Online Supplementary Appendices to "Varying–Coefficient Panel Data Models with Nonstationarity and Partially Observed Factor Structure"

CHAOHUA DONG*, JITI GAO⋆ AND BIN PENG†

*Zhongnan University of Economics and Law

⋆Monash University and †Deakin University

This online supplementary file includes four sections: Appendix A includes preliminary lemmas and their proofs of the direct estimation method; Appendix B presents the PCA–based approach, and states the correspondingly asymptotic properties with their proofs; Appendix C compares both methods, and provides extensive numerical studies; Appendix D further discusses some extensions.

Before proceeding future, recall that S_β, S_{β_ℓ} and S_{γ_ℓ} are selection matrices, and have been defined in the main text. We further define some variables which will be repeatedly used throughout this file. Let $\widetilde{\Psi}_{NT} = \mathrm{diag}\{\frac{1}{\sqrt{NT^2}}I_{md_x}, \frac{1}{\sqrt{NT}}I_{nd_v}\}$, $\Psi_{NT} = \mathrm{diag}\left\{\frac{1}{\sqrt{NT}}I_{md_x}, \frac{1}{\sqrt{N}}I_{nd_v}\right\}$, and $\Psi_T = \mathrm{diag}\{\frac{1}{\sqrt{T}}I_{md_x}, I_{nd_v}\}$. $O(1)$ stands for a constant, and may be different at each appearance.

Appendix A

A.1 Preliminary Lemmas

Lemma A.1. *Let Assumptions 1–4 hold. As $(N,T)\to(\infty,\infty)$, the following results hold:*

1. $\left\|\frac{1}{NT}\sum_i [H_m(r_{it})H_m'(r_{it})]\otimes[x_{it}x_{it}'] - \tau_t\Sigma_m\right\| = o_P(1)$,
2. $\left\|\frac{1}{N}\sum_i \mathcal{H}_n(v_i)\mathcal{H}_n'(v_i) - \Sigma_n\right\| = o_P(1)$,
3. $\left\|\frac{1}{NT^{1/2}}\sum_i [H_m(r_{it})\otimes x_{it}]\mathcal{H}_n'(v_i)\right\| = o_P(1)$,
4. $\left\|(\sum_i Q_{it}Q_{it}')^{-1}\sum_i Q_{it}f_{0t}'\Delta_{\gamma_0}(v_i)\right\| = O_P(\tau_t^{-\frac{1}{2}}n^{-\mu_1})$,
5. $\left\|(\sum_i Q_{it}Q_{it}')^{-1}\sum_i Q_{it}x_{it}'\Delta_{\beta_0}(r_{it})\right\| = O_P(\sqrt{T}m^{-\mu_2})$,

where $\tau_t = t/T$.

Lemma A.2. *Let Assumptions 1–4 hold. As $(N,T)\to(\infty,\infty)$, the following results hold:*

1. $\left\|\frac{1}{NT^2}\sum_{i,t}[H_m(r_{it})H_m'(r_{it})]\otimes[x_{it}x_{it}'] - \frac{1}{2}\Sigma_m\right\| = o_P(1)$,
2. $\left\|\frac{1}{NT^{3/2}}\sum_{i,t}[H_m(r_{it})\otimes x_{it}]f_{0t}'\mathscr{H}_n'(v_i)\right\| = o_P(1)$,
3. $\left\|\frac{1}{NT^{3/2}}\sum_{i,t}[H_m(r_{it})\otimes x_{it}]e_{it}\right\| = O_P\left(\sqrt{\frac{m}{NT}}\right)$,
4. $\left\|\frac{1}{NT}\sum_{i,t}\mathscr{H}_n(v_i)f_{0t}e_{it}\right\| = O_P\left(\sqrt{\frac{n}{NT}}\right)$,
5. $\left\|\left(\sum_{i,t}Q_{it}Q_{it}'\right)^{-1}\sum_{i,t}Q_{it}f_{0t}'\Delta_{\gamma_0}(v_i)\right\| = O_P(n^{-\mu_1})$,
6. $\left\|\left(\sum_{i,t}Q_{it}Q_{it}'\right)^{-1}\sum_{i,t}Q_{it}x_{it}'\Delta_{\beta_0}(r_{it})\right\| = O_P(\sqrt{T}m^{-\mu_2})$

A.2 Proofs of the Direct Estimation Method

Before we prove Lemmas A.1 and A.2, we would like to make some remarks about the approach used in the derivations. In the proof of Lemma A.1, the mutual independence between x_{it} and r_{it} is fully employed. In addition, the fact that the assumption on cross–sectional independence simplifies the detailed derivations.

In the proof of Lemma A.2, we are able to fully explore the cross–sectional independence on x_{it} to be able to establish results for convergence in probability and convergence in moments. An intuitive explanation is that we take the average over both i and t, so after some tedious algebra we can evaluate the moments involved. Without the availability of the cross–sectional independence, as in the pure integrated time series case, we would only be able to establish results for convergence in distribution.

Poof of Lemma A.1:

(1). Write

$$\begin{aligned}E\left[\frac{1}{NT}\sum_i [H_m(r_{it})H_m'(r_{it})]\otimes[x_{it}x_{it}']\right] &= \frac{1}{N}\sum_i E[H_m(r_{it})H_m'(r_{it})]\otimes\frac{E\left[x_{it}x_{it}'\right]}{T}\\ &= E[H_m(r_{it})H_m'(r_{it})]\otimes\left[DD'\tau_t\right]\cdot(1+o(1)),\end{aligned}$$

where the second equality follows from a straightforward calculation (e.g., Dong and Linton, 2018). Below, we consider the second moment. Recall that we have denoted $H_{it} = H_m(r_{it})H_m'(r_{it})$ and $X_{it} = x_{it}x_{it}'$, and let H_{it,l_1l_2} and X_{it,k_1k_2} stand for the $(l_1,l_2)^{th}$ and $(k_1,k_2)^{th}$ elements of H_{it} and X_{it}, respectively. Then we are able to write

$$\begin{aligned}& E\left\|\frac{1}{NT}\sum_i H_{it}\otimes X_{it} - \frac{1}{NT}\sum_i E\left[H_{it}\otimes X_{it}\right]\right\|^2\\ =& \frac{1}{N^2T^2}\sum_i E\left\|H_{it}\otimes X_{it} - E[H_{it}]\otimes E[X_{it}]\right\|^2\\ &+\frac{1}{N^2T^2}\sum_{i\neq j}\sum_{l_1,l_2=1}^{m}\sum_{k_1,k_2=1}^{d_x}\\ &\quad E\left[\left(H_{it,l_1l_2}X_{it,k_1k_2} - E[H_{it,l_1l_2}]E[X_{it,k_1k_2}]\right)\left(H_{jt,l_1l_2}X_{jt,k_1k_2} - E[H_{jt,l_1l_2}]E[X_{jt,k_1k_2}]\right)\right]\\ \le& O(1)\frac{1}{N^2T^2}\sum_i E\left\|H_{it}\otimes X_{it}\right\|^2\\ &+\frac{1}{N^2T^2}\sum_{i\neq j}\sum_{l_1,l_2=1}^{m}\sum_{k_1,k_2=1}^{d_x} E[X_{it,k_1k_2}]E[X_{jt,k_1k_2}]E\left[\left(H_{it,l_1l_2} - E[H_{it,l_1l_2}]\right)\left(H_{jt,l_1l_2} - E[H_{jt,l_1l_2}]\right)\right]\\ =& O(1)\frac{m^2}{N},\end{aligned}$$

where the last equality follows from Assumption 2.2. Thus, the first result of this lemma follows.

(2). Similar to the proof of (1), the second result follows immediately.

(3). Write

$$\left\|\frac{1}{NT^{1/2}}\sum_i [H_m(r_{it})\otimes x_{it}]\mathcal{H}_n'(v_i)\right\|$$

$$= \left\| \frac{1}{N} \sum_i \left[H_m(r_{it}) \otimes \frac{\sum_{s=1}^t w_{is}}{\sqrt{T}} \right] \mathcal{H}_n'(v_i) \right\| + o_P(1) = o_P(1),$$

where the last equality follows from a procedure similar to (1) of this lemma using $E[\sum_{s=1}^t w_{is}] = 0$, and w_{it} being independent of r_{it} and v_i.

(4). Let $Q_t = (Q_{1t}, \ldots, Q_{Nt})'$ and $\Delta_t = (f_{0t}'\Delta_{\gamma_0}(v_1), \ldots, f_{0t}'\Delta_{\gamma_0}(v_N))'$ for notational simplicity. Note that by Assumption 2.3, we have

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^N E\|\Delta_{\gamma_0}(v_i)\|^2 &= \sum_{\ell=1}^{d_v} \int \Delta_{\gamma_{0\ell}}^2(w) f_{v_\ell}(w) dw = \sum_{\ell=1}^{d_v} \int \Delta_{\gamma_{0\ell}}^2(w)\pi(w) \frac{f_{v_\ell}(w)}{\pi(w)} dw \\ &\le O(1) \sum_{\ell=1}^{d_v} \int \Delta_{\gamma_{0\ell}}^2(w)\pi(w) dw = O(n^{-2\mu_1}), \end{aligned}$$

where $\Delta_{\gamma_{0\ell}}(w)$ stands for the ℓ^{th} element of $\Delta_{\gamma_0}(w)$, and the first inequality and the last equality follow from Assumption 2.3.

Then write

$$\begin{aligned} &\left\| (Q_t'Q_t)^{-1} Q_t'\Delta_t \right\|^2 = \Delta_t' Q_t \Psi_{NT} \left(\Psi_{NT} Q_t'Q_t \Psi_{NT} \right)^{-1} \Psi_{NT} \left(Q_t'Q_t \right)^{-1} Q_t'\Delta_t \\ &\le \lambda_{\max} \left\{ \Psi_{NT} \left(\Psi_{NT} Q_t'Q_t \Psi_{NT} \right)^{-1} \Psi_{NT} \right\} \cdot \Delta_t' Q_t \left(Q_t'Q_t \right)^{-1} Q_t'\Delta_t \\ &\le \lambda_{\max}\{\Psi_{NT}\} \cdot \lambda_{\min}^{-1} \left\{ \Psi_{NT} Q_t'Q_t \Psi_{NT} \right\} \cdot \lambda_{\max}\{\Psi_{NT}\} \cdot \Delta_t' Q_t \left(Q_t'Q_t \right)^{-1} Q_t'\Delta_t \\ &\le \lambda_{\min}^{-1} \left\{ \Psi_{NT} Q_t'Q_t \Psi_{NT} \right\} \cdot \lambda_{\max}(Q_t \left(Q_t'Q_t \right)^{-1} Q_t') \cdot \|\Delta_t\|^2 / N \\ &\le \lambda_{\min}^{-1} \left\{ \Psi_{NT} Q_t'Q_t \Psi_{NT} \right\} \cdot O_P(n^{-2\mu_1}) \end{aligned} \tag{A.1}$$

where the first inequality follows from the exercise 5 on page 267 of Magnus and Neudecker (2007), and the last step follows from $\|\Delta_t\|^2 / N = O_P(n^{-2\mu_1})$ by Assumption 2.3. Note that $\lambda_{\min}^{-1} \left\{ \Psi_{NT} Q_t'Q_t \Psi_{NT} \right\} = O_P(\tau_t^{-1})$ by the first two results of this lemma. Thus, the result follows immediately.

(5). By Assumptions 2.1 and 2.3, simple calculation gives $\frac{1}{N}\sum_i |x_{it}'\Delta_{\beta_0}(r_{it})|^2 = O_P(t \cdot m^{-2\mu_2})$. Then by the same procedure as in (A.1), the result follows. ∎

Proof of Lemma A.2:

(1)–(2). The first two results follow by procedures similar to (1) and (3) of Lemma A.1, but accounting for cross–sectional and time dimensions simultaneously. The number $\frac{1}{2}$ of the first result comes from $\frac{1}{T}\sum_{t=1}^T \tau_t \to \int_0^1 w\, dw = \frac{1}{2}$.

(3). Write

$$\begin{aligned} &E \left\| \frac{1}{NT^{3/2}} \sum_{i,t} [H_m(r_{it}) \otimes x_{it}] e_{it} \right\|^2 \\ &= \frac{1}{N^2T^3} \sum_{i,j=1}^N \sum_{t=1}^T E\big[e_{it} e_{jt} [H_m(r_{it}) \otimes x_{it}]' [H_m(r_{jt}) \otimes x_{jt}] \big] \\ &= O(1) \frac{m}{N^2T^3} \sum_{i,j=1}^N \sum_{t=1}^T \sigma_{ij} E[\|x_{i_1t}\| \cdot \|x_{i_2t}\|] = O(1) \frac{m}{NT}, \end{aligned} \tag{A.2}$$

where the second equality follows from Assumptions 2.3 and 3; the third equality follows from the fact that $E\|x_{it}\|^2/T = O(1)$, and Assumption 3. Thus, the result follows.

(4). Similar to (3), the result follows.

(5)–(6). These two results can be proved in exactly the same way as (A.1). Thus, the details are omitted. ■

Proof of Lemma 2.1:

By Lemma A.1 and Lemma A.2, we obtain that

$$\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Psi_T Q_{it} Q_{it} \Psi_T - \mathbb{Q} \right\| = o_P(1)$$

where $\mathbb{Q} = \mathrm{diag}\left\{\frac{1}{2}\Sigma_m, \Sigma_n\right\}$. Then (1) and (3) of this lemma are obvious. In addition, we provide a more generalized version of proof for (1) and (3) later when deriving Lemma 2.2, so we omit the details here.

We now take a look at the result (2). Note that the next equation always holds.

$$\begin{aligned}\|\widetilde{C}_{\gamma_\ell,t}\|^2 - f_{0t,\ell}^2 &= \|f_{0t,\ell} C_{\gamma_{0\ell}} + S_{\gamma_\ell} U_{Nt} + S_{\gamma_\ell} \Delta_{Nt}\|^2 - f_{0t,\ell}^2 \\ &= f_{0t,\ell}^2 \|\gamma_{0\ell}\|_{L^2}^2 - f_{0t,\ell}^2 + \text{residuals},\end{aligned}$$

where

$$\begin{aligned}U_{Nt} &= \left(\sum_{i=1}^{N} Q_{it} Q_{it}'\right)^{-1} \sum_{i=1}^{N} Q_{it} e_{it}, \\ \Delta_{Nt} &= \left(\sum_{i=1}^{N} Q_{it} Q_{it}'\right)^{-1} \sum_{i=1}^{N} Q_{it} (x_{it}' \Delta_{\beta_0}(r_{it}) + f_{0t}' \Delta_{\gamma_0}(v_i)).\end{aligned}$$

The result follows from Lemma A.1. The proof is now complete. ■

Proof of Theorem 2.1:

(1). Write

$$\begin{aligned}&\frac{\sqrt{NT^2}}{\|H_m(r)\|}(\widetilde{\beta}_m(r) - \beta_0(r)) \\ = \ &\frac{\sqrt{NT^2}}{\|H_m(r)\|}[H_m'(r) \otimes I_{d_x}](\widetilde{C}_\beta - C_{\beta_0}) + \frac{\sqrt{NT^2}}{\|H_m(r)\|}\Delta_{\beta_0}(r) \\ = \ &\frac{\sqrt{NT^2}}{\|H_m(r)\|}[H_m'(r) \otimes I_{d_x}] S_\beta \Psi_T \mathbb{Q}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Psi_T Q_{it} e_{it} + o_P(1) \\ = \ &\frac{1}{\sqrt{NT\|H_m(r)\|^2}}[H_m'(r) \otimes I_{d_x}] S_\beta \mathbb{Q}^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \Psi_T Q_{it} e_{it} + o_P(1) \\ \to_D \ &N(0, \widetilde{\Sigma}_\beta)\end{aligned}$$

where the second equality follows from $\frac{NT^3}{m^{2\mu_2}} \to 0$, $\frac{NT^2}{n^{2\mu_1}} \to 0$, Lemma A.1 and Lemma A.2; and the last step follows by verifying Lemma B.1 of Chen et al. (2012*b*), and the value of $\widetilde{\Sigma}_\beta$ follows from

$$\widetilde{\Sigma}_\beta = \lim_{N,T,m} \frac{1}{NT\|H_m(r)\|^2} \sum_{t=1}^{T} \sum_{i=1}^{N} [H_m'(r) \otimes I_{d_x}] S_\beta \sigma_e^2 \mathbb{Q}^{-1} E[\Psi_T Q_{it} Q_{it}' \Psi_T] \mathbb{Q}^{-1} S_\beta' [H_m(r) \otimes I_{d_x}]$$
$$+ \lim_{N,T,m} \frac{1}{NT\|H_m(r)\|^2} \sum_{t=1}^{T} \sum_{i\neq j} [H_m'(r) \otimes I_{d_x}] S_\beta \sigma_{ij} \mathbb{Q}^{-1} E[\Psi_T Q_{it} Q_{jt}' \Psi_T] \mathbb{Q}^{-1} S_\beta' [H_m(r) \otimes I_{d_x}]$$
$$= \lim_m \frac{\sigma_e^2}{\|H_m(r)\|^2} [H_m'(r) \otimes I_{d_x}] S_\beta \mathbb{Q}^{-1} S_\beta' [H_m(r) \otimes I_{d_x}]$$
$$= \lim_m \frac{2\sigma_e^2}{\|H_m(r)\|^2} [H_m'(r) \otimes I_{d_x}] \{E[H_m(r_{it}) H_m'(r_{it})] \otimes (DD')\}^{-1} [H_m(r) \otimes I_{d_x}]$$

using Lemma A.1 and Lemma A.2.

(2). Write

$$\sqrt{N}(\widetilde{f}_{t,\ell} - f_{0t,\ell}) = \frac{\sqrt{N}}{\widetilde{f}_{t,\ell} + f_{0t,\ell}} (\|\widetilde{C}_{\gamma_\ell,t}\|^2 - f_{0t,\ell}^2)$$
$$= \frac{\sqrt{N}}{\widetilde{f}_{t,\ell} + f_{0t,\ell}} (\|f_{0t,\ell} C_{\gamma_{0\ell}} + S_{\gamma_\ell} U_{Nt} + S_{\gamma_\ell} \Delta_{Nt}\|^2 - f_{0t,\ell}^2)$$
$$= \frac{\sqrt{N}}{\widetilde{f}_{t,\ell} + f_{0t,\ell}} (\|f_{0t,\ell} C_{\gamma_{0\ell}}\|^2 + 2 f_{0t,\ell} C_{\gamma_{0\ell}}' S_{\gamma_\ell} U_{Nt} - f_{0t,\ell}^2) + o_P(1)$$
$$= \frac{\sqrt{N}}{\widetilde{f}_{t,\ell} + f_{0t,\ell}} 2 f_{0t,\ell} C_{\gamma_{0\ell}}' S_{\gamma_\ell} \Psi_T \mathbb{Q}_t^{-1} \frac{1}{N} \sum_{i=1}^{N} \Psi_T Q_{it} e_{it} + o_P(1)$$
$$= C_{\gamma_{0\ell}}' S_{\gamma_\ell} \mathbb{Q}_t^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Psi_T Q_{it} e_{it} + o_P(1) \to_D N(0, \sigma_{f_\ell}^2),$$

where $\mathbb{Q}_t = \mathrm{diag}\{\tau_t \Sigma_m, \Sigma_n\}$; the third equality is due to the fact that all the other terms are negligible by Lemma A.1; and the last step follows from the same procedure as Lemma A.1 of Chen et al. (2012*a*), and the value $\sigma_{f_\ell}^2$ follows from

$$\sigma_{f_\ell}^2 = \lim_{N,n} \frac{1}{N} \sum_{i=1}^{N} C_{\gamma_{0\ell}}' S_{\gamma_\ell} \sigma_e^2 \mathbb{Q}_t^{-1} E[\Psi_T Q_{it} Q_{it}' \Psi_T] \mathbb{Q}_t^{-1} S_{\gamma_\ell}' C_{\gamma_{0\ell}}$$
$$+ \lim_{N,n} \frac{1}{N} \sum_{i\neq j} C_{\gamma_{0\ell}}' S_{\gamma_\ell} \sigma_{ij} \mathbb{Q}_t^{-1} E[\Psi_T Q_{it} Q_{jt}' \Psi_T] \mathbb{Q}_t^{-1} S_{\gamma_\ell}' C_{\gamma_{0\ell}}$$
$$= \lim_n \sigma_e^2 C_{\gamma_{0\ell}}' S_{\gamma_\ell} \mathbb{Q}_t^{-1} S_{\gamma_\ell}' C_{\gamma_{0\ell}} + \lim_{N,n} \frac{\sigma_{ij}}{N} \sum_{i\neq j} C_{\gamma_{0\ell}}' S_{\gamma_\ell} \mathbb{Q}_t^{-1} E[\Psi_T Q_{it} Q_{jt}' \Psi_T] \mathbb{Q}_t^{-1} S_{\gamma_\ell}' C_{\gamma_{0\ell}}$$
$$= \lim_n \sigma_e^2 C_{\gamma_{0\ell}}' S_\ell \Sigma_n^{-1} S_\ell' C_{\gamma_{0\ell}} + \lim_{N,n} \frac{1}{N} \sum_{i\neq j} \sigma_{ij} C_{\gamma_{0\ell}}' S_\ell \Sigma_n^{-1} E[\mathcal{H}_n(v_i) \mathcal{H}_n'(v_j)] \Sigma_n^{-1} S_\ell' C_{\gamma_{0\ell}}.$$

(3). We now turn to the asymptotic distribution associated with $\widetilde{\gamma}_\ell(w)$.

$$\frac{\sqrt{NT}}{\|H_n(w)\|} (\widetilde{\gamma}_\ell(w) - \gamma_{0\ell}(w)) = \frac{\sqrt{NT}}{\|H_n(w)\|} H_n'(w) (\widetilde{C}_{\gamma_\ell} - C_{\gamma_{0\ell}}) + o_P(1)$$
$$= \frac{\sqrt{NT}}{\|H_n(w)\|} H_n'(w) \frac{1}{\|S_{\gamma_\ell} \widetilde{\mathbb{C}}\|} (S_{\gamma_\ell} \widetilde{\mathbb{C}} - \bar{f}_{0,\ell} C_{\gamma_{0\ell}}) + \frac{\sqrt{NT}}{\|H_n(w)\|} \frac{\bar{f}_{0,\ell} - \|S_{\gamma_\ell} \widetilde{\mathbb{C}}\|}{\|S_{\gamma_\ell} \widetilde{\mathbb{C}}\|} H_n'(w) C_{\gamma_{0\ell}} + o_P(1)$$
$$= \frac{1}{|f_\ell^*| \cdot \|H_n(w)\| \sqrt{NT}} H_n'(w) S_{\gamma_\ell} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{Q}^{-1} \Psi_T Q_{it} e_{it} + o_P(1) \to_D N(0, \widetilde{\sigma}_{\gamma_\ell}^2),$$

where $\bar{f}_{0,\ell} = \frac{1}{T} \sum_{t=1}^{T} f_{0t,\ell}$; in the second equality we utilize the identification condition; the third equality follows from $\bar{f}_{0,\ell} - \|S_{\gamma_\ell}\widetilde{\mathbb{C}}\| = O_P(1/\sqrt{NT})$ that can be derived by a procedure similar to those for (2) of this lemma; and the rest steps are similar to (1) of this theorem, and the value of $\widetilde{\sigma}^2_{\gamma_\ell}$ follows from

$$\begin{aligned}
\widetilde{\sigma}^2_{\gamma_\ell} &= \lim_{N,T,n} \frac{1}{NT\|H_n(w)\|^2|f^*_\ell|^2} \sum_{t=1}^{T}\sum_{i=1}^{N} H_n'(w)S_{\gamma_\ell}\sigma_e^2\mathbb{Q}^{-1}E[\Psi_T Q_{it}Q_{it}'\Psi_T]\mathbb{Q}^{-1}S_{\gamma_\ell}'H_n(w) \\
&\quad + \lim_{N,T,n} \frac{1}{NT\|H_n(w)\|^2|f^*_\ell|^2} \sum_{t=1}^{T}\sum_{i\neq j} H_n'(w)S_{\gamma_\ell}\sigma_{ij}\mathbb{Q}^{-1}E[\Psi_T Q_{it}Q_{jt}'\Psi_T]\mathbb{Q}^{-1}S_{\gamma_\ell}'H_n(w) \\
&= \lim_{n} \frac{\sigma_e^2}{\|H_n(w)\|^2|f^*_\ell|^2} H_n'(w)S_\ell\Sigma_n^{-1}S_\ell'H_n(w) \\
&\quad + \lim_{N,n} \frac{1}{N\|H_n(w)\|^2|f^*_\ell|^2} \sum_{i\neq j}\sigma_{ij}H_n'(w)S_\ell\Sigma_n^{-1}E[\mathcal{H}_n(v_i)\mathcal{H}_n'(v_j)]\Sigma_n^{-1}S_\ell'H_n(w).
\end{aligned}$$

The proof is now complete. ■

Proof of Lemma 2.2:

(1). Note that by condition $1 \le d_x^* + d_v^* < d_x + d_v$, we allow for $d_x^* = 0$ or $d_x^* = d_x$. Similarly, $d_v^* = 0$ or $d_v^* = d_v$ is allowed. Without loss of generality, we assume that $1 \le d_x^* < d_x$ and $1 \le d_v^* < d_v$ in the following proof for notational simplicity.

