NT}\sum_{i=1}^{N}F_o\gamma_{o,i}\gamma_{o,i}'F_o'\right]\hat{F} \\
&:= I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F}) + \cdots + I_{5NT}(\hat{\theta}, \hat{g}_k, \hat{F}) + I_{6NT}(\hat{F}) + \cdots + I_{9NT}(\hat{F}),
\end{aligned}
$$

where the definitions of $I_{1NT}(\theta, g, F)$ to $I_{5NT}(\theta, g, F)$ and $I_{6NT}(F)$ to $I_{9NT}(F)$ are obvious.

Note that $I_{9NT}(\hat{F}) = F_o(\Gamma_o'\Gamma_o/N)(F_o'\hat{F}/T)$. Thus, we can write

$$
\begin{aligned}
&\hat{F}V_{NT} - F_o(\Gamma_o'\Gamma_o/N)(F_o'\hat{F}/T) \\
&= I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F}) + \cdots + I_{5NT}(\hat{\theta}, \hat{g}_k, \hat{F}) + I_{6NT}(\hat{F}) + \cdots + I_{8NT}(\hat{F}).
\end{aligned} \tag{9.5}
$$

Right multiplying each side of (9.5) by $(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}$, we obtain

$$
\begin{aligned}
&\hat{F}V_{NT}(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1} - F_o \\
&= \left[I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F}) + \cdots + I_{8NT}(\hat{F})\right](F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}.
\end{aligned} \tag{9.6}
$$

If V_{NT} is non-singular, then $V_{NT}(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}$ is equal to Q^{-1}. We now examine each term on the right hand side of (9.6) and show that V_{NT} is non-singular. Write

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\left\|\hat{F}V_{NT}(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1} - F_o\right\| \\
&\leq \frac{1}{\sqrt{T}}\left[\|I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F})\| + \cdots + \|I_{8NT}(\hat{F})\|\right]\cdot\left\|(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}\right\|.
\end{aligned}
$$

It is easy to show that $(F_o'\hat{F}/T)^{-1} = O_P(1)$ and $(\Gamma_o'\Gamma_o/N)^{-1} = O_P(1)$. Then we just need to focus on $\frac{1}{\sqrt{T}}\|I_{jNT}(\hat{\theta}, \hat{g}_k, \hat{F})\|$ with $j = 1, 2, \ldots, 5$ and $\frac{1}{\sqrt{T}}\|I_{jNT}(\hat{F})\|$ with $j = 1, 2, \ldots, 8$.

For $I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F})$, we have

$$
\frac{1}{\sqrt{T}}\left\|I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F})\right\| \leq \frac{\sqrt{m}}{NT}\sum_{i=1}^{N}\|\phi_i[\theta_o, g_o] - \phi_i[\theta, g]\|^2\Big|_{(\theta,g)=(\hat{\theta},\hat{g}_k)}
$$

$$
\begin{aligned}
&= \sqrt{m}E|g_o(x'\theta_o) - g(x'\theta)|^2|_{(\theta,g)=(\hat{\theta},\hat{g}_k)} \cdot (1 + o_P(1)) \\
&= O_P(\|(\hat{\theta},\hat{g}_k) - (\theta_o, g_o)\|_w^2),
\end{aligned}
$$

where the first equality follows from Lemma A2 of Newey and Powell (2003) by applying the same procedure as used for (1) of Lemma A2; and the second equality follows from *Step 3* of (1) of Lemma A2.

For $I_{2NT}(\hat{\theta},\hat{g}_k,\hat{F})$, write

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\left\|I_{2NT}(\hat{\theta},\hat{g}_k,\hat{F})\right\| \leq \frac{\sqrt{m}}{NT}\sum_{i=1}^{N}\left\|\left(\phi_i[\theta_o,g_o] - \phi_i[\theta,g]\right)(F_o\gamma_{o,i})'\right\|\Big|_{(\theta,g)=(\hat{\theta},\hat{g}_k)} \\
&\leq \sqrt{m}\Big\{\frac{1}{NT}\sum_{i=1}^{N}\|\phi_i[\theta_o,g_o] - \phi_i[\theta,g]\|^2\Big\}^{1/2}\Big|_{(\theta,g)=(\hat{\theta},\hat{g}_k)}\Big\{\frac{1}{NT}\sum_{i=1}^{N}\|F_o\gamma_{o,i}\|^2\Big\}^{1/2} \\
&= O(1)\left\{E|g_o(x'\theta_o) - g(x'\theta)|^2\right\}^{1/2}\big|_{(\theta,g)=(\hat{\theta},\hat{g}_k)} \cdot (1 + o_P(1)) \\
&= O_P(\|(\hat{\theta},\hat{g}_k) - (\theta_o, g_o)\|_w),
\end{aligned}
$$

where the second inequality follows from Cauchy-Schwarz inequality; the first equality follows from Lemma A2 of Newey and Powell (2003) by applying the same procedure as used for (1) of Lemma A2 and the fact that $\frac{1}{NT}\sum_{i=1}^{N}\|F_o\gamma_{o,i}\|^2 = O_P(1)$; and the second equality follows from *Step 3* of (1) of Lemma A2. Similarly, we have $\frac{1}{\sqrt{T}}\left\|I_{3NT}(\hat{\theta},\hat{g}_k,\hat{F})\right\| = O_P(\|(\hat{\theta},\hat{g}_k) - (\theta_o, g_o)\|_w)$.

For $I_{4NT}(\hat{\theta},\hat{g}_k,\hat{F})$, write

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\left\|I_{4NT}(\hat{\theta},\hat{g}_k,\hat{F})\right\| \leq \frac{\sqrt{m}}{NT}\sum_{i=1}^{N}\left\|\left(\phi_i[\theta_o,g_o] - \phi_i[\theta,g]\right)\varepsilon_i'\right\|\Big|_{(\theta,g)=(\hat{\theta},\hat{g}_k)} \\
&\leq \sqrt{m}\left\{\frac{1}{NT}\sum_{i=1}^{N}\|\phi_i[\theta_o,g_o] - \phi_i[\theta,g]\|^2\right\}^{1/2}\Bigg|_{(\theta,g)=(\hat{\theta},\hat{g}_k)}\left\{\frac{1}{NT}\sum_{i=1}^{N}\|\varepsilon_i\|^2\right\}^{1/2} \\
&= O(1)\left\{E|g_o(x'\theta_o) - g(x'\theta)|^2\right\}^{1/2}\big|_{(\theta,g)=(\hat{\theta},\hat{g}_k)} \cdot (1 + o_P(1)) \\
&= O_P(\|(\hat{\theta},\hat{g}_k) - (\theta_o, g_o)\|_w),
\end{aligned}
$$

where the second inequality follows from Cauchy-Schwarz inequality; the first equality follows from Lemma A2 of Newey and Powell (2003) by applying the same procedure as used for (1) of Lemma A2 and the fact that $\frac{1}{NT}\sum_{i=1}^{N}\|\varepsilon_i\|^2 = O_P(1)$; and the second equality follows from *Step 3* of (1) of Lemma A2. Similarly, we have $\frac{1}{\sqrt{T}}\left\|I_{5NT}(\hat{\theta},\hat{g}_k,\hat{F})\right\| = O_P(\|(\hat{\theta},\hat{g}_k) - (\theta_o, g_o)\|_w)$.

For $I_{6NT}(\hat{F})$, write

$$
E\left\|\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'\right\|^2 = \sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{jt}\varepsilon_{js}]
$$

$$
\begin{aligned}
&= \sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{N^2T^2}\left(\sum_{i=1}^{N}E[\varepsilon_{it}^2\varepsilon_{is}^2]+\sum_{i\neq j}E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{jt}\varepsilon_{js}]\right)\\
&= \frac{1}{N^2T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}E[\varepsilon_{it}^2\varepsilon_{is}^2]+\sum_{t=1}^{T}\sum_{i\neq j}E[\varepsilon_{it}^2\varepsilon_{jt}^2]+\sum_{t\neq s}\sum_{i\neq j}E[\varepsilon_{it}\varepsilon_{is}\varepsilon_{jt}\varepsilon_{js}]\right)\\
&= O\left(\frac{1}{N}\right)+O\left(\frac{1}{T}\right)+O\left(\frac{1}{NT}\right)=O\left(\eta_{NT}^2\right),
\end{aligned}
$$

where the fourth equality follows from Assumption 1.3. We immediately obtain $\frac{1}{\sqrt{T}}\|I_{6NT}(\hat{F})\| = O_P\left(\eta_{NT}\right)$.

For $I_{7NT}(\hat{F})$ and $I_{8NT}(\hat{F})$, write

$$
\begin{aligned}
E\left\|\frac{1}{NT}\sum_{i=1}^{N}F_o\gamma_{o,i}\varepsilon_i'\right\|^2 &= \sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E[f_{o,t}'\gamma_{o,i}\varepsilon_{is}f_{o,t}'\gamma_{o,j}\varepsilon_{js}]\\
&= \sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E[f_{o,t}'\gamma_{o,i}f_{o,t}'\gamma_{o,j}]E[\varepsilon_{is}\varepsilon_{js}]\\
&\leq \sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{E\|f_{o,t}\|^4E\|\gamma_{o,i}\|^4E\|f_{o,t}\|^4E\|\gamma_{o,j}\|^4\right\}^{1/4}|E[\varepsilon_{is}\varepsilon_{js}]|\\
&\leq O(1)\frac{1}{N^2T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E[\varepsilon_{is}\varepsilon_{js}]| = O\left(\frac{1}{N}\right),
\end{aligned}
$$

where the last equality follows from Assumption 1.3. We then immediately obtain $\frac{1}{\sqrt{T}}\|I_{7NT}(\hat{F})\| = \frac{1}{\sqrt{T}}\|I_{8NT}(\hat{F})\| = O_P\left(\frac{1}{\sqrt{N}}\right)$.

Based on the above analysis and by left multiplying (9.5) by $\hat{F}'/T$, we obtain

$$
V_{NT}-(\hat{F}'F_o/T)(\Gamma_o'\Gamma_o/N)(F_o'\hat{F}/T) = T^{-1}\hat{F}'\left[I_{1NT}(\hat{\theta},\hat{g}_k,\hat{F})+\cdots+I_{8NT}(\hat{F})\right] = o(1).
$$

Thus,

$$
V_{NT} = (\hat{F}'F_o/T)(\Gamma_o'\Gamma_o/N)(F_o'\hat{F}/T)+o_P(1).
$$

When proving Theorem 3.1, we have shown that $F_o'\hat{F}/T$ is non-singular in probability one. This implies that V_{NT} is invertible in probability one. We now left multiply (9.5) by F_o'/T to obtain

$$
(F_o'\hat{F}/T)V_{NT} = (F_o'F_o/T)(\Gamma_o'\Gamma_o/N)(F_o'\hat{F}/T)+o_P(1)
$$

based on the above analysis. The above equality shows that the columns of $F_o'\hat{F}/T$ are the (non-normalized) eigenvectors of the matrix $(F_o'F_o/T)(\Gamma_o'\Gamma_o/N)$, and V_{NT} consists of the eigenvalues of the same matrix (in the limit). Thus, $V_{NT}\to_P V$, where V is $m\times m$ and consists of the m eigenvalues of the matrix $\Sigma_F\Sigma_\Gamma$.

(2). Based on the above analysis, (9.6) can be rewritten as

$$\frac{1}{\sqrt{T}}\|\hat{F}Q^{-1} - F_o\| = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P(\eta_{NT}),$$

where $Q^{-1} = V_{NT}(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}$.

(3). According to (9.6),

$$\frac{1}{T}F_o'(\hat{F} - F_oQ) = \frac{1}{T}F_o'\left[I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F}) + \cdots + I_{8NT}(\hat{F})\right]V_{NT}^{-1}.$$

By (1) of this lemma, we know $V_{NT}^{-1} = O_P(1)$. Therefore, in the following analysis we just need to focus on $\frac{1}{T}F_o'\left[I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F}) + \cdots + I_{8NT}(\hat{F})\right]$. Furthermore, by (1) of this lemma, it is easy to show that

$$\left\|\frac{1}{T}F_o'\left[I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F}) + \cdots + I_{5NT}(\hat{\theta}, \hat{g}_k, \hat{F})\right]\right\| = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w).$$

We now consider $\left\|\frac{1}{T}F_o'I_{6NT}(\hat{F})\right\|$ below. For the term $\frac{1}{NT}\sum_{i=1}^N\left\|\varepsilon_i'\hat{F}\right\|^2$, we have

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^N\left\|\varepsilon_i'\hat{F}\right\|^2 &\le \frac{2}{NT}\sum_{i=1}^N\|\varepsilon_i'F_oQ\|^2 + \frac{2}{NT}\sum_{i=1}^N\left\|\varepsilon_i'\left(\hat{F} - F_oQ\right)\right\|^2 \\
&\le \frac{2}{NT}\sum_{i=1}^N\|\varepsilon_i'F_o\|^2\|Q\|^2 + \frac{2}{NT}\mathrm{tr}\left[\left(\hat{F} - F_oQ\right)'\varepsilon'\varepsilon\left(\hat{F} - F_oQ\right)\right] \\
&\le O_P(1) + \frac{1}{N}\|\varepsilon\|_{\mathrm{sp}}^2\frac{1}{T}\left\|\hat{F} - F_oQ\right\|^2 \\
&= O_P(1) + O_P(1 + T/N)\,O_P\left(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w^2 + \eta_{NT}^2\right)
\end{aligned}$$

by the result (2) of this lemma. It then gives that

$$\begin{aligned}
&\left\|\frac{1}{T}F_o'I_{6NT}(\hat{F})\right\| = \frac{1}{NT^2}\sum_{i=1}^N\|F_o'\varepsilon_i\|\cdot\left\|\varepsilon_i'\hat{F}\right\| \\
&\le \frac{1}{T}\left(\frac{1}{NT}\sum_{i=1}^N\|F_o'\varepsilon_i\|^2\right)^{1/2}\left(\frac{1}{NT}\sum_{i=1}^N\left\|\varepsilon_i'\hat{F}\right\|^2\right)^{1/2} \\
&= \frac{1}{T} + \frac{1}{T}O_P\left(1 + \sqrt{T/N}\right)O_P\left(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w + \eta_{NT}\right) \\
&= O_P\left(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w + \eta_{NT}^2\right)
\end{aligned}$$

as long as $\sqrt{T/N}\eta_{NT} = O(1)$, which is sufficient here.

For $\left\|\frac{1}{T}F_o'I_{7NT}(\hat{F})\right\|$, we have

$$\left\|\frac{1}{T}F_o'I_{7NT}(\hat{F})\right\| \leq \left\|\frac{1}{T}F_o'F_o\right\| \cdot \left\|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{o,i}\varepsilon_i'(\hat{F}-F_oQ)\right\| + \left\|\frac{1}{T}F_o'F_o\right\| \cdot \left\|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{o,i}\varepsilon_i'F_oQ\right\|$$
$$\leq \left\|\frac{1}{T}F_o'F_o\right\| \cdot \left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\gamma_{o,i}\varepsilon_i'\right\| \cdot \frac{1}{\sqrt{T}}\|\hat{F}-F_oQ\| + \left\|\frac{1}{T}F_o'F_o\right\| \cdot \left\|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{o,i}\varepsilon_i'F_o\right\| \cdot \|Q\|.$$

By Assumption 1.4, we know $\left\|\frac{1}{T}F_o'F_o\right\| = O_P(1)$. By results (1) and (2) of this lemma, we have $\|Q\| = O_P(1)$ and $\frac{1}{\sqrt{T}}\|\hat{F}-F_oQ\| = O_P(\|(\hat{\theta},\hat{g}_k)-(\theta_o,g_o)\|_w) + O_P(\eta_{NT})$. Therefore, we will focus on $\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\gamma_{o,i}\varepsilon_i'\right\|$ and $\left\|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{o,i}\varepsilon_i'F_o\right\|$ below. For notational simplicity, we temporarily assume that both $\gamma_{o,i}$ and $f_{o,t}$ are scalars. Write

$$E\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\gamma_{o,i}\varepsilon_i\right\|^2 = \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}E[\gamma_{o,i}\gamma_{o,j}]E[\varepsilon_{it}\varepsilon_{jt}]$$
$$\leq O(1)\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}|E[\varepsilon_{it}\varepsilon_{jt}]| = O\left(\frac{1}{N}\right)$$

and

$$E\left\|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{o,i}\varepsilon_i'F_o\right\|^2 = \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\gamma_{o,i}f_{o,t}\gamma_{o,j}f_{o,s}]E[\varepsilon_{it}\varepsilon_{js}]$$
$$\leq O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{js}]| = O\left(\frac{1}{NT}\right),$$

which immediately yields

$$\left\|\frac{1}{T}F_o'I_{7NT}(\hat{F})\right\| = O_P(\|(\hat{\theta},\hat{g}_k)-(\theta_o,g_o)\|_w \cdot \eta_{NT}) + O_P\left(\eta_{NT}^2\right)$$

Similarly, we can show $\left\|\frac{1}{T}F_o'I_{8NT}(\hat{F})\right\| = O_P(\|(\hat{\theta},\hat{g}_k)-(\theta_o,g_o)\|_w \cdot \eta_{NT}) + O_P\left(\eta_{NT}^2\right)$.