We consider $\mathbb{C} = \mathbb{C}_0 + U$, where $\mathbb{C}_0 = (C_{\beta_0}', \bar{f}_{0,1}C_{\gamma_{01}}', \ldots, \bar{f}_{0,d_v}C_{\gamma_{0d_v}}')'$, $\bar{f}_{0,\ell} = \frac{1}{T}\sum_{t=1}^T f_{0t,\ell}$, $S_{\beta_\ell}\mathbb{C} = S_{\beta_\ell}\mathbb{C}_0 + U_m\frac{1}{\sqrt{NT^2}}$, $S_{\gamma_\ell}\mathbb{C} = S_{\gamma_\ell}\mathbb{C}_0 + U_n\frac{1}{\sqrt{NT}}$, and $\|U_m\| = b\sqrt{m}$ and $\|U_n\| = b\sqrt{n}$, and b is a large positive constant. Obviously, U is made of $U_m\frac{1}{\sqrt{NT^2}}$ and $U_n\frac{1}{\sqrt{NT}}$. We show that for any given $\epsilon > 0$, there exists a large constant b such that

$$\liminf_{N,T} \Pr\left\{\inf_{\|U_m\|=b\sqrt{m},\ \|U_n\|=b\sqrt{n}} \Upsilon_{NT}(\mathbb{C}) > \Upsilon_{NT}(\mathbb{C}_0)\right\} \ge 1-\epsilon, \tag{A.3}$$

which implies with a probability of at least $1-\epsilon$ that there exists a local minimum satisfying that $\|S_{\beta_\ell}(\widehat{\mathbb{C}} - \mathbb{C}_0)\| \le b\sqrt{\frac{m}{NT^2}}$ and $\|S_{\gamma_\ell}(\widehat{\mathbb{C}} - \mathbb{C}_0)\| = b\sqrt{\frac{n}{NT}}$. The above argument is in the same spirit as in the proof for Lemma A.1 of Wang and Xia (2009), wherein a kernel version is studied under the i.i.d. assumption.

$$\begin{aligned}
&\Upsilon_{NT}(\mathbb{C}) - \Upsilon_{NT}(\mathbb{C}_0) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - Q_{it}'\mathbb{C}_0 - Q_{it}'U)^2 + \sum_{\ell=1}^{d_x^*}\rho_{\beta_\ell}\left\|S_{\beta_\ell}\mathbb{C}_0 + U_m\frac{1}{\sqrt{NT^2}}\right\| + \sum_{\ell=1}^{d_v^*}\rho_{\gamma_\ell}\left\|S_{\gamma_\ell}\mathbb{C}_0 + U_n\frac{1}{\sqrt{NT}}\right\| \\
&\quad + \sum_{\ell=d_x^*+1}^{d_x}\frac{\rho_{\beta_\ell}}{\sqrt{NT^2}}\|U_m\| + \sum_{\ell=d_v^*+1}^{d_v}\frac{\rho_{\gamma_\ell}}{\sqrt{NT}}\|U_n\| \\
&\quad - \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - Q_{it}'\mathbb{C}_0)^2 - \sum_{\ell=1}^{d_x^*}\rho_{\beta_\ell}\|S_{\beta_\ell}\mathbb{C}_0\| - \sum_{\ell=1}^{d_v^*}\rho_{\gamma_\ell}\|S_{\gamma_\ell}\mathbb{C}_0\| \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}(Q_{it}'U)^2 - 2\sum_{i=1}^{N}\sum_{t=1}^{T}Q_{it}'U(y_{it} - Q_{it}'\mathbb{C}_0) + \sum_{\ell=d_x^*+1}^{d_x}\frac{\rho_{\beta_\ell}}{\sqrt{NT^2}}\|U_m\| + \sum_{\ell=d_v^*+1}^{d_v}\frac{\rho_{\gamma_\ell}}{\sqrt{NT}}\|U_n\| \\
&\quad + \sum_{\ell=1}^{d_x^*}\rho_{\beta_\ell}\left(\left\|S_{\beta_\ell}\mathbb{C}_0 + U_m\frac{1}{\sqrt{NT^2}}\right\| - \|S_{\beta_\ell}\mathbb{C}_0\|\right) + \sum_{\ell=1}^{d_v^*}\rho_{\gamma_\ell}\left(\left\|S_{\gamma_\ell}\mathbb{C}_0 + U_n\frac{1}{\sqrt{NT}}\right\| - \|S_{\gamma_\ell}\mathbb{C}_0\|\right) \\
&\ge \sum_{i=1}^{N}\sum_{t=1}^{T}U'\widetilde{\Psi}_{NT}^{-1}\widetilde{\Psi}_{NT}Q_{it}Q_{it}'\widetilde{\Psi}_{NT}\widetilde{\Psi}_{NT}^{-1}U - 2\sum_{i=1}^{N}\sum_{t=1}^{T}Q_{it}'\widetilde{\Psi}_{NT}\widetilde{\Psi}_{NT}^{-1}U(\Delta_{it} + e_{it})
\end{aligned}$$

$$+\sum_{\ell=1}^{d_x^*}\rho_{\beta_\ell}\left(\left\|S_{\beta_\ell}\mathbb{C}_0+U_m\frac{1}{\sqrt{NT^2}}\right\|-\|S_{\beta_\ell}\mathbb{C}_0\|\right)+\sum_{\ell=1}^{d_v^*}\rho_{\gamma_\ell}\left(\left\|S_{\gamma_\ell}\mathbb{C}_0+U_n\frac{1}{\sqrt{NT}}\right\|-\|S_{\gamma_\ell}\mathbb{C}_0\|\right)$$
$$\geq O(1)\|U'\widetilde{\Psi}_{NT}^{-1}\|^2-A_{NT}\widetilde{\Psi}_{NT}^{-1}U-\sum_{\ell=1}^{d_x^*}\rho_{\beta_\ell}\left\|U_m\frac{1}{\sqrt{NT^2}}\right\|\cdot\|C_{\beta_\ell}^*\|-\sum_{\ell=1}^{d_v^*}\rho_{\gamma_\ell}\left\|U_n\frac{1}{\sqrt{NT}}\right\|\cdot\|C_{\gamma_\ell}^*\|, \quad \text{(A.4)}$$

where $\Delta_{it}=x_{it}'\Delta_{\beta_0}(r_{it})+f_{0t}'\Delta_{\gamma_0}(v_i)$, $A_{NT}=2\sum_{i=1}^N\sum_{t=1}^T Q_{it}'\widetilde{\Psi}_{NT}(\Delta_{it}+e_{it})$, $C_{\beta_\ell}^*$ lies between $S_{\beta_\ell}\mathbb{C}_0$ and $S_{\beta_\ell}\mathbb{C}_0+U_m\frac{1}{\sqrt{NT^2}}$, and $C_{\gamma_\ell}^*$ lies between $S_{\gamma_\ell}\mathbb{C}_0$ and $S_{\gamma_\ell}\mathbb{C}_0+U_n\frac{1}{\sqrt{NT}}$.

Note $\|A_{NT}\|=O_P(\sqrt{m+n})$ by Lemma A.2, and by construction, $\|U'\widetilde{\Psi}_{NT}^{-1}\|=O(\sqrt{m+n})$. In connection with that Assumption 5.1 and the right hand side of (A.4) is in quadratic form of U, we know that $\Upsilon_{NT}(\mathbb{C})-\Upsilon_{NT}(\mathbb{C}_0)\geq 0$ with probability approaching one. Thus, (A.3) holds. The rest proofs are straightforward, so omitted.

(2). For simplicity, we show that $\Pr(\|S_{\beta_\ell}\overline{\mathbb{C}}\|=0)\to 1$ with $\forall\ell\in\mathcal{A}_\beta^\dagger$ only. The proofs for $S_{\gamma_\ell}\overline{\mathbb{C}}$ with $\forall\ell\in\mathcal{A}_\gamma^\dagger$ are the same. If $\|S_{\beta_\ell}\overline{\mathbb{C}}\|\neq 0$, $\overline{\mathbb{C}}$ must satisfy the following equation

$$0=\frac{\partial\Upsilon_{NT}(\overline{\mathbb{C}})}{\partial S_{\beta_\ell}\mathbb{C}}=-2A_1+A_2, \quad \text{(A.5)}$$

where $A_1=\sum_{i=1}^N\sum_{t=1}^T H_m(r_{it})x_{it,\ell}(y_{it}-Q_{it}'\overline{\mathbb{C}})$ and $A_2=\frac{\rho_{\beta_\ell}}{\|S_{\beta_\ell}\overline{\mathbb{C}}\|}S_{\beta_\ell}\overline{\mathbb{C}}$.

For A_1, we write

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T H_m(r_{it})\frac{x_{it,\ell}}{\sqrt{T}}(y_{it}-Q_{it}'\overline{\mathbb{C}})$$
$$=\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T H_m(r_{it})\frac{x_{it,\ell}}{\sqrt{T}}e_{it}+\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T H_m(r_{it})\frac{x_{it,\ell}}{\sqrt{T}}Q_{it}'(\mathbb{C}_0-\overline{\mathbb{C}}),$$

where it is easy to know that $\left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T H_m(r_{it})\frac{x_{it,\ell}}{\sqrt{T}}e_{it}\right\|=O_P\left(\sqrt{\frac{m}{NT}}\right)$ by Lemma A.2. Thus, we focus on the rest term and write

$$\left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T H_m(r_{it})\frac{x_{it,\ell}}{\sqrt{T}}Q_{it}'(\mathbb{C}_0-\overline{\mathbb{C}})\right\|$$
$$\leq\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{\|H_m(r_{it})\|^2\|x_{it,\ell}\|^2}{T}\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\mathbb{C}_0-\overline{\mathbb{C}})'\Psi_T^{-1}\Psi_TQ_{it}Q_{it}'\Psi_T\Psi_T^{-1}(\mathbb{C}_0-\overline{\mathbb{C}})\right\}^{1/2}$$
$$\leq O_P(\sqrt{m}\|\Psi_T^{-1}(\mathbb{C}_0-\overline{\mathbb{C}})\|)=O_P\left(\sqrt{\frac{m(m+n)}{NT}}\right),$$

where the second inequality follows from Lemma A.2 and the first result of this lemma. Therefore, we conclude that $\|A_1\|=O_P\left(\sqrt{\frac{m(m+n)}{NT}}\right)$.

On the other hand,

$$\left\|\frac{1}{NT^{3/2}}A_2\right\|\geq\frac{1}{NT^{3/2}}\rho_\beta^\dagger\geq a_0\sqrt{\frac{m(m+n)}{NT}}$$

by Assumption 5. Therefore, $\Pr(\|A_1\|<\|A_2\|)\to 1$, which implies that, with a probability tending to 1, (A.5) does not hold. The above analysis implies that $S_{\beta_\ell}\overline{\mathbb{C}}$ must be located at a place where the objective function is not differentiable with respect to $S_{\beta_\ell}\mathbb{C}$. Since the objective function is not differentiable with respect to $S_{\beta_\ell}\mathbb{C}$ only at the origin, we immediately obtain that $\Pr(\|S_{\beta_\ell}\overline{\mathbb{C}}\| = 0) \to 1$ with $\forall \ell \in \mathcal{A}_\beta^\dagger$. The proof is then complete. ■

Proof of Lemma 2.3:

Again, without loss of generality, we assume that $1 \le d_x^* < d_x$ and $1 \le d_v^* < d_v$ in the following proof for notational simplicity. After some simple algebra, we can obtain the first derivative of $\Upsilon_{NT}(\mathbb{C})$ with respect to $\mathbb{C}$. Then it is easy to know that $\overline{\mathbb{C}}^*$ must be the solution of the following equation:

$$2\sum_{i=1}^N\sum_{t=1}^T Q_{it}^*(y_{it} - Q_{it}^{*\prime}\overline{\mathbb{C}}^*) + P^*\overline{\mathbb{C}}^* = 0,$$

where the definition of Q_{it}^* should be obvious, $P^* = \mathrm{diag}\{P_\beta^*, P_\gamma^*\}$, and

$$\begin{aligned} P_\beta^* &= I_m \otimes \mathrm{diag}\{\rho_{\beta_1}\|S_{\beta_1}\overline{\mathbb{C}}^*\|^{-1}, \ldots, \rho_{\beta_{d_x^*}}\|S_{\beta_{d_x^*}}\overline{\mathbb{C}}^*\|^{-1}\}, \\ P_\gamma^* &= \mathrm{diag}\{\rho_{\gamma_1}\|S_{\gamma_1}\overline{\mathbb{C}}^*\|^{-1}I_n, \ldots, \rho_{\gamma_{d_v^*}}\|S_{\gamma_{d_v^*}}\overline{\mathbb{C}}^*\|^{-1}I_n\}. \end{aligned}$$

It implies that $\overline{\mathbb{C}}^*$ must have the form

$$\overline{\mathbb{C}}^* = \left(\sum_{i=1}^N\sum_{t=1}^T Q_{it}^*Q_{it}^{*\prime} + \frac{P^*}{2}\right)^{-1}\sum_{i=1}^N\sum_{t=1}^T Q_{it}^*y_{it}.$$

Thus, consider

$$\overline{\mathbb{C}}^* - \overline{\mathbb{C}}_{\text{ora}} = \widetilde{\Psi}_{NT}^*\Sigma_{NT}\widetilde{\Psi}_{NT}^*\left(\sum_{i=1}^N\sum_{t=1}^T Q_{it}^*y_{it}\right),$$

where $\widetilde{\Psi}_{NT}^* = \mathrm{diag}\{\frac{1}{\sqrt{NT^2}}I_{md_x^*}, \frac{1}{\sqrt{NT}}I_{nd_v^*}\}$, and

$$\begin{aligned} \Sigma_{NT} = &\left(\widetilde{\Psi}_{NT}^*\sum_{i=1}^N\sum_{t=1}^T Q_{it}^*Q_{it}^{*\prime}\widetilde{\Psi}_{NT}^* + \widetilde{\Psi}_{NT}^*\frac{P^*}{2}\widetilde{\Psi}_{NT}^*\right)^{-1} \\ &- \left(\widetilde{\Psi}_{NT}^*\sum_{i=1}^N\sum_{t=1}^T Q_{it}^*Q_{it}^{*\prime}\widetilde{\Psi}_{NT}^*\right)^{-1}. \end{aligned}$$

By Assumption 2.2 and Lemma A.3 of Dong et al. (2018), it is easy to know that the rate of $\|\Sigma_{NT}\|$ converging to 0 is the same as

$$\begin{aligned} &\left\|\widetilde{\Psi}_{NT}^*\sum_{i=1}^N\sum_{t=1}^T Q_{it}^*Q_{it}^{*\prime}\widetilde{\Psi}_{NT}^* + \widetilde{\Psi}_{NT}^*\frac{P^*}{2}\widetilde{\Psi}_{NT}^* - \widetilde{\Psi}_{NT}^*\sum_{i=1}^N\sum_{t=1}^T Q_{it}^*Q_{it}^{*\prime}\widetilde{\Psi}_{NT}^*\right\| \\ = &\left\|\widetilde{\Psi}_{NT}^*\frac{P^*}{2}\widetilde{\Psi}_{NT}^*\right\| = O\left(\frac{\rho_\beta^*\sqrt{m}}{NT^2} + \frac{\rho_\gamma^*\sqrt{n}}{NT}\right). \end{aligned}$$

Then we know that $\|\overline{\mathbb{C}}^* - \overline{\mathbb{C}}_{\text{ora}}\| = O_P\left(\frac{\rho_\beta^*\sqrt{m}}{NT^2} + \frac{\rho_\gamma^*\sqrt{n}}{NT}\right)$. The proof is complete. ■

Proof of Theorem 2.2:

1). Same as the proof of Lemma 2.2, by condition $1 \le d_x^* + d_v^* < d_x + d_v$, we allow for $d_x^* = 0$ or $d_x^* = d_x$. Similarly, $d_v^* = 0$ or $d_v^* = d_v$ is allowed. Without loss of generality, we assume that $1 \le d_x^* < d_x$ and $1 \le d_v^* < d_v$ in the following proof for notational simplicity.

In what follows, we prove $\Pr(S_{\widehat{\rho},\beta} = \mathcal{A}_\beta^*) \to 1$, and $\Pr(S_{\widehat{\rho},\gamma} = \mathcal{A}_\gamma^*) \to 1$ can be proved similarly. Before proceeding further, we introduce some notations to facilitate the development. For an arbitrary model S, we say it is under–fitted if it misses at least one variable with a nonzero coefficient (under–fitted case allows for including redundant regressors); it is over–fitted if S not only includes all relevant variables but also includes at least one redundant regressor. Then, according to whether the model S_ρ is under fitted, correctly fitted, or over fitted, we create three mutually exclusive sets A^-, $A^0 = \left\{\rho \in \mathbb{R}^{d_x+d_v} : S_{\rho,\beta} = \mathcal{A}_\beta^*\right\}$ and $A^+ = \left\{\rho \in \mathbb{R}^{d_x+d_v} : S_{\rho,\beta} \supset \mathcal{A}_\beta^*, S_{\rho,\beta} \neq \mathcal{A}_\beta^*\right\}$. Let $\widetilde{\mathbb{C}}$ be the unregularized estimator as in (2.4) of the main text, and there is a sequence $\{\rho_{NT}\}$ that ensures the conditions required by Lemma 2.2 hold. Let $\overline{\mathbb{C}}_{\rho_{NT}}$ denote the estimator obtained by implementing (2.7) of the main text using ρ_{NT}.

Case 1 — In this case, we consider under-fitted models. Without losing generality, we assume that only one variable is missing, so we assume that $\|S_{\beta_\ell}\overline{\mathbb{C}}_{\widehat{\rho}}\| \neq 0$ for $\ell = 1, \ldots, d_x^* - 1$ are obtained from the under–fitted model and $\|S_{\beta_\ell}\overline{\mathbb{C}}_{\widehat{\rho}}\|$ for $\ell \geq d_x^*$ are 0. It does not matter whether the sparsity of $\gamma_0(\cdot)$ is correctly identified or not.

We then write

$$\begin{aligned}
\mathrm{SSR}_{\widehat{\rho}} &= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (y_{it} - Q_{it}'\overline{\mathbb{C}}_{\widehat{\rho}})^2. \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (y_{it} - Q_{it}'\widetilde{\mathbb{C}} + Q_{it}'\widetilde{\mathbb{C}} - Q_{it}'\overline{\mathbb{C}}_{\widehat{\lambda}})^2 \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (y_{it} - Q_{it}'\widetilde{\mathbb{C}})^2 + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (Q_{it}'\widetilde{\mathbb{C}} - Q_{it}'\overline{\mathbb{C}}_{\widehat{\rho}})^2 \\
&\quad + \frac{2}{N}\sum_{i=1}^N\sum_{t=1}^T (\widetilde{\mathbb{C}} - \overline{\mathbb{C}}_{\widehat{\rho}})' Q_{it}(y_{it} - Q_{it}'\widetilde{\mathbb{C}}) \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (y_{it} - Q_{it}'\widetilde{\mathbb{C}})^2 + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (Q_{it}'\widetilde{\mathbb{C}} - Q_{it}'\overline{\mathbb{C}}_{\widehat{\rho}})^2 \\
&:= \mathrm{SSR}_1 + \mathrm{SSR}_{\widehat{\rho}}^2,
\end{aligned}$$

where the fourth equality is due to the construction of the unregularized estimator.

We now consider $\mathrm{SSR}_{\widehat{\rho}}^2$ and write

$$\begin{aligned}
\mathrm{SSR}_{\widehat{\rho}}^2 &= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (\widetilde{\mathbb{C}} - \overline{\mathbb{C}}_{\widehat{\rho}})'\Psi_T^{-1}\Psi_T Q_{it}Q_{it}'\Psi_T\Psi_T^{-1}(\widetilde{\mathbb{C}} - \overline{\mathbb{C}}_{\widehat{\rho}}) \\
&\geq O(1)\|\Psi_T^{-1}(\widetilde{\mathbb{C}} - \overline{\mathbb{C}}_{\widehat{\rho}})\|^2 + o_P(1) \\
&= O(1)\left\|\sqrt{T}S_{\beta_{d_x^*}}\widetilde{\mathbb{C}}\right\|^2 + o_P(1)
\end{aligned}$$

where Ψ_T is defined in the beginning of this file.

Similarly, we can obtain that $\mathrm{SSR}_{\rho_{NT}} := \mathrm{SSR}_1 + \mathrm{SSR}_{\rho_{NT}}^2$, where

$$\begin{aligned}
\mathrm{SSR}_{\rho_{NT}}^2 &= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (\widetilde{\mathbb{C}} - \overline{\mathbb{C}}_{\rho_{NT}})'\Psi_T^{-1}\Psi_T Q_{it}Q_{it}'\Psi_T\Psi_T^{-1}(\widetilde{\mathbb{C}} - \overline{\mathbb{C}}_{\rho_{NT}}) \\
&\leq O(1)\|\Psi_T^{-1}(\widetilde{\mathbb{C}} - \overline{\mathbb{C}}_{\rho_{NT}})\|^2 + o_P(1) \\
&\leq O(1)\|\Psi_T^{-1}(\widetilde{\mathbb{C}} - \mathbb{C}_0)\|^2 + O(1)\|\Psi_T^{-1}(\mathbb{C}_0 - \overline{\mathbb{C}}_{\rho_{NT}})\|^2 = o_P(1),
\end{aligned}$$

where the last step follows from Lemma 2.1 and Lemma 2.2, and $\mathbb{C}_0$ is defined in the proof of Lemma 2.2.