Based on the above analysis,

$$\left\|\frac{1}{T}F_o'(\hat{F}-F_oQ)\right\| = O_P(\|(\hat{\theta},\hat{g}_k)-(\theta_o,g_o)\|_w) + O_P\left(\eta_{NT}^2\right) \tag{9.7}$$

and

$$\left\|\frac{1}{T}\hat{F}'(\hat{F}-F_oQ)\right\| = \left\|\frac{1}{T}(\hat{F}-F_oQ+F_oQ)'(\hat{F}-F_oQ)\right\|$$
$$\leq \left\|\frac{1}{T}(\hat{F}-F_oQ)'(\hat{F}-F_oQ)\right\| + \|Q\| \cdot \left\|\frac{1}{T}F_o'(\hat{F}-F_oQ)\right\|$$

$$= O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P\left(\eta_{NT}^2\right). \tag{9.8}$$

(4). Note (9.7) and (9.8) can be respectively expressed as

$$\frac{1}{T}F_o'\hat{F} - \frac{1}{T}F_o'F_oQ = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P\left(\eta_{NT}^2\right)$$

and

$$I_m - \frac{1}{T}\hat{F}'F_oQ = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P\left(\eta_{NT}^2\right).$$

Simple algebra further gives

$$\frac{1}{T}Q'F_o'\hat{F} - \frac{1}{T}Q'F_o'F_oQ = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P\left(\eta_{NT}^2\right)$$

and

$$I_m - \frac{1}{T}H'F_o'\hat{F} = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P\left(\eta_{NT}^2\right).$$

Summing up both equations above, we obtain

$$I_m - \frac{1}{T}Q'F_o'F_oQ = O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P\left(\eta_{NT}^2\right). \tag{9.9}$$

Note when proving Theorem 3.1, we have showed that

$$\left\|P_{\hat{F}} - P_{F_o}\right\|^2 = \mathrm{tr}\left[(P_{\hat{F}} - P_{F_o})^2\right] = 2\mathrm{tr}\left[I_m - \hat{F}'P_{F_o}\hat{F}/T\right] \tag{9.10}$$

and, when proving this lemma, we have shown that

$$\frac{F_o'\hat{F}}{T} = \frac{F_o'F_o}{T}Q + O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P\left(\eta_{NT}^2\right).$$

Therefore, we can write

$$\hat{F}'P_{F_o}\hat{F}/T = Q'\left(\frac{F_o'F_o}{T}\right)Q + O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P\left(\eta_{NT}^2\right).$$

In connection with (9.9), we then obtain that

$$\hat{F}'P_{F_o}\hat{F}/T = I_m + O_P(\|(\hat{\theta}, \hat{g}_k) - (\theta_o, g_o)\|_w) + O_P\left(\eta_{NT}^2\right).$$

Then the proof of the second result of this lemma is complete. ■

Proof of Theorem 3.2:

Having proved Lemma 3.1, we now turn to investigating $\sqrt{NT}(\hat{\theta}-\theta_o)$. By (3.8), we can show that $\frac{\partial S_{NT}(\hat{\theta},\hat{C},\hat{F})}{\partial\theta}=0$. Following Yu and Ruppert (2002), we just need to focus on the following equation

$$0=\frac{\partial S_{NT}}{\partial\theta}\Big|_{(\theta,C,F)=(\hat{\theta},\hat{C},\hat{F})}=\frac{\partial S_{NT}}{\partial\theta}\Big|_{(\theta,C,F)=(\theta_o,\hat{C},\hat{F})}+\frac{\partial S_{NT}}{\partial\theta\partial\theta'}\Big|_{(\theta,C,F)=(\tilde{\theta},\hat{C},\hat{F})}(\hat{\theta}-\theta_o),$$

where $\tilde{\theta}$ lies between θ_o and $\hat{\theta}$, and

$$\begin{aligned}
\frac{\partial S_{NT}(\theta,C,F)}{\partial\theta} &= -\frac{2}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta_o,g_k^{(1)}]'M_F(Y_i-\phi_i[\theta,g_k]),\\
\frac{\partial S_{NT}(\theta,C,F)}{\partial\theta\partial\theta'} &= \frac{2}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta,g_k^{(1)}]'M_F\psi_{1i}[\theta,g_k^{(1)}]\\
&\quad -\frac{2}{NT}\sum_{i=1}^{N}\psi_{2i}[\theta,g_k^{(2)}]'\ \{M_F\otimes I_d\}\ \{(Y_i-\phi_i[\theta,g_k])\otimes I_d\},\\
\underset{T\times d}{\psi_{1i}[\theta,g]} &= (g(x_{i1}'\theta)x_{i1},\ldots,g(x_{iT}'\theta)x_{iT})',\\
\underset{Td\times d}{\psi_{2i}[\theta,g]} &= (g(x_{i1}'\theta)x_{i1}x_{i1}',\ldots,g(x_{iT}'\theta)x_{iT}x_{iT}')'. \qquad (9.11)
\end{aligned}$$

Since $\tilde{\theta}$ lies between θ_o and $\hat{\theta}$, it is easy to show $\|(\tilde{\theta},\hat{g}_k)-(\theta_o,g_o)\|_w\to_P 0$ by Theorem 3.1. Thus, it is reasonable to focus on a sufficiently small neighborhood of (θ_o,g_o) in the following proof. Then the rest of the analysis is similar to the arguments of (2.19)-(2.21) of Amemiya (1993), and part (b) of Lemma B.1 of Yu and Ruppert (2002). Note that one can easily extend the arguments (2.19)-(2.21) of Amemiya (1993) to the current setting by treating β and $\|\cdot\|$ of Amemiya (1993) as (θ,g) and $\|\cdot\|_w$ of this paper respectively, which becomes exactly the same as Lemma A.2 of Newey and Powell (2003). Following the same spirit, Assumption 3 then allows us to simplify the analysis by focusing on the expectation. Also, in the following analysis, we will repeatedly use $g_k(w)=C'\mathcal{H}(w)$ defined in (3.2).

We consider $\frac{\partial S_{NT}}{\partial\theta\partial\theta'}\Big|_{(\theta,C,F)=(\tilde{\theta},\hat{C},\hat{F})}$ first. Write

$$\begin{aligned}
&\frac{\partial S_{NT}}{\partial\theta\partial\theta'}\Big|_{(\theta,C,F)=(\tilde{\theta},\hat{C},\hat{F})}\\
&=\left\{\frac{2}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta,g_k^{(1)}]'M_F\psi_{1i}[\theta,g_k^{(1)}]\right\}\Big|_{(\theta,C,F)=(\tilde{\theta},\hat{C},\hat{F})}\\
&\quad-\left\{\frac{2}{NT}\sum_{i=1}^{N}\psi_{2i}[\theta,g_k^{(2)}]'\ \{M_F\otimes I_d\}\ \{(Y_i-\phi_i[\theta,g_k])\otimes I_d\}\right\}\Big|_{(\theta,C,F)=(\tilde{\theta},\hat{C},\hat{F})}\\
&:=2A_{1NT}\big|_{(\theta,C,F)=(\tilde{\theta},\hat{C},\hat{F})}-2A_{2NT}\big|_{(\theta,C,F)=(\tilde{\theta},\hat{C},\hat{F})},
\end{aligned}$$

where the definitions of A_{1NT} and A_{2NT} are obvious. We then examine A_{1NT} and A_{2NT} respectively.

Step 1: For A_{1NT}, write

$$\begin{aligned}A_{1NT} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_k^{(1)}(x_{it}'\theta)\right)^2 x_{it}x_{it}' - \frac{1}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta, g_k^{(1)}]' P_F \psi_{1i}[\theta, g_k^{(1)}] \\ &:= A_{1NT,1} - A_{1NT,2}.\end{aligned}$$

For the first term on the right hand side above, write

$$\begin{aligned}A_{1NT,1} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_o^{(1)}(x_{it}'\theta_o)\right)^2 x_{it}x_{it}' \\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_k^{(1)}(x_{it}'\theta)\right)^2 x_{it}x_{it}' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_{o,k}^{(1)}(x_{it}'\theta)\right)^2 x_{it}x_{it}' \\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_{o,k}^{(1)}(x_{it}'\theta)\right)^2 x_{it}x_{it}' - \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_o^{(1)}(x_{it}'\theta)\right)^2 x_{it}x_{it}' \\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_o^{(1)}(x_{it}'\theta)\right)^2 x_{it}x_{it}' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_o^{(1)}(x_{it}'\theta_o)\right)^2 x_{it}x_{it}' \\ &:= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_o^{(1)}(x_{it}'\theta_o)\right)^2 x_{it}x_{it}' + A_{1NT,11} + A_{1NT,12} + A_{1NT,13}, \qquad (9.12)\end{aligned}$$

where $g_{o,k}$ is defined in (3.3); and the definitions of $A_{1NT,11}$, $A_{1NT,12}$ and $A_{1NT,13}$ are obvious. By (10) of Lemma A2, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_o^{(1)}(x_{it}'\theta_o)\right)^2 x_{it}x_{it}' = V_1 + o_P(1).$$

We then focus on $A_{1NT,11}$ to $A_{1NT,13}$ respectively. For $A_{1NT,11}$, write

$$\begin{aligned}&\|A_{1NT,11}\|\,|_{(\theta,C)=(\tilde\theta,\hat C)} \\ &= \left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_k^{(1)}(x_{it}'\theta)\right)^2 x_{it}x_{it}' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_{o,k}^{(1)}(x_{it}'\theta)\right)^2 x_{it}x_{it}'\right\|\,\Big|_{(\theta,C)=(\tilde\theta,\hat C)} \\ &= \left\|E\left[\left(g_k^{(1)}(x'\theta)\right)^2 xx'\right] - E\left[\left(g_{o,k}^{(1)}(x'\theta)\right)^2 xx'\right]\right\|\,\Big|_{(\theta,C)=(\tilde\theta,\hat C)} + o_P(1) \\ &= \left\|E\left[(C-C_o)'\dot{\mathcal{H}}(x'\theta)\dot{\mathcal{H}}(x'\theta)'(C+C_o)xx'\right]\right\|\,\Big|_{(\theta,C)=(\tilde\theta,\hat C)} + o_P(1) \\ &\le \left\{E\left|(C-C_o)'\dot{\mathcal{H}}(x'\theta)\right|^2 E\left\|\dot{\mathcal{H}}(x'\theta)'(C+C_o)xx'\right\|^2\right\}^{1/2}\Big|_{(\theta,C)=(\tilde\theta,\hat C)} + o_P(1) \\ &\le \left\{(C-C_o)'E\left[\dot{\mathcal{H}}(x'\theta)\dot{\mathcal{H}}(x'\theta)'\right](C-C_o)\right\}^{1/2}\end{aligned}$$

$$
\cdot \left\{ 2E \left\| C'\dot{\mathcal{H}}(x'\theta)xx' \right\|^2 + 2E \left\| C_o'\dot{\mathcal{H}}(x'\theta)xx' \right\|^2 \right\}^{1/2} \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} + o_P(1)
$$

$$
\leq O(1) \left\{ (C - C_o)' E \left[\dot{\mathcal{H}}(x'\theta)\dot{\mathcal{H}}(x'\theta)' \right] (C - C_o) \right\}^{1/2} \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} + o_P(1)
$$

$$
\leq O(1)\|\hat{C} - C_o\| + o_P(1) \leq O(1)\|\hat{g}_k - g_o\|_{L^2} + o_P(1) = o_P(1),
$$

where the second equality follows from (10) of Lemma A2; the first inequality follows from Cauchy-Schwarz inequality; the third inequality follows from Assumption 2.1; the fourth inequality follows from Assumption 3.2; and the fifth inequality follows from the definition of $\| \cdot \|_{L^2}$.

For $A_{1NT,12}$, write

$$
\begin{aligned}
\|A_{1NT,12}\| \, |_{\theta=\tilde{\theta}} &= \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(g_{o,k}^{(1)}(x_{it}'\theta) \right)^2 x_{it}x_{it}' - \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(g_o^{(1)}(x_{it}'\theta) \right)^2 x_{it}x_{it}' \right\| \Big|_{\theta=\tilde{\theta}} \\
&= \left\| E \left[\left\{ \left(g_{o,k}^{(1)}(x'\theta) \right)^2 - \left(g_o^{(1)}(x'\theta) \right)^2 \right\} xx' \right] \right\| \Big|_{\theta=\tilde{\theta}} + o_P(1) \\
&= \left\| E \left[\left(g_{o,k}^{(1)}(x'\theta) - g_o^{(1)}(x'\theta) \right) \left(g_{o,k}^{(1)}(x'\theta) + g_o^{(1)}(x'\theta) \right) xx' \right] \right\| \Big|_{\theta=\tilde{\theta}} + o_P(1) \\
&= \left\| E \left[\delta_{o,k}^{(1)}(x'\theta) \left(g_{o,k}^{(1)}(x'\theta) + g_o^{(1)}(x'\theta) \right) xx' \right] \right\| \Big|_{\theta=\tilde{\theta}} + o_P(1) \\
&\leq O(1) E \left\| \delta_{o,k}^{(1)}(x'\theta)xx' \right\| \Big|_{\theta=\tilde{\theta}} + o_P(1) \leq O\left(k^{-\frac{r}{2}+\frac{11}{12}} \right) E\|x\|^2 + o_P(1) \\
&\leq O\left(k^{-\frac{r}{2}+\frac{11}{12}} \right) + o_P(1) = o_P(1),
\end{aligned}
$$

where the second equality follows from (10) of Lemma A2; the first inequality follows from Assumption 2.1; the second inequality follows from (10) of Lemma A1; and the third inequality follows from Assumption 1.2.

For $A_{1NT,13}$, write

$$
\begin{aligned}
\|A_{1NT,13}\| \, |_{\theta=\tilde{\theta}} &= \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(g_o^{(1)}(x_{it}'\tilde{\theta}) \right)^2 x_{it}x_{it}' - \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(g_o^{(1)}(x_{it}'\theta_o) \right)^2 x_{it}x_{it}' \right\| \Big|_{\theta=\tilde{\theta}} \\
&= \left\| E \left[\left(g_o^{(1)}(x'\theta) \right)^2 xx' \right] - E \left[\left(g_o^{(1)}(x'\theta_o) \right)^2 xx' \right] \right\| \Big|_{\theta=\tilde{\theta}} + o_P(1) \\
&= \left\| E \left[\left(g_o^{(1)}(x'\theta) - g_o^{(1)}(x'\theta_o) \right) \left(g_o^{(1)}(x'\theta) + g_o^{(1)}(x'\theta_o) \right) xx' \right] \right\| \Big|_{\theta=\tilde{\theta}} + o_P(1) \\
&= \left\| E \left[g_o^{(2)}(x'\theta^*) \left(x'\theta - x'\theta_o \right) \left(g_o^{(1)}(x'\theta_1) + g_o^{(1)}(x'\theta_o) \right) xx' \right] \right\| \Big|_{\theta=\tilde{\theta}} + o_P(1) \\
&\leq O(1)\|\tilde{\theta} - \theta_o\| E\|x\|^3 + o_P(1) = o_P(1),
\end{aligned}
$$

where θ^* lies between θ and θ_o; the second equality follows from (10) of Lemma A2; the fourth equality follows from Mean Value Theorem; the inequality follows from Assumption 2.1; and the last equality follows from Theorem 3.1.

Based on the above analysis, it is easy to show $A_{1NT,1}|_{(\theta,C)=(\tilde{\theta},\hat{C})} \to_P V_1$.

We now focus on $A_{1NT,2}$.

$$\begin{aligned}
A_{1NT,2} &= \frac{1}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta,g_k^{(1)}]'P_F\psi_{1i}[\theta,g_k^{(1)}] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta,g_k^{(1)}]'P_{F_o}\psi_{1i}[\theta,g_k^{(1)}] + \frac{1}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta,g_k^{(1)}]'\left(P_F-P_{F_o}\right)\psi_{1i}[\theta,g_k^{(1)}] \\
&:= A_{1NT,21}+A_{1NT,22}.
\end{aligned}$$

For $A_{1NT,21}$, write

$$\begin{aligned}
A_{1NT,21} &= \frac{1}{N}\sum_{i=1}^{N}\left(\psi_{1i}[\theta,g_k^{(1)}]'F_o/T\right)\Sigma_F^{-1}(1+o_P(1))\left(F_o'\psi_{1i}[\theta,g_k^{(1)}])/T\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}x_{it}g_k^{(1)}(x_{it}'\theta)f_{o,t}'\right)\Sigma_F^{-1}(1+o_P(1))\left(\frac{1}{T}\sum_{t=1}^{T}f_{o,t}g_k^{(1)}(x_{it}'\theta)x_{it}'\right),
\end{aligned}$$

where the first equality follows from (2.3). Further note that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}x_{it}g_k^{(1)}(x_{it}'\theta)f_{o,t}' &= \frac{1}{T}\sum_{t=1}^{T}x_{it}g_k^{(1)}(x_{it}'\theta)f_{o,t}' - E\left[xg_k^{(1)}(x'\theta)f_o'\right] \\
&\quad +E\left[xg_k^{(1)}(x'\theta)f_o'\right] - E\left[xg_{o,k}^{(1)}(x'\theta)f_o'\right] \\
&\quad +E\left[xg_{o,k}^{(1)}(x'\theta)f_o'\right] - E\left[xg_o^{(1)}(x'\theta)f_o'\right] \\
&\quad +E\left[xg_o^{(1)}(x'\theta)f_o'\right] - E\left[xg_o^{(1)}(x'\theta_o)f_o'\right] \\
&\quad +E\left[xg_o^{(1)}(x'\theta_o)f_o'\right] \\
&:= B_{1iT}+B_2+B_3+B_4+E\left[xg_o^{(1)}(x'\theta_o)f_o'\right],
\end{aligned}$$

where the definitions of B_{1iT}, B_2, B_3 and B_4 are obvious. By Assumption 3.1, we know that $\max_{1\le i\le N}\|B_{1iT}\||_{(\theta,C)=(\tilde\theta,\hat C)} = o_P(1)$. As with $A_{1NT,11}$ to $A_{1NT,13}$, it is easy to show that $\|B_2\||_{\theta=\tilde\theta}=o_P(1)$, $\|B_3\||_{\theta=\tilde\theta}=o_P(1)$ and $\|B_4\|=o_P(1)$. Then, we immediately obtain that

$$A_{1NT,21}|_{(\theta,C)=(\tilde\theta,\hat C)} = V_2\Sigma_F^{-1}V_2'+o_P(1). \tag{9.13}$$

For $A_{1NT,22}$, we have $\|A_{1NT,22}\|\le\|P_F-P_{F_o}\|\frac{1}{NT}\sum_{i=1}^{N}\left\|\psi_{1i}[\theta,g_k^{(1)}]\right\|^2$. Note that it is easy to show that $\frac{1}{NT}\sum_{i=1}^{N}\left\|\psi_{1i}[\theta,g_k^{(1)}]\right\|^2=O_P(1)$ uniformly. In connection with Theorem 3.1, we immediately obtain $\|A_{1NT,22}\|\,|_{(\theta,C,F)=(\tilde\theta,\hat C,\hat F)}=o_P(1)$.

Therefore, we have shown that $A_{1NT,2}|_{(\theta,C,F)=(\tilde\theta,\hat C,\hat F)}=V_2\Sigma_F^{-1}V_2'+o_P(1)$. In connection with the result $A_{1NT,1}|_{(\theta,C)=(\tilde\theta,\hat C)}\to_P V_1$, we obtain $A_{1NT}|_{(\theta,C,F)=(\tilde\theta,\hat C,\hat F)}=V_1-V_2\Sigma_F^{-1}V_2'+o_P(1)$.

Step 2: Turing to A_{2NT}, write

$$\begin{aligned}
A_{2NT} &= \frac{1}{NT}\sum_{i=1}^{N}\psi_{2i}[\theta,g_k^{(2)}]'\ \{M_F\otimes I_d\}\ \{(Y_i-\phi_i[\theta,g_k])\otimes I_d\} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\psi_{2i}[\theta,g_k^{(2)}]'\ \{M_F\otimes I_d\}\ \{(\varepsilon_i+\phi_i[\theta_o,g_o]-\phi_i[\theta,g_k])\otimes I_d\} \\
&\quad +\frac{1}{NT}\sum_{i=1}^{N}\psi_{2i}[\theta,g_k^{(2)}]'\ \{M_F\otimes I_d\}\ \{(F_o\gamma_i-FQ^{-1}\gamma_i)\otimes I_d\} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}x_{it}'g_k^{(2)}(x_{it}'\theta)\left(\varepsilon_{it}+g_o(x_{it}'\theta_o)-g_k(x_{it}'\theta)\right) \\
&\quad -\frac{1}{NT}\sum_{i=1}^{N}\psi_{2i}[\theta,g_k^{(2)}]'\ \{P_F\otimes I_d\}\ \{(\varepsilon_i+\phi_i[\theta_o,g_o]-\phi_i[\theta,g_k])\otimes I_d\} \\
&\quad +\frac{1}{NT}\sum_{i=1}^{N}\psi_{2i}[\theta,g_k^{(2)}]'\ \{M_F\otimes I_d\}\ \{(F_o\gamma_i-FQ^{-1}\gamma_i)\otimes I_d\} \\
&:= A_{2NT,1}-A_{2NT,2}+A_{2NT,3},
\end{aligned}$$

where the definitions of $A_{2NT,1}$ to $A_{2NT,3}$ are obvious.

By Lemma 3.1, it is easy to show that $A_{2NT,3}|_{(\theta,C,F)=(\tilde{\theta},\hat{C},\hat{F})}=o_P(1)$.

For $A_{2NT,1}$, write

$$\begin{aligned}
A_{2NT,1} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}'+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_o(x_{it}'\theta_o)-g_k(x_{it}'\theta)\right)g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}' \\
&:= A_{2NT,11}+A_{2NT,12}.
\end{aligned}$$

By (11) of Lemma A2, it is straightforward to obtain $A_{2NT,11}=o_P(1)$ uniformly.

For $A_{2NT,12}$, write

$$\begin{aligned}
A_{2NT,12} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}g_o(x_{it}'\theta_o)g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}'-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}g_o(x_i'\theta)g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}' \\
&\quad +\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}g_o(x_{it}'\theta)g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}'-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}g_{o,k}(x_{it}'\theta)g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}' \\
&\quad +\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}g_{o,k}(x_{it}'\theta)g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}'-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}g_k(x_{it}'\theta)g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}' \\
&:= D_{1NT}+D_{2NT}+D_{3NT},
\end{aligned}$$

where the definitions of D_{1NT} to D_{3NT} are obvious.

For D_{1NT}, write

$$
\begin{aligned}
&\|D_{1NT}\||_{(\theta,C)=(\tilde{\theta},\hat{C})} \\
&= \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g_o(x_{it}'\theta_o) g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g_o(x_{it}'\theta) g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}' \right\| \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} \\
&= \left\| E\left[\left(g_o(x'\theta_o) - g_o(x'\theta)\right) g_k^{(2)}(x'\theta)xx' \right] \right\| \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} + o_P(1) \\
&= \left\| E\left[g_o^{(1)}(x'\theta^*)(x'\theta_o - x'\theta) g_k^{(2)}(x'\theta)xx' \right] \right\| \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} + o_P(1) \\
&\le O(1)\|\theta_o - \tilde{\theta}\| E\|x\|^3 + o_P(1) = o_P(1),
\end{aligned}
$$

where θ^* lies between θ and θ_o; the second equality follows from Assumption 3.1; the third equality follows from Mean Value Theorem; the first inequality follows from Assumption 2.1; and the last equality follows from Theorem 3.1 by noting that $\tilde{\theta}$ lies between $\hat{\theta}$ and θ_o.

For D_{2NT}, write

$$
\begin{aligned}
&\|D_{2NT}\| \,|_{(\theta,C)=(\tilde{\theta},\hat{C})} \\
&= \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g_o(x_{it}'\theta) g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g_{o,k}(x_{it}'\theta) g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}' \right\| \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} \\
&= \left\| E\left[\delta_{o,k}(x'\theta) g_k^{(2)}(x'\theta)xx' \right] \right\| \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} + o_P(1) \le O(1)k^{-\frac{r}{2}+\frac{5}{12}} E\|x\|^2 = o_P(1),
\end{aligned}
$$

where the second equality follows from Assumption 3.1; and the inequality follows from (2) of Lemma A1 and Assumption 2.1.

For D_{3NT}, write

$$
\begin{aligned}
&\|D_{3NT}\||_{(\theta,C)=(\tilde{\theta},\hat{C})} \\
&= \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g_{o,k}(x_{it}'\theta) g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g_k(x_{it}'\theta) g_k^{(2)}(x_{it}'\theta)x_{it}x_{it}' \right\| \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} \\
&= \left\| E\left[\mathcal{H}(x'\theta)'C_o \ddot{\mathcal{H}}(x'\theta)'Cxx' \right] - E\left[\mathcal{H}(x'\theta)'C \ddot{\mathcal{H}}(x'\theta)'Cxx' \right] \right\| \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} + o_P(1) \\
&= \left\| E\left[\left(\mathcal{H}(x'\theta)'C_o - \mathcal{H}(x'\theta)'C\right) \ddot{\mathcal{H}}(x'\theta)'Cxx' \right] \right\| \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} + o_P(1) \\
&\le \left\{ (C_o - C)'E\left[\mathcal{H}(x'\theta)\mathcal{H}(x'\theta)'\right](C_o - C) \cdot E\left\| g_k^{(2)}(x'\theta)xx' \right\|^2 \right\}^{1/2} \Big|_{(\theta,C)=(\tilde{\theta},\hat{C})} + o_P(1) \\
&\le O(1)\left\{E\|x\|^4\right\}^{1/2} \|C_o - \hat{C}\| + o_P(1) \le O(1)\|\hat{g}_k - g_o\|_{L^2} + o_P(1) = o_P(1),
\end{aligned}
$$

where the second equality follows from Assumption 3.1; the first inequality follows from Cauchy-Schwarz inequality; the second inequality follows from Assumptions 2.1 and 3.2; the third inequality follows from the fact that $\|\hat{g}_k - g_o\|_{L^2}^2 = \|\hat{C} - C_o\|^2 + \|\delta_{o,k}\|_{L^2}^2$; and the last equality follows from Theorem 3.1.

When analyzing D_{1NT} to D_{3NT}, we have shown that $A_{2NT,12}|_{(\theta,C)=(\tilde{\theta},\hat{C})} = o_P(1)$. In connection with $A_{2NT,11} = o_P(1)$ uniformly, we obtain that $A_{2NT,1}|_{(\theta,C)=(\tilde{\theta},\hat{C})} = o_P(1)$.

For $A_{2NT,2}$,

$$A_{2NT,2} = \frac{1}{NT}\sum_{i=1}^{N}\psi_{2i}[\theta, g_k^{(2)}]' \ \{P_F(\varepsilon_i + \phi_i[\theta_o, g_o] - \phi_i[\theta, g_k]) \otimes I_d\}$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\psi_{2i}[\theta, g_k^{(2)}]' \ \{P_F\varepsilon_i \otimes I_d\} + \frac{1}{NT}\sum_{i=1}^{N}\psi_{2i}[\theta, g_k^{(2)}]' \ \{P_F(\phi_i[\theta_o, g_o] - \phi_i[\theta, g_k]) \otimes I_d\}.$$

By a procedure similar to that used for (4) and (12) of Lemma A2, it is easy to show that $A_{2NT,2}|_{(\theta,C,F)=(\tilde{\theta},\hat{C},\hat{F})}$.

Based on the analysis of *Steps 1-2*, we have shown that $\frac{\partial S_{NT}}{\partial\theta\partial\theta'}\Big|_{(\theta,C,F)=(\tilde{\theta},\hat{C},\hat{F})} \to_P 2V_*$, where $V_* = V_1 - V_2\Sigma_F^{-1}V_2'$.

We now focus on $\frac{\partial S_{NT}}{\partial\theta}\Big|_{(\theta,C,F)=(\theta_o,\hat{C},\hat{F})}$.

$$\frac{\partial S_{NT}}{\partial\theta}\Big|_{(\theta,C,F)=(\theta_o,\hat{C},\hat{F})} = -\frac{2}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta_o, g_k^{(1)}]'M_F(Y_i - \phi_i[\theta_o, g_k])\Big|_{(C,F)=(\hat{C},\hat{F})}$$

$$= \Big\{ -\frac{2}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta_o, g_o^{(1)}]'M_{F_o}\varepsilon_i$$

$$-\frac{2}{NT}\sum_{i=1}^{N}\left(\psi_{1i}[\theta_o, g_k^{(1)}] - \psi_{1i}[\theta_o, g_o^{(1)}]\right)' M_{F_o}\varepsilon_i$$

$$+\frac{2}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta_o, g_k^{(1)}]'(P_F - P_{F_o})\varepsilon_i$$

$$-\frac{2}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta_o, g_k^{(1)}]'M_F(\phi_i[\theta_o, g_o] - \phi_i[\theta_o, g_k])$$

$$-\frac{2}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta_o, g_k^{(1)}]'M_F F_o\gamma_i\Big\}\Big|_{(C,F)=(\hat{C},\hat{F})}$$

$$= 2\left(-\mathcal{A}_{1NT} - \mathcal{A}_{2NT} + \mathcal{A}_{3NT} - \mathcal{A}_{4NT} - \mathcal{A}_{5NT}\right)\big|_{(C,F)=(\hat{C},\hat{F})},$$

where the definitions of $\mathcal{A}_{1NT}$ to $\mathcal{A}_{5NT}$ are obvious.

For $\mathcal{A}_{1NT}$, it is easy to show that $\sqrt{NT}\mathcal{A}_{1NT} = N(0, \tilde{\Sigma})$ by Assumption 3.4.

For $\mathcal{A}_{2NT}$, write

$$\mathcal{A}_{2NT} = \frac{1}{NT}\sum_{i=1}^{N}\left(\psi_{1i}[\theta_o, g_k^{(1)}] - \psi_{1i}[\theta_o, g_o^{(1)}]\right)' M_{F_o}\varepsilon_i$$

$$
\begin{aligned}
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_k^{(1)}(x_{it}'\theta_o) - g_o^{(1)}(x_{it}'\theta_o)\right) x_{it}\varepsilon_{it} \\
&\quad -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(g_k^{(1)}(x_{it}'\theta_o) - g_o^{(1)}(x_{it}'\theta_o)\right) x_{it}f_{o,t}'\left(\frac{1}{T}\sum_{t=1}^{T} f_{o,t}f_{o,t}'\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} f_{o,t}\varepsilon_{it} \\
&= \mathcal{A}_{2NT,1} - \mathcal{A}_{2NT,2},
\end{aligned}
$$

where the definitions of $\mathcal{A}_{2NT,1}$ and $\mathcal{A}_{2NT,2}$ are obvious.

By (10) of Lemma A1 and (1) of Theorem 3.1, we obtain $\|\hat{g}_k^{(1)} - g_o^{(1)}\|_{L^2} = o_P(1)$. In connection with (1) and (2) of Lemma A3, we immediately obtain that $\mathcal{A}_{2NT,1}|_{C=\hat{C}} = o_P\left(\frac{1}{\sqrt{NT}}\right)$ and $\mathcal{A}_{2NT,2}|_{C=\hat{C}} = o_P\left(\frac{1}{\sqrt{NT}}\right)$ respectively, which in turn yields $\mathcal{A}_{2NT}|_{C=\hat{C}} = o_P\left(\frac{1}{\sqrt{NT}}\right)$. By (3) of Lemma A3, it is straightforward to show that $\mathcal{A}_{3NT}|_{(C,F)=(\hat{C},\hat{F})} = o_P\left(\frac{1}{\sqrt{NT}}\right)$.