Note that simple algebra shows that $\mathrm{SSR}_1 \to_P \sigma_e^2$. Based on the analysis on $\mathrm{SSR}_{\widehat{\rho}}^2$ and $\mathrm{SSR}_{\rho_{NT}}^2$, we then can conclude that

$$\Pr\left(\inf_{\widehat{\rho}\in A^-} \mathrm{BIC}_{\widehat{\rho}} > \mathrm{BIC}_{\rho_{NT}}\right) \to 1.$$

Case 2 — In this case, we consider over–fitted models. Consider $\forall \widehat{\rho} \in A^+$ and recall that $\overline{\mathbb{C}}_{\widehat{\rho}}$ determines $S_{\widehat{\rho},\beta}$. Having considered Case 1, we then assume the sparsity of γ_0 is either correctly identified or over identified here. Under such a model $S_{\widehat{\rho},\beta}$, we can define another unregularized estimator $\check{\mathbb{C}}_{\widehat{\rho}}$ as

$$\check{\mathbb{C}}_{\widehat{\rho}} = \underset{\mathbb{C}}{\operatorname{argmin}} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (y_{it} - Q_{it}'\mathbb{C})^2,$$

where $\|S_{\beta_\ell}\mathbb{C}\| = 0$ with $\forall \ell \notin S_{\widehat{\rho},\beta}$. Since $\check{\mathbb{C}}_{\widehat{\rho}}$ is the unregularized estimator under the model determined by $S_{\widehat{\rho},\beta}$, we obtain immediately that $\mathrm{SSR}_{\widehat{\rho}} \geq \mathrm{SSR}_{S_{\widehat{\rho},\beta}}$, where

$$\mathrm{SSR}_{S_{\widehat{\rho},\beta}} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (y_{it} - Q_{it}'\check{\mathbb{C}}_{\widehat{\rho}})^2.$$

It follows that

$$\begin{aligned}
&\ln \mathrm{SSR}_{\widehat{\rho}} - \ln \mathrm{SSR}_1 \geq \ln \mathrm{SSR}_{S_{\widehat{\rho},\beta}} - \ln \mathrm{SSR}_1 \\
&= \ln\left\{\frac{\mathrm{SSR}_1}{\mathrm{SSR}_1} + \frac{1}{NT\cdot \mathrm{SSR}_1}\sum_{i=1}^N\sum_{t=1}^T (\widetilde{\mathbb{C}} - \check{\mathbb{C}}_{\widehat{\rho}})'\Psi_T^{-1}\Psi_T Q_{it}Q_{it}'\Psi_T\Psi_T^{-1}(\widetilde{\mathbb{C}} - \check{\mathbb{C}}_{\widehat{\rho}})\right\} \\
&\geq -\frac{O(1)}{NT\cdot \mathrm{SSR}_1}\sum_{i=1}^N\sum_{t=1}^T (\widetilde{\mathbb{C}} - \check{\mathbb{C}}_{\widehat{\rho}})'\Psi_T^{-1}\Psi_T Q_{it}Q_{it}'\Psi_T\Psi_T^{-1}(\widetilde{\mathbb{C}} - \check{\mathbb{C}}_{\widehat{\rho}}) \geq -\frac{O(1)}{\mathrm{SSR}_1}\|\Psi_T^{-1}(\widetilde{\mathbb{C}} - \check{\mathbb{C}}_{\widehat{\rho}})\|^2 \\
&\geq -\frac{O(1)}{\mathrm{SSR}_1}\|\Psi_T^{-1}(\widetilde{\mathbb{C}} - \mathbb{C}_0)\|^2 - \frac{O(1)}{\mathrm{SSR}_1}\|\Psi_T^{-1}(\mathbb{C}_0 - \check{\mathbb{C}}_{\widehat{\rho}})\|^2 \\
&\geq -\left|O_P\left(\frac{m+n}{NT}\right)\right|,
\end{aligned}$$

where the third inequality follows from Assumption 2.2, and the last step follows from Lemma 2.1 and Lemma 2.2.

Similarly, we can obtain that $\ln \mathrm{SSR}_{\rho_{NT}} - \ln \mathrm{SSR}_1 = O_P\left(\frac{m+n}{NT}\right)$. Thus, we obtain

$$\ln \mathrm{SSR}_{\widehat{\rho}} - \ln \mathrm{SSR}_{\rho_{NT}} \geq -\left|O_P\left(\frac{m+n}{NT}\right)\right|.$$

We then write

$$\inf_{\widehat{\rho}\in A^+} \mathrm{BIC}_{\widehat{\rho}} - \mathrm{BIC}_{\rho_{NT}} = \ln \mathrm{SSR}_{\widehat{\rho}} - \ln \mathrm{SSR}_{\rho_{NT}} + (\mathrm{df}_{\widehat{\rho}} - \mathrm{df}_{\rho_{NT}})\frac{(m+n)\ln(NT)}{NT}.$$

By Lemma 2.2, we know that $\Pr(\mathrm{df}_{\rho_{NT}} \to d_x^* + d_v^*) = 1$. Since $\widehat{\rho} \in A^+$ and we assume that γ_0 is either correctly identified or over identified, we must have that $\Pr(\mathrm{df}_{\rho_{NT}} \geq d_x^* + d_v^* + 1) \to 1$. Then it is clear that

$$\Pr\left(\inf_{\widehat{\rho}\in A^+} \mathrm{BIC}_{\widehat{\rho}} > \mathrm{BIC}_{\rho_{NT}}\right) \to 1.$$

Combining Cases 1 and 2, we obtain that $\Pr\left(\inf_{\widehat{\rho}\in A^-\cup A^+}\mathrm{BIC}_{\widehat{\rho}} > \mathrm{BIC}_{\rho_{NT}}\right) \to 1$, which in turn implies $\Pr(S_{\widehat{\rho},\beta} = \mathcal{A}_\beta^*) \to 1$. Similarly, we can show that $\Pr(S_{\widehat{\rho},\gamma} = \mathcal{A}_\gamma^*) \to 1$. The proof is complete. ■

Proof of Corollary 2.2:

The proof of Corollary 2.2 is almost identical to those given for Theorem 2.1, but requires one to account for the rate of divergence of f_{0t}. ■

Appendix B

In this appendix, we adopt the PCA approach initially proposed in Bai (2009) to estimate unobserved factors. As explained in Sections 4.1 and 4.2 of Bai et al. (2009), if both x_{it} and f_{0t} require different normalizers across the time dimension (e.g., $x_{it} \sim I(1)$ and $f_{0t} \sim I(0)$), certain technical challenges would occur when conducting the estimation. We aim to provide a simpler method to solve these issues, so we focus on the case where f_{0t} is stationary. Moreover, we further show some possible extensions using the PCA–based approach in Appendix D of this supplementary file.

B.1 Estimation via PCA

We still focus on model (1.3) of the main text, and firstly state the necessary assumptions.

Assumption B.1.

1. *(a) Let $\{\varepsilon_{ij} \mid i \in \mathbb{Z}^+, j \in \mathbb{Z}\}$ be an array of d_x–dimensional independent and identically distributed (i.i.d.) random variables over i and j, and let $\{\varepsilon_{ij}\}$ be independent of $\{r_{it}, v_i, f_{0t}\}$. Moreover, $E[\varepsilon_{11}] = 0$, $E[\varepsilon_{11}\varepsilon_{11}'] = I_{d_x}$, $E\|\varepsilon_{11}\|^q < \infty$ for some $q > 4$, and the characteristic function of ε_{11} is integrable.*

   (b) For each $i \geq 1$, let $x_{it} = x_{i,t-1} + w_{it}$, where $\max_{i\geq 1}\|x_{i0}\| = O_P(1)$, and w_{it} is a linear process given by $w_{it} = \sum_{j=0}^{\infty} D_j\varepsilon_{i,t-j}$. In addition, $\{D_j \mid j \in \mathbb{Z}\}$ is a sequence of deterministic matrices such that (1) $D_0 = I_{d_x}$, (2) $\sum_{j=0}^{\infty} j\|D_j\| < \infty$, and (3) $D := \sum_{j=0}^{\infty} D_j$ is of full rank.

2. *Let $\{e_{i1},\ldots,e_{iT}; r_{i1},\ldots,r_{iT}; v_i\}$ be identically distributed across i, and let $r_t = (r_{1t},\ldots,r_{NT})'$ be strictly stationary across t. Let $H_{it} = H_m(r_{it})H_m'(r_{it})$ with H_{it,l_1l_2} standing for the $(l_1,l_2)^{th}$ element of H_{it}, where $1 \leq l_1, l_2 \leq m$. Let $\mathcal{H}_i = \mathcal{H}_n(v_i)\mathcal{H}_n'(v_i)$ with $\mathcal{H}_{i,l_1l_2}$ standing for the $(l_1,l_2)^{th}$ element of $\mathcal{H}_i$, where $1 \leq l_1, l_2 \leq nd_v$.*

   (a) $\max_{t\geq 1}\sum_{i\neq j} E\left[\left(H_{it,l_1l_2} - E[H_{it,l_1l_2}]\right)\left(H_{jt,l_1l_2} - E[H_{jt,l_1l_2}]\right)\right] = O(N)$ uniformly for l_1, l_2, and $\sum_{i\neq j} E\left[\left(\mathcal{H}_{i,l_1l_2} - E[\mathcal{H}_{i,l_1l_2}]\right)\left(\mathcal{H}_{j,l_1l_2} - E[\mathcal{H}_{j,l_1l_2}]\right)\right] = O(N)$ uniformly for l_1, l_2;

   (b) $\sum_{i\neq j}\sum_{t\neq s} E\left[\left(H_{it,l_1l_2} - E[H_{it,l_1l_2}]\right)\left(H_{js,l_1l_2} - E[H_{js,l_1l_2}]\right)\right] = O(NT)$ uniformly for l_1, l_2;

   (c) $0 < \lambda_{\min}(E[H_m(r_{11})H_m'(r_{11})]) \leq \lambda_{\max}(E[H_m(r_{11})H_m'(r_{11})]) < \infty$ uniformly in m;

   (d) $0 < \lambda_{\min}(E[\mathscr{H}_n(v_1)\mathscr{H}_n'(v_1)]) \leq \lambda_{\max}(E[\mathscr{H}_n(v_1)\mathscr{H}_n'(v_1)]) < \infty$ uniformly in n.

3. *(a) $\max_{1\leq \ell\leq d_x}\{\sum_{j=m}^{\infty} b_{0\ell,j}^2\}^{1/2} = O(m^{-\mu_2})$ and $\max_{1\leq \ell\leq d_v}\{\sum_{j=n}^{\infty} c_{0\ell,j}^2\}^{1/2} = O(n^{-\mu_1})$, where $b_{0\ell,j} = \int_{\mathcal{R}} \beta_{0\ell}(w)h_j(w)\pi(w)dw$, $c_{0\ell,j} = \int_{\mathcal{R}} \gamma_{0\ell}(w)h_j(w)\pi(w)dw$, and μ_1 and μ_2 are two positive constants;*

(b) *Let $f_r(w)$ be the density function of r_{it}, and $f_{v_\ell}(w)$ be the density function of $v_{i,\ell}$, where $v_{i,\ell}$ stands for the ℓ^{th} element of v_i for $\ell = 1, \ldots, d_v$. Suppose that $\sup_{w\in\mathcal{R}} f_r(w)/\pi(w) < \infty$ and $\sup_{w\in\mathcal{R}} f_{v_\ell}(w)/\pi(w) < \infty$ for $\ell = 1, \ldots, d_v$.*

(c) $\frac{m^2}{NT} \to 0$ *and* $Tm^{-2\mu_2} \to 0$.

4. *Let $\frac{1}{T}\sum_{t=1}^T f_{0t}f_{0t}' \to_P \Sigma_F > 0$ and $\max_{t\geq 1} E\|f_{0t}\|^4 < \infty$. Let $\frac{1}{N}\sum_{i=1}^N \gamma_0(v_i)\gamma_0'(v_i) \to_P \Sigma_\Gamma$ and $E\|\gamma_0(v_1)\|^4 < \infty$.*

Assumption B.2. *Suppose e_t and the filtration $\mathcal{B}_{NT,t} = \sigma(x_j, r_j, e_{j-1}; f_{01}, \ldots, f_{0T}; v_1, \ldots, v_N \,|\, j \leq t+1)$ form a martingale difference sequence such that almost surely $E[e_t \,|\, \mathcal{B}_{NT,t-1}] = 0$ and $E[e_te_t' \,|\, \mathcal{B}_{NT,t-1}] = \Sigma_e = \{\sigma_{ij}\}_{N\times N}$, where $e_t = (e_{1t}, \ldots, e_{Nt})'$, and x_t and r_t are defined similarly. In addition, let $\sigma_{ii} = \sigma_e^2$ for $i \geq 1$, and suppose that $\sum_{i\neq j}|\sigma_{ij}| = O(N)$ and $\max_{i,t} E[e_{it}^4 \,|\, \mathcal{B}_{NT,t-1}] < \infty$.*

The Assumption B.1 is the combination of Assumption 2 and Assumption 4 of the main text with minor modifications. The Assumption B.2 is more restrictive than Assumption 3, because the filtration $\mathcal{F}_t$ includes more variables.

Note that under Assumptions B.1 and B.2 in hand, we can always recover $\beta_0(\cdot)$ regardless of the availability of f_{0t}'s, which will help us tackle the aforementioned technical issue. By virtue of the series expansion for $\beta_0(\cdot)$, we have the OLS estimator of C_{β_0} as follows:

$$\ddot{C}_\beta = \left(\sum_{i=1}^N Z_i'Z_i\right)^{-1}\sum_{i=1}^N Z_i'Y_i, \tag{B.1}$$

where $Z_i = (H_m(r_{i1}) \otimes x_{i1}, \ldots, H_m(r_{iT}) \otimes x_{iT})'$ and $Y_i = (y_{i1}, \ldots, y_{iT})'$. The approximation rate of $\ddot{C}_\beta$ to C_{β_0} is summarized by the next lemma.

Lemma B.1. *Let Assumptions B.1 and B.2 hold. As $(N,T) \to (\infty, \infty)$, $\|\ddot{C}_\beta - C_{\beta_0}\| = O_P\left(\frac{1}{\sqrt{T}}\right) + O_P\left(m^{-\mu_2}\right)$.*

Lemma B.1 allows us to narrow down the set that C_{β_0} belongs to as follows.

$$\mathsf{B}_T := \left\{C \mid \|C - \ddot{C}_\beta\| \leq \omega_0 T^{-\frac{1}{2}}\right\},$$

where ω_0 is a sufficiently large constant. The aim of defining B_T is to eschew the annoyance that $\{x_{i1}, \ldots, x_{iT}\}$ and $\{f_{01}, \ldots, f_{0T}\}$ require different normalizers when deriving asymptotic properties.

We now proceed to full estimation, and rewrite our model in matrix notation as

$$Y_i = \phi_i[\beta_0] + F_0\gamma_0(v_i) + e_i, \tag{B.2}$$

where $\phi_i[\beta] := (x_{i1}'\beta(r_{i1}), \ldots, x_{iT}'\beta(r_{iT}))'$ for $\forall \beta(\cdot) = (\beta_1(\cdot), \ldots, \beta_{d_x}(\cdot))'$, and F_0 and e_i are defined accordingly. Moreover, let $\Gamma_0 = (\gamma_0(v_1), \ldots, \gamma_0(v_N))'$ for later use. Left–multiplying M_{F_0} on both sides of (B.2) gives $M_{F_0}(Y_i - \phi_i[\beta_0]) = M_{F_0}e_i$. To estimate C_{β_0} and F_0, we thus define the objective function:

$$R_{NT}(C_\beta, F) = \frac{1}{NT}\sum_{i=1}^N (Y_i - \phi_i[\beta_m])'\, M_F\, (Y_i - \phi_i[\beta_m]),$$

where $\beta_m(r) = [H_m'(r) \otimes I_{d_x}]C_\beta$. The estimators of (C_{β_0}, F_0) are obtained by

$$(\widehat{C}_\beta, \widehat{F}) = \underset{(C_\beta, F) \in \mathsf{B}_T \times \mathsf{D}_F}{\operatorname{argmin}} R_{NT}(C_\beta, F), \tag{B.3}$$

where $\mathsf{D}_F := \left\{F \mid \frac{1}{T}F'F = I_{d_v}\right\}$. The restriction $\widehat{F} \in \mathsf{D}_F$ is for solving the identification issue of the factor model (e.g., Bai, 2009). The estimator of $\beta_0(r)$ is correspondingly defined as $\widehat{\beta}_m(r) = [H_m'(r) \otimes I_{d_x}]\widehat{C}_\beta$.

Following the same arguments as in Bai (2009, p. 1236), (B.3) can be decomposed into the following two expressions:

$$\begin{aligned} \widehat{C}_\beta &= \underset{C_\beta \in \mathsf{B}_T}{\operatorname{argmin}} \frac{1}{NT} \sum_{i=1}^N \left(Y_i - \phi_i[\beta_m]\right)' M_{\widehat{F}} \left(Y_i - \phi_i[\beta_m]\right), \\ \frac{1}{NT} \sum_{i=1}^N & (Y_i - \phi_i[\widehat{\beta}_m])(Y_i - \phi_i[\widehat{\beta}_m])'\widehat{F} = \widehat{F} V_{NT}, \end{aligned} \tag{B.4}$$

where V_{NT} is a diagonal matrix with the diagonal being the d_v largest eigenvalues of

$$\frac{1}{NT} \sum_{i=1}^N (Y_i - \phi_i[\widehat{\beta}_m])(Y_i - \phi_i[\widehat{\beta}_m])'$$

arranged in descending order. Consequently, a routine estimator of Γ_0 would be

$$\frac{1}{T}(Y_1 - \phi_1[\widehat{\beta}_m]. \ldots, Y_N - \phi_N[\widehat{\beta}_m]))'\widehat{F},$$

which, however, does not reveal the information of the loading function. Thus, using (B.2), we establish the estimator of $\gamma_0(\cdot)$ as follows.

$$\widehat{C}_\gamma = \left[\sum_{i=1}^N \mathscr{H}_n(v_i)\mathscr{H}_n(v_i)'\right]^{-1} \sum_{i=1}^N \mathscr{H}_n(v_i) \left\{\frac{1}{T}\widehat{F}'(Y_i - \phi_i[\widehat{\beta}_m])\right\}, \tag{B.5}$$

which gives the estimator of $\gamma_0(v)$ by $\widehat{\gamma}_n(v) = \mathscr{H}_n'(v)\widehat{C}_\gamma$.

Numerically, we just need to implement an iterative procedure to obtain $\widehat{C}_\beta$ and $\widehat{F}$ by (B.4). Afterwards, we can implement (B.5). We refer interested readers to Jiang et al. (2017), where the algorithm for the linear panel data setting with interactive fixed effects has been studied carefully. In order to start the iteration, we can use (B.1) as an initial estimate in practice. "fmincon" function of MATLAB provides an easy way to set up the restriction B_T.

B.2 Asymptotic Properties

To derive the consistency, we impose the following assumption.

Assumption B.3. *Let* $\inf_{F \in \mathsf{D}_F} \lambda_{\min}(\Omega_\dagger(F)) \geq A_1 > 0$ *uniformly, where*

$$\begin{aligned} \Omega_\dagger(F) &= \frac{1}{NT^2} \left\{\Omega_1(F) - \Omega_2'(F)\left[(\Gamma_0'\Gamma_0) \otimes I_T\right]^{-1} \Omega_2(F)\right\}, \\ \Omega_1(F) &= \sum_{i=1}^N Z_i' M_F Z_i, \quad \Omega_2(F) = \sum_{i=1}^N \gamma_0(v_i) \otimes (M_F Z_i). \end{aligned}$$

Assumption B.3 ensures that the estimators given in (B.3) are well defined, and is equivalent to Assumption A of Bai (2009). We are now ready to summarize the consistency and some useful rates in the following lemma.

Lemma B.2. *Let Assumptions B.1–B.3 hold. As* $(N,T) \to (\infty,\infty)$*,*

1. $\|\widehat{\beta}_m - \beta_0\|_{L^2} = o_P\left(\frac{1}{\sqrt{T}}\right)$,
2. $\|P_{\widehat{F}} - P_{F_0}\| = o_P(1)$,
3. $V_{NT} \to_P V$,
4. $\frac{1}{\sqrt{T}}\|\widehat{F}\Pi_{NT}^{-1} - F_0\| = O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$,
5. $\left\|\frac{1}{T}\widehat{F}'(\widehat{F} - F_0\Pi_{NT})\right\| = O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right)$,
6. $\|P_{\widehat{F}} - P_{F_0}\|^2 = O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right)$,

where $\Pi_{NT}^{-1} = V_{NT}(F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}$*, and* V *is a* $d_v \times d_v$ *diagonal matrix consisting of the eigenvalues of* $\Sigma_F\Sigma_\Gamma$*.*

Having established Lemma B.2, we provide the rates of convergence associated to (B.3).

Lemma B.3. *Let Assumptions B.1–B.3 hold. In addition, let* $\frac{N}{T} \to \nu$ *with* $0 \le \nu < \infty$*. As* $(N,T) \to (\infty,\infty)$*,*

1. $\|\widehat{\beta}_m - \beta_0\|_{L^2} = O_P\left(\sqrt{\frac{m}{NT^2}}\right) + O_P\left(m^{-\mu_2}\right)$,
2. $\|P_{\widehat{F}} - P_{F_0}\| = O_P\left(\sqrt[4]{\frac{m}{NT}}\right) + O_P\left(\sqrt[4]{Tm^{-2\mu_2}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right)$.

To establish the normality, we impose some extra assumptions.

Assumption B.4.

1. *Let* $\mathcal{F}_{Nts}^{\dagger} = \sigma(r_{1t},\ldots,r_{Nt}; r_{1s},\ldots,r_{Ns})$*. Suppose that* $E[f_{0t}'f_{0s}\,|\,\mathcal{F}_{Nts}^{\dagger}] = a_{ts}$ *a.s. for* $t \neq s$*, and* $\sum_{t\neq s}|a_{ts}| = O(T)$*.*
2. *Let* $\mathcal{F}_{ij,T}^{*} = \sigma(x_{it}, x_{jt}, r_{it}, r_{jt}, f_{0t}; v_1,\ldots,v_N\,|\,t \le T)$*. Suppose that* $E[e_{it}e_{jt}\,|\,\mathcal{F}_{ij,T}^{*}] = \sigma_{ij}$ *a.s. for* $i \neq j$ *and* $\sum_{i\neq j}|\sigma_{ij}| = O(N)$*.*

Assumption B.4 further imposes more restrictions on the unknown factors and error terms in order to ensure that the estimator $\widehat{\beta}_m$ given by (B.3) is not asymptotically biased in the sense of Theorem 3 of Bai (2009). The current requirements of Assumption B.4 are in the same spirit as Connor et al. (2012, Eq. 3 and Eq. 20) and Jiang et al. (2017, pp. 21–22). Without this assumption, some other types of conditions are needed to achieve asymptotic normality. For example, one can require $N/T \to \kappa$ with $0 < \kappa < \infty$ and establish the normality with a bias as in Theorem 3 of Bai (2009).