Based on the above analysis, the result follows. ■

Proof of Theorem 3.3:

For notational simplicity, denote

$$
\begin{aligned}
&\underset{NT\times k}{\mathcal{H}_{NT}(\theta)} = (\mathcal{H}_1(\theta)', \ldots, \mathcal{H}_N(\theta)')', \quad \mathcal{H}_i(\theta) = \left(\mathcal{H}(x_{i1}'\theta), \ldots, \mathcal{H}(x_{iT}'\theta)\right)', \\
&\underset{NT\times 1}{\tilde{\delta}_{NT}} = (\delta_{o,k,1}(\theta_o)', \ldots, \delta_{o,k,N}(\theta_o)')', \quad \delta_{o,k,i}(\theta_o) = \left(\delta_{o,k}(x_{i1}'\theta_o), \ldots, \delta_{o,k}(x_{iT}'\theta_o)\right)', \\
&W(\theta) = (I_N \otimes M_{F_o})\,\mathcal{H}_{NT}(\theta)\left[\mathcal{H}_{NT}(\theta)'\,(I_N \otimes M_{F_o})\,\mathcal{H}_{NT}(\theta)\right]^{-1}\mathcal{H}_{NT}(\theta)'\,(I_N \otimes M_{F_o}),
\end{aligned}
$$

where for $i = 1, \ldots, N$.

Noting that $W(\theta)$ is symmetric and idempotent uniformly in θ, we can show that $\lambda_{max}(W(\theta)) = 1$ uniformly. Also, by (14) of Lemma A2 and Assumption 4.1, we have

$$
\lambda_{min}\left(\frac{1}{NT}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}\mathcal{H}_i(\hat{\theta})\right) \geq \frac{1}{2}\rho_1 > 0 \tag{9.14}
$$

w.p.a. 1.

Simple algebra shows that

$$
\|\hat{g}_k - g_o\|_{L^2}^2 = \|\hat{C} - C_o\|^2 + \|\delta_{o,k}\|_{L^2}^2, \tag{9.15}
$$

where C_o and $\delta_{o,k}$ are defined in (3.3). It is easy to show that $\|\delta_{o,k}\|_{L^2}^2 = O_P(k^{-r})$ by Assumption 2.1 and (3) of Lemma A1. Thus, we will focus on $\|\hat{C} - C_o\|$ in what follows.

By (3.8), we know

$$
0 = \frac{\partial S_{NT}}{\partial C}\Big|_{(\theta,C,F)=(\hat{\theta},\hat{C},\hat{F})} = -\frac{2}{NT}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{\hat{F}}(Y_i - \phi_i[\hat{\theta}, \hat{g}_k]),
$$

Further write

$$0 = \frac{2}{NT}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}(Y_i - \phi_i[\hat{\theta},\hat{g}_k]) + \frac{2}{NT}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'(M_{\hat{F}} - M_{F_o})(Y_i - \phi_i[\hat{\theta},\hat{g}_k]),$$

which gives

$$\begin{aligned}\hat{C} &= \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o} Y_i\right)\\ &\quad + \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'(M_{\hat{F}} - M_{F_o})(Y_i - \phi_i[\hat{\theta},\hat{g}_k])\right). \qquad (9.16)\end{aligned}$$

We then obtain

$$\begin{aligned}\hat{C} - C_o &= \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'(M_{F_o} + M_{\hat{F}} - M_{F_o})\varepsilon_i\\ &\quad + \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\delta_{o,k,i}(\theta_o)\\ &\quad + \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\left(\phi_i[\theta_o, g_{o,k}] - \phi_i[\hat{\theta}, g_{o,k}]\right)\\ &\quad + \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'(M_{\hat{F}} - M_{F_o})F_o\gamma_{o,i}\\ &\quad + \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'(M_{\hat{F}} - M_{F_o})\left(\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta},\hat{g}_k]\right)\\ &= \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\varepsilon_i(1 + o_P(1))\\ &\quad + \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\delta_{o,k,i}(\theta_o)\\ &\quad + \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\left(\phi_i[\theta_o, g_{o,k}] - \phi_i[\hat{\theta}, g_{o,k}]\right)\\ &\quad + \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{\hat{F}}(F_o - \hat{F}Q^{-1})\gamma_{o,i}\\ &\quad + \left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})' M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'(M_{\hat{F}} - M_{F_o})\left(\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta},\hat{g}_k]\right)\\ &:= A_{1NT} + A_{2NT} + A_{3NT} + A_{4NT} + A_{5NT}.\end{aligned}$$

where the definitions of A_{1NT} to A_{5NT} are obvious.

By (9.14) and (13) of Lemma A2, we immediately obtain that $\|A_{1NT}\| = O_P\left(\sqrt{k}\eta_{NT}\right)$.

For A_{2NT}, write

$$\|A_{2NT}\|^2 = \left\| \left[\mathcal{H}_{NT}(\hat{\theta})'(I_N \otimes M_{F_o})\mathcal{H}_{NT}(\hat{\theta})\right]^{-1} \mathcal{H}_{NT}(\hat{\theta})'(I_N \otimes M_{F_o})\tilde{\delta}_{NT} \right\|^2$$

$$= \tilde{\delta}'_{NT}(I_N \otimes M_{F_o})\mathcal{H}_{NT}(\hat{\theta}) \left[\mathcal{H}_{NT}(\hat{\theta})'(I_N \otimes M_{F_o})\mathcal{H}_{NT}(\hat{\theta})/(NT)\right]^{-1}$$

$$\cdot \left[\mathcal{H}_{NT}(\hat{\theta})' \left(I_N \otimes M_{F_o}\right) \mathcal{H}_{NT}(\hat{\theta})\right]^{-1} \mathcal{H}_{NT}(\hat{\theta})' \left(I_N \otimes M_{F_o}\right) \tilde{\delta}_{NT}/(NT)$$

$$\leq \left(\lambda_{min}\left\{\mathcal{H}_{NT}(\hat{\theta})' \left(I_N \otimes M_{F_o}\right) \mathcal{H}_{NT}(\hat{\theta})/(NT)\right\}\right)^{-1}$$

$$\cdot \tilde{\delta}'_{NT} \left(I_N \otimes M_{F_o}\right) \mathcal{H}_{NT}(\hat{\theta}) \left[\mathcal{H}_{NT}(\hat{\theta})' \left(I_N \otimes M_{F_o}\right) \mathcal{H}_{NT}(\hat{\theta})\right]^{-1} \mathcal{H}_{NT}(\hat{\theta})' \left(I_N \otimes M_{F_o}\right) \tilde{\delta}_{NT}/(NT)$$

$$\leq \left(\lambda_{min}\left\{\mathcal{H}_{NT}(\hat{\theta})' \left(I_N \otimes M_{F_o}\right) \mathcal{H}_{NT}(\hat{\theta})/(NT)\right\}\right)^{-1} \cdot \lambda_{max}(W(\hat{\theta})) \cdot \left\|\tilde{\delta}_{NT}\right\|^2/(NT)$$

$$\leq O_P(1) \cdot \lambda_{max}(W(\hat{\theta})) \cdot \left\|\tilde{\delta}_{NT}\right\|^2/(NT) \leq O_P(1) \cdot \left\|\tilde{\delta}_{NT}\right\|^2/(NT),$$

where the first inequality follows from exercise 5 on page 267 of Magnus and Neudecker (2007) (or (A.6) in Su and Jin (2012)); and the third inequality follows (9.14). Moreover, it is easy to show that $\left\|\tilde{\delta}_{NT}\right\|^2/(NT) = O_P(k^{-r})$. Then we obtain that $\|A_{2NT}\| = O_P(k^{-r/2})$.

We now focus on A_{4NT} below and write

$$A_{4NT} = \left(\sum_{i=1}^N \mathcal{H}_i(\hat{\theta})'M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1} \sum_{i=1}^N \mathcal{H}_i(\hat{\theta})'(M_{F_o} + M_{\hat{F}} - M_{F_o})(F_o - \hat{F}Q^{-1})\gamma_{o,i}$$

$$= \left(\sum_{i=1}^N \mathcal{H}_i(\hat{\theta})'M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1} \sum_{i=1}^N \mathcal{H}_i(\hat{\theta})'M_{F_o}(F_o - \hat{F}Q^{-1})\gamma_{o,i}(1 + o(1))$$

$$= -\left(\sum_{i=1}^N \mathcal{H}_i(\hat{\theta})'M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}$$

$$\times \sum_{i=1}^N \mathcal{H}_i(\hat{\theta})'M_{F_o}(I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F}) + \cdots + I_{8NT}(\hat{F}))(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}\gamma_{o,i}(1 + o(1)),$$

where $I_{jNT}(\cdot)$ for $j = 1, \ldots, 8$ are defined in the proof of Lemma 3.1. By following the same procedure as used in the proof of Lemma 3.1, it is easy to show that for $j = 6, 7, 8$

$$\left\| \left(\sum_{i=1}^N \mathcal{H}_i(\hat{\theta})'M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1} \sum_{i=1}^N \mathcal{H}_i(\hat{\theta})'M_{F_o}I_{jNT}(\hat{F}) \cdot (F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}\gamma_{o,i} \right\| = O_P(k^{1/2}\eta^2_{NT}).$$

We then focus on $I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F})$ to $I_{5NT}(\hat{\theta}, \hat{g}_k, \hat{F})$ and use $I_{2NT}(\hat{\theta}, \hat{g}_k, \hat{F})$ as an example below.

$$I_{2NT}(\hat{\theta}, \hat{g}_k, \hat{F}) = \left[\frac{1}{NT}\sum_{i=1}^N \left(\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta}, \hat{g}_k]\right)(F_o\gamma_{o,i})'\right]\hat{F}$$

$$
\begin{aligned}
&= \left[\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\theta_o, g_{o,k}] - \phi_i[\theta_o, \hat{g}_k]\right)(F_o\gamma_{o,i})'\right]\hat{F} + \left[\frac{1}{NT}\sum_{i=1}^{N}\phi_i[\theta_o, \delta_{o,k}]\,(F_o\gamma_{o,i})'\right]\hat{F} \\
&\quad + \left[\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\theta_o, \hat{g}_k] - \phi_i[\hat{\theta}, \hat{g}_k]\right)(F_o\gamma_{o,i})'\right]\hat{F} \\
&= I_{2NT,1}(\hat{g}_k, \hat{F}) + I_{2NT,2}(\hat{F}) + I_{2NT,3}(\hat{\theta}, \hat{g}_k, \hat{F}),
\end{aligned}
$$

where the definitions of $I_{2NT,1}(\cdot)$-$I_{2NT,3}(\cdot)$ are obvious. Similar to the analysis for A_{2NT}, it is easy to show that

$$
\begin{aligned}
&\left\|\left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}I_{2NT,2}(\hat{F})\cdot(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}\gamma_{o,i}\right\| \\
&= O_P(k^{-r/2}), \\
&\left\|\left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}I_{2NT,3}(\hat{\theta}, \hat{g}_k, \hat{F})\cdot(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}\gamma_{o,i}\right\| \\
&= O_P(\|\hat{\theta} - \theta_o\|).
\end{aligned}
$$

Then we just need to focus on $I_{2NT,1}(\hat{g}_k, \hat{F})$. Write

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}I_{2NT,1}(\hat{g}_k, \hat{F})\cdot(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}\gamma_{o,i} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}\left[\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\theta_o, g_{o,k}] - \phi_i[\theta_o, \hat{g}_k]\right)(F_o\gamma_{o,i})'\right]\hat{F}(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}\gamma_{o,i} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}\left[\frac{1}{NT}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})\gamma_{o,i}'(\Gamma_o'\Gamma_o/N)^{-1}\gamma_{o,i}\right](C_o - \hat{C}),
\end{aligned}
$$

Note that we can show that $\left|\left[\frac{1}{NT}\sum_{i=1}^{N}\mathcal{H}_i(\theta)\gamma_{o,i}'(\Gamma_o'\Gamma_o/N)^{-1}\gamma_{o,i}\right](C_o - \hat{C})\right| = O_P(\hat{C} - C_o)$ uniformly. Therefore, we have

$$
\begin{aligned}
&\left\|\left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}I_{2NT,1}(\hat{g}_k, \hat{F})\cdot(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}\gamma_{o,i}\right\| \\
&= O_P(\|\hat{C} - C_o\|),
\end{aligned}
$$

which implies

$$
\left(\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}\mathcal{H}_i(\hat{\theta})\right)^{-1}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'M_{F_o}I_{2NT,1}(\hat{g}_k, \hat{F})\cdot(F_o'\hat{F}/T)^{-1}(\Gamma_o'\Gamma_o/N)^{-1}\gamma_{o,i}
$$

$$= O_P(1)(\hat{C} - C_o).$$

Then we can go through the same procedure for $I_{1NT}(\hat{\theta}, \hat{g}_k, \hat{F})$ to $I_{5NT}(\hat{\theta}, \hat{g}_k, \hat{F})$. Similar to the analysis for A_{2NT}, we can obtain $\|A_{3NT}\| = O_P(\|\hat{\theta}-\theta_o\|) = O_P\left(\frac{1}{\sqrt{NT}}\right)+O_P\left(\Pi_{NT1}\right)+O_P\left(\Pi_{NT2}\right)$. It is easy to show that $\Pi_{NT1} = O_P(1)(\hat{C} - C_o)$. The derivation of Π_{NT2} is the same as that for A_{4NT}. For A_{5NT}, we have

$$\begin{aligned}
&\left\|\frac{1}{NT}\sum_{i=1}^{N}\mathcal{H}_i(\hat{\theta})'\left(M_{\hat{F}} - M_{F_o}\right)\left(\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta}, \hat{g}_k]\right)\right\| \\
&\leq \frac{\|P_{\hat{F}} - P_{F_o}\|}{NT}\sum_{i=1}^{N}\left\|\mathcal{H}_i(\hat{\theta})\right\| \cdot \left\|\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta}, \hat{g}_k]\right\| \\
&\leq \|P_{\hat{F}} - P_{F_o}\|\left(\frac{1}{NT}\sum_{i=1}^{N}\left\|\mathcal{H}_i(\hat{\theta})\right\|^2\right)^{1/2}\left(\frac{1}{NT}\sum_{i=1}^{N}\left\|\phi_i[\theta_o, g_o] - \phi_i[\hat{\theta}, \hat{g}_k]\right\|^2\right)^{1/2} \\
&= O_P\left(\sqrt{k}\|P_{\hat{F}} - P_{F_o}\| \cdot \|(\theta_o, g_o) - (\hat{\theta}, \hat{g}_k)\|_w\right).
\end{aligned}$$

Based on the above analysis and after some rearrangement, we then obtain

$$\|\hat{C} - C_o\| = O_P\left(\sqrt{k}\eta_{NT}\right) + O_P(k^{-r/2}),$$

which further implies

$$\|\hat{g} - g_o\|_{L^2} = O_P\left(\sqrt{k}\eta_{NT}\right) + O_P(k^{-r/2}).$$

Then the proof is complete. ■

Proofs of the results in Section 4 are similar to those given above, and thus provided in the supplementary file of this paper.

References

Amemiya, T. (1993), 'Non-linear regression model', *Handbook of Econometrics, Chapter 6* **1**, 333–389.

Bai, J. (2003), 'Inferential theory for factor models of large dimensions', *Econometrica* **71**(1), 135–171.

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* **77**(4), 1229–1279.

Bai, J., Kao, C. and Ng, S. (2009), 'Panel cointegration with global stochastic trends', *Journal of Econometrics* **149**(1), 82–99.

Berger, A., Demsetz, R. and Strahan, P. (1999), 'The consolidation of the financial services industry: Causes, consequences, and implications for the future', *Journal of Banking and Finance* **23**, 135–194.

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

Bernstein, D. S. (2009), *Matrix Mathematics: Theory, Facts, and Formulas*, Princeton University Press.

Bradley, R. C. (2005), 'Basic properties of strong mixing conditions: A survey and some open questions', *Probability Surveys* **2**(2), 107–144.

Brasili, A. and Vulpes, G. (2008), 'Banking integration and co-movements in EU banks' fragility', *In Lars Jonung (eds), Building the Financial Foundations of the Euro. Routledge* .

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

Chen, J., Gao, J. and Li, D. (2012*a*), 'A new diagnostic test for cross–section uncorrelatedness in nonparametric panel data models', *Econometric Theory* **28**(5), 1–20.