Theorem B.1. *Let Assumptions B.1–B.4 hold. Additionally, let* $\frac{mN}{T} \to 0$*,* $\frac{mT}{N^2} \to 0$*, and* $\frac{NT^2}{m^{2\mu_2}} \to 0$*. For* $\forall r \in \mathcal{R}$*, as* $(N,T) \to (\infty,\infty)$*,*

$$\frac{\sqrt{NT^2}}{\|H_m(r)\|}(\widehat{\beta}_m(r) - \beta_0(r)) \to_D N(0,\widetilde{\Sigma}_\beta),$$

where $\widetilde{\Sigma}_\beta$ *is the same as defined in Theorem 2.1 of the main text.*

Similar comments to those for Theorem 2.1.1 of the main text may be made here. The conditions on T, N, m and μ_2 are seemingly a bit complicated. Nevertheless, they are reasonable and easily satisfied. Consider $\frac{mN}{T} \to 0$, $\frac{mT}{N^2} \to 0$, and $\frac{NT^2}{m^{2\mu_2}} \to 0$ in the body of Theorem B.1 as an example. Let $N = \lfloor T^{b_1} \rfloor$ and $m = \lfloor T^{b_2} \rfloor$ with $b_1, b_2 > 0$. Then the conditions are fulfilled if $b_1 + b_2 < 1$, $1 + b_2 < 2b_1$ and $2 + b_1 < 2\mu_2 b_2$. Thus, it renders that

$$0.5 < b_1 < 1, \quad (2 + b_1)/(2\mu_2) < b_2 < 0.5, \quad \text{and} \quad \mu_2 \geq 3.$$

Note that many harsh conditions on m, n, μ_1 and μ_2 kick in only when deriving the asymptotic normality, and they are unnecessary for asymptotic consistency.

In what follows, we consider the estimation on the factor structure.

Assumption B.5. *Let $F_0 \in \mathsf{D}_F$ and let $\frac{\Gamma_0'\Gamma_0}{N}$ be a $d_v \times d_v$ diagonal matrix with distinct entries a.s. Moreover, suppose that $\sqrt{N} \sup_{i \geq 1} \sup_{F \in \{F \mid \frac{1}{\sqrt{T}}\|F - F_0\| \leq \epsilon\}} \left\| \frac{1}{T} F' e_i \right\| = O_P\left(\frac{\sqrt{N} \ln N}{\sqrt{T}} \right)$, where ϵ is a sufficiently small positive number.*

The first condition of this assumption serves the purpose of identifying both $\gamma_0(\cdot)$ and F_0, and is similar to Assumptions 3.2 and 4.1 of Fan et al. (2016). The second condition of this Assumption can be easily verified. We then present the rates of convergence associated with factors and the loading functions in the next lemma.

Lemma B.4. *Let Assumptions B.1–B.3 and B.5 hold. In addition, let $\frac{N}{T} \to \nu$ with $0 \leq \nu < \infty$. As $(N, T) \to (\infty, \infty)$,*

1. $\frac{1}{\sqrt{T}} \|\widehat{F} - F_0\| = O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\sqrt{T} m^{-\mu_2}\right)$,
2. $\|\widehat{\gamma}_n - \gamma_0\|_{L^2} = O_P\left(\sqrt{\frac{m}{NT}}\right) + O_P\left(\frac{\sqrt{n}}{N}\right) + O_P(\max\{\sqrt{T} m^{-\mu_2}, n^{-\mu_1}\})$.

Lemma B.4 helps us further obtain the next theorem.

Theorem B.2. *Let Assumptions B.1–B.3 and B.5 hold. Suppose that $\frac{N(\ln N)^2}{T} \to 0$ and $\frac{NT}{m^{2\mu_2}} \to 0$. As $(N, T) \to (\infty, \infty)$,*

1. *For any given t, $\sqrt{N}(\widehat{f}_t - f_{0t}) \to_D N(0, \Sigma_\Gamma^{-1} \Sigma_\Gamma^* \Sigma_\Gamma^{-1})$, where $\widehat{f}_t$ denotes the t^{th} column of $\widehat{F}'$, and $\Sigma_\Gamma^* = \lim_N \frac{1}{N} \sum_{i,j=1}^N \sigma_{ij} E[\gamma_0(v_i)\gamma_0'(v_j)]$.*
2. $\|\widehat{\gamma}_n - \gamma_0\|_{L^2} = O_P\left(\sqrt{\frac{m}{NT}}\right) + O_P\left(\frac{\sqrt{n}}{N}\right) + O_P(\max\{\sqrt{T} m^{-\mu_2}, n^{-\mu_1}\})$.

Relevant comments similar to those for Theorem 2.1.2 and discussion on σ_{ij}'s of Section 5 of the main text may be made here for Theorem B.2.1. The asymptotic distribution in the first result of Theorem B.2 is consistent with Theorem 1 of Bai and Ng (2013), wherein a factor model without regressors is considered. Due to plugging $\widehat{\beta}_m(\cdot)$ and $\widehat{F}$ in (B.5), unlike in Theorem 2.1.3, it seems impossible to establish an asymptotic normality for $\widehat{\gamma}_n(v)$ for the PCA approach.

B.3 Preliminary Lemmas of PCA Method

Lemma B.5. *Let Assumptions B.1 and B.2 hold. As* $(N,T)\to(\infty,\infty)$,

1. $\|\frac{1}{NT}e'e\|=o_P(1)$ *and* $\|\frac{1}{NT}ee'\|=o_P(1)$, *in which* $e=(e_1,\ldots,e_N)'$,
2. $\sup_{F\in\mathsf{D}_F}\frac{1}{NT}\sum_i e_i'P_Fe_i=o_P(1)$,
3. $\sup_{F\in\mathsf{D}_F}\left|\frac{1}{NT}\sum_i\gamma_0'(v_i)F_0'M_Fe_i\right|=o_P(1)$,
4. $\sup_{(C_\beta,F)\in\mathsf{B}_T\times\mathsf{D}_F}\left|\frac{1}{NT}\sum_i(\phi_i[\beta_{0,m}]-\phi_i[\beta_m])'M_Fe_i\right|=o_P(1)$,
5. $\sup_{F\in\mathsf{D}_F}\left|\frac{1}{NT}\sum_i\phi_i[\Delta_{\beta_0}]'M_Fe_i\right|=o_P(1)$,
6. $\sup_{F\in\mathsf{D}_F}\left|\frac{1}{NT}\sum_i\phi_i[\Delta_{\beta_0}]'M_F\phi_i[\Delta_{\beta_0}]\right|=o_P(1)$,
7. $\sup_{F\in\mathsf{D}_F}\left|\frac{1}{NT}\sum_i\phi_i[\Delta_{\beta_0}]'M_FF_0\gamma_0(v_i)\right|=o_P(1)$,
8. $\sup_{(C_\beta,F)\in\mathsf{B}_T\times\mathsf{D}_F}\left|\frac{1}{NT}\sum_i\phi_i\left[\Delta_{\beta_0}\right]'M_F(\phi_i\left[\beta_m\right]-\phi_i\left[\beta_{0,m}\right])\right|=o_P(1)$.

Lemma B.6. *Let Assumptions B.1–B.3 and B.5 hold. In addition, let* $\frac{N}{T}\to\nu$ *with* $0\le\nu<\infty$. *As* $(N,T)\to(\infty,\infty)$,

1. $\|\Pi_{NT}-I_{d_v}\|=O_P\left(\sqrt{m/(NT)}+\sqrt{T/m^{2\mu_2}}+\frac{1}{N}\right)$,
2. $\left\|\frac{1}{T}F_0'\widehat{F}-I_{d_v}\right\|=O_P\left(\sqrt{m/(NT)}+\sqrt{T/m^{2\mu_2}}+\frac{1}{N}\right)$.

B.4 Proofs of PCA-based Method

Proof of Lemma B.1:

Simply algebra gives

$$\ddot{C}_\beta-C_{\beta_0}=\left(\sum_{i=1}^NZ_i'Z_i\right)^{-1}\sum_{i=1}^NZ_i'e_i+\left(\sum_{i=1}^NZ_i'Z_i\right)^{-1}\sum_{i=1}^NZ_i'\Delta_i,$$

where Δ_i is defined accordingly.

By Lemma A.2, $\|(\sum_{i=1}^NZ_i'Z_i)^{-1}\sum_{i=1}^NZ_i'e_i\|=O_P\left(\sqrt{\frac{m}{NT}}\right)$, so consider $(\sum_{i=1}^NZ_i'Z_i)^{-1}\sum_{i=1}^NZ_i'\Delta_i$ below. Note that it is easy to show

$$\begin{aligned}\frac{1}{NT^2}\sum_{i=1}^N\|\Delta_i\|^2&\le\frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T(f_{0t}'\gamma_0(v_i))^2+\frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T(x_{it}'\Delta_{\beta_0}(r_{it}))^2\\&=O_P\left(\frac{1}{T}+m^{-2\mu_2}\right).\end{aligned}$$

By a development similar to (A.1), we immediately obtain

$$\left\|\left(\sum_{i=1}^{N} Z_i' Z_i\right)^{-1} \sum_{i=1}^{N} Z_i' \Delta_i\right\| = O_P\left(\frac{1}{\sqrt{T}}\right)$$

Then the result follows. ■

Proof of Lemma B.5:

(1). Firstly, write

$$\begin{aligned}
\frac{1}{N^2T^2} E\|e'e\|^2 &= E\left\|\frac{1}{NT}\sum_{i=1}^{N} e_i e_i'\right\|^2 = \frac{1}{N^2T^2}\sum_{t,s=1}^{T}\left(\sum_{i=1}^{N} E[e_{it}^2 e_{is}^2] + \sum_{i\neq j} E[e_{it}e_{is}e_{jt}e_{js}]\right) \\
&= \frac{1}{N^2T^2}\sum_{t=1}^{T}\left(\sum_{i=1}^{N} E[e_{it}^4] + \sum_{i\neq j} E[(e_{it}e_{jt} - \sigma_{ij})^2]\right) \\
&\quad + \frac{1}{N^2T^2}\sum_{t\neq s}\left(\sum_{i=1}^{N} E[e_{it}^2 e_{is}^2] + \sum_{i\neq j} E[(e_{it}e_{jt} - \sigma_{ij})(e_{is}e_{js} - \sigma_{ij})]\right) + \frac{1}{N^2}\sum_{i\neq j}\sigma_{ij}^2 \\
&= O(1)\frac{1}{N} + O(1)\frac{1}{T},
\end{aligned}$$

where the fifth equality follows from Assumption B.2. Thus, $\frac{1}{NT}\|e'e\| = O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$.

Similarly, we can write

$$\begin{aligned}
E\left\|\frac{1}{NT} ee'\right\|^2 &= \left\{E\left[\frac{1}{NT} e_i' e_j\right]^2\right\}_{N\times N} = \sum_{i,j=1}^{N}\frac{1}{N^2T^2}\sum_{t,s=1}^{T} E\left[e_{it}e_{jt}e_{is}e_{js}\right] \\
&= \sum_{t,s=1}^{T}\frac{1}{N^2T^2}\sum_{i,j=1}^{N} E\left[e_{it}e_{jt}e_{is}e_{js}\right] = O\left(\frac{1}{N}\right) + O\left(\frac{1}{T}\right).
\end{aligned}$$

(2). Write

$$\begin{aligned}
\sup_{F\in \mathsf{D}_F}\frac{1}{NT}\sum_{i=1}^{N} e_i' P_F e_i &= \sup_{F\in \mathsf{D}_F}\frac{1}{NT}\mathrm{tr}\left(P_F e'e\right) \leq \sup_{F\in \mathsf{D}_F}\frac{d_v}{NT}\|P_F\|_{\mathrm{sp}}\|e'e\|_{\mathrm{sp}} \\
&\leq \sup_{F\in \mathsf{D}_F}\frac{d_v}{NT}\|P_F\|_{\mathrm{sp}}\|e'e\| = o_P(1),
\end{aligned}$$

where the first inequality follows from the fact that $|\mathrm{tr}\,(A)| \leq \mathrm{rank}\,(A)\,\|A\|_{\mathrm{sp}}$; and the second equality follows from (1) of this lemma.

(3). Write

$$\begin{aligned}
&\sup_{F\in \mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^{N}\gamma_0(v_i)' F_0' M_F e_i\right| = \sup_{F\in \mathsf{D}_F}\left|\frac{1}{NT}\mathrm{tr}\left(F_0' M_F e' \Gamma_0\right)\right| \leq \sup_{F\in \mathsf{D}_F}\frac{d_v}{NT}\left\|F_0' M_F e' \Gamma_0\right\|_{\mathrm{sp}} \\
&\leq \sup_{F\in \mathsf{D}_F}\frac{d_v}{NT}\|F_0\|_{\mathrm{sp}}\|M_F\|_{\mathrm{sp}}\|\Gamma_0\|_{\mathrm{sp}}\|e\|_{\mathrm{sp}} = \sup_{F\in \mathsf{D}_F}\frac{d_v}{NT}\|F_0\|_{\mathrm{sp}}\|M_F\|_{\mathrm{sp}}\|\Gamma_0\|_{\mathrm{sp}}\left\|ee'\right\|_{\mathrm{sp}}^{1/2} \\
&\leq \sup_{F\in \mathsf{D}_F}\frac{d_v}{\sqrt{NT}}\|F_0\|_{\mathrm{sp}}\|M_F\|_{\mathrm{sp}}\|\Gamma_0\|_{\mathrm{sp}}\left(\frac{1}{NT}\left\|ee'\right\|\right)^{1/2} = o_P\left(1\right),
\end{aligned}$$

where the first inequality follows from the fact that $|\mathrm{tr}\,(A)| \leq \mathrm{rank}\,(A)\,\|A\|_{\mathrm{sp}}$; the second equality follows from Fact 5.10.18 of Bernstein (2005); and the last equality follows from (1) of this lemma.

(4). Write

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\beta_{0,m}]-\phi_i[\beta_m]\right)' M_F e_i \\
=\;&\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\beta_{0,m}]-\phi_i[\beta_m]\right)' e_i+\frac{1}{NT}\sum_{i=1}^{N}\left(\phi_i[\beta_{0,m}]-\phi_i[\beta_m]\right)' P_F e_i \\
:=\;&\Lambda_1+\Lambda_2.
\end{aligned}
$$

Note that Λ_1 can be written as

$$
\Lambda_1=\sqrt{T}(C_{\beta_0}-C_\beta)'\frac{1}{NT^{3/2}}\sum_{i=1}^{N}Z_i'e_i.
$$

It is easy to know that by Assumptions B.1 and B.2, $E\left\|\frac{1}{NT^{3/2}}\sum_{i=1}^{N}Z_i'e_i\right\|^2=O\left(\frac{m}{NT}\right)$. In connection with the construction of B_T, we obtain that

$$
\sup_{C_\beta\in\mathsf{B}_T}\|\Lambda_1\|=\sqrt{T}\|C_{\beta_0}-C_\beta\|\cdot\left\|\frac{1}{NT^{3/2}}\sum_{i=1}^{N}Z_i'e_i\right\|=O\left(\sqrt{\frac{m}{NT}}\right).
$$

In order to consider Λ_2, let $\Delta b=(\phi_1[\beta_{0,m}]-\phi_1[\beta_m],\ldots,\phi_N[\beta_{0,m}]-\phi_N[\beta_m])$. Note that

$$
\begin{aligned}
\sup_{C_\beta\in\mathsf{B}_T}\frac{1}{NT}\|\Delta b\|^2&=\sup_{C_\beta\in\mathsf{B}_T}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[x_{it}'\{\beta_m(r_{it})-\beta_{0,m}(r_{it})\}\right]^2\\
&=O_P(1)\sup_{C_\beta\in\mathsf{B}_T}T\|\beta_m-\beta_{0,m}\|_{L^2}^2=O_P(1),
\end{aligned}
$$

where the second equality follows from (1) of Lemma A.2 and Assumption B.1.2, and the third equality follows from the construction of B_T. Then we are able to write

$$
\begin{aligned}
&\sup_{(C_\beta,F)\in\mathsf{B}_T\times\mathsf{D}_F}|\Lambda_2|=\sup_{(C_\beta,F)\in\mathsf{B}_T\times\mathsf{D}_F}\left|\frac{1}{NT}\mathrm{tr}\left(P_Fe'\Delta b\right)\right|\\
\le\;&\frac{d_v}{NT}\sup_{(C_\beta,F)\in\mathsf{B}_T\times\mathsf{D}_F}\|P_Fe'\Delta b\|_{\mathrm{sp}}\le\sup_{(C_\beta,F)\in\mathsf{B}_T\times\mathsf{D}_F}\frac{d_v}{NT}\|P_F\|_{\mathrm{sp}}\|e\|_{\mathrm{sp}}\|\Delta b\|_{\mathrm{sp}}\\
=\;&\sup_{(C_\beta,F)\in\mathsf{B}_T\times\mathsf{D}_F}d_v\|P_F\|_{\mathrm{sp}}\left(\frac{1}{NT}\|ee'\|\right)^{1/2}\left(\frac{1}{\sqrt{NT}}\|\Delta b\|\right)=o_P(1),
\end{aligned}
$$

where the second equality follows from Fact 5.10.18 of Bernstein (2005), and the last step follows from (1) of this lemma and $\sup_{C_\beta\in\mathsf{B}_T}\frac{1}{NT}\|\Delta b\|^2=O_P(1)$ shown above.

We now can conclude that $\sup_{(C_\beta,F)\in\mathsf{B}_T\times\mathsf{D}_F}|\Lambda_2|=o_P(1)$. Then the result follows.

(5). Let $\Delta=(\phi_1[\Delta_{\beta_0}],\ldots,\phi_N[\Delta_{\beta_0}])'$. The proof is similar to that for (4) of this lemma except that we need to use the fact that

$$
\frac{1}{NT}\|\Delta\|^2=\frac{1}{NT}\sum_{i=1}^{N}\|\phi_i[\Delta_{\beta_0}]\|^2=O_P\left(Tm^{-2\mu_2}\right)=o_P(1).
$$

(6). Note that

$$
\sup_{F\in\mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i[\Delta_{\beta_0}]'M_F\phi_i[\Delta_{\beta_0}]\right|\le\frac{1}{NT}\sum_{i=1}^{N}\|\phi_i[\Delta_{\beta_0}]\|^2=o_P(1).
$$

(7). Write

$$
\sup_{F\in \mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i\left[\Delta_{\beta_0}\right]' M_F F_0\gamma_0(v_i)\right| = \sup_{F\in \mathsf{D}_F}\left|\frac{1}{NT}\mathrm{tr}\left(M_F F_0\Gamma_0'\Delta\right)\right|
$$

$$
\leq \frac{d_v}{NT}\sup_{F\in \mathsf{D}_F}\|M_F\|_{\mathrm{sp}}\|F_0\|_{\mathrm{sp}}\|\Gamma_0\|_{\mathrm{sp}}\|\Delta\|_{\mathrm{sp}} = o_P(1),
$$

where $\Delta = (\phi_1\left[\Delta_{\beta_0}\right],\ldots,\phi_N\left[\Delta_{\beta_0}\right])'$, and the second equality follows from that $\frac{1}{NT}\|\Delta\|^2 = o_P(1)$ as in (5) of this lemma.

(8). Write

$$
\sup_{(C_\beta,F)\in \mathsf{B}_T\times \mathsf{D}_F}\left|\frac{1}{NT}\sum_{i=1}^{N}\phi_i\left[\Delta_{\beta_0}\right]' M_F\left\{\phi_i\left[\beta_m\right]-\phi_i\left[\beta_{0,m}\right]\right\}\right|
$$

$$
\leq T\cdot\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\|\phi_i\left[\Delta_{\beta_0}\right]\|^2\right\}^{1/2}\cdot\sup_{C_\beta\in \mathsf{B}_T}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\|\phi_i\left[\beta_m\right]-\phi_i\left[\beta_{0,m}\right]\|^2\right\}^{1/2}
$$

$$
= T\cdot O_P(m^{-\mu_2})\cdot O_P(T^{-\frac{1}{2}}) = o_P(1),
$$

where the first inequality follows from Cauchy–Schwarz inequality; and the last equality follows from Assumption B.1 and the construction of B_T. ■

Proof of Lemma B.2:

Firstly, we define some variables: $\Delta\phi_i[\beta_m] = \phi_i[\beta_{0,m}] - \phi_i[\beta_m]$ for $i \geq 1$, and $\xi_F = \mathrm{vec}\,(M_F F_0)$ for $\forall F \in \mathsf{D}_F$. In addition, let $A_1 = \frac{1}{NT}\sum_{i=1}^{N} Z_i' M_F Z_i$, $A_2 = \frac{1}{NT}(\Gamma_0'\Gamma_0)\otimes I_T$, and $A_3 = \frac{1}{NT}\sum_{i=1}^{N}\gamma_0(v_i)\otimes(M_F Z_i)$. We are now ready to start the proof.

(1). By Lemma B.5, it is straightforward to obtain that

$$
\begin{aligned}
& R_{NT}(C_\beta, F) - R_{NT}(C_{\beta_0}, F_0) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\left(\Delta\phi_i[\beta_m] + F_0\gamma(v_i)\right)' M_F\left(\Delta\phi_i[\beta_m] + F_0\gamma(v_i)\right) + o_P(1) \\
&= (C_{\beta_0} - C_\beta)'\frac{1}{NT}\sum_{i=1}^{N} Z_i' M_F Z_i (C_{\beta_0} - C_\beta) + \frac{1}{NT}\mathrm{tr}\left(M_F F_0\Gamma_0'\Gamma_0 F_0' M_F\right) \\
&\quad + 2(C_{\beta_0} - C_\beta)'\frac{1}{NT}\sum_{i=1}^{N} Z_i' M_F F_0\gamma_0(v_i) + o_P(1) \\
&= (C_{\beta_0} - C_\beta)' A_1 (C_{\beta_0} - C_\beta) + \xi_F' A_2\xi_F + 2(C_{\beta_0} - C_\beta)' A_3'\xi_F + o_P(1) \\
&= \sqrt{T}(C_{\beta_0} - C_\beta)'\left(\frac{A_1 - A_3' A_2^{-1} A_3}{T}\right)\sqrt{T}(C_{\beta_0} - C_\beta) \\
&\quad + [\xi_F' + (C_{\beta_0} - C_\beta)' A_3' A_2^{-1}] A_2 [\xi_F + A_2^{-1} A_3 (C_{\beta_0} - C_\beta)] + o_P(1) \\
&= \sqrt{T}(C_{\beta_0} - C_\beta)'\Omega_\dagger(F)\sqrt{T}(C_{\beta_0} - C_\beta) \\
&\quad + [\xi_F' + (C_{\beta_0} - C_\beta)' A_3' A_2^{-1}] A_2 [\xi_F + A_2^{-1} A_3 (C_{\beta_0} - C_\beta)] + o_P(1),
\end{aligned}
$$

where $\Omega_\dagger(F)$ has been defined in Assumption B.3. Then by the same arguments as in Bai (2009, p. 1265), we obtain that $\sqrt{T}\|C_{\beta_0} - \widehat{C}_\beta\| = o_P(1)$. Therefore, further write

$$
\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2} = \sqrt{T}\|\widehat{\beta}_m - \beta_{0,m}\|_{L^2} + \sqrt{T}\|\Delta_{\beta_0}\|_{L^2}
$$

$$= \sqrt{T}\|\widehat{C}_\beta - C_{\beta_0}\| + \sqrt{T}\|\Delta_{\beta_0}\|_{L^2} = o_P(1),$$

where the last step follows from $\sqrt{T}\|C_{\beta_0} - \widehat{C}_\beta\| = o_P(1)$ and Assumption B.1.3.

(2). By the first result of this lemma, we further obtain that

$$0 \geq R_{NT}(\widehat{C}_\beta, \widehat{F}) - R_{NT}(C_{\beta_0}, F_0) = \frac{1}{NT}\mathrm{tr}\left[\left(F_0' M_{\widehat{F}} F_0\right)\left(\Gamma_0'\Gamma_0\right)\right] + o_P(1),$$

which indicates that $\frac{1}{NT}\mathrm{tr}\left[\left(F_0' M_{\widehat{F}} F_0\right)\left(\Gamma_0'\Gamma_0\right)\right] = o_P(1)$. As in Bai (2009, p. 1265), we can further conclude that $\frac{1}{T}\mathrm{tr}\left(F_0' M_{\widehat{F}} F_0\right) = o_P(1)$, $\frac{1}{T}\widehat{F}' F_0$ is invertible with probability approaching one, and $\left\|P_{\widehat{F}} - P_{F_0}\right\| = o_P(1)$. Then the second result follows.

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

$$\begin{aligned}
\widehat{F} V_{NT} &= \frac{1}{NT}\sum_{i=1}^N (Y_i - \phi_i[\widehat{\beta}_m])(Y_i - \phi_i[\widehat{\beta}_m])'\widehat{F} \\
&= \frac{1}{NT}\sum_{i=1}^N (\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m])(\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m])'\widehat{F} \\
&\quad + \frac{1}{NT}\sum_{i=1}^N (\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m])\left(F_0\gamma_0(v_i)\right)'\widehat{F} + \frac{1}{NT}\sum_{i=1}^N (F_0\gamma_0(v_i))(\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m])'\widehat{F} \\
&\quad + \frac{1}{NT}\sum_{i=1}^N (\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m])e_i'\widehat{F} + \frac{1}{NT}\sum_{i=1}^N e_i(\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m])'\widehat{F} \\
&\quad + \frac{1}{NT}\sum_{i=1}^N e_i e_i'\widehat{F} + \frac{1}{NT}\sum_{i=1}^N F_0\gamma_0(v_i)e_i'\widehat{F} + \frac{1}{NT}\sum_{i=1}^N e_i\gamma_0(v_i)'F_0'\widehat{F} \\
&\quad + \frac{1}{NT}\sum_{i=1}^N F_0\gamma_0(v_i)\gamma_0(v_i)'F_0'\widehat{F} \\
&:= I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{5NT}(\widehat{\beta}_m, \widehat{F}) + I_{6NT}(\widehat{F}) + \cdots + I_{9NT}(\widehat{F}),
\end{aligned}$$

where the definitions of $I_{1NT}(\beta, F)$ to $I_{5NT}(\beta, F)$ and $I_{6NT}(F)$ to $I_{9NT}(F)$ should be obvious.

Note that $I_{9NT}(\widehat{F}) = F_0(\Gamma_0'\Gamma_0/N)(F_0'\widehat{F}/T)$. Thus, we can write

$$\begin{aligned}
&\widehat{F}V_{NT} - F_0(\Gamma_0'\Gamma_0/N)(F_0'\widehat{F}/T) \\
&= I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{5NT}(\widehat{\beta}_m, \widehat{F}) + I_{6NT}(\widehat{F}) + \cdots + I_{8NT}(\widehat{F}).
\end{aligned} \tag{B.6}$$

Right multiplying each side of (B.6) by $(F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}$, we obtain

$$\begin{aligned}
&\widehat{F}V_{NT}(F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1} - F_0 \\
&= \left[I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{8NT}(\widehat{F})\right](F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}.
\end{aligned} \tag{B.7}$$

We examine each term on the right hand side of (B.7) and show that V_{NT} is non–singular. Write

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\left\|\widehat{F}V_{NT}(F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1} - F_0\right\| \\
&\leq \frac{1}{\sqrt{T}}\left[\|I_{1NT}(\widehat{\beta}_m, \widehat{F})\| + \cdots + \|I_{8NT}(\widehat{F})\|\right]\cdot\|(F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}\| \\
&\leq O_P(1)\frac{1}{\sqrt{T}}\left[\|I_{1NT}(\widehat{\beta}_m, \widehat{F})\| + \cdots + \|I_{8NT}(\widehat{F})\|\right].
\end{aligned} \tag{B.8}$$

Thus, we focus on each term on the right hand side of (B.8).

For $I_{1NT}(\widehat{\beta}_m, \widehat{F})$, we have

$$\begin{aligned}\frac{1}{\sqrt{T}}\|I_{1NT}(\widehat{\beta}_m, \widehat{F})\| &\leq \frac{\sqrt{d_v}}{NT}\sum_{i=1}^{N}\|\phi_i[\beta_{0,m}] - \phi_i[\widehat{\beta}_m]\|^2 + \frac{\sqrt{d_v}}{NT}\sum_{i=1}^{N}\|\phi_i[\Delta_{\beta_0}]\|^2 \\ &= O_P(T\|\widehat{C}_\beta - C_{\beta_0}\|^2) + O_P(Tm^{-2\mu_2}) = o_P(1),\end{aligned}$$

where the first equality follows from the (1) of Lemma A.2 and the proof for (5) of Lemma B.5; and the second equality follows from (1) of this lemma.

For $I_{2NT}(\widehat{\beta}_m, \widehat{F})$, write

$$\begin{aligned}\frac{1}{\sqrt{T}}\|I_{2NT}(\widehat{\beta}_m, \widehat{F})\| &\leq \sqrt{d_v}\Big\{\frac{1}{NT}\sum_{i=1}^{N}\|\phi_i[\beta_0] - \phi_i[\widehat{\beta}_m]\|^2\Big\}^{1/2}\Big\{\frac{1}{NT}\sum_{i=1}^{N}\|F_0\gamma_0(v_i)\|^2\Big\}^{1/2} \\ &= O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) = o_P(1),\end{aligned}$$

where the first inequality follows from Cauchy–Schwarz inequality; and the last line follows from the same arguments given for $I_{1NT}(\widehat{\beta}_m, \widehat{F})$ and the fact that $\frac{1}{NT}\sum_{i=1}^{N}\|F_0\gamma_0(v_i)\|^2 = O_P(1)$.

Similar to the development for $\frac{1}{\sqrt{T}}\|I_{2NT}(\widehat{\beta}_m, \widehat{F})\|$, we have for $j = 3, 4, 5$,

$$\frac{1}{\sqrt{T}}\|I_{jNT}(\widehat{\beta}_m, \widehat{F})\| = O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) = o_P(1).$$

By (1) of Lemma B.5 and $\frac{1}{\sqrt{T}}\|\widehat{F}\| = O(1)$, we obtain

$$\frac{1}{\sqrt{T}}\|I_{6NT}(\widehat{F})\| = O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right).$$

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

$$\begin{aligned}E\left\|\frac{1}{NT}\sum_{i=1}^{N}F_0\gamma_0(v_i)e_i'\right\|^2 &= \sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E[f_{0t}'\gamma_0(v_i)e_{is}f_{0t}'\gamma_0(v_j)e_{js}] \\ &\leq O(1)\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|\sigma_{ij}| \\ &\leq O(1)\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|\sigma_{ij}| = O\left(\frac{1}{N}\right),\end{aligned}$$

where the first inequality follows from Assumption B.2, and Assumption B.1.4. We then can conclude that

$$\frac{1}{\sqrt{T}}\|I_{7NT}(\widehat{F})\| = \frac{1}{\sqrt{T}}\|I_{8NT}(\widehat{F})\| = O_P\left(\frac{1}{\sqrt{N}}\right).$$

Based on the above analysis and by left multiplying (B.6) by $\widehat{F}'/T$, we obtain

$$V_{NT} - (\widehat{F}'F_0/T)(\Gamma_0'\Gamma_0/N)(F_0'\widehat{F}/T) = \frac{1}{T}\widehat{F}'\left[I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{8NT}(\widehat{F})\right] = o_P(1).$$

Thus, $V_{NT} = (\widehat{F}'F_0/T)(\Gamma_0'\Gamma_0/N)(F_0'\widehat{F}/T) + o_P(1)$. When proving the second result of this lemma, we have shown that $F_0'\widehat{F}/T$ is non–singular with probability approaching one, which implies that V_{NT} is invertible with probability approaching one. We now left multiply (B.6) by F_0'/T to obtain

$$(F_0'\widehat{F}/T)V_{NT} = (F_0'F_0/T)(\Gamma_0'\Gamma_0/N)(F_0'\widehat{F}/T) + o_P(1)$$

based on the above analysis. It shows that the columns of $F_0'\widehat{F}/T$ are the (non-normalized) eigenvectors of the matrix $(F_0'F_0/T)(\Gamma_0'\Gamma_0/N)$, and V_{NT} consists of the eigenvalues of the same matrix (in the limit). Thus, the third result of this lemma follows.

(4). According to the above analysis, (B.8) can be summarized by

$$\frac{1}{\sqrt{T}}\|\widehat{F}\Pi_{NT}^{-1} - F_0\| = O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right).$$

(5). According to (B.7),

$$\frac{1}{T}F_0'(\widehat{F} - F_0\Pi_{NT}) = \frac{1}{T}F_0'[I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{8NT}(\widehat{F})]V_{NT}^{-1}.$$

Note that $V_{NT}^{-1} = O_P(1)$, so we focus on $\frac{1}{T}F_0'[I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{8NT}(\widehat{F})]$ below.

By the proof given for the first result of this lemma, it is easy to show that

$$\frac{1}{T}\|F_0'[I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{5NT}(\widehat{\beta}_m, \widehat{F})]\| = O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}).$$

We now consider $\|\frac{1}{T}F_0'I_{6NT}(\widehat{F})\|$. Write

$$\frac{1}{T}\|F_0'I_{6NT}(\widehat{F})\| \le \frac{1}{T}\left(\frac{1}{NT}\sum_{i=1}^N \|F_0'e_i\|^2\right)^{1/2}\left(\frac{1}{NT}\sum_{i=1}^N \|e_i'\widehat{F}\|^2\right)^{1/2}.$$

Note that $\frac{1}{NT}\sum_{i=1}^N \|F_0'e_i\|^2 = O_P(1)$. For $\frac{1}{NT}\sum_{i=1}^N \|e_i'\widehat{F}\|^2$, write

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^N \|e_i'\widehat{F}\|^2 &\le \frac{2}{NT}\sum_{i=1}^N \left\|e_i'F_0\Pi_{NT}\right\|^2 + \frac{2}{NT}\sum_{i=1}^N \mathrm{tr}\left\{e_i'(\widehat{F} - F_0\Pi_{NT})(\widehat{F} - F_0\Pi_{NT})'e_i\right\} \\ &= \frac{2}{NT}\sum_{i=1}^N \left\|e_i'F_0\Pi_{NT}\right\|^2 + \frac{2}{NT}\mathrm{tr}\left\{(\widehat{F} - F_0\Pi_{NT})(\widehat{F} - F_0\Pi_{NT})'e'e\right\} \\ &\le O_P(1) + O(1)\frac{1}{N}\|e'e\|\frac{1}{T}\|\widehat{F} - F_0\Pi_{NT}\|^2,\end{aligned}$$

where e has been defined in (1) of Lemma B.5. In connection with (1) of Lemma B.5 and the result (4) of this lemma, it then gives that

$$\begin{aligned}\frac{1}{T}\|F_0'I_{6NT}(\widehat{F})\| &= O_P(1)\frac{1}{T} + O_P\left(\frac{1}{\sqrt{T}}\right)\left\{\frac{1}{NT}\|e'e\|\frac{1}{T}\|\widehat{F} - F_0\Pi_{NT}\|^2\right\}^{1/2} \\ &= O_P(1)\frac{1}{T} + O_P\left(\frac{1}{\sqrt{T}}\right)O_P\left(\frac{1}{\sqrt[4]{N}} + \frac{1}{\sqrt[4]{T}}\right)O_P\left(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2} + \frac{1}{\sqrt{N}}\right),\end{aligned}$$

where the second equality follows from (1) of Lemma B.5 and the second result of this lemma.

For $\frac{1}{T}\|F_0'I_{7NT}(\widehat{F})\|$, we have

$$\begin{aligned}\frac{1}{T}\|F_0'I_{7NT}(\widehat{F})\| \le{}& \frac{1}{T}\|F_0'F_0\| \cdot \left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^N \gamma_0(v_i)e_i'\right\| \cdot \frac{1}{\sqrt{T}}\|\widehat{F} - F_0\Pi_{NT}\| \\ &+\frac{1}{T}\|F_0'F_0\| \cdot \left\|\frac{1}{NT}\sum_{i=1}^N \gamma_0(v_i)e_i'F_0\right\| \cdot \|\Pi_{NT}\|.\end{aligned}$$

By Assumption B.1.4, $\left\|\frac{1}{T}F_0'F_0\right\| = O_P(1)$. Also, $\|\Pi_{NT}\|$ and $\frac{1}{\sqrt{T}}\|\widehat{F} - F_0\Pi_{NT}\|$ have been studied in results (3) and (4) of this lemma respectively. Therefore, focus on $\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^N \gamma_0(v_i)e_i'\right\|$ and $\left\|\frac{1}{NT}\sum_{i=1}^N \gamma_0(v_i)e_i'F_0\right\|$ below. Write

$$E\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^N \gamma_0(v_i)e_i'\right\|^2 = \frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T E[\gamma_0'(v_i)\gamma_0(v_j)e_{it}e_{jt}] \leq O\left(\frac{1}{N}\right) \tag{B.9}$$

and similar to (A.2)

$$E\left\|\frac{1}{NT}\sum_{i=1}^N \gamma_0(v_i)e_i'F_0\right\|^2 = O\left(\frac{1}{NT}\right), \tag{B.10}$$

which immediately yields

$$\begin{aligned}\frac{1}{T}\|F_0'I_{7NT}(\widehat{F})\| &= O_P\left(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}\cdot\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{\sqrt{NT}}\right)\\ &\leq O_P(T\|\widehat{\beta}_m - \beta_0\|_{L^2}^2) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right).\end{aligned}$$

Similarly, $\frac{1}{T}\|F_0'I_{8NT}(\widehat{F})\| = O_P(T\|\widehat{\beta}_m - \beta_0\|_{L^2}^2) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right)$.

Based on the above analysis, we have

$$\frac{1}{T}\|F_0'(\widehat{F} - F_0\Pi_{NT})\| = O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right), \tag{B.11}$$

which further indicates

$$\begin{aligned}\frac{1}{T}\|\widehat{F}'(\widehat{F} - F_0\Pi_{NT})\| &\leq \frac{1}{T}\|(\widehat{F} - F_0\Pi_{NT})'(\widehat{F} - F_0\Pi_{NT})\| + \|\Pi_{NT}\|\cdot\frac{1}{T}\|F_0'(\widehat{F} - F_0\Pi_{NT})\|\\ &= O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right).\end{aligned} \tag{B.12}$$

(6). Note that (B.11) and (B.12) indicate

$$\frac{1}{T}\Pi_{NT}'F_0'\widehat{F} - \frac{1}{T}\Pi_{NT}'F_0'F_0\Pi_{NT} = O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right)$$

and

$$I_{d_v} - \frac{1}{T}\Pi_{NT}'F_0'\widehat{F} = O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right).$$

Summing up the above two equations yields

$$I_{d_v} - \frac{1}{T}\Pi_{NT}'F_0'F_0\Pi_{NT} = O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right). \tag{B.13}$$

Note that it is easy to show that

$$\begin{aligned}\left\|P_{\widehat{F}} - P_{F_0}\right\|^2 &= \mathrm{tr}\left[(P_{\widehat{F}} - P_{F_0})^2\right] = \mathrm{tr}\left[P_{\widehat{F}} - P_{\widehat{F}}P_{F_0} - P_{F_0}P_{\widehat{F}} + P_{F_0}\right]\\ &= \mathrm{tr}\left[I_{d_v}\right] - 2\cdot\mathrm{tr}\left[P_{\widehat{F}}P_{F_0}\right] + \mathrm{tr}\left[I_{d_v}\right] = 2\cdot\mathrm{tr}[I_{d_v} - \widehat{F}'P_{F_0}\widehat{F}/T]\end{aligned}$$

and, when proving this lemma, we have shown that

$$\frac{F_0'\widehat{F}}{T} = \frac{F_0'F_0}{T}\Pi_{NT} + O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right).$$

Therefore, we can write

$$\widehat{F}'P_{F_0}\widehat{F}/T = \Pi_{NT}'\left(\frac{F_0'F_0}{T}\right)\Pi_{NT} + O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right),$$

which in connection with (B.13) gives

$$\widehat{F}'P_{F_0}\widehat{F}/T = I_{d_v} + O_P(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}) + O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{T}\right).$$

Then the proof of the last result of this lemma is completed. ■

Proof of Lemma B.3:

(1). By the first equation of (B.4), we write

$$\begin{aligned}
\widehat{C}_\beta - C_{\beta_0} &= \left(\sum_{i=1}^N Z_i'M_{\widehat{F}}Z_i\right)^{-1}\sum_{i=1}^N Z_i'M_{\widehat{F}}e_i + \left(\sum_{i=1}^N Z_i'M_{\widehat{F}}Z_i\right)^{-1}\sum_{i=1}^N Z_i'M_{\widehat{F}}F_0\gamma_0(v_i) \\
&\quad + \left(\sum_{i=1}^N Z_i'M_{\widehat{F}}Z_i\right)^{-1}\sum_{i=1}^N Z_i'M_{\widehat{F}}\phi_i[\Delta_{\beta_0}] \\
&:= \Lambda_1 + \Lambda_2 + \Lambda_3,
\end{aligned}$$

where the definitions of Λ_1, Λ_2 and Λ_3 should be obvious. By Lemma A.2 and Lemma B.2, it is easy to know that

$$\frac{1}{NT^2}\sum_{i=1}^N Z_i'M_{\widehat{F}}Z_i = \frac{1}{NT^2}\sum_{i=1}^N Z_i'M_{F_0}Z_i + o_P(1) = \frac{1}{2}\Sigma_m + o_P(1).$$

Similar to (A.1), we obtain $\|\Lambda_3\| = O_P\left(m^{-\mu_2}\right)$. In the following, we focus on studying Λ_2 at first, and then turn to Λ_1.

In the following, let $\Xi_{NT} = (F_0'\widehat{F}/T)^{-1}(\Gamma_0'\Gamma_0/N)^{-1}$ for simplicity, so $\Pi_{NT}^{-1} = V_{NT}\Xi_{NT}$. We now start our investigation on Λ_2, and write

$$\begin{aligned}
\frac{1}{NT^2}\sum_{i=1}^N Z_i'M_{\widehat{F}}F_0\gamma_0(v_i) &= -\frac{1}{NT^2}\sum_{i=1}^N Z_i'M_{\widehat{F}}(\widehat{F}\Pi_{NT}^{-1} - F_0)\gamma_0(v_i) \\
&= -\frac{1}{NT^2}\sum_{i=1}^N Z_i'M_{\widehat{F}}\left[I_{1NT}(\widehat{\beta}_m, \widehat{F}) + \cdots + I_{8NT}(\widehat{F})\right]\Xi_{NT}\gamma_0(v_i) \\
&:= -(J_{1NT} + \cdots + J_{8NT}),
\end{aligned}$$

where the second equality follows from (B.7); and the definitions of J_{1NT} to J_{8NT} should be obvious. In view of the decomposition of J_{2NT} below, it is actually easy to show that $\|J_{1NT}\| = o_P(\|\widehat{C}_\beta - C_{\beta_0}\|)$. Thus, we start from J_{2NT} and write

$$\begin{aligned}
J_{2NT} &= \frac{1}{NT^2}\sum_{i=1}^N Z_i'M_{\widehat{F}}\frac{1}{NT}\sum_{j=1}^N(\phi_j[\beta_{0,m}] - \phi_j[\widehat{\beta}_m])\left(F_0\gamma_0(v_j)\right)'\widehat{F}\Xi_{NT}\gamma_0(v_i) \\
&\quad + \frac{1}{NT^2}\sum_{i=1}^N Z_i'M_{\widehat{F}}\frac{1}{NT}\sum_{j=1}^N\phi_j[\Delta_{\beta_0}]\left(F_0\gamma_0(v_j)\right)'\widehat{F}\Xi_{NT}\gamma_0(v_i)
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{j=1}^N Z_i' M_{\widehat{F}} Z_j \gamma_0(v_j)'\Big(\frac{F_0'\widehat{F}}{T}\Big)\Big(\frac{F_0'\widehat{F}}{T}\Big)^{-1}\Big(\frac{\Gamma_0'\Gamma_0}{N}\Big)^{-1}\gamma_0(v_i)(\widehat{C}_\beta - C_{\beta_0})\\
&\quad + \frac{1}{NT^2}\sum_{i=1}^N Z_i' M_{\widehat{F}} \frac{1}{NT}\sum_{j=1}^N \phi_j[\Delta_{\beta_0}]\left(F_0\gamma_0(v_j)\right)'\widehat{F}\Xi_{NT}\gamma_0(v_i)\\
&= -\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{j=1}^N Z_i' M_{\widehat{F}} Z_j \gamma_0(v_j)'\Big(\frac{\Gamma_0'\Gamma_0}{N}\Big)^{-1}\gamma_0(v_i)(\widehat{C}_\beta - C_{\beta_0})\\
&\quad + \frac{1}{NT^2}\sum_{i=1}^N Z_i' M_{\widehat{F}} \frac{1}{NT}\sum_{j=1}^N \phi_j[\Delta_{\beta_0}]\left(F_0\gamma_0(v_j)\right)'\widehat{F}\Xi_{NT}\gamma_0(v_i)\\
&:= J_{2NT,1} + J_{2NT,2}.
\end{aligned}
$$

By a derivation similar to (A.1), $\|(\sum_{i=1}^N Z_i' M_{\widehat{F}} Z_i)^{-1} NT^2 J_{2NT,2}\| = O_P\left(m^{-\mu_2}\right)$, so negligible. We will further study $J_{2NT,1}$ later.