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

Chen, X. (2007), 'Large sample sieve estimation of semi-nonparametric models', *Handbook of Econometrics, Chapter 76* **6**, 5549–5632.

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

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

Dong, C. and Gao, J. (2014), 'Specification testing in structural nonparametric cointegration', *http://ssrn.com/abstract=2379928* .

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

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

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

Gao, J. (2007), *Nonlinear Time Series: Semiparametric and Nonparametric Methods*, CRC Press.

Gao, J., Tong, H. and Wolff, R. (2002), 'Model specification tests in nonparametric stochastic regression models', *Journal of Multivariate Analysis* **83**, 324–359.

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

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

Houston, J. F. and Stiroh, K. J. (2006), 'Three decades of financial sector risk', *Federal Reserve Bank of New York Staff Report No. 248* .

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

Jones, K. and Critchfield, T. (2005), 'Consolidation in the U.S. banking industry: Is the long, strange trip about to end?', *JFDIC Bank. Rev.* **17**, 31–61.

Kapetanios, G., Pesaran, M. H. and Yamagata, T. (2011), 'Panels with non-stationary multifactor error structures', *Journal of Econometrics* **160**(2), 326–348.

Lee, J. and Robinson, P. (2016), 'Series estimation under cross-sectional dependence', *Journal of Econometrics* **190**(1), 1–17.

Li, D., Qian, J. and Su, L. (2016), 'Panel data models with interactive fixed effects and multiple structural breaks', *Journal of the American Statistical Association* p. forthcoming.

Li, Q., Ouyang, D. and Racine, J. S. (2013), 'Categorical semiparametrics varying-coefficient models', *Journal of Applied Econometrics* **28**(4), 551–579.

Li, Q. and Racine, J. S. (2010), 'Smooth varying-coefficient estimation and inference for qualitative and quantitative data', *Econometric Theory* **26**(6), 1607–1637.

Lu, X. and Su, L. (2016), 'Shrinkage estimation of dynamic panel data models with interactive fixed effects', *Journal of Econometrics* **190**, 148–175.

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

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

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

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

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

Sealey, C. and Lindley, J. (1977), 'Inputs, outputs, and a theory of production and cost at depository financial institutions', *Journal of Finance* **32**, 1251–1266.

Stiroh, K. (2000), 'How did bank holding companies prosper in the 1990s?', *Journal of Banking and Finance* **24**, 1703–1745.

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

Su, L., Jin, S. and Zhang, Y. (2015), 'Specification test for panel data models with interactive fixed effects', *Journal of Econometrics* **190**, 222–244.

Sun, Y., Carroll, R. J. and Li, D. (2009), 'Semiparametric sstimation of fixed effects panel data varying coefficient models', *Advances in Econometrics* **25**, 101–130.

Wheelock, D. and Wilson, P. (2012), 'Do large banks have lower costs? New estimates of returns to scale for U.S. banks', *Journal of Money, Credit and Banking* **44**(1), 171–199.

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

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

Supplementary Appendix to "Semiparametric Single-Index Panel Data Models with Interactive Fixed Effects: Theory and Practice"

(NOT for publication)

GUOHUA FENG[§], BIN PENG[†], LIANGJUN SU[⋆] AND THOMAS TAO YANG[*]

[§]University of North Texas, [†]University of Technology Sydney, [⋆]Singapore Management University and [*]Australian National University

In this supplementary document, we provide the proofs of Lemmas A2-A4 and 4.1, and Theorem 4.1.

Appendix B

Note that when no misunderstanding arises, we may suppress the subscript indexes i and t to simplify notations. In this supplementary document, ρ_1, $O(1)$ and A always denote constants and may be different at each appearance. Due to the identification restrictions given in Assumption 1.4, we have $\frac{1}{T}\|F\|^2 = m$, which will be repeatedly used in the following analysis. Throughout this supplementary document, $\eta_{NT} = \frac{1}{\min\{\sqrt{N},\sqrt{T}\}}$ as defined in Lemma 3.1 of the main text.

Proofs of Section 3

Lemma A2. *Let $\phi_i[\theta, g] = (g(x_{i1}'\theta), \ldots, g(x_{iT}'\theta))'$ with $i = 1, \ldots, N$. We consider the following limits on 3-fold Cartesian product space formed by $\Theta \times \Im_1 \times D_F$. Under Assumptions 1 and 2, as $(N, T) \to (\infty, \infty)$,*

1. $\sup_{(\theta,g)} |L_{NT}(\theta, g) - L(\theta, g)| = o_P(1)$, where $L(\theta, g)$ is defined in Assumption 2 and

$$L_{NT}(\theta, g) = \frac{1}{NT}\sum_{i=1}^{N} (\phi_i[\theta, g] - \phi_i[\theta_o, g_o])' M_{F_o} (\phi_i[\theta, g] - \phi_i[\theta_o, g_o]);$$

2. $\sup_F \left\| \frac{1}{NT}\sum_{i=1}^{N} \varepsilon_i' P_F \varepsilon_i \right\| = o_P(1)$;

- Guohua Feng, University of North Texas, U.S.A. Email: Guohua.Feng@unt.edu.
- Bin Peng, University of Technology Sydney, Australia. Email: Bin.Peng@uts.edu.au.
- Liangjun Su, Singapore Management University, Singapore. Email: ljsu@smu.edu.sg.
- Thomas Tao Yang, Australian National University, Australia. Email: Tao.Yang@anu.edu.au.

3. $\sup_F \left\| \frac{1}{NT} \sum_{i=1}^N \gamma_{o,i}' F_o' M_F \varepsilon_i \right\| = o_P(1)$;

4. $\sup_{(\theta,g,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \phi_i[\theta, g]' M_F \varepsilon_i \right\| = o_P(1)$;

5. $\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \phi_i[\theta, \delta_{o,k}]' M_F \varepsilon_i \right\| = o_P(1)$;

6. $\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \phi_i[\theta, \delta_{o,k}]' M_F \phi_i[\theta, \delta_{o,k}] \right\| = o_P(1)$;

7. $\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \phi_i[\theta, \delta_{o,k}]' M_F F_o \gamma_{o,i} \right\| = o_P(1)$;

8. $\sup_{(\theta_1,\theta_2,g,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \phi_i[\theta_1, \delta_{o,k}]' M_F \phi_i[\theta_2, g] \right\| = o_P(1)$;

9. $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it}^2 = \sigma_\varepsilon^2 + O_P\left(\frac{1}{\sqrt{NT}}\right)$;

10. $\sup_{(\theta,g)} \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left(g^{(1)}(x_{it}'\theta)\right)^2 x_{it}x_{it}' - E\left[\left(g^{(1)}(x_{11}'\theta)\right)^2 x_{11}x_{11}'\right] \right\| = o_P(1)$;

11. $\sup_{(\theta,g)} \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it} g^{(2)}(x_{it}'\theta) x_{it}x_{it}' \right\| = o_P(1)$;

12. $\sup_F \left| \frac{1}{NT} \sum_{i=1}^N (\phi_i[\theta_o, g_o] - \phi_i[\theta, g])' M_F F_o \gamma_{o,i} \right| = O_P(\|(\theta_o, g_o) - (\theta, g)\|_w)$ for $\forall(\theta, g)$ in a sufficient small neighborhood of (θ_o, g_o).

   In addition, assume that Assumption 4 holds for the following results. Then

13. $\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^N \mathcal{H}_i(\theta)' M_F \varepsilon_i \right\| = O_P(\sqrt{k}\eta_{NT})$, where $\mathcal{H}_i(\theta) = (\mathcal{H}(x_{i1}'\theta), \ldots, \mathcal{H}(x_{iT}'\theta))'$ for $i = 1, \ldots, N$;

14. $\sup_\theta \left\| \frac{1}{NT} \sum_{i=1}^N \mathcal{H}_i(\theta)' M_{F_o} \mathcal{H}_i(\theta) - \Upsilon(\theta) \right\| = o_P(1)$, where $\mathcal{H}_i(\theta) = (\mathcal{H}(x_{i1}'\theta), \ldots, \mathcal{H}(x_{iT}'\theta))'$ for $i = 1, \ldots, N$.

Proof of Lemma A2:

(1). In what follows, we use Lemma A2 of Newey and Powell (2003) to prove this lemma. Write

$$\begin{aligned} L_{NT}(\theta, g) &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [\Delta g(x_{it}'\theta)]^2 - \frac{1}{N} \sum_{i=1}^N A_{iT}' \left(\frac{1}{T} F_o' F_o\right)^{-1} A_{iT} \\ &:= L_{1NT} - L_{2NT}, \end{aligned}$$

where $A_{iT} = \frac{1}{T} \sum_{t=1}^T \Delta g(x_{it}'\theta) f_{o,t}$ with $\Delta g(x_{it}'\theta) = g(x_{it}'\theta) - g_o(x_{it}'\theta_o)$, and the definitions of L_{1NT} and L_{2NT} are obvious.

We start with L_{1NT}.

Step 1: By Assumptions 1.4 and 2.2, we know that $\Theta \times \Im_1$ is a compact set with respect to norm $\| \cdot \|_w$.

Step 2: We now prove that for $\forall(\theta, g) \in \Theta \times \Im_1$, $L_{1NT}(\theta, g) \to_P L_1(\theta, g)$, where $L_1(\theta, g) = E[\Delta g(x_{11}'\theta)]^2$. Write

$$
\begin{aligned}
& E\left|L_{1NT}(\theta, g) - L_1(\theta, g)\right|^2 \\
&= \frac{1}{N^2T^2} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left\{\left[\Delta g(x_{it}'\theta)^2 - L_1(\theta, g)\right]\left[\Delta g(x_{js}'\theta)^2 - L_1(\theta, g)\right]\right\} \\
&\leq O(1)\frac{1}{N^2T^2} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} c_{\nu_1}\left(\alpha_{ij}(|t-s|)\right)^{\nu_1/(4+\nu_1)} \\
&\quad \left(E\left[\left(g_o(x_{it}'\theta_o) - g(x_{it}'\theta)\right)^{4+\nu_1}\right] \cdot E\left[\left(g_o(x_{js}'\theta_o) - g(x_{it}'\theta)\right)^{4+\nu_1}\right]\right)^{2/(4+\nu_1)} \\
&\leq O(1)\frac{1}{N^2T^2} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \left(\alpha_{ij}(|t-s|)\right)^{\nu_1/(4+\nu_1)} = O\left(\frac{1}{NT}\right),
\end{aligned}
\tag{1}
$$

where $c_{\nu_1} = 2^{(4+2\nu_1)/(4+\nu_1)} \cdot (4+\nu_1)/\nu_1$; the first inequality is due to the Davydov inequality; and the second inequality follows from the fact that g is uniformly bounded. Therefore, we have proved that, for $\forall(\theta, g) \in \Theta \times \Im_1$, $L_{1NT}(\theta, g) = L_1(\theta, g) + O_P\left(\frac{1}{\sqrt{NT}}\right)$.

Step 3: By *Step 2*, for $\forall(\theta_1, g_1)$, $(\theta_2, g_2) \in \Theta \times \Im_1$, we can write

$$
\left|L_{1NT}(\theta_1, g_1) - L_{1NT}(\theta_2, g_2)\right| = (1 + o_P(1)) \cdot \left|L_1(\theta_1, g_1) - L_1(\theta_2, g_2)\right|. \tag{2}
$$

To verify the third condition of Lemma A2 of Newey and Powell (2003), we just need to prove

$$
\left|L_1(\theta_1, g_1) - L_1(\theta_2, g_2)\right| \leq O(1) \cdot \|(\theta_1, g_1) - (\theta_2, g_2)\|_w.
$$

Thus, write

$$
\left|L_1(\theta_1, g_1) - L_1(\theta_2, g_2)\right| \leq \left|L_1(\theta_1, g_1) - L_1(\theta_1, g_2)\right| + \left|L_1(\theta_1, g_2) - L_1(\theta_2, g_2)\right|. \tag{3}
$$

For the first term on right hand side (RHS) of (3), write

$$
\begin{aligned}
& \left|L_1(\theta_1, g_1) - L_1(\theta_1, g_2)\right| \\
&= \left|E\left[\left(g_o(x'\theta_o) - g_1(x'\theta_1)\right)^2 - \left(g_o(x'\theta_o) - g_2(x'\theta_1)\right)^2\right]\right| \\
&= \left|E\left[\left(g_2(x'\theta_1) - g_1(x'\theta_1)\right) \cdot \left(2g_o(x'\theta_o) - g_1(x'\theta_1) - g_2(x'\theta_1)\right)\right]\right| \\
&\leq \left\{E\left[g_2(x'\theta_1) - g_1(x'\theta_1)\right]^2 \cdot E\left[2g_o(x'\theta_o) - g_1(x'\theta_1) - g_2(x'\theta_1)\right]^2\right\}^{1/2},
\end{aligned}
$$

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

$$
E\left[g_2(x'\theta_1) - g_1(x'\theta_1)\right]^2 \quad \text{and} \quad E\left[2g_o(x'\theta_o) - g_1(x'\theta_1) - g_2(x'\theta_1)\right]^2
$$

respectively. For $E\left[g_2(x'\theta_1) - g_1(x'\theta_1)\right]^2$, write

$$E\left[g_2(x'\theta_1)-g_1(x'\theta_1)\right]^2=\int(g_1(w)-g_2(w))^2f_{\theta_1}(w)dw$$
$$\leq O(1)\int(g_1(w)-g_2(w))^2dw=O(1)\|g_1-g_2\|_{L^2}^2,$$

where f_{θ_1} defines the density function of $z=x'\theta_1$; and the inequality follows from Assumption 2.1. Also, note that

$$E[g(x'\theta)]^2=\int(g(w))^2f_\theta(w)dw\leq O(1)\int(g(w))^2dw=O(1)\|g\|_{L^2}^2\leq O(1)$$

uniformly, where $f_\theta(w)$ defines the density function of $z=x'\theta$; the first inequality follows from Assumption 2.1; and the last inequality follows from the fact that $\Im_1$ is a compact set. Thus, we have

$$E\left[2g_o(x'\theta_o)-g_2(x'\theta_1)-g_1(x'\theta_1)\right]^2\leq 8E\left[g_o(x'\theta_o)\right]^2+4E\left[g_2(x'\theta_1)\right]^2+4E\left[g_1(x'\theta_1)\right]^2\leq O(1)$$

uniformly. Hence, we have shown

$$|L_1(\theta_1,g_1)-L_1(\theta_1,g_2)|\leq O(1)\|g_1-g_2\|_{L^2}. \tag{4}$$

We now consider the second term on RHS of (3).

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

Applying the same procedure as above, it is easy to show $E\left[2g_o(x'\theta_o)-g_2(x'\theta_1)-g_2(x'\theta_2)\right]^2$ is bounded uniformly on $\Theta\times\Im_1$, so we just need to focus on $E\left[g_2(x'\theta_2)-g_2(x'\theta_1)\right]^2$. Write

$$E\left[g_2(x'\theta_2)-g_2(x'\theta_1)\right]^2=E\left[(\theta_2-\theta_1)'xx'(\theta_2-\theta_1)\left\{g_2^{(1)}(x'\theta^*)\right\}^2\right]$$
$$\leq\|\theta_2-\theta_1\|^2E\left[\|x\|g_2^{(1)}(x'\theta^*)\right]^2\leq O(1)\|\theta_2-\theta_1\|^2E\|x\|^2\leq O(1)\|\theta_2-\theta_1\|^2,$$

where θ^* lies between θ_1 and θ_2; and the second inequality follows from Assumption 2.1. Then we know

$$|L_1(\theta_1,g_2)-L_1(\theta_2,g_2)|\leq O(1)\|\theta_2-\theta_1\|. \tag{5}$$

By (2)-(5), we immediately obtain

$$|L_{1NT}(\theta_1,g_1)-L_{1NT}(\theta_2,g_2)|\leq O_P(1)\left\|(\theta_1,g_1)-(\theta_2,g_2)\right\|_w,$$

which verifies the third condition of Lemma A2 of Newey and Powell (2003).

Based on *Steps 1-3*, we have shown that $L_{1NT}(\theta,g) \to_P L_1(\theta,g)$ uniformly in (θ,g).

Similarly, we can show that $L_{2NT} \to_P E[\Delta g(x_{11}'\theta)f_{o,1}']\Sigma_F^{-1}E[\Delta g(x_{11}'\theta)f_{o,1}]$ uniformly in (θ,g). Then the proof is complete.