For J_{3NT}, write

$$
\begin{aligned}
J_{3NT} &= \frac{1}{NT^2}\sum_{i=1}^N Z_i' M_{\widehat{F}} \frac{1}{NT}\sum_{j=1}^N F_0\gamma_0(v_j)(\phi_j[\beta_0] - \phi_j[\widehat{\beta}_m])'\widehat{F}\Xi_{NT}\gamma_0(v_i)\\
&= -\frac{1}{NT^2}\sum_{i=1}^N Z_i' M_{\widehat{F}}(\widehat{F}\Pi_{NT}^{-1} - F_0)\frac{1}{NT}\sum_{j=1}^N \gamma_0(v_j)(\phi_j[\beta_0] - \phi_j[\widehat{\beta}_m])'\widehat{F}\Xi_{NT}\gamma_0(v_i)\\
&:= \frac{1}{NT^2}\sum_{i=1}^N Z_i' M_{\widehat{F}} J_{3NT,i},
\end{aligned}
$$

where the definition of $J_{3NT,i}$ is obvious. By the analysis similar to (A.1), we just need to focus on $\frac{1}{NT^2}\sum_{i=1}^N \|J_{3NT,i}\|^2$ in order to show $\|(\sum_{i=1}^N Z_i' M_{\widehat{F}} Z_i)^{-1} NT^2 J_{3NT}\| = o_P(\|\widehat{C}_\beta - C_{\beta_0}\|)$. Thus, write

$$
\begin{aligned}
\frac{1}{NT^2}\sum_{i=1}^N \|J_{3NT,i}\|^2 &\le \frac{\|\widehat{F}\Pi_{NT}^{-1} - F_0\|^2}{NT^2}\sum_{i=1}^N \left\|\frac{1}{NT}\sum_{j=1}^N \gamma_0(v_j)(\phi_j[\beta_0] - \phi_j[\widehat{\beta}_m])'\right\|^2 \|\widehat{F}\Xi_{NT}\gamma_0(v_i)\|^2\\
&\le O_P(1)\frac{1}{T}\|\widehat{F}\Pi_{NT}^{-1} - F_0\|^2\left(\frac{1}{N}\sum_{j=1}^N\left\{\frac{1}{T^2}\|\phi_j[\beta_0] - \phi_j[\widehat{\beta}_m]\|^2\right\}^{1/2}\right)^2\\
&= o_P(\|\widehat{C}_\beta - C_{\beta_0}\|^2),
\end{aligned}
$$

where the second inequality follows from Assumption B.1.4, $\Xi_{NT} = O_P(1)$ and $\frac{1}{\sqrt{T}}\|\widehat{F}\| = O(1)$; and the equality follows from $\frac{1}{\sqrt{T}}\|\widehat{F}\Pi_{NT}^{-1} - F_0\| = o_P(1)$. Thus,

$$
\left\|\left(\sum_{i=1}^N Z_i' M_{\widehat{F}} Z_i\right)^{-1} NT^2 J_{3NT}\right\| = o_P(\|\widehat{C}_\beta - C_{\beta_0}\|).
$$

For J_{4NT}, write

$$
\begin{aligned}
J_{4NT} &= \frac{1}{N^2T^3}\sum_{i=1}^N\sum_{j=1}^N Z_i' M_{\widehat{F}}(\phi_j[\beta_{0,m}] - \phi_j[\widehat{\beta}_m])e_j' F_0\Pi_{NT}\Xi_{NT}\gamma_0(v_i)\\
&\quad + \frac{1}{N^2T^3}\sum_{i=1}^N\sum_{j=1}^N Z_i' M_{\widehat{F}}\phi_j[\Delta_{\beta_0}]e_j' F_0\Pi_{NT}\Xi_{NT}\gamma_0(v_i)
\end{aligned}
$$

$$+\frac{1}{N^2T^3}\sum_{i=1}^{N}\sum_{j=1}^{N} Z_i'M_{\widehat{F}}(\phi_j[\beta_0]-\phi_j[\widehat{\beta}_m])e_j'(\widehat{F}-F_0\Pi_{NT})\Xi_{NT}\gamma_0(v_i)$$
$$:= J_{4NT,1}+J_{4NT,2}+J_{4NT,3}.$$

For $J_{4NT,1}$, write

$$\begin{aligned}\|J_{4NT,1}\| &= \left\|\frac{1}{N^2T^3}\sum_{i=1}^{N}\sum_{j=1}^{N} Z_i'M_{\widehat{F}}Z_j(\widehat{C}_\beta-C_{\beta_0})e_j'F_0\Pi_{NT}\Xi_{NT}\gamma_0(v_i)\right\| \\ &\leq O_P(1)\frac{1}{NT}\sum_{i=1}^{N}\left\|\frac{Z_i'}{\sqrt{T}}M_{\widehat{F}}\right\|\cdot\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{Z_j}{\sqrt{T}}\right\|\frac{1}{T}\|e_j'F_0\|\cdot\|\widehat{C}_\beta-C_{\beta_0}\| \\ &\leq \frac{1}{T}O_P(\sqrt{mT})\cdot O_P(\sqrt{mT})\cdot O_P(T^{-1/2})\cdot\|\widehat{C}_\beta-C_{\beta_0}\| \\ &= o_P(\|\widehat{C}_\beta-C_{\beta_0}\|).\end{aligned}$$

Thus, $\|J_{4NT,1}\|$ is negligible. Similarly, we can show both $\|J_{4NT,2}\|$ and $\|J_{4NT,3}\|$ are negligible by accounting for $\frac{1}{T}\|\phi_j[\Delta_{\beta_0}]\|^2=O_P(Tm^{-2\mu_2})$ and $\frac{1}{\sqrt{T}}\|\widehat{F}\Pi_{NT}^{-1}-F_0\|=o_P(1)$, respectively. Analogous to the derivation of J_{3NT} and J_{4NT}, we can conclude that $\|J_{5NT}\|$ is negligible.

Below, we take a careful look at J_{6NT}. According to Assumption B.3, let $\Omega_e=E[e_ie_i']=\sigma_e^2I_T$ for notational simplicity. Thus, write

$$\begin{aligned}J_{6NT} &= \frac{1}{NT^3}\sum_{i=1}^{N}Z_i'M_{\widehat{F}}\Omega_e\widehat{F}\Xi_{NT}\gamma_0(v_i)+\frac{1}{NT^2}\sum_{i=1}^{N}Z_i'M_{\widehat{F}}\frac{1}{NT}\sum_{j=1}^{N}\left(e_je_j'-\Omega_e\right)\widehat{F}\Xi_{NT}\gamma_0(v_i) \\ &:= J_{6NT,1}+J_{6NT,2}.\end{aligned}$$

We focus on $J_{6NT,2}$ at first.

$$\begin{aligned}J_{6NT,2} &= \frac{1}{N^2T^3}\sum_{i=1}^{N}\sum_{j=1}^{N}Z_i'\left(e_je_j'-\Omega_e\right)\widehat{F}\Xi_{NT}\gamma_0(v_i) \\ &\quad+\frac{1}{N^2T^3}\sum_{i=1}^{N}\sum_{j=1}^{N}Z_i'P_{\widehat{F}}\left(e_je_j'-\Omega_e\right)\widehat{F}\Xi_{NT}\gamma_0(v_i) \\ &:= J_{6NT,21}+J_{6NT,22}.\end{aligned}$$

Further decompose $J_{6NT,21}$ as

$$\begin{aligned}J_{6NT,21} &= \frac{1}{N^2T^3}\sum_{i=1}^{N}\sum_{j=1}^{N}Z_i'\left(e_je_j'-\Omega_e\right)F_0\Pi_{NT}\Xi_{NT}\gamma_0(v_i) \\ &\quad+\frac{1}{N^2T^3}\sum_{i=1}^{N}\sum_{j=1}^{N}Z_i'\left(e_je_j'-\Omega_e\right)(\widehat{F}-F_0\Pi_{NT})\Xi_{NT}\gamma_0(v_i) \\ &:= J_{6NT,211}+J_{6NT,212}.\end{aligned}$$

Let $\Omega_{e,ts}$ be the $(t,s)^{th}$ element of Ω_e, and $a_{is}=\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{t=1}^{T}\frac{z_{it}}{\sqrt{T}}(e_{jt}e_{js}-\Omega_{e,ts})$. Then by a development similar to Jiang et al. (2017, pp. 30–31), we obtain that

$$\|J_{6NT,211}\| = \left\|\frac{1}{N^2T^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}z_{it}\left(e_{jt}e_{js}-\Omega_{e,ts}\right)f_{0s}'\Pi_{NT}\Xi_{NT}\gamma_0(v_i)\right\|$$

$$= \frac{1}{N^{\frac{1}{2}}T} \cdot \left\| \frac{1}{N}\sum_{i=1}^{N} \left(\frac{1}{N^{\frac{1}{2}}T^2} \sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} z_{it}\left(e_{jt}e_{js} - \Omega_{e,ts}\right) f_{0s}' \right) \Pi_{NT}\Xi_{NT}\gamma_0(v_i) \right\|$$
$$= o_P\left(\frac{m^{\frac{1}{2}}}{N^{\frac{1}{2}}T} \right)$$

and

$$\|J_{6NT,212}\| \leq \frac{1}{N^{\frac{3}{2}}T^2} \sum_{i=1}^{N}\sum_{s=1}^{T} \left\| a_{is}(\widehat{f}_s' - f_{0s}'\Pi_{NT})\Xi_{NT}\gamma_0(v_i) \right\|$$
$$\leq \frac{1}{N^{\frac{1}{2}}T}\frac{1}{N}\sum_{i=1}^{N} \left\{ \frac{1}{T}\sum_{s=1}^{T}\|a_{is}\|^2 \right\}^{1/2} \left\{ \frac{1}{T}\sum_{s=1}^{T} \left\| \widehat{f}_s' - f_{0s}'\Pi_{NT} \right\|^2 \right\}^{1/2} \|\Xi_{NT}\gamma_0(v_i)\|$$
$$\leq O_P(1)\frac{\sqrt{m}}{N^{\frac{1}{2}}T}\left(\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2} + \frac{1}{\sqrt{N}} \right) = O_P(1)\frac{\sqrt{m}\|\widehat{\beta}_m - \beta_0\|_{L^2}}{\sqrt{NT}} + O_P\left(\frac{\sqrt{m}}{NT} \right).$$

Thus, we can conclude that

$$\|J_{6NT,21}\| = o_P\left(\frac{m^{\frac{1}{2}}}{N^{\frac{1}{2}}T} \right) + O_P\left(\frac{\sqrt{m}\|\widehat{\beta}_m - \beta_0\|_{L^2}}{\sqrt{NT}} \right) + O_P\left(\frac{\sqrt{m}}{NT} \right).$$

Similarly, we can obtain

$$\|J_{6NT,22}\| \leq O(1)\frac{1}{T^{\frac{1}{2}}} \cdot \frac{1}{N}\sum_{i=1}^{N} \left\| \frac{Z_i'\widehat{F}}{T^{\frac{3}{2}}} \right\| \left\| \frac{1}{NT^2}\sum_{j=1}^{N} \widehat{F}'\left(e_je_j' - \Omega_e\right)\widehat{F} \right\|$$
$$= o_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + o_P\left(\frac{m^{\frac{1}{2}}}{N^{\frac{1}{2}}T} \right).$$

Therefore, $\|J_{6NT,2}\| = o_P(\|\widehat{\beta}_m - \beta_0\|_{L^2}) + o_P\left(\frac{m^{\frac{1}{2}}}{N^{\frac{1}{2}}T} \right)$. We will consider $J_{6NT,1}$ together with $J_{2NT,1}$ and J_{8NT} later on.

We now have only one term J_{7NT} left to consider.

$$J_{7NT} = \frac{1}{NT^2}\sum_{i=1}^{N} Z_i' M_{\widehat{F}}(F_0 - \widehat{F}\Pi_{NT}^{-1})\frac{1}{NT}\sum_{j=1}^{N}\gamma_0(v_j)e_j'\widehat{F}\Xi_{NT}\gamma_0(v_i).$$

Notice that

$$\frac{1}{NT}\sum_{j=1}^{N}\gamma_0(v_j)e_j'\widehat{F} = \frac{1}{NT}\sum_{j=1}^{N}\gamma_0(v_j)e_j'F_0 + \frac{1}{NT}\sum_{j=1}^{N}\gamma_0(v_j)e_j'(F_0 - \widehat{F}\Pi_{NT}^{-1})$$
$$= O_P\left(\frac{1}{\sqrt{NT}} \right) + \left\| \frac{1}{N\sqrt{T}}\sum_{j=1}^{N}\gamma_0(v_j)e_j' \right\| \frac{1}{\sqrt{T}}\|F_0 - \widehat{F}\Pi_{NT}^{-1}\|$$
$$= O_P\left(\frac{1}{\sqrt{NT}} \right) + O_P\left(\frac{1}{\sqrt{N}} \right)\frac{1}{\sqrt{T}}\|F_0 - \widehat{F}\Pi_{NT}^{-1}\|,$$

where the second equality follows from (B.10), and the third equality follows from (B.9). Then, similar to the arguments for $J6$ of Bai (2009, pp. 1271–1272), $\|J_{7NT}\| = o_P\left(\frac{m^{\frac{1}{2}}}{N^{\frac{1}{2}}T} \right) + o_P(\|\widehat{C}_\beta - C_{\beta_0}\|)$.

Based on the above analysis, we have

$$\widehat{C}_\beta - C_{\beta_0} + \Sigma_m^{-1}J_{2NT,1}$$

$$
\begin{aligned}
&= -\Sigma_m^{-1}\left\{\frac{1}{NT^2}\sum_{i=1}^N Z_i' M_{\widehat{F}} e_i + J_{6NT,1} + J_{8NT}\right\} + \text{negligible terms}\\
&= -\Sigma_m^{-1}\cdot\frac{1}{NT^2}\sum_{i=1}^N\left\{Z_i' M_{\widehat{F}} + \frac{1}{N}\sum_{j=1}^N Z_j' M_{\widehat{F}}\gamma_0(v_j)'(\Gamma_0'\Gamma_0/N)^{-1}\gamma_0(v_i)\right\} e_i\\
&\quad -\Sigma_m^{-1}\cdot J_{6NT,1} + \text{negligible terms}.
\end{aligned}
$$

Further organise the above equation, we have

$$
\begin{aligned}
\widehat{C}_\beta - C_{\beta_0} &= -A_{1NT}^{-1}\Sigma_m^{-1}\cdot\frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\left\{\frac{Z_i'}{\sqrt{T}}M_{\widehat{F}} + A_{3,i}\right\} e_i\\
&\quad -A_{1NT}^{-1}\Sigma_m^{-1}\cdot J_{6NT,1} + \text{negligible terms},
\end{aligned}
$$

where

$$
\begin{aligned}
A_{1NT} &= I_{md_x} - \Sigma_m^{-1}A_{2NT}\cdot(1+o_P(1)),\\
A_{2NT} &= \frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N\frac{Z_i'}{\sqrt{T}}M_{\widehat{F}}\frac{Z_j}{\sqrt{T}}\gamma_0(v_j)'\left(\frac{\Gamma_0'\Gamma_0}{N}\right)^{-1}\gamma_0(v_i),\\
A_{3,i} &= \frac{1}{N}\sum_{j=1}^N\frac{Z_j'}{\sqrt{T}}M_{\widehat{F}}\gamma_0(v_j)'(\Gamma_0'\Gamma_0/N)^{-1}\gamma_0(v_i).
\end{aligned}
$$

Note that

$$
\frac{N^{\frac{1}{2}}T}{m^{\frac{1}{2}}}J_{6NT,1} = \frac{\sqrt{N}}{\sqrt{mT}}\cdot\frac{1}{NT}\sum_{i=1}^N\frac{Z_i'}{\sqrt{T}}M_{\widehat{F}}\Omega_e\widehat{F}\Xi_{NT}\gamma_0(v_i) = O_P\left(\sqrt{\frac{N}{T}}\right) = O_P(1),
$$

where the last equality follows from the condition in the body of this lemma. Thus, we obtain that $\|J_{6NT,1}\| = O_P(1)\frac{m^{\frac{1}{2}}}{N^{\frac{1}{2}}T}$. Moreover, it is easy to show

$$
\frac{1}{NT^{3/2}}\sum_{i=1}^N\left\{\frac{Z_i'}{\sqrt{T}}M_{\widehat{F}} + A_{3,i}\right\} e_i = O_P\left(\frac{m^{\frac{1}{2}}}{N^{\frac{1}{2}}T}\right)
$$

Thus, the first result of this lemma follows.

(2). In connection with (6) of Lemma B.2 and (1) of this lemma, we immediately obtain that

$$
\begin{aligned}
\|P_{\widehat{F}} - P_{F_0}\| &= O_P\left(\sqrt{\sqrt{T}\|\widehat{\beta}_m - \beta_0\|_{L^2}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)\\
&= O_P\left(\sqrt[4]{m/(NT)}\right) + O_P\left(\sqrt[4]{T/m^{2\mu_2}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}
$$

The proof is now complete. ■

Proof of Theorem B.1:

Recall that we have denoted A_{1NT}, A_{2NT} and $A_{3,i}$ in the proof of Lemma B.3, and we will keep using these notations in what follows. By the definition of $\widehat{\beta}_m$, write

$$
\frac{N^{\frac{1}{2}}T}{m^{\frac{1}{2}}}(\widehat{\beta}_m(r) - \beta_0(r))
$$

$$
\begin{aligned}
&= \frac{N^{\frac{1}{2}}T}{m^{\frac{1}{2}}}\left[H_m'(r)\otimes I_{d_x}\right](\widehat{C}_\beta - C_{\beta_0}) + o_P(1)\\
&= -\frac{N^{\frac{1}{2}}T}{m^{\frac{1}{2}}}\left[H_m'(r)\otimes I_{d_x}\right]A_{1NT}^{-1}\Sigma_m^{-1}\cdot\frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\left\{\frac{Z_i'}{\sqrt{T}}M_{\widehat{F}} + A_{3,i}\right\}e_i + o_P(1)\\
&:= \Lambda + o_P(1),
\end{aligned}
$$

where the first equality follows from $\|\Delta_{\beta_0}(r)\| = O_P(\sqrt{T}m^{-\mu_2})$ and the condition $\frac{NT^2}{m^{2\mu_2}} \to 0$; and the second equality follows from the proof of Lemma B.3, and the conditions $\frac{mN}{T} \to 0$ and $\frac{NT^2}{m^{2\mu_2}} \to 0$.

We then just need to consider Λ. Start from $\frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N \frac{Z_i'}{\sqrt{T}}M_{\widehat{F}}e_i$.

$$
\begin{aligned}
\frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\frac{Z_i'}{\sqrt{T}}M_{\widehat{F}}e_i &= \frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\frac{Z_i'}{\sqrt{T}}M_{F_0}e_i + \frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\frac{Z_i'}{\sqrt{T}}(M_{\widehat{F}} - M_{F_0})e_i\\
&= \frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\frac{Z_i'}{\sqrt{T}}M_{F_0}e_i - \frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\frac{Z_i'}{\sqrt{T}}(P_{\widehat{F}} - P_{F_0})e_i\\
&:= D_1 - D_2.
\end{aligned}
$$

Firstly, we show $\left\|\frac{N^{\frac{1}{2}}T}{m^{\frac{1}{2}}}\left[H_m'(r)\otimes I_{d_x}\right]A_{1NT}^{-1}\Sigma_m^{-1}D_2\right\| = o_P(1)$. Let $U_{iT} = \frac{Z_i}{\sqrt{T}}$ and let $U_{iT,j}$ be the j^{th} column of U_{iT}. Write

$$
\begin{aligned}
D_2 &= \frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N U_{iT}'\left(\frac{\widehat{F}\widehat{F}'}{T} - P_{F_0}\right)e_i\\
&= \frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\frac{U_{iT}'(\widehat{F} - F_0\Pi_{NT})}{T}\Pi_{NT}'F_0'e_i + \frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\frac{U_{iT}'(\widehat{F} - F_0\Pi_{NT})}{T}(\widehat{F} - F_0\Pi_{NT})'e_i\\
&\quad + \frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\frac{U_{iT}'F_0\Pi_{NT}}{T}(\widehat{F} - F_0\Pi_{NT})'e_i + \frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N\frac{U_{iT}'F_0}{T}[\Pi_{NT}\Pi_{NT}' - (F_0'F_0/T)^{-1}]F_0'e_i\\
&:= D_{21} + D_{22} + D_{23} + D_{24},
\end{aligned}
$$

where the definitions of D_{21} to D_{24} should be obvious.

In the following, we let $D_{2\ell,j}$ be the j^{th} row of $D_{2\ell}$ for $\ell = 1, 2, 3, 4$. Thus, for D_{21}, consider

$$
\begin{aligned}
\|D_{21,j}\| &\le \left\|\frac{1}{NT^{\frac{3}{2}}}\sum_{i=1}^N(e_i'F_0)\otimes\frac{U_{iT,j}'}{\sqrt{T}}\right\|\cdot\left\|\frac{1}{\sqrt{T}}\,\mathrm{vec}\left[(\widehat{F} - F_0\Pi_{NT})\Pi_{NT}'\right]\right\|\\
&= O_P\left(\frac{1}{N^{\frac{1}{2}}T}\right)\frac{1}{\sqrt{T}}\|\widehat{F} - F_0\Pi_{NT}\|,
\end{aligned}
$$

where the equality follows from the development similar to (B.10). Summing up over j for $D_{21,j}$, we obtain that $\|D_{21}\| = O_P\left(\frac{m^{\frac{1}{2}}}{N^{\frac{1}{2}}T}\right)\frac{1}{\sqrt{T}}\|\widehat{F} - F_0\Pi_{NT}\|$.

For D_{22}, write

$$
\begin{aligned}
\|D_{22,j}\| &\le \left\|\frac{1}{NT}\sum_{i=1}^N e_i'\otimes\frac{U_{iT,j}'}{\sqrt{T}}\right\|\cdot\left\|\frac{1}{T}\,\mathrm{vec}\left[(\widehat{F} - F_0\Pi_{NT})(\widehat{F} - F_0\Pi_{NT})'\right]\right\|\\
&= O_P\left(\frac{1}{\sqrt{NT}}\right)\frac{1}{T}\|\widehat{F} - F_0\Pi_{NT}\|^2,
\end{aligned}
$$

where the equality follows from the development similar to (B.9). Summing $D_{22,j}$ up over j, we obtain that $\|D_{22}\| = O_P\left(\sqrt{\frac{m}{NT}}\right)\frac{1}{T}\|\widehat{F} - F_0\Pi_{NT}\|^2$.

For D_{23}, write

$$\|D_{23,j}\| \leq \left\| \frac{1}{NT} \sum_{i=1}^{N} e_i' \otimes \frac{U_{iT,j}' F_0}{T} \right\| \cdot \left\| \frac{1}{\sqrt{T}} \operatorname{vec}\left[\Pi_{NT} (\widehat{F} - F_0 \Pi_{NT})' \right] \right\|.$$

Note that

$$\begin{aligned}
& E \left\| \frac{1}{NT} \sum_{i=1}^{N} e_i' \otimes \frac{U_{iT,j}' F_0}{T} \right\|^2 = \frac{1}{N^2 T^4} \sum_{s=1}^{T} E \left\| \sum_{i=1}^{N} e_{is} \sum_{t=1}^{T} \frac{z_{it,j}}{\sqrt{T}} f_{0t}' \right\|^2 \\
&= \frac{1}{N^2 T^3} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} E \left[\left(\sum_{t=1}^{T} \frac{z_{i_1 t,j}}{\sqrt{T}} f_{0t}' \right) \left(\sum_{t=1}^{T} \frac{z_{i_2 t,j}}{\sqrt{T}} f_{0t} \right) \right] \sigma_{i_1 i_2} \\
&= \frac{1}{N^2 T^3} \sum_{t=1}^{T} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} E \left[\frac{z_{i_1 t,j}}{\sqrt{T}} \frac{z_{i_2 t,j}}{\sqrt{T}} E \left[\|f_{0t}\|^2 \mid R_{N,tt} \right] \right] \sigma_{i_1 i_2} \\
&\quad + \frac{2}{N^2 T^3} \sum_{t_1 > t_2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} E \left[\frac{z_{i_1 t_1,j}}{\sqrt{T}} \frac{z_{i_2 t_2,j}}{\sqrt{T}} E \left[f_{0t_1}' f_{0t_2} \mid R_{N,t_1 t_2} \right] \right] \sigma_{i_1 i_2} \\
&\leq O(1) \frac{2}{N^2 T^3} \sum_{t_1 \geq t_2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} |a_{t_1 t_2}| \cdot |\sigma_{i_1 i_2}| = O(1) \frac{1}{NT^2},
\end{aligned}$$

where $z_{it,j}$ stands for the j^{th} element of z_{it}, and the second and fourth equalities follow from Assumption B.4.

For D_{24}, write

$$\begin{aligned}
\|D_{24,j}\| &\leq \left\| \frac{1}{NT^{\frac{3}{2}}} \sum_{i=1}^{N} (e_i' F_0) \otimes \frac{U_{iT,j}' F_0}{T} \right\| \cdot \left\| \Pi_{NT} \Pi_{NT}' - (F_0' F_0 / T)^{-1} \right\| \\
&= O_P \left(\frac{1}{N^{\frac{1}{2}} T} \right) \left\| \Pi_{NT} \Pi_{NT}' - (F_0' F_0 / T)^{-1} \right\|,
\end{aligned}$$

where the equality follows from the development similar to (B.10). Summing $D_{24,j}$ up over j, we obtain that $\|D_{24}\| = O_P \left(\frac{m^{\frac{1}{2}}}{N^{\frac{1}{2}} T} \right) \left\| \Pi_{NT} \Pi_{NT}' - (F_0' F_0 / T)^{-1} \right\|$.

Based on the analyses of D_{21} to D_{24}, we obtain

$$\begin{aligned}
\frac{N^{\frac{1}{2}} T}{m^{\frac{1}{2}}} \|D_2\| &= O_P(1) \frac{1}{\sqrt{T}} \|\widehat{F} - F_0 \Pi_{NT}\| + O_P(1) \left\| \Pi_{NT} \Pi_{NT}' - (F_0' F_0 / T)^{-1} \right\| \\
&\quad + O_P(1) \sqrt{T} \cdot \frac{1}{T} \|\widehat{F} - F_0 \Pi_{NT}\|^2,
\end{aligned}$$

which further gives $\left\| \frac{N^{\frac{1}{2}} T}{m^{\frac{1}{2}}} \left[H_m'(r) \otimes I_{d_x} \right] A_{1NT}^{-1} \Sigma_m^{-1} D_2 \right\| = o_P(1)$ given the condition $\frac{mT}{N^2} \to 0$.

Similarly, we obtain

$$\begin{aligned}
& \left\| \frac{N^{\frac{1}{2}} T}{m^{\frac{1}{2}}} \left[H_m'(r) \otimes I_{d_x} \right] A_{1NT}^{-1} \Sigma_m^{-1} \frac{1}{NT^{3/2}} \sum_{i=1}^{N} A_{3,i} e_i \right. \\
& \left. - \frac{N^{\frac{1}{2}} T}{m^{\frac{1}{2}}} \left[H_m'(r) \otimes I_{d_x} \right] A_{1NT}^{-1} \Sigma_m^{-1} \frac{1}{NT^{3/2}} \sum_{i=1}^{N} \widetilde{A}_{3,i} e_i \right\| = o_P(1),
\end{aligned}$$

where $\widetilde{A}_{3,i} = \frac{1}{N} \sum_{j=1}^{N} \frac{Z_j'}{\sqrt{T}} M_{F_0} \gamma_0(v_j)' (\Gamma_0' \Gamma_0 / N)^{-1} \gamma_0(v_i)$.

Finally, we just need to focus on

$$
\begin{aligned}
\Lambda &= \frac{N^{\frac{1}{2}}T}{m^{\frac{1}{2}}}\left[H_m'(r)\otimes I_{d_x}\right]\widetilde{A}_{1NT}^{-1}\Sigma_m^{-1}\cdot\frac{1}{NT^{3/2}}\sum_{i=1}^N\left\{\frac{Z_i'}{\sqrt{T}}M_{F_0}+\widetilde{A}_{3,i}\right\}e_i+o_P(1)\\
&=\frac{1}{\sqrt{NTm}}\left[H_m'(r)\otimes I_{d_x}\right]\widetilde{A}_{1NT}^{-1}\Sigma_m^{-1}\cdot\sum_{i=1}^N\left\{\frac{Z_i'}{\sqrt{T}}M_{F_0}+\widetilde{A}_{3,i}\right\}e_i+o_P(1)
\end{aligned}
$$

where $\widetilde{A}_{1NT}=I_{md_x}-\Sigma_m^{-1}\widetilde{A}_{2NT}$ and $\widetilde{A}_{2NT}=\frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N\frac{Z_i'}{\sqrt{T}}M_{F_0}\frac{Z_j}{\sqrt{T}}\gamma_0(v_j)'\Sigma_\Gamma^{-1}\gamma_0(v_i)$. Then the rest proof of the normality follows from Lemma A.2 and verifying Lemma B.1 of Chen et al. (2012*b*), so omitted. ■

Proof of Lemma B.6:

(1). Recall that $\Pi_{NT}=\frac{\Gamma_0'\Gamma_0}{N}\cdot\frac{F_0'\widehat{F}}{T}\cdot V_{NT}^{-1}$, where V_{NT} is defined in Lemma B.2. Thus, by (B.11) and Assumption B.5, we obtain that

$$
\frac{1}{T}F_0'\widehat{F}=\Pi_{NT}+O_P\left(\sqrt{m/(NT)}+\sqrt{T/m^{2\mu_2}}+\frac{1}{N}\right). \tag{B.14}
$$

Bringing (B.14) in $\Pi_{NT}=\frac{\Gamma_0'\Gamma_0}{N}\cdot\frac{F_0'\widehat{F}}{T}\cdot V_{NT}^{-1}$, we obtain that

$$
\Pi_{NT}=\frac{\Gamma_0'\Gamma_0}{N}\cdot\Pi_{NT}\cdot V_{NT}^{-1}+O_P\left(\sqrt{m/(NT)}+\sqrt{T/m^{2\mu_2}}+\frac{1}{N}\right),
$$

which gives

$$
\Pi_{NT}\cdot V_{NT}=\frac{\Gamma_0'\Gamma_0}{N}\cdot\Pi_{NT}+O_P\left(\sqrt{m/(NT)}+\sqrt{T/m^{2\mu_2}}+\frac{1}{N}\right).
$$

Given the conditions in the body of this lemma, the rest proof is identical to that given for Proposition C.3 of the supplementary file of Fan et al. (2016).

(2). Using (B.14) and the first result of this lemma, the second result follows immediately. ■

Proof of Lemma B.4:

(1). By Lemma B.2 and Lemma B.6,

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\|\widehat{F}-F_0\| &\le \frac{1}{\sqrt{T}}\|\widehat{F}(I_{d_v}-\Pi_{NT}^{-1})\|+\frac{1}{\sqrt{T}}\|\widehat{F}\Pi_{NT}^{-1}-F_0\|\\
&=O_P\left(\sqrt{T/m^{2\mu_2}}+\frac{1}{\sqrt{N}}\right).
\end{aligned}
$$

(2). Expand $\widehat{C}_\gamma$ as follows.

$$
\begin{aligned}
\widehat{C}_\gamma-C_{\gamma_0} &= \left[\sum_{i=1}^N\mathscr{H}_n(v_i)\mathscr{H}_n'(v_i)\right]^{-1}\sum_{i=1}^N\mathscr{H}_n(v_i)\left\{\frac{1}{T}\widehat{F}'F_0-I_{d_v}\right\}\mathscr{H}_n'(v_i)C_{\gamma_0}\\
&+\left[\sum_{i=1}^N\mathscr{H}_n(v_i)\mathscr{H}_n'(v_i)\right]^{-1}\frac{1}{T}\sum_{i=1}^N\mathscr{H}_n(v_i)\widehat{F}'F_0\Delta_{\gamma_0}(v_i)\\
&+\left[\sum_{i=1}^N\mathscr{H}_n(v_i)\mathscr{H}_n'(v_i)\right]^{-1}\frac{1}{T}\sum_{i=1}^N\mathscr{H}_n(v_i)\widehat{F}'e_i\\
&+\left[\sum_{i=1}^N\mathscr{H}_n(v_i)\mathscr{H}_n'(v_i)\right]^{-1}\frac{1}{T}\sum_{i=1}^N\mathscr{H}_n(v_i)\widehat{F}'(\phi_i[\beta_0]-\phi_i[\widehat{\beta}_m])
\end{aligned}
$$

$$:= \Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4.$$

Start from Λ_3 and write

$$\begin{aligned}\Lambda_3 \; &= \left[\sum_{i=1}^N \mathscr{H}_n(v_i)\mathscr{H}_n'(v_i)\right]^{-1} \frac{1}{T}\sum_{i=1}^N \mathscr{H}_n(v_i)(\widehat{F}-F_0\Pi_{NT})'e_i \\ &\quad + \left[\sum_{i=1}^N \mathscr{H}_n(v_i)\mathscr{H}_n'(v_i)\right]^{-1} \frac{1}{T}\sum_{i=1}^N \mathscr{H}_n(v_i)\Pi_{NT}'F_0'e_i \\ &:= \Lambda_{31} + \Lambda_{32}.\end{aligned}$$

For Λ_{31},

$$\begin{aligned}\left\|\frac{1}{NT}\sum_{i=1}^N \mathscr{H}_n(v_i)(\widehat{F}-F_0\Pi_{NT})'e_i\right\| &\le \left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^N e_i'\otimes \mathscr{H}_n(v_i)\right\|\cdot\frac{1}{\sqrt{T}}\|\widehat{F}-F_0\Pi_{NT}\| \\ &= O_P\left(\sqrt{\frac{n}{N}}\right)\cdot\frac{1}{\sqrt{T}}\left\|\widehat{F}-F_0\Pi_{NT}\right\| \\ &= O_P\left(\sqrt{\frac{nT}{Nm^{2\mu_2}}}+\frac{\sqrt{n}}{N}\right),\end{aligned}$$

where the first equality follows from a development similar to (B.9) by accounting for the dimension of $\mathscr{H}_n(\cdot)$; and the second equality follows from Lemma B.2. In connection with (2) of Lemma A.2, we obtain that $\|\Lambda_{31}\| = O_P\left(\sqrt{\frac{nT}{Nm^{2\mu_2}}}+\frac{\sqrt{n}}{N}\right)$.

For Λ_{32}, by Lemma B.6, we have

$$\frac{1}{NT}\sum_{i=1}^N \mathscr{H}_n(v_i)\Pi_{NT}'F_0'e_i = \frac{1}{NT}\sum_{i=1}^N \mathscr{H}_n(v_i)F_0'e_i\cdot(1+o_P(1)).$$

Thus, we just focus on $\frac{1}{NT}\sum_{i=1}^N \mathscr{H}_n(v_i)F_0'e_i$ below, and write

$$E\left\|\frac{1}{NT}\sum_{i=1}^N \mathscr{H}_n(v_i)F_0'e_i\right\|^2 = \frac{1}{N^2T^2}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T E[f_{0t}'\mathscr{H}_n'(v_i)\mathscr{H}_n(v_j)f_{0s}e_{it}e_{js}] = O(1)\frac{n}{NT},$$

where the second equality follows from the development similar to (A.2). Thus, $\|\Lambda_{32}\| = O_P\left(\sqrt{\frac{n}{NT}}\right)$.

Therefore, we can conclude that $\|\Lambda_3\| = O_P\left(\frac{\sqrt{n}}{N}\right) + O_P\left(\sqrt{\frac{nT}{Nm^{2\mu_2}}}\right)$.

Similar to (A.1), we just need to consider the next term for Λ_2.

$$\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\widehat{F}'F_0\Delta_{\gamma_0}(v_i)\right\|^2 \le O_P(1)\frac{1}{N}\sum_{i=1}^N\|\Delta_{\gamma_0}(v_i)\|^2 = O_P(n^{-2\mu_1}).$$

Thus, $\|\Lambda_2\| = O_P(n^{-\mu_1})$. Again, similar to (A.1), we just need to consider the next term for Λ_4.

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}(\phi_i[\beta_0]-\phi_i[\widehat{\beta}_m])'\widehat{F}\right\|^2 \\ &\le \frac{2}{NT^2}\sum_{i=1}^N\|(\phi_i[\beta_0]-\phi_i[\widehat{\beta}_m])'(\widehat{F}-F_0\Pi_{NT})\|^2 + \frac{2}{NT^2}\sum_{i=1}^N\|(\phi_i[\beta_0]-\phi_i[\widehat{\beta}_m])'F_0\|^2 \\ &\le \frac{2\|\widehat{F}-F_0\Pi_{NT}\|^2}{NT^2}\sum_{i=1}^N\|\phi_i[\beta_0]-\phi_i[\widehat{\beta}_m]\|^2 + \frac{2\|F_0\|^2}{NT^2}\sum_{i=1}^N\|\phi_i[\beta_0]-\phi_i[\widehat{\beta}_m]\|^2\end{aligned}$$

$$:= 2\Lambda_{41} + 2\Lambda_{42}.$$

It is easy to know Λ_{42} is the leading term, and $\Lambda_{42} = O_P(T\|\widehat{\beta}_m - \beta_0\|_{L^2}^2) = O_P\left(\frac{m}{NT}\right) + O_P\left(Tm^{-2\mu_2}\right)$. Therefore, $\|\Lambda_4\| = O_P\left(\sqrt{\frac{m}{NT}}\right) + O_P(\sqrt{T}m^{-\mu_2})$.

Similarly, for Λ_1, we just need to consider

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\left\{\frac{1}{T}\widehat{F}'F_0 - I_{d_v}\right\}\mathscr{H}_n'(v_i)C_{\gamma_0}\right\|^2 \leq \left\|\frac{1}{T}\widehat{F}'F_0 - I_{d_v}\right\|^2 \cdot \frac{1}{N}\sum_{i=1}^{N}\|\mathscr{H}_n'(v_i)C_{\gamma_0}\|^2$$

$$= O_P(1)\left\|\frac{1}{T}\widehat{F}'F_0 - I_{d_v}\right\|^2 = O_P\left(\frac{m}{NT}\right) + O_P\left(Tm^{-2\mu_2}\right) + O_P\left(\frac{1}{N^2}\right),$$

where the second equality follows from (2) of Lemma B.6. Thus,

$$\|\Lambda_1\| = O_P\left(\sqrt{\frac{m}{NT}}\right) + O_P(\sqrt{T}m^{-\mu_2}) + O_P\left(\frac{1}{N}\right).$$

Based on the above arguments, the second result follows. ■

Proof of Theorem B.2:

(1). For each fixed t, we consider the asymptotic distribution of $\sqrt{N}(\widehat{f}_t - f_{0t})$. By (B.7), we write

$$\sqrt{N}(\widehat{f}_t - f_{0t}) = \sqrt{N}(\widehat{f}_t - \Pi_{NT}^{-1}\widehat{f}_t) + \sqrt{N}(\Pi_{NT}^{-1}\widehat{f}_t - f_{0t}) = \sqrt{N}(\Pi_{NT}^{-1}\widehat{f}_t - f_{0t}) + o_P(1)$$

$$= \sqrt{N}(\Gamma_0'\Gamma_0/N)^{-1}(\widehat{F}'F_0/T)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\left(\widehat{F}'e_ie_{it} + \widehat{F}'e_i\gamma_0'(v_i)f_{0t} + \widehat{F}'F_0\gamma_0(v_i)e_{it}\right) + o_P(1)$$

$$= \sqrt{N}(\Gamma_0'\Gamma_0/N)^{-1}\frac{1}{N}\sum_{i=1}^{N}\gamma_0(v_i)e_{it} + o_P(1) \to_D N(0, \Sigma_\Gamma^{-1}\Sigma_\Gamma^*\Sigma_\Gamma^{-1}),$$

where the second equality follows from Lemma B.6 and $\frac{NT}{m^{2\mu_2}} \to 0$; the third equality follows from the proof of Lemma B.2; the fourth equality follows from Assumption B.5; and the last step follows from procedures similar to those under (A.44) of Chen et al. (2012*b*). Thus, the result follows.

(2). By Lemma B.4, the second result follows. ■

Appendix C

In this Appendix, we comment on both methods, and then provide some numerical simulations to further compare the finite sample performance of both methods.

Having established Theorem 2.1 of the main text and Theorem B.1, it is easy to see that the direct estimation method and the PCA method are asymptotically equivalent in terms of the estimation on $\beta_0(\cdot)$, as the asymptotic covariances associated with the estimators of $\beta_0(\cdot)$ are identical. Moreover, similar to Corollary 2.1 of the main text, we can obtain that

$$\sqrt{\frac{NT^2}{2\widehat{\sigma}_e^2}}\widehat{\widehat{\Sigma}}_\beta^{-\frac{1}{2}}(\widehat{\beta}_m(r) - \beta_0(r)) \to_D N(0, I_{d_x}),$$

where $\widehat{\sigma}_e^2 = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(y_{it} - x_{it}'\widehat{\beta}_m(r_{it}) - \widehat{f}_t'\widehat{\gamma}_n(v_i)\right)^2$, and $\widehat{\widehat{\Sigma}}_\beta$ is defined in Corollary 2.1 of the main text.

To compare both methods in general, firstly, we believe that in terms of the numerical implementation, the direct estimation method would be preferred, as the PCA–based approach always requires an algorithm involving iterations, which may yield some inconsistent estimates under some circumstances (cf., Jiang et al., 2017). Second, to identify both the factors and loading functions, the method of the main text requires less restrictive conditions, which seems to be obvious in view of Assumption B.5. As a consequence, the theoretical development of the direct estimation method is more straightforward. Finally, we point out that as mentioned by Connor et al. (2012), it is of interest to allow the variables of the loading functions to change over both i and t. The direct estimation method can obviously be employed with minor modifications on the notations. However, our experience suggests that the PCA–based approach may no longer be working, as we cannot project out the factor structure when v is indexed by both i and t.

C.1 Numerical Simulations

Below, we implement simulation studies to compare the direct estimation method of the main text (referred to as MDE hereafter) and the PCA–based approach of Appendix B (referred to as MPCA hereafter). The data generating process is identical to Section 3 of the main text, and still consider Case 1 to Case 3. We let $d_x^* = d_x = 1$ and $d_v^* = d_v = 2$, which is assumed to be known already.

Apart from reporting RMSE associated to β_{01}, we also report the next measurement to compare the estimate on the factor structure of each method. For each generated dataset, we calculate

$$se_{\gamma' f} = \frac{1}{NT}\|\widehat{W} - F_0\Gamma_0'\|^2,$$

where for MDE, $\widehat{W} = \widetilde{C}_{1:T}\mathcal{H}_{1:N}'$, $\widetilde{C}_{1:T} = (\widetilde{C}_1, \ldots, \widetilde{C}_T)'$, $\widetilde{C}_t$ is defined by (2.5) of the main text, and $\mathcal{H}_{1:N} = (\mathcal{H}_n(v_1), \ldots, \mathcal{H}_n(v_N))'$; for MPCA, $\widehat{W} = \widehat{F}\widehat{\Gamma}'$, $\widehat{\Gamma} = \frac{1}{T}(Y_1 - \phi_1[\widehat{\beta}_m], \ldots, Y_N - \phi_N[\widehat{\beta}_m])'\widehat{F}$.

We summarize the results in Table C.1 below. Several facts are revealed. Both methods are almost identical in terms of the estimation on β_{01} regardless whether there is a trending in f_{0t}. The differences are negligible, as it is down to the third decimal. For MDE, as T increases, the RMSEs associated to the factor structure tends to remain at the same level, which matches the second result of Theorem 2.1 and Corollary 2.2. For MPCA, although $\frac{1}{T}\widehat{F}'\widehat{F}$ implies the estimates of f_{0t}'s follow a stationary process implicitly, it seems that the estimates on both β_0 and the factor structure are not affected too much even for Case 2 and Case 3. Theoretically supporting this point may lead to another research paper which we wish to consider in the future study.

Table C.1: Results of MDE and MPCA

			β_{01}			Factor Structure		
		$N \setminus T$	80	160	240	80	160	240
MDE	Case 1	80	0.050	0.025	0.019	0.288	0.288	0.333
		160	0.034	0.020	0.013	0.202	0.232	0.232
		240	0.032	0.016	0.011	0.189	0.189	0.189
	Case 2	80	0.070	0.042	0.034	0.289	0.289	0.332
		160	0.047	0.031	0.024	0.202	0.233	0.232
		240	0.043	0.025	0.019	0.189	0.189	0.189
	Case 3	80	0.055	0.033	0.029	0.288	0.288	0.333
		160	0.040	0.025	0.019	0.203	0.233	0.232
		240	0.036	0.020	0.015	0.189	0.189	0.189
MPCA	Case 1	80	0.058	0.028	0.020	0.289	0.248	0.228
		160	0.039	0.022	0.014	0.242	0.191	0.169
		240	0.036	0.018	0.012	0.218	0.167	0.147
	Case 2	80	0.058	0.029	0.021	0.288	0.247	0.228
		160	0.040	0.023	0.015	0.242	0.191	0.169
		240	0.037	0.018	0.012	0.218	0.167	0.146
	Case 3	80	0.057	0.028	0.021	0.288	0.248	0.227
		160	0.040	0.023	0.015	0.242	0.191	0.169
		240	0.036	0.018	0.012	0.217	0.167	0.146

To further compare both methods under different scenarios, we modify the above DGP slightly as follows:

1. EC 1: The DGP of f_{0t}'s is modified as $f_{0t} = \rho_f f_{0,t-1} + U(1,2)$, and let $\rho_f = 0.2, 0.8$.
2. EC 2: The DGP of e_{it}'s is modified as $e_{it} = N(0, \sigma_t^2)$ where $\sigma_t^2 = \frac{1}{N}\sum_{i=1}^N \|x_{it}\|^2/t$.
3. EC 3: The DGP of v_i's is modified as $v_i = N(0, 4\sigma_i^2)$ where $\sigma_i^2 = \frac{1}{T^2}\sum_{t=1}^T \|x_{it}\|^2$.

The results are summarized in Table C.2. Again, the same pattern remains. Both methods are almost identical in terms of the estimation on β_{01}. For MDE, as T increases, the RMSEs associated to the factor structure tends to remain at the same level. Given the number of factors is known, it is not very clear to claim which method is obviously better than the other one numerically. However, when the number of factors is unknown, MDE allows us to identify which unobservable factor(s) can be removed from the system precisely as discussed in the main text.

Table C.2: Results of Extra Cases

			β_{01}			Factor Structure		
		N\T	80	160	240	80	160	240
EC1 ($\rho_f = 0.2$)	MDE	80	0.050	0.025	0.019	0.289	0.289	0.333
		160	0.034	0.020	0.013	0.202	0.232	0.232
		240	0.032	0.016	0.011	0.189	0.189	0.189
	MPCA	80	0.056	0.028	0.021	0.294	0.250	0.229
		160	0.038	0.022	0.015	0.244	0.192	0.170
		240	0.035	0.018	0.012	0.219	0.167	0.147
EC1 ($\rho_f = 0.8$)	MDE	80	0.052	0.026	0.020	0.289	0.289	0.333
		160	0.036	0.021	0.014	0.202	0.232	0.232
		240	0.033	0.017	0.011	0.189	0.189	0.189
	MPCA	80	0.058	0.028	0.021	0.264	0.216	0.199
		160	0.039	0.022	0.015	0.216	0.171	0.153
		240	0.036	0.018	0.012	0.198	0.153	0.136
EC2	MDE	80	0.007	0.004	0.003	0.006	0.004	0.004
		160	0.005	0.003	0.002	0.004	0.004	0.003
		240	0.005	0.003	0.002	0.004	0.003	0.002
	MPCA	80	0.014	0.010	0.008	0.169	0.172	0.173
		160	0.008	0.006	0.004	0.167	0.169	0.170
		240	0.007	0.004	0.003	0.166	0.168	0.168
EC3	MDE	80	0.050	0.025	0.019	0.295	0.296	0.344
		160	0.035	0.020	0.013	0.205	0.237	0.237
		240	0.032	0.017	0.011	0.191	0.192	0.191
	MPCA	80	0.056	0.028	0.021	0.299	0.259	0.242
		160	0.040	0.022	0.015	0.254	0.212	0.195
		240	0.036	0.018	0.012	0.235	0.192	0.173

Appendix D

We further discuss some possible extensions in this section.

D.1 Estimation on Σ_e of Assumption 3

If we can provide a consistent estimator of Σ_e under some conditions using certain norm (e.g., spectral norm, Frobenius norm, etc.), then we are able to make inferences based on results (2) and (3) of Theorem 2.1. The question then becomes how to estimate a high dimensional covariance matrix Σ_e (e.g., Fan et al., 2013; Chen and Leng, 2016).