(2). Write

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i' P_F \varepsilon_i &= \frac{1}{NT}\mathrm{tr}\left(P_F\varepsilon'\varepsilon\right) \le \frac{1}{NT}\left\|\varepsilon\right\|_{\mathrm{sp}}^2 \mathrm{tr}\left(P_F\right) \\ &= \frac{m}{NT}O_P(\max\{N,T\}) = O_P(\max\{N^{-1},T^{-1}\}) \text{ uniformly in } F.\end{aligned}$$

The proof is then complete.

(3). Write

$$\begin{aligned}\left|\frac{1}{NT}\sum_{i=1}^{N}\gamma_{o,i}'F_o'M_F\varepsilon_i\right| &= \left|\frac{1}{NT}\mathrm{tr}\left(F_o'M_F\varepsilon'\Gamma_o\right)\right| \le \frac{m}{NT}\left\|F_o'M_F\varepsilon'\Gamma_o\right\|_{\mathrm{sp}} \\ &\le \frac{m}{NT}\left\|F_o\right\|_{\mathrm{sp}}\left\|M_F\right\|_{\mathrm{sp}}\left\|\varepsilon\right\|_{\mathrm{sp}}\left\|\Gamma_o\right\|_{\mathrm{sp}} \\ &= \frac{m}{NT}O_P(\sqrt{T})\cdot 1\cdot O_P(\max\{\sqrt{N},\sqrt{T}\})O_P(\sqrt{N}) \\ &= O_P(\eta_{NT}),\end{aligned}$$

uniformly in F, where the first inequality follows from the fact that $\left|\mathrm{tr}\left(A\right)\right| \le \mathrm{rank}(A)\left\|A\right\|_{\mathrm{sp}}$.

(4). Let $\Phi \equiv \Phi\left(\theta,g\right) \equiv \left(\phi_1\left[\theta,g\right],\ldots,\phi_N\left[\theta,g\right]\right)'$. Write

$$\begin{aligned}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i\left[\theta,g\right]' M_F\varepsilon_i\right| &= \frac{1}{NT}\left|\mathrm{tr}\left(\varepsilon'\Phi\right) - \mathrm{tr}\left(P_F\varepsilon'\Phi\right)\right| \\ &\le \frac{1}{NT}\left\{\left\|\varepsilon\right\|_{\mathrm{sp}}\left\|\Phi\right\| + m\left\|P_F\right\|_{\mathrm{sp}}\left\|\varepsilon\right\|_{\mathrm{sp}}\left\|\Phi\right\|\right\} \\ &\le \frac{1}{NT}O_P(\max\{\sqrt{N},\sqrt{T}\})O_P(\sqrt{NT}) \\ &= O_P(\eta_{NT})\end{aligned}$$

uniformly, provided that $\left\|\Phi\left(\theta,g\right)\right\| = O_P(\sqrt{NT})$ uniformly in (θ,g), which can be easily verified by following similar arguments as used in the proof of (1) of this lemma.

(5). Let $\Delta\left(\theta\right) = \left(\delta_1\left(\theta\right),\ldots,\delta_N\left(\theta\right)\right)'$, where $\delta_i\left(\theta\right) = \left(\delta_{o,k}\left(x_{i1}'\theta\right),\ldots,\delta_{o,k}\left(x_{iT}'\theta\right)\right)'$. The proof is similar to that of (4) except that we need to use the fact that

$$\frac{1}{NT}\left\|\Delta\left(\theta\right)\right\|^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_{o,k}^2\left(x_{it}'\theta\right) = o_P\left(1\right),$$

uniformly in θ, where the uniform result follows from Assumption 2.1, (3) of Lemma A1 and Lemma A2 of Newey and Powell (2003).

(6). Applying a procedure similar to that used for (5), we have

$$\begin{aligned}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i\left[\theta,\delta_{o,k}\right]' M_F\phi_i\left[\theta,\delta_{o,k}\right]\right| &\leq \left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i\left[\theta,\delta_{o,k}\right]'\phi_i\left[\theta,\delta_{o,k}\right]\right| \\ &= \frac{1}{NT}\left\|\Delta\left(\theta\right)\right\|^2 = o_P\left(1\right),\end{aligned}$$

uniformly, where $\Delta(\theta)$ is defined in (5) of this lemma.

(7). Let $\Phi_k\left(\theta\right) = \left(\phi_1\left[\theta,\delta_{o,k}\right],\ldots,\phi_N\left[\theta,\delta_{o,k}\right]\right)'$.

$$\begin{aligned}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i\left[\theta,\delta_{o,k}\right]' M_F F_o\gamma_{o,i}\right| &= \left|\frac{1}{NT}\mathrm{tr}\left(M_F F_o\Gamma_o'\Phi_k\left(\theta\right)'\right)\right| \\ &\leq \frac{m}{NT}\left\|M_F\right\|_{\mathrm{sp}}\left\|F_o\right\|_{\mathrm{sp}}\left\|\Gamma_o\right\|_{\mathrm{sp}}\left\|\Phi_k\left(\theta\right)\right\|_{\mathrm{sp}} \\ &= \frac{m}{NT}\cdot O_P(\sqrt{T})O_P(\sqrt{N})o_P(\sqrt{NT}) \\ &= o_P\left(1\right),\end{aligned}$$

uniformly, where the second equality follows from that $\frac{1}{NT}\left\|\Phi_k\left(\theta\right)\right\|_{\mathrm{sp}}^2 = o_P(1)$ uniformly in θ as (5) of this lemma.

(8). By Cauchy-Schwarz inequality, we have

$$\begin{aligned}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i\left[\theta_1,\delta_{o,k}\right]' M_F\phi_i\left[\theta_2,g\right]\right| &\leq \left\{\frac{1}{NT}\sum_{i=1}^{N}\phi_i\left[\theta_1,\delta_{o,k}\right]' M_F\phi_i\left[\theta_1,\delta_{o,k}\right]\right\}^{1/2} \\ &\quad\times\left\{\frac{1}{NT}\sum_{i=1}^{N}\phi_i\left[\theta_2,g\right]'\phi_i\left[\theta_2,g\right]\right\}^{1/2} \\ &= o_P\left(1\right)O_P(1) = o_P\left(1\right),\end{aligned}$$

uniformly, where the first equality follows from the arguments in (4) and (5) of this lemma.

(9). Write

$$\begin{aligned}&E\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2-\sigma_\varepsilon^2\right|^2 = \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\{\left[\varepsilon_{it}^2-\sigma_\varepsilon^2\right]\left[\varepsilon_{js}^2-\sigma_\varepsilon^2\right]\right\} \\ &\leq \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}c_{\nu_1}\left(\alpha_{ij}(|t-s|)\right)^{\frac{\nu_1}{4+\nu_1}}\left(E|\varepsilon_{it}|^{4+\nu_1}\right)^{\frac{2}{4+\nu_1}}\left(E|\varepsilon_{js}|^{4+\nu_1}\right)^{\frac{2}{4+\nu_1}} \\ &\leq O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\alpha_{ij}(|t-s|)\right)^{\nu_1/(4+\nu_1)} = O\left(\frac{1}{NT}\right),\end{aligned}$$

where $c_{\nu_1} = 2^{(4+2\nu_1)/(4+\nu_1)}\cdot(4+\nu_1)/\nu_1$; the first inequality is due to the Davydov inequality; and the second inequality follows from Assumption 1.2. Thus, the result follows.

(10)-(11). The proofs of these two results are similar to that given for result (1) of this lemma.

(12). By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\left| \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\theta_o, g_o] - \phi_i[\theta, g])' M_F F_o \gamma_{o,i} \right| \\
&\leq \left\{ \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\theta_o, g_o] - \phi_i[\theta, g])' (\phi_i[\theta_o, g_o] - \phi_i[\theta, g]) \right\}^{1/2} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \gamma_{o,i}' F_o' M_F F_o \gamma_{o,i} \right\}^{1/2} \\
&= O_P(\|(\theta_o, g_o) - (\theta, g)\|_w) O_P(1) = O_P(\|(\theta_o, g_o) - (\theta, g)\|_w),
\end{aligned}
$$

as $\frac{1}{NT}\sum_{i=1}^{N} \gamma_{o,i}' F_o' M_F F_o \gamma_{o,i} \leq \frac{1}{NT}\sum_{i=1}^{N} \gamma_{o,i}' F_o' F_o \gamma_{o,i} = O_P(1)$ by Markov inequality and

$$
\begin{aligned}
&E \left| \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\theta_o, g_o] - \phi_i[\theta, g])' (\phi_i[\theta_o, g_o] - \phi_i[\theta, g]) \right| \\
&= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} E\left[g_o(x_{it}'\theta_o) - g(x_{it}'\theta) \right]^2 = O(\|(\theta_o, g_o) - (\theta, g)\|_w^2).
\end{aligned}
$$

(13). Let $\mathcal{H}[j] \equiv (\mathcal{H}_{1,j}(\theta), \ldots, \mathcal{H}_{N,j}(\theta))'$, where $\mathcal{H}_{i,j}(\theta)$ presents the j^{th} column of $\mathcal{H}_i(\theta)$ for $j = 1, \ldots, k$. Write

$$
\begin{aligned}
&\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^{N} \mathcal{H}_i(\theta)' M_F \varepsilon_i \right\| = \sup_{(\theta,F)} \left\{ \sum_{j=1}^{k} \left| \frac{1}{NT} \sum_{i=1}^{N} \mathcal{H}[j]' M_F \varepsilon_i \right|^2 \right\}^{1/2} \\
&= \sup_{(\theta,F)} \left\{ \sum_{j=1}^{k} \frac{1}{N^2T^2} \left| \mathrm{tr}\left(\varepsilon' \mathcal{H}[j]\right) - \mathrm{tr}\left(P_F \varepsilon' \mathcal{H}[j]\right) \right|^2 \right\}^{1/2} \\
&\leq \sup_{(\theta,F)} \left\{ \sum_{j=1}^{k} \frac{2}{N^2T^2} \left\{ \|\varepsilon\|_{\mathrm{sp}}^2 \|\mathcal{H}[j]\|^2 + m^2 \|P_F\|_{\mathrm{sp}}^2 \|\varepsilon\|_{\mathrm{sp}}^2 \|\mathcal{H}[j]\|^2 \right\} \right\}^{1/2} \\
&\leq \frac{\sqrt{k}}{NT} O_P(\max\{\sqrt{N}, \sqrt{T}\}) O_P(\sqrt{NT}) = O_P(\sqrt{k}\eta_{NT}) \qquad (6)
\end{aligned}
$$

uniformly, provided that $\|\mathcal{H}[j]\| = O_P(\sqrt{NT})$ uniformly by Lemma A1.

Therefore, the result follows.

(14). Write

$$
\begin{aligned}
&\frac{1}{NT} \sum_{i=1}^{N} \mathcal{H}_i(\theta)' M_{F_o} \mathcal{H}_i(\theta) - \Upsilon(\theta) \\
&= \frac{1}{NT} \sum_{i=1}^{N} \mathcal{H}_i(\theta)' \mathcal{H}_i(\theta) - \Upsilon_1(\theta) + \frac{1}{NT} \sum_{i=1}^{N} \mathcal{H}_i(\theta)' P_{F_o} \mathcal{H}_i(\theta) - \Upsilon_2(\theta).
\end{aligned}
$$

where $\Upsilon(\theta) = \Upsilon_1(\theta) + \Upsilon_2(\theta)$ is defined in Assumption 4 with $\Upsilon_1(\theta) = E[\mathcal{H}(x_{11}'\theta)\mathcal{H}(x_{11}'\theta)']$ and $\Upsilon_2(\theta) = E[\mathcal{H}(x_{11}'\theta)f_{o,1}']\Sigma_F^{-1}E[f_{o,1}\mathcal{H}(x_{11}'\theta)']$.

Consider $\frac{1}{NT}\sum_{i=1}^N \mathcal{H}_i(\theta)'\mathcal{H}_i(\theta) - \Upsilon_1(\theta)$ first. Let $\Upsilon_{1,uv}(\theta)$ denote the $(u,v)^{th}$ element of $\Upsilon_1(\theta)$ and $\Upsilon_{1NT,uv}(\theta) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{H}_u(x_{it}'\theta)\mathbb{H}_v(x_{it}'\theta)$ with $1 \le u, v \le k$, where $\mathbb{H}_v(w)$ is denoted in (3.2) of the main file.

Note that $\{\mathbb{H}_v, v = 0, 1, 2, \ldots\}$ is a sequence of known, uniformly bounded and integrable functions. Thus, following the same procedure as used for proving Theorem 2 of Jennrich (1969) or (B.10)-(B.18) of Chen et al. (2012), we can show that

$$\sup_\theta \left\| \frac{1}{NT}\sum_{i=1}^N \mathbb{H}_u(x_{it}'\theta)\mathbb{H}_v(x_{it}'\theta) - \Upsilon_{1,uv}(\theta) \right\| = O_P\left(\sqrt{\frac{\ln(NT)}{NT}}\right) = o_P(1).$$

Applying the same procedure as used for (6), we have

$$\sup_\theta \left\| \frac{1}{NT}\sum_{i=1}^N \mathcal{H}_i(\theta)'\mathcal{H}_i(\theta) - \Upsilon_1(\theta) \right\| = O_P\left(\sqrt{\frac{k^2\ln(NT)}{NT}}\right) = o_P(1),$$

where the last equality follows from Assumption 4.2.

Similarly, it is easy to know

$$\sup_\theta \left\| \frac{1}{T}\mathcal{H}_i(\theta)'F_o - E[\mathcal{H}(x_{11}'\theta)f_{o,1}'] \right\| = O_P\left(\sqrt{\frac{k\ln(T)}{T}}\right) = o(1).$$

Based on the above analysis and Assumption 4.2, the result then follows immediately. ■

Lemma A3. *Let* $\phi_i[\theta, g] = (g(x_{i1}'\theta), \ldots, g(x_{iT}'\theta))'$ *and* $\psi_{1i}[\theta, g] = (g(x_{i1}'\theta)x_{i1}, \ldots, g(x_{iT}'\theta)x_{iT})'$ *with* $i = 1, \ldots, N$. *We consider the following limits on 3-fold Cartesian product space formed by* $\Theta \times \Im_1 \times D_F$. *Under Assumptions 1-3,*

1. $\left\| \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \left(g^{(1)}(x_{it}'\theta_o) - g_o^{(1)}(x_{it}'\theta_o)\right) x_{it}\varepsilon_{it} \right\| = o_P\left(\frac{1}{\sqrt{NT}}\right)$, as $\|g^{(1)} - g_o^{(1)}\|_{L^2} = o(1)$;
2. $\frac{1}{NT}\sum_{i=1}^N \left(\psi_{1i}[\theta_o, g^{(1)}] - \psi_{1i}[\theta_o, g_o^{(1)}]\right)' P_{F_o}\varepsilon_i = o_P\left(\frac{1}{\sqrt{NT}}\right)$, as $\|g^{(1)} - g_o^{(1)}\|_{L^2} = o(1)$;
3. $\left\| \frac{1}{NT}\sum_{i=1}^N \psi_{1i}[\theta_o, g_o^{(1)}]' P_T \varepsilon_i \right\| = o_P\left(\frac{1}{\sqrt{NT}}\right)$, as $\|P_T\| = o(1)$.

Proof of Lemma A3:

(1). Write

$$\begin{aligned}
&E\left\| \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \left(g^{(1)}(x_{it}'\theta_o) - g_o^{(1)}(x_{it}'\theta_o)\right) x_{it}\varepsilon_{it} \right\|^2 \\
&= \frac{1}{N^2T^2}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T E\left[\left(g^{(1)}(x_{it}'\theta_o) - g_o^{(1)}(x_{it}'\theta_o)\right) x_{it}\varepsilon_{it}\left(g^{(1)}(x_{js}'\theta_o) - g_o^{(1)}(x_{js}'\theta_o)\right) x_{js}'\varepsilon_{js}\right]
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{js}]| \\
&\quad \left\{E\left\|g^{(1)}(x_{it}'\theta_o)-g_o^{(1)}(x_{it}'\theta_o)\right\|^4 E\|x_{it}\|^4 E\left\|g^{(1)}(x_{js}'\theta_o)-g_o^{(1)}(x_{js}'\theta_o)\right\|^4 E\|x_{js}\|^4\right\}^{1/4} \\
&\leq O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{js}]| \\
&\quad \left\{E\left\|g^{(1)}(x_{it}'\theta_o)-g_o^{(1)}(x_{it}'\theta_o)\right\|^2 E\left\|g^{(1)}(x_{js}'\theta_o)-g_o^{(1)}(x_{js}'\theta_o)\right\|^2\right\}^{1/4} \\
&\leq O(\|g^{(1)}-g_o^{(1)}\|_{L^2})\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[\varepsilon_{it}\varepsilon_{js}]| = O\left(\frac{\|g^{(1)}-g_o^{(1)}\|_{L^2}}{NT}\right),
\end{aligned}
$$

where the first inequality is obtained using Cauchy-Schwarz inequality twice; the second inequality follows from Assumptions 1.2 and 2.1; and the third inequality is obtained by applying a procedure similar o that used for (1) of Lemma A2.

(2). Write

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\left(g^{(1)}(x_{it}'\theta_o)-g_o^{(1)}(x_{it}'\theta_o)\right)x_{it}f_{o,t}'\right]\left[\frac{1}{T}\sum_{t=1}^{T}f_{o,t}f_{o,t}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}f_{o,t}\varepsilon_{it}\right] \\
&=\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\left(g^{(1)}(x_{it}'\theta_o)-g_o^{(1)}(x_{it}'\theta_o)\right)x_{it}f_{o,t}'\right]\left[\Sigma_F^{-1}(1+o_P(1))\right]\left[\frac{1}{T}\sum_{t=1}^{T}f_{o,t}\varepsilon_{it}\right].
\end{aligned}
$$

Then consider the leading term below.

$$
\begin{aligned}
&E\left\|\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\left(g^{(1)}(x_{it}'\theta_o)-g_o^{(1)}(x_{it}'\theta_o)\right)x_{it}f_{o,t}'\right]\Sigma_F^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}f_{o,t}\varepsilon_{it}\right]\right\|^2 \\
&=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left\{\left[\frac{1}{T}\sum_{t=1}^{T}f_{o,t}'\varepsilon_{it}\right]\Sigma_F^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\left(g^{(1)}(x_{it}'\theta_o)-g_o^{(1)}(x_{it}'\theta_o)\right)f_{o,t}x_{it}'\right]\right. \\
&\quad\left.\left[\frac{1}{T}\sum_{t=1}^{T}\left(g^{(1)}(x_{jt}'\theta_o)-g_o^{(1)}(x_{jt}'\theta_o)\right)x_{jt}f_{o,t}'\right]\Sigma_F^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}f_{o,t}\varepsilon_{jt}\right]\right\} \\
&\leq\frac{1}{N^2T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\left|E\left[\varepsilon_{it_1}\varepsilon_{jt_2}\right]\right| \\
&\quad\cdot\left|E\left[\left(g^{(1)}(x_{is_1}'\theta_o)-g_o^{(1)}(x_{is_1}'\theta_o)\right)x_{is_1}'f_{o,s_1}'\Sigma_F^{-1}f_{o,t_1}\left(g^{(1)}(x_{js_2}'\theta_o)-g_o^{(1)}(x_{js_2}'\theta_o)\right)x_{js_2}f_{o,s_2}'\Sigma_F^{-1}f_{o,t_2}\right]\right| \\
&\leq\frac{1}{N^2T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\left|E\left[\varepsilon_{it_1}\varepsilon_{jt_2}\right]\right| \\
&\quad\cdot\left\{E\left[\left(g^{(1)}(x_{is_1}'\theta_o)-g_o^{(1)}(x_{is_1}'\theta_o)\right)\left(g^{(1)}(x_{js_2}'\theta_o)-g_o^{(1)}(x_{js_2}'\theta_o)\right)\right]^2\right\}^{1/2}
\end{aligned}
$$

$$\cdot \left\{ E\left[x_{is_1}' f_{o,s_1}' \Sigma_F^{-1} f_{o,t_1} x_{js_2} f_{o,s_2}' \Sigma_F^{-1} f_{o,t_2} \right]^2 \right\}^{1/2}$$

$$\leq o(1)\frac{1}{N^2T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}|E\left[\varepsilon_{it_1}\varepsilon_{jt_2}\right]| = o\left(\frac{1}{NT}\right),$$

where the third inequality follows from Assumptions 2.1 and 3.3; and the last equality is obtained by applying the same procedure as used for (1) of this supplementary document.

Therefore, the result follows.

(3). Let p_{ts} denote the $(t,s)^{th}$ element of P_T with $1 \leq t, s \leq T$. Then write

$$E\left\| \frac{1}{NT}\sum_{i=1}^{N}\psi_{1i}[\theta_o, g_o^{(1)}]' P_T \varepsilon_i \right\|^2$$

$$= E\left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} g_o^{(1)}(x_{it}'\theta_o) x_{it} p_{ts}\varepsilon_{is} \right\|^2$$

$$= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{s_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_2=1}^{T} E\left[g_o^{(1)}(x_{it_1}'\theta_o) x_{it_1}' p_{t_1s_1}\varepsilon_{is_1} g_o^{(1)}(x_{jt_2}'\theta_o) x_{jt_2} p_{t_2s_2}\varepsilon_{js_2}\right]$$

$$\leq O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{s_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_2=1}^{T}|E[\varepsilon_{is_1}\varepsilon_{js_2}]| E\left[\|x_{it_1}\||p_{t_1s_1}|\|x_{jt_2}\||p_{t_2s_2}|\right]$$

$$\leq O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{s_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_2=1}^{T}|E[\varepsilon_{is_1}\varepsilon_{js_2}]| \left\{E\|x_{it_1}\|^4 E\|x_{jt_2}\|^4\right\}^{1/4}\left\{E[|p_{t_1s_1}|^2|p_{t_2s_2}|^2]\right\}^{1/2}$$

$$\leq o(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}|E[\varepsilon_{is_1}\varepsilon_{js_2}]| = o\left(\frac{1}{NT}\right),$$

where the first inequality follows from Assumptions 1.3 and 2.1; the second inequality follows from Cauchy-Schwarz inequality; and the last equality is obtained by applying a procedure similar to that used for (1) of this file. ■

Proofs of Section 4

Lemma A4. *Let $\phi_i[\theta, g] = (g(x_{i1}'\theta), \ldots, g(x_{iT}'\theta))'$ with $i = 1, \ldots, N$. We consider the following limits on 3-fold Cartesian product space formed by $\Theta \times \Im_2 \times D_F$. Under Assumptions 1 and 2*, as $(N,T) \to (\infty, \infty)$,*

1. $\sup_{(\theta,g)} |L_{NT}(\theta, g) - L(\theta, g)| = o_P(1)$, where $L(\theta, g)$ and $L_{NT}(\theta, g)$ have the same form as that defined Lemma A2;
2. $\sup_F \left\| \frac{1}{NT}\sum_{i=1}^N \varepsilon_i' P_F \varepsilon_i \right\| = o_P(1)$;
3. $\sup_F \left\| \frac{1}{NT}\sum_{i=1}^N \gamma_{o,i}' F_o' M_F \varepsilon_i \right\| = o_P(1)$;

4. $\sup_{(\theta,g,F)} \left\| \frac{1}{NT} \sum_{i=1}^{N} \phi_i[\theta, g]' M_F \varepsilon_i \right\| = o_P(1)$;

5. $\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^{N} \phi_i[\theta, \delta_{o,k}]' M_F \varepsilon_i \right\| = o_P(1)$;

6. $\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^{N} \phi_i[\theta, \delta_{o,k}]' M_F \phi_i[\theta, \delta_{o,k}] \right\| = o_P(1)$;

7. $\sup_{(\theta,F)} \left\| \frac{1}{NT} \sum_{i=1}^{N} \phi_i[\theta, \delta_{o,k}]' M_F F_o \gamma_{o,i} \right\| = o_P(1)$;

8. $\sup_{(\theta_1,\theta_2,g,F)} \left\| \frac{1}{NT} \sum_{i=1}^{N} \phi_i[\theta_1, \delta_{o,k}]' M_F \phi_i[\theta_2, g] \right\| = o_P(1)$;

9. $\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^2 = \sigma_\varepsilon^2 + O_P\left(\frac{1}{\sqrt{NT}}\right)$;

10. $\sup_F \left| \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\theta_o, g_o] - \phi_i[\theta, g])' M_F F_o \gamma_{o,i} \right| = O_P(\|(\theta_o, g_o) - (\theta, g)\|_w)$ for $\forall(\theta, g)$ in a sufficient small neighbourhood of (θ_o, g_o).

Proof of Lemma A4:

(1). In what follows, we use Lemma A2 of Newey and Powell (2003) to prove this lemma.

$$
\begin{aligned}
L_{NT}(\theta, g) \;&=\; \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [\Delta g(x_{it}'\theta)]^2 - \frac{1}{N} \sum_{i=1}^{N} A_{iT}' \left(\frac{1}{T} F_o' F_o\right)^{-1} A_{iT} \\
&:= L_{1NT} - L_{2NT},
\end{aligned}
$$

where $A_{iT} = \frac{1}{T} \sum_{t=1}^{T} \Delta g(x_{it}'\theta) f_{o,t}$ with $\Delta g(x_{it}'\theta) = g(x_{it}'\theta) - g_o(x_{it}'\theta_o)$; and the definitions of L_{1NT} and L_{2NT} should are obvious. Note that by Assumption 2*.2, we have, for $\forall(\theta, g) \in \Theta \times \Im_2$, $L_{NT}(\theta, g) = L(\theta, g) + o_P(1)$.

Step 1: By Assumptions 1.4 and 2*.2, we know that $\Theta \times \Im_2$ is a compact set with respect to norm $\|\cdot\|_{\widetilde{w}}$.

Step 2: This step is analogous to *Step 2* in the proof of (1) of Lemma A2. In order to verify the third condition of Lemma A2 of Newey and Powell (2003), we just need to show that $|L(\theta_1, g_1) - L(\theta_2, g_2)| \leq O(1) \cdot \|(\theta_1, g_1) - (\theta_2, g_2)\|_{\widetilde{w}}$. Start from $L_1(\theta, g) = E[\Delta g(x'\theta)]^2$. Write

$$
|L_1(\theta_1, g_1) - L_1(\theta_2, g_2)| \leq |L_1(\theta_1, g_1) - L_1(\theta_1, g_2)| + |L_1(\theta_1, g_2) - L_1(\theta_2, g_2)|\,. \tag{7}
$$

For the first term on right hand side (RHS) of (7), write

$$
\begin{aligned}
&|L_1(\theta_1, g_1) - L_1(\theta_1, g_2)| \\
&= \left| E\left[\left(g_o(x'\theta_o) - g_1(x'\theta_1)\right)^2 - \left(g_o(x'\theta_o) - g_2(x'\theta_1)\right)^2 \right] \right| \\
&= \left| E\left[\left(g_2(x'\theta_1) - g_1(x'\theta_1)\right) \cdot \left(2g_o(x'\theta_o) - g_1(x'\theta_1) - g_2(x'\theta_1)\right) \right] \right| \\
&\leq \left\{ E\left[g_2(x'\theta_1) - g_1(x'\theta_1)\right]^2 \cdot E\left[2g_o(x'\theta_o) - g_1(x'\theta_1) - g_2(x'\theta_1)\right]^2 \right\}^{1/2},
\end{aligned}
$$

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

$$E\left[g_2(x'\theta_1)-g_1(x'\theta_1)\right]^2 \quad \text{and} \quad E\left[2g_o(x'\theta_o)-g_1(x'\theta_1)-g_2(x'\theta_1)\right]^2$$

respectively. For $E\left[g_2(x'\theta_1)-g_1(x'\theta_1)\right]^2$, write

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

where the inequality follows from Assumption 2*.1. For $\forall(\theta,g)\in\Theta\times\Im_2$,

$$E[g(x'\theta)]^2 = \int (g(w))^2 f_\theta(w)dw \le O(1)\int (g(w))^2\exp(-w^2)dw = O(1)\|g\|^2_{\widetilde{L}^2} \le O(1),$$

where the first inequality follows from Assumption 2*.1, and the last inequality follows from $\Im_2$ being a compact set. Thus, we have

$$E\left[2g_o(x'\theta_o)-g_2(x'\theta_1)-g_1(x'\theta_1)\right]^2 \le 8E\left[g_o(x'\theta_o)\right]^2 + 4E\left[g_2(x'\theta_1)\right]^2 + 4E\left[g_1(x'\theta_1)\right]^2 \le O(1).$$

Hence, we have shown

$$|L_1(\theta_1,g_1)-L_1(\theta_1,g_2)| \le O(1)\|g_1-g_2\|_{\widetilde{L}^2}. \tag{8}$$

We now consider the second term on RHS of (3).

$$\begin{aligned}
&|L_1(\theta_1,g_2)-L_1(\theta_2,g_2)| \\
&= \left|E\left[\left(g_o(x'\theta_o)-g_2(x'\theta_1)\right)^2 - \left(g_o(x'\theta_o)-g_2(x'\theta_2)\right)^2\right]\right| \\
&= \left|E\left[\left(g_2(x'\theta_2)-g_2(x'\theta_1)\right)\cdot\left(2g_o(x'\theta_o)-g_2(x'\theta_1)-g_2(x'\theta_2)\right)\right]\right| \\
&\le \left\{E\left[g_2(x'\theta_2)-g_2(x'\theta_1)\right]^2 \cdot E\left[2g_o(x'\theta_o)-g_2(x'\theta_1)-g_2(x'\theta_2)\right]^2\right\}^{1/2}.
\end{aligned}$$

Applying a procedure as above, it is easy to show $E\left[2g_o(x'\theta_o)-g_2(x'\theta_1)-g_2(x'\theta_2)\right]^2$ is bounded uniformly on $\Theta\times\Im_2$, so we just need to focus on $E\left[g_2(x'\theta_2)-g_2(x'\theta_1)\right]^2$. Write

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

where θ^* lies between θ_1 and θ_2; and the second inequality follows from Assumption 2*.1. Then we have

$$|L_1(\theta_1,g_2)-L_1(\theta_2,g_2)| \le O(1)\|\theta_2-\theta_1\|. \tag{9}$$

By (7)-(9), we immediately obtain

$$|L_1(\theta_1, g_1) - L_1(\theta_2, g_2)| \leq O(1) \|(\theta_1, g_1) - (\theta_2, g_2)\|_{\widetilde{w}},$$

which verifies the third condition of Lemma A2 of Newey and Powell (2003).

Similarly, we can prove the continuity of $L_2(\theta, g) = E[\Delta g(x'\theta) f_o']\Sigma_F^{-1} E[f_o \Delta g(x'\theta)]$.

With *Steps 1 and 2*, the proof is complete.

(2)-(9). Noting that $\frac{1}{NT}\sum_{i=1}^N \|\phi_i[\theta, g]\|^2 = O(1)$ uniformly in (θ, g), we can prove (2)-(9) by using a procedure similar to that used for Lemma A2.

(10). Write

$$\begin{aligned}
& E\left|\frac{1}{NT}\sum_{i=1}^N (\phi_i[\theta_o, g_o] - \phi_i[\theta, g])' M_F F_o \gamma_{o,i}\right| \\
\leq & \frac{1}{NT}\sum_{i=1}^N E\left[\left\{(\phi_i[\theta_o, g_o] - \phi_i[\theta, g])' M_F (\phi_i[\theta_o, g_o] - \phi_i[\theta, g])\right\}^{1/2} \left\{\gamma_{o,i}' F_o' M_F F_o \gamma_{o,i}\right\}^{1/2}\right] \\
\leq & \frac{1}{N}\sum_{i=1}^N E\left[\left\{\frac{1}{T}(\phi_i[\theta_o, g_o] - \phi_i[\theta, g])'(\phi_i[\theta_o, g_o] - \phi_i[\theta, g])\right\}^{1/2} \left\{\frac{1}{T}\gamma_{o,i}' F_o' \ F_o \gamma_{o,i}\right\}^{1/2}\right] \\
\leq & \frac{1}{N}\sum_{i=1}^N \left\{E\left[\frac{1}{T}(\phi_i[\theta_o, g_o] - \phi_i[\theta, g])'(\phi_i[\theta_o, g_o] - \phi_i[\theta, g])\right] E\left[\frac{1}{T}\gamma_{o,i}' F_o' \ F_o \gamma_{o,i}\right]\right\}^{1/2} \\
\leq & \left\{\frac{1}{N}\sum_{i=1}^N E\left[\frac{1}{T}(\phi_i[\theta_o, g_o] - \phi_i[\theta, g])'(\phi_i[\theta_o, g_o] - \phi_i[\theta, g])\right]\right\}^{1/2} \left\{\frac{1}{N}\sum_{i=1}^N E\left[\frac{1}{T}\gamma_{o,i}' F_o' \ F_o \gamma_{o,i}\right]\right\}^{1/2} \\
\leq & O(1)\left\{\frac{1}{N}\sum_{i=1}^N E\left[\frac{1}{T}(\phi_i[\theta_o, g_o] - \phi_i[\theta, g])'(\phi_i[\theta_o, g_o] - \phi_i[\theta, g])\right]\right\}^{1/2} \\
\leq & O(1)\left\{E[g_o(x'\theta_o) - g(x'\theta)]^2\right\}^{1/2} = O(\|(\theta_o, g_o) - (\theta, g)\|_w),
\end{aligned}$$

where the first inequality follows from Exercise 1 of Magnus and Neudecker (2007); the second inequality follows from the fact $\lambda_{max}(M_F) = 1$; the third and forth inequalities follow from Cauchy-Schwarz inequality; and the last equality follows from the proof of result (1) of this lemma.

Based on the above discussions, the result follows. ■

Proof of Theorem 4.1:

(1). Expanding $S_{NT}(\theta, C, F) - S_{NT}(\theta_o, C_o, F_o)$, we have

$$\begin{aligned}
& \widetilde{S}_{NT}(\theta, C, F) - \widetilde{S}_{NT}(\theta_o, C_o, F_o) \\
= & \frac{1}{NT}\sum_{i=1}^N \left\{(Y_i - \phi_i[\theta, g_k])' M_F (Y_i - \phi_i[\theta, g_k]) - (Y_i - \phi_i[\theta_o, g_{o,k}])' M_{F_o} (Y_i - \phi_i[\theta_o, g_{o,k}])\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{NT}\sum_{i=1}^{N}\left\{(Y_i - \phi_i[\theta, g_k])' M_F (Y_i - \phi_i[\theta, g_k]) - (\phi_i[\theta_o, \delta_{o,k}] + \varepsilon_i)' M_{F_o} (\phi_i[\theta_o, \delta_{o,k}] + \varepsilon_i)\right\} \\
&= \tilde{\mathbb{S}}_{NT}(\theta, C, F) \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_i' (M_F - M_{F_o})\varepsilon_i + \frac{2}{NT}\sum_{i=1}^{N}\gamma_{o,i}' F_o' M_F \varepsilon_i + \frac{2}{NT}\sum_{i=1}^{N}(\phi_i[\theta_o, g_{o,k}] - \phi_i[\theta, g_k])' M_F \varepsilon_i \\
&\quad + \frac{2}{NT}\sum_{i=1}^{N}\phi_i[\theta_o, \delta_{o,k}]' (M_F - M_{F_o})\varepsilon_i + \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\phi_i[\theta_o, \delta_{o,k}]' (M_F - M_{F_o})\phi_i[\theta_o, \delta_{o,k}] \\
&\quad + \frac{2}{NT}\sum_{i=1}^{N}(\phi_i[\theta_o, g_{o,k}] - \phi_i[\theta, g_k])' M_F \phi_i[\theta_o, \delta_{o,k}] + \frac{2}{NT}\sum_{i=1}^{N}\gamma_{o,i}' F_o' M_F \phi_i[\theta_o, \delta_{o,k}],
\end{aligned}
$$

where g_k is defined in (4.2), $g_{o,k}$ and $\delta_{o,k}$ are defined in (4.3), and

$$
\tilde{\mathbb{S}}_{NT}(\theta, C, F) = \frac{1}{NT}\sum_{i=1}^{N}(\phi_i[\theta_o, g_{o,k}] - \phi_i[\theta, g_k] + F_o\gamma_i)' M_F (\phi_i[\theta_o, g_{o,k}] - \phi_i[\theta, g_k] + F_o\gamma_i).
$$

By (2)-(9) of Lemma A4, we immediately obtain

$$
\widetilde{S}_{NT}(\theta, C, F) - \widetilde{S}_{NT}(\theta_o, C_o, F_o) = \tilde{\mathbb{S}}_{NT}(\theta, C, F) + o_P(1). \tag{10}
$$

Let $b_i = \phi_i[\theta, g_k] - \phi_i[\theta_o, g_{o,k}]$ and $b = (b_1', \ldots, b_N')'$. As in the proof of Lemma A4, we can readily argue that $(NT)^{-1}b'b = O(1)$ uniformly in (θ, C). Let $\eta = \text{vec}(M_F F_o)$, $A_1 = I_N \otimes M_F$, $A_2 = (\Gamma_o'\Gamma_o) \otimes I_T$, and $A_3 = (\gamma_{o,1} \otimes I_T, \ldots, \gamma_{o,N} \otimes I_T)$. Then

$$
\begin{aligned}
\tilde{\mathbb{S}}_{NT}(\theta, C, F) &= \frac{1}{NT}\sum_{i=1}^{N}b_i' M_F b_i + \frac{1}{NT}\sum_{i=1}^{N}\gamma_{o,i}' F_o' M_F F_o \gamma_{o,i} - \frac{2}{NT}\sum_{i=1}^{N}b_i' M_F F_o \gamma_{o,i} \\
&= \frac{1}{NT}\sum_{i=1}^{N}b_i' M_F b_i + \frac{1}{NT}\text{tr}\left(M_F F_o \Gamma_o' \Gamma_o F_o' M_F\right) - \frac{2}{NT}\sum_{i=1}^{N}\text{tr}\left(M_F F_o \gamma_{o,i} b_i'\right) \\
&= \frac{1}{NT}b' A_1 b + \frac{1}{NT}\eta' A_2 \eta - \frac{2}{NT}\eta' \sum_{i=1}^{N}(\gamma_{o,i} \otimes I_T)\, b_i \\
&= \frac{1}{NT}b' A_1 b + \frac{1}{NT}\eta' A_2 \eta - \frac{2}{NT}b' A_3' \eta \\
&= \frac{1}{NT}b' A_1 b + \frac{1}{NT}\vartheta' A_2 \vartheta - \frac{1}{NT}b' A_3' A_2^{-1} A_3 b,
\end{aligned} \tag{11}
$$

where $\vartheta = \eta - A_2^{-1}A_3 b$, and the third equality follows from the fact that

$$
\text{tr}(B_1 B_2 B_3) = \text{vec}(B_1)'(B_2 \otimes I)\,\text{vec}(B_3') \text{ and } \text{tr}(B_1 B_2 B_3 B_4) = \text{vec}(B_1)'(B_2 \otimes B_4')\,\text{vec}(B_3')
$$

for any conformable matrices B_1, B_2, B_3, B_4 and an identity matrix I (see, e.g., Bernstein (2009, p. 253)). We now argue that the last term in (11) is $o_P(1)$ uniformly in b. Observe that

$(NT)^{-1}b'A_3'A_2^{-1}A_3b \leq \lambda_{\max}\left(A_3'A_2^{-1}A_3\right)(NT)^{-1}b'b = o_P(1)$ for any $(NT)^{-1}b'b = O(1)$ provided $\lambda_{\max}\left(A_3'A_2^{-1}A_3\right) = o_P(1)$. Note that $\lambda_{\max}\left(A_3'A_2^{-1}A_3\right) \leq [\lambda_{\min}(A_2/N)]^{-1}\lambda_{\max}\left(N^{-1}A_3'A_3\right) = \underline{c}_\Gamma^{-1}\lambda_{\max}\left(N^{-1}A_3'A_3\right)$ where $\underline{c}_\Gamma^{-1} \equiv [\lambda_{\min}(\Gamma_o'\Gamma_o/N)]^{-1} = O_P(1)$. Define the upper block-triangular matrix

$$C_1 = \begin{pmatrix} \gamma_{o,1}'\gamma_{o,1}I_T & \gamma_{o,1}'\gamma_{o,2}I_T & \cdots & \gamma_{o,1}'\gamma_{o,N}I_T \\ 0 & \gamma_{o,2}'\gamma_{o,2}I_T & \cdots & \gamma_{o,2}'\gamma_{o,N}I_T \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_{o,N}'\gamma_{o,N}I_T \end{pmatrix}.$$

Noting that the $NT \times NT$ matrix $A_3'A_3$ has a typical $T \times T$ block submatrix $T^{-1}\gamma_{o,i}'\gamma_{o,j}I_T$, we have $A_3'A_3 = C_1 + C_1' - C_d$ where $C_d = \text{diag}(\gamma_{o,1}'\gamma_{o,1}I_T, \ldots, \gamma_{o,N}'\gamma_{o,N}I_T)$. By the fact that the eigenvalues of a block upper/lower triangular matrix are the combined eigenvalues of its diagonal block matrices, Weyl's inequality, and Assumption 1.4, we have

$$\begin{aligned} \lambda_{\max}\left(N^{-1}A_3'A_3\right) &\leq N^{-1}\left\{2\lambda_{\max}(C_1) - \lambda_{\min}(C_d)\right\} \leq 2N^{-1}\max_{1\leq i\leq N}\|\gamma_{o,i}\|^2 \\ &= N^{-1}o_P(N^{1/2}) = o_P(1), \end{aligned} \tag{12}$$

where the first equality follows from the fact that $\max_{1\leq i\leq N}\|\gamma_{o,i}\|^2 = o_P(N^{1/2})$ under Assumption 1.4 by the Markov inequality. It follows that $\lambda_{\max}\left(A_3'A_2^{-1}A_3\right) = o_P(1)$ and $\frac{1}{NT}b'A_3'A_2^{-1}A_3b = o_P(1)$ uniformly in b. This, in conjunction with (10)-(11) and the fact that $S_{NT}(\hat\theta, \hat C, \hat F) - S_{NT}(\theta_o, C_o, F_o) \leq 0$, implies that

$$(NT)^{-1}\hat b'\hat A_1\hat b = (NT)^{-1}\sum_{i=1}^N \hat b_i'M_{\hat F}\hat b_i = o_P(1), \tag{13}$$

where $\hat A_1 = I_N \otimes M_{\hat F}$, $\hat b = (\hat b_1', \ldots, \hat b_N')'$, and $\hat b_i = \phi_i[\hat\theta, \hat g_k] - \phi_i[\theta_o, g_{o,k}]$.

By (10), (11), (13), and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} 0 &\geq \widetilde S_{NT}(\hat\theta, \hat C, \hat F) - \widetilde S_{NT}(\theta_o, C_o, F_o) \\ &= \frac{1}{NT}\hat b'\hat A_1\hat b + \frac{1}{NT}\text{tr}\left[\left(F_o'M_{\hat F}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right] - \frac{2}{NT}\sum_{i=1}^N \hat b_i'M_{\hat F}F_o\gamma_{o,i} + o_P(1) \\ &\geq \frac{1}{N}\hat b'\hat A_1\hat b + \frac{1}{NT}\text{tr}\left[\left(F_o'M_{\hat F}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right] \\ &\quad -2\left\{\frac{1}{N}\hat b'\hat A_1\hat b\right\}^{1/2}\left\{\frac{1}{NT}\text{tr}\left[\left(F_o'M_{\hat F}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right]\right\}^{1/2} + o_P(1) \\ &= o_P(1) + \frac{1}{NT}\text{tr}\left[\left(F_o'M_{\hat F}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right] - 2o_P(1)\left\{\frac{1}{NT}\text{tr}\left[\left(F_o'M_{\hat F}F_o\right)\left(\Gamma_o'\Gamma_o\right)\right]\right\}^{1/2}. \end{aligned}$$

It follows that $\frac{1}{NT}\text{tr}\left[\left(F_o' M_{\hat{F}} F_o\right)\left(\Gamma_o'\Gamma_o\right)\right] = o_P(1)$. As in Bai (2009, p.1265), this further implies that $\frac{1}{T}\text{tr}\left(F_o' M_{\hat{F}} F_o\right) = o_P(1)$, $\frac{1}{T}\hat{F}' F_o$ is invertible, and $\left\|P_{\hat{F}} - P_{F_o}\right\| = o_P(1)$.

(2). By (13) and noting that

$$\left|(NT)^{-1}\sum_{i=1}^{N}\hat{b}_i'\left(M_{F_o} - M_{\hat{F}}\right)\hat{b}_i\right| \leq \left\|P_{\hat{F}} - P_{F_o}\right\|(NT)^{-1}\sum_{i=1}^{N}\hat{b}_i'\hat{b}_i = o_P(1),$$

we have

$$o_P(1) = (NT)^{-1}\sum_{i=1}^{N}\hat{b}_i' M_{\hat{F}}\hat{b}_i = (NT)^{-1}\sum_{i=1}^{N}\hat{b}_i' M_{F_o}\hat{b}_i - o_P(1).$$

Then $(NT)^{-1}\sum_{i=1}^{N}\hat{b}_i' M_{F_o}\hat{b}_i = o_P(1)$. Noting $\hat{b}_i = \phi_i[\hat{\theta},\hat{g}_k] - \phi_i[\theta_o, g_{o,k}] = \Big(\phi_i[\hat{\theta},\hat{g}_k] - \phi_i[\theta_o, g_o]\Big) + \phi_i[\theta_o, \delta_{o,k}]$, we have

$$\begin{aligned}
o_P(1) &= (NT)^{-1}\sum_{i=1}^{N}\hat{b}_i' M_{F_o}\hat{b}_i \\
&= (NT)^{-1}\sum_{i=1}^{N}\Big(\phi_i[\hat{\theta},\hat{g}_k] - \phi_i[\theta_o, g_o]\Big)' M_{F_o}\Big(\phi_i[\hat{\theta},\hat{g}_k] - \phi_i[\theta_o, g_o]\Big) \\
&\quad + (NT)^{-1}\sum_{i=1}^{N}\Big(\phi_i[\hat{\theta},\hat{g}_k] - \phi_i[\theta_o, g_o]\Big)' M_{F_o}\phi_i[\theta_o, \delta_{o,k}] \\
&\quad + (NT)^{-1}\sum_{i=1}^{N}\left(\phi_i[\theta_o, \delta_{o,k}]\right)' M_{F_o}\left(\phi_i[\theta_o, \delta_{o,k}]\right) \\
&= (NT)^{-1}\sum_{i=1}^{N}\Big(\phi_i[\hat{\theta},\hat{g}_k] - \phi_i[\theta_o, g_o]\Big)' M_{F_o}\Big(\phi_i[\hat{\theta},\hat{g}_k] - \phi_i[\theta_o, g_o]\Big) + o_P(1) \\
&= L(\hat{\theta},\hat{g}_k) + o_P(1)
\end{aligned}$$

where the third equality follows from (6) and (8) of Lemma A4; and the fourth equality follows from (1) of Lemma A4. Consequently, $L(\hat{\theta},\hat{g}_k) = o_P(1)$.

Note that we have shown $L(\theta, g)$ is continuous with respect to $\|\cdot\|_{\widetilde{w}}$ on $\Theta \times \Im_2$ in the proof of (1) of Lemma A4, which indicates that, for $\forall(\theta, C)$ satisfying $\|(\theta, C) - (\theta_o, C_o)\| \geq \epsilon > 0$, we have $L(\theta, g_k)$ does not converge to zero for $g_k(\cdot) = C'\mathsf{H}(\cdot)$. In other words, if $\|(\hat{\theta},\hat{g}_k) - (\theta_o, g_o)\|_{\widetilde{w}} \not\to_P 0$, we have $L(\hat{\theta},\hat{g}_k) \neq o_P(1)$. See Bai (2009, p. 1265) and Newey and Powell (2003, p. 1576) for a similar argument. Therefore, we have proved $\|(\hat{\theta},\hat{g}_k) - (\theta_o, g_o)\|_{\widetilde{w}} \to_P 0$.

The proof is now complete. ■

Proof of Lemma 4.1:

The proof is similar to that given for Lemma 3.1, thus is omitted. ■

References

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* **77**(4), 1229–1279.

Bernstein, D. S. (2009), *Matrix Mathematics: Theory, Facts, and Formulas*, Princeton University Press.

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

Jennrich, R. I. (1969), 'Asymptotic properties of non-linear least squares estimators', *Annals of Mathematical Statistics* **40**, 633–643.

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

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