Suppose that we know $e_t = (e_{1t}, \ldots, e_{Nt})'$ for $t \geq 1$. A naive estimator of Σ_e would be $\frac{1}{T}\sum_{t=1}^{T} e_t e_t'$ provided $\frac{N^2}{T} \to 0$. Since e_t's are unobservable, we modify the naive estimator as

$$\widetilde{\Sigma}_e = \{\widetilde{\sigma}_{ij}\}_{N\times N} = \frac{1}{T}\sum_{t=1}^{T}(Y_t - Q_t'\widetilde{\mathbb{C}})(Y_t - Q_t'\widetilde{\mathbb{C}})',$$

where $Y_t = (y_{1t}, \ldots, y_{Nt})'$ and $Q_t = (Q_{1t}, \ldots, Q_{Nt})'$. To relax the restriction $\frac{N^2}{T} \to 0$, we can apply the generalised shrinkage technique to $\widetilde{\Sigma}_e$ as in Fan et al. (2013), forcing very small off–diagonal entries

$\widetilde{\sigma}_{ij}$ to be zero. Let $s_\varphi(\cdot)$ be a shrinkage function satisfying the following restrictions: (i) $|s_\varphi(w)| \leq |w|$ for $w \in \mathbb{R}$; (ii) $s_\varphi(w) = 0$ if $|w| \leq \varphi$; (iii) $|s_\varphi(w) - w| \leq \varphi$, where φ is a tuning parameter. The shrinkage function satisfying the above three restrictions covers some commonly used thresholdings in the literature, e.g., the hard thresholding, the soft thresholding and the SCAD function. Thus, the final form of the estimator of Σ_e is

$$\widehat{\Sigma}_e = \{\widehat{\sigma}_{ij}\}_{N\times N}, \quad \widehat{\sigma}_{ij} = \begin{cases} \widetilde{\sigma}_{ii}, & i = j, \\ s_\varphi(\widetilde{\sigma}_{ij}), & i \neq j. \end{cases} \tag{D.1}$$

The investigation on (D.1) can be done by following Fan et al. (2013) and Chen and Leng (2016), and it may lead to another research paper.

D.2 Alternative Methods for Factors Selection

We comment on some possible alternative methods for selecting the factors.

First, one may adopt a PCA–based approach as shown in the online supplementary file of this paper, and consider the ratio criterion studied in Lam and Yao (2012) and Ahn and Horenstein (2013). Specifically, we define $\widehat{\lambda}_j$ as the j^{th} largest eigenvalue of the estimated sample covariance matrix

$$\frac{1}{N}\sum_{i=1}^{N}(Y_i - \phi_i[\widehat{\beta}_m^J])(Y_i - \phi_i[\widehat{\beta}_m^J])', \tag{D.2}$$

where $Y_i = (y_{i1}, \ldots, y_{iT})'$, $\phi_i[\beta] := (x_{i1}'\beta(r_{i1}), \ldots, x_{iT}'\beta(r_{iT}))'$, and $\widehat{\beta}_m^J$ is obtained from the PCA–based approach assuming that the number of factors is a pre–specified fixed positive integer J. We then estimate the number of factors by

$$\widehat{r} = \operatorname*{argmin}_{j\in\{1,2,\ldots,J-1\}} \frac{\widehat{\lambda}_{j+1}}{\widehat{\lambda}_j}.$$

Note that slightly over–identifying the number of factors usually does not have any serious impact on consistency and rates of convergence of the subsequent estimation (Fan et al., 2013; Moon and Weidner, 2015). That is why $\widehat{\beta}_m^J$ is adopted in (D.2). After identifying the number of factors by $\widehat{r}$, we can update our estimate on $\beta_0(\cdot)$ again. Similarly, we may consider using the criterion provided in Bai and Ng (2002). However, for the PAC–based approach, it seems that the information of partially observed factor structure is not fully utilized, so we can only identify the number of relevant $v_{i,\ell}$'s without knowing which can be removed from the system.

Alternatively, one may follow Sun et al. (2016) to test

$$\mathsf{H}_0 : \Pr\{\gamma_{0\ell}(w) = 0\} = 1 \quad \text{v.s.} \quad \mathsf{H}_1 : \Pr\{\gamma_{0\ell}(w) \neq 0\} > 0 \tag{D.3}$$

for $\ell = 1, \ldots, d_v$. To achieve good finite sample performance, one may need to further consider a residual–based bootstrap method as suggested in the conclusion of their paper. Such a methodology has also been used in Su and Chen (2013) and Su et al. (2015) under stationary panel data settings. Moreover, the approach of (D.3) further leads to the next discussion.

D.3 Constancy Test

As pointed out by one referee, testing the constancy of the coefficient functions is of interest. Su et al. (2015) adopt the residual based test to exam whether $y_{it} = m(x_{it}) + \gamma_{0i}' f_{0t} + e_{it}$ posses a parametric form. In another work, Sun et al. (2016) use the residual based test to construct a nonparametric test on a varying coefficient model with integrated time series. Some relevant studies also include but not limited to Fan and Li (1996), Dong and Gao (2018), etc.

Following the same spirit, we can test the constancy of the coefficients of model (1.3) in several different ways. For example,

$$\begin{aligned} &\mathsf{H}_0: \quad \Pr\{\beta_0(\cdot) \equiv b_0\} = 1 \quad \text{for some } b_0 \in \mathbb{R}^{d_x}, \\ &\mathsf{H}_1: \quad \Pr\{\beta_0(\cdot) \equiv b\} < 1 \quad \text{for all } b \in \mathbb{R}^{d_x}, \end{aligned} \tag{D.4}$$

or

$$\begin{aligned} &\mathsf{H}_0: \quad \Pr\{\gamma_0(\cdot) \equiv r_0\} = 1 \quad \text{for some } r_0 \in \mathbb{R}^{d_v}, \\ &\mathsf{H}_1: \quad \Pr\{\gamma_0(\cdot) \equiv r\} < 1 \quad \text{for all } r \in \mathbb{R}^{d_v}, \end{aligned} \tag{D.5}$$

or

$$\begin{aligned} &\mathsf{H}_0: \quad \Pr\{\beta_0(\cdot) \equiv b_0 \text{ and } \gamma_0(\cdot) \equiv r_0\} = 1 \quad \text{for some } b_0 \in \mathbb{R}^{d_x} \text{ and some } r_0 \in \mathbb{R}^{d_v}, \\ &\mathsf{H}_1: \quad \Pr\{\beta_0(\cdot) \equiv b \text{ or } \gamma_0(\cdot) \equiv r\} < 1 \quad \text{for all } b \in \mathbb{R}^{d_x} \text{ or all } r \in \mathbb{R}^{d_v}. \end{aligned} \tag{D.6}$$

Under the null of (D.4), the estimation and testing procedure will be very similar to Su et al. (2015), but one needs to account for the nonstationarity of x_{it}. On the other hand, the study of Sun et al. (2016) sheds a light on how to incorporate the integrated regressors for the varying–coefficient models. Under the null hypotheses (D.5) and (D.6), model (1.3) reduces to parametric/semiparametric panel data models with time effects respectively. A test can still be established similarly.

Generally speaking, (D.4)–(D.6) are more challenging than they look like, as the theoretical development involves unobservable factors and integrated time series, which essentially requires some considerable new developments. Thus, it should be left for future research. We however will provide some detailed development on the testing issues with the corresponding discussions and simulation studies below.

For the time being, assume that $\{f_{0t}\}$ is observable and $\{e_{it}\}$ is independent of $\{x_{it}, v_i, f_{0t}\}$ for simplicity. Consider (D.6) as an example. Under the null of (D.6), we can write model (1.3) of the main text as

$$y_{it} = x_{it}'\beta_0 + f_{0t}'\gamma_0 + e_{it} := z_{it}'\theta_0 + e_{it}, \tag{D.7}$$

where $\theta_0 = (\beta_0', \gamma_0')'$ and z_{it} is defined accordingly. To facilitate development, we further introduce some notations. Let $D_T = \mathrm{diag}\{\frac{1}{\sqrt{T}} I_{d_x}, I_{d_v}\}$, and $D_{NT} = \mathrm{diag}\{\frac{1}{\sqrt{NT^2/2}} I_{d_x}, \frac{1}{\sqrt{NT/2}} I_{d_v}\}$. Suppose that $\mathsf{S}_1 = \{j \mid j \text{ is odd and } 1 \le j \le T\}$ and $\mathsf{S}_2 = \{j \mid j \text{ is even and } 2 \le j \le T\}$. Let $t \neq s \in \mathsf{S}_j$ read as $t \in \mathsf{S}_j$, $s \in \mathsf{S}_j$ and $t \neq s$, where $j = 1, 2$.

We construct our statistic as follows:

$$L_{NT} = \frac{NT/2 \cdot L_{1NT} - B_{NT}}{\widehat{\sigma}_e^2 \sqrt{L_{2NT}}},$$

where

$$L_{1NT} = \frac{1}{N^2(T/2)^2} \sum_{i,j=1}^{N} \sum_{t \neq s \in \mathsf{S}_2} \widehat{e}_{it} \widehat{e}_{js} (D_T z_{it})'(D_T z_{js}),$$

$$L_{2NT} = \frac{2}{N^2(T/2)^2} \sum_{i,j=1}^{N} \Big\{ \sum_{t \neq s \in \mathsf{S}_2} + \sum_{t \in \mathsf{S}_1} \sum_{s \in \mathsf{S}_2} + \sum_{t \neq s \in \mathsf{S}_1} \Big\} \left[(D_T z_{it})'(D_T z_{js})\right]^2,$$

$$B_{NT} = \frac{\widehat{\sigma}_e^2}{NT/2} \sum_{i=1}^{N} \sum_{t \in \mathsf{S}_1} \|D_T z_{it}\|^2, \quad \widehat{\sigma}_e^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{e}_{it}^2, \quad \widehat{e}_{it} = y_{it} - z_{it}' \widehat{\theta}_{\mathsf{S}_1}, \quad \widetilde{e}_{it} = y_{it} - z_{it}' \widetilde{\theta},$$

$$\widehat{\theta}_{\mathsf{S}_1} = \left(\sum_{i=1}^{N} \sum_{t \in \mathsf{S}_1} z_{it} z_{it}' \right)^{-1} \sum_{i=1}^{N} \sum_{t \in \mathsf{S}_1} z_{it} y_{it}, \quad \widetilde{\theta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} z_{it}' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} y_{it}.$$

Below, we shall show that as $(N, T) \to (\infty, \infty)$,

$$L_{NT} \to_D N(0, 1). \tag{D.8}$$

Before proceeding further, we make a few comments. The construction of L_{NT} does not involve a nonparametric kernel as in Su et al. (2015) and Sun et al. (2016), so we can consider it as an improvement and simplification. As a consequence, it allows us to avoid a sensitive question "bandwidth selection" in practice. When establishing L_{NT}, the sample split is due to a technical challenge raised in the theoretical development, as using full sample will cause some crucial values cancelling with each other asymptotically. Below, we start our development.

For L_{2NT}, it is easy to know that

$$\begin{aligned}
\operatorname*{plim}_{N,T} L_{2NT} &= \lim_{N,T} \frac{2}{N^2(T/2)^2} \sum_{i,j=1}^{N} \sum_{t \neq s \in \mathsf{S}_2} E\left[(D_T z_{it})'(D_T z_{js})\right]^2 \\
&\quad + \lim_{N,T} \frac{2}{N^2(T/2)^2} \sum_{i,j=1}^{N} \sum_{t \in \mathsf{S}_1} \sum_{s \in \mathsf{S}_2} E\left[(D_T z_{it})'(D_T z_{js})\right]^2 \\
&\quad + \lim_{N,T} \frac{2}{N^2(T/2)^2} \sum_{i,j=1}^{N} \sum_{t \neq s \in \mathsf{S}_1} E\left[(D_T z_{it})'(D_T z_{js})\right]^2 \\
&:= \mathsf{z}_1^2 + \mathsf{z}_2^2 + \mathsf{z}_3^2,
\end{aligned}$$

where the definition of z_j^2 for $j = 1, 2, 3$ should be obvious.

For L_{1NT}, write

$$\begin{aligned}
L_{1NT} &= \frac{1}{N^2(T/2)^2} \sum_{i,j=1}^{N} \sum_{t \neq s \in \mathsf{S}_2} [e_{it} + z_{it}'(\theta_0 - \widehat{\theta}_{\mathsf{S}_1})][e_{js} + z_{js}'(\theta_0 - \widehat{\theta}_{\mathsf{S}_1})](D_T z_{it})'(D_T z_{js}) \\
&= \frac{1}{N^2(T/2)^2} \sum_{i,j=1}^{N} \sum_{t \neq s \in \mathsf{S}_2} e_{it} e_{js} (D_T z_{it})'(D_T z_{js}) \\
&\quad + \frac{2}{N^2(T/2)^2} (\theta_0 - \widehat{\theta}_{\mathsf{S}_1})' \sum_{i,j=1}^{N} \sum_{t \neq s \in \mathsf{S}_2} e_{it} z_{js} (D_T z_{it})'(D_T z_{js})
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{N^2(T/2)^2}(\theta_0-\widehat{\theta}_{\mathsf{S}_1})'\sum_{i,j=1}^{N}\sum_{t\neq s\in\mathsf{S}_2} z_{it}z_{js}'(D_Tz_{it})'(D_Tz_{js})(\theta_0-\widehat{\theta}_{\mathsf{S}_1})\\
:=\ & L_{1NT,1}+2L_{1NT,2}+L_{1NT,3}.
\end{aligned}
$$

Note that by some routine practice, one can show that

$$
NT/2\cdot L_{1NT,1}\to_D N(0,\sigma_{L_1}^2),
$$

where

$$
\begin{aligned}
\sigma_{L_1}^2 &= \lim_{N,T}\frac{1}{N^2(T/2)^2}E\left[\sum_{i,j=1}^{N}\sum_{t\neq s\in\mathsf{S}_2} e_{it}e_{js}(D_Tz_{it})'(D_Tz_{js})\right]^2\\
&= \lim_{N,T}\frac{1}{N^2(T/2)^2}\sum_{i_1,j_1=1}^{N}\sum_{t_1\neq s_1\in\mathsf{S}_2}\sum_{i_2,j_2=1}^{N}\sum_{t_2\neq s_2\in\mathsf{S}_2}E\Big[e_{i_1t_1}e_{j_1s_1}e_{i_2t_2}e_{j_2s_2}\\
&\quad\cdot(D_Tz_{i_1t_1})'(D_Tz_{j_1s_1})(D_Tz_{i_2t_2})'(D_Tz_{j_2s_2})\Big]\\
&= \lim_{N,T}\frac{2}{N^2(T/2)^2}\sum_{i_1,j_1=1}^{N}\sum_{i_2,j_2=1}^{N}\sum_{t_1\neq s_1\in\mathsf{S}_2}E\Big[e_{i_1t_1}e_{j_1s_1}e_{i_2t_1}e_{j_2s_1}\\
&\quad\cdot(D_Tz_{i_1t_1})'(D_Tz_{j_1s_1})(D_Tz_{i_2t_1})'(D_Tz_{j_2s_1})\Big]\\
&= \lim_{N,T}\frac{2\sigma_e^4}{N^2(T/2)^2}\sum_{i_1,j_1=1}^{N}\sum_{t_1\neq s_1\in\mathsf{S}_2}E\left[(D_Tz_{i_1t_1})'(D_Tz_{j_1s_1})\right]^2=\sigma_e^4\mathsf{z}_1^2
\end{aligned}
$$

in which the third equality follows from the martingale difference condition across t; and the fourth equality requires e_{it} being independent across i.

We now consider $L_{1NT,2}$.

$$
\begin{aligned}
L_{1NT,2} &= (\theta_0-\widehat{\theta}_{\mathsf{S}_1})'\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t\neq s\in\mathsf{S}_2} e_{it}z_{js}(D_Tz_{it})'(D_Tz_{js})\\
&= -\left\{\left(\sum_{k=1}^{N}\sum_{\ell\in\mathsf{S}_1} z_{k\ell}z_{k\ell}'\right)^{-1}\sum_{k=1}^{N}\sum_{\ell\in\mathsf{S}_1} z_{k\ell}e_{k\ell}\right\}'\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t\neq s\in\mathsf{S}_2} e_{it}z_{js}(D_Tz_{it})'(D_Tz_{js})\\
&= -\left\{D_{NT}\left(D_{NT}\sum_{k=1}^{N}\sum_{\ell\in\mathsf{S}_1} z_{k\ell}z_{k\ell}'D_{NT}\right)^{-1}D_{NT}\sum_{k=1}^{N}\sum_{\ell\in\mathsf{S}_1} z_{k\ell}e_{k\ell}\right\}'\\
&\quad\cdot\frac{1}{N^2(T/2)^2}\sum_{i,j=1}^{N}\sum_{t\neq s\in\mathsf{S}_2} e_{it}z_{js}(D_Tz_{it})'(D_Tz_{js}).
\end{aligned}
$$

For simplicity, suppose that for $j=1,2$

$$
D_{NT}\sum_{k=1}^{N}\sum_{\ell\in\mathsf{S}_j} z_{k\ell}z_{k\ell}'D_{NT}\to_P I_{d_x+d_v}.
$$

Then we can further write that

$$
L_{1NT,2}=-\frac{1}{NT/2}\sum_{k=1}^{N}\sum_{\ell\in\mathsf{S}_1}(D_Tz_{k\ell})'e_{k\ell}\cdot\frac{1}{N^2(T/2)^2}\sum_{i,j=1}^{N}\sum_{t\neq s\in\mathsf{S}_2} e_{it}(D_Tz_{js}z_{js}'D_T)(D_Tz_{it})\cdot(1+o_P(1))
$$

$$
\begin{aligned}
&= -\frac{1}{NT/2}\sum_{k=1}^{N}\sum_{\ell\in \mathsf{S}_1}(D_T z_{k\ell})' e_{k\ell}\cdot \frac{1}{NT/2}\sum_{i=1}^{N}\sum_{t\in\mathsf{S}_2} e_{it}(D_T z_{it})\cdot(1+o_P(1))\\
&= -\frac{1}{N^2(T/2)^2}\sum_{k=1}^{N}\sum_{\ell\in\mathsf{S}_1}\sum_{i=1}^{N}\sum_{t\in\mathsf{S}_2} e_{k\ell}e_{it}(D_T z_{k\ell})'(D_T z_{it})\cdot(1+o_P(1)).
\end{aligned}
$$

Following the same procedure as $L_{1NT,1}$, we know that

$$
NT/2\cdot L_{1NT,2}\to_D N(0,\sigma^2_{L_2}),
$$

where

$$
\begin{aligned}
\sigma^2_{L_2} &= \lim_{N,T}\frac{1}{N^2(T/2)^2}\sum_{i,j=1}^{N}\sum_{t\in\mathsf{S}_1}\sum_{s\in\mathsf{S}_2} E\left[e_{it}e_{js}(D_T z_{it})'(D_T z_{js})\right]^2\\
&= \lim_{N,T}\frac{\sigma_e^4}{N^2(T/2)^2}\sum_{i,j=1}^{N}\sum_{t\in\mathsf{S}_1}\sum_{s\in\mathsf{S}_2} E\left[(D_T z_{it})'(D_T z_{js})\right]^2 = \sigma_e^4\mathsf{z}_2^2.
\end{aligned}
$$

For $L_{1NT,3}$, write

$$
\begin{aligned}
L_{1NT,3} &= \frac{1}{N^2(T/2)^2}(\theta_0-\widehat{\theta}_{\mathsf{S}_1})'\left\{\sum_{i,j=1}^{N}\sum_{t\neq s\in\mathsf{S}_2} z_{it}z_{js}'(D_T z_{it})'(D_T z_{js})\right\}(\theta_0-\widehat{\theta}_{\mathsf{S}_1})\\
&= \frac{1}{N^2(T/2)^2}\sum_{i,j=1}^{N}\sum_{t,s\in\mathsf{S}_1} e_{it}e_{js}(D_T z_{it})'(D_T z_{js})\cdot(1+o_P(1))\\
&:= \widetilde{L}_{1NT,3}\cdot(1+o_P(1)),
\end{aligned}
$$

where the second equality follows from a development similar to $L_{1NT,2}$ after replacing $\theta_0-\widehat{\theta}_{\mathsf{S}_1}$ with its definition. Note that

$$
E[NT/2\cdot\widetilde{L}_{1NT,3}] = \frac{\sigma_e^2}{N(T/2)}\sum_{i=1}^{N}\sum_{t\in\mathsf{S}_1} E\|D_T z_{it}\|^2
$$

and

$$
\begin{aligned}
E[NT/2\cdot\widetilde{L}_{1NT,3}]^2 &= \frac{1}{N^2(T/2)^2}\sum_{i_1,j_1=1}^{N}\sum_{t_1,s_1\in\mathsf{S}_1}\sum_{i_2,j_2=1}^{N}\sum_{t_2,s_2\in\mathsf{S}_1} E\Big[e_{i_1t_1}e_{j_1s_1}e_{i_2t_2}e_{j_2s_2}\\
&\quad\cdot(D_T z_{i_1t_1})'(D_T z_{j_1s_1})(D_T z_{i_2t_2})'(D_T z_{j_2s_2})\Big]\\
&= \frac{1}{N^2(T/2)^2}\sum_{i_1,j_1=1}^{N}\sum_{t_1\in\mathsf{S}_1}\sum_{i_2,j_2=1}^{N}\sum_{t_2\in\mathsf{S}_1} E\Big[e_{i_1t_1}e_{j_1t_1}e_{i_2t_2}e_{j_2t_2}\\
&\quad\cdot(D_T z_{i_1t_1})'(D_T z_{j_1t_1})(D_T z_{i_2t_2})'(D_T z_{j_2t_2})\Big]\\
&\quad+\frac{2}{N^2(T/2)^2}\sum_{i_1,j_1=1}^{N}\sum_{i_2,j_2=1}^{N}\sum_{t_1\neq s_1\in\mathsf{S}_1} E\Big[e_{i_1t_1}e_{j_1s_1}e_{i_2t_1}e_{j_2s_1}\\
&\quad\cdot(D_T z_{i_1t_1})'(D_T z_{j_1s_1})(D_T z_{i_2t_1})'(D_T z_{j_2s_1})\Big]\\
&= \frac{\sigma_e^4}{N^2(T/2)^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{t_1\neq t_2\in\mathsf{S}_1} E[\|D_T z_{i_1t_1}\|^2\|D_T z_{i_2t_2}\|^2]
\end{aligned}
$$

$$+\frac{2\sigma_e^4}{N^2(T/2)^2}\sum_{i_1,j_1=1}^{N}\sum_{t_1\neq s_1\in \mathsf{S}_1}E[(D_T z_{i_1t_1})'(D_T z_{j_1s_1})]^2+o(1).$$

Similar to the development of $L_{1NT,1}$, we obtain that

$$NT/2\cdot L_{1NT,3}-B_{NT}\to_D N(0,\sigma_{L_3}^2),$$

where

$$\sigma_{L_3}^2=\lim_{N,T}\frac{2\sigma_e^4}{N^2(T/2)^2}\sum_{i_1,j_1=1}^{N}\sum_{t_1\neq s_1\in \mathsf{S}_1}E[(D_T z_{i_1t_1})'(D_T z_{j_1s_1})]^2=\sigma_e^4\mathsf{z}_3^2$$

in which we need to assume that z_{it} is independence across i. Note that z_{it} being independence across i is not really important, and it can be relaxed by modifying the form of B_{NT}.

Based on the above development, we conclude that (D.8) follows. Note that the development of (D.8) requires the martingale difference condition and cross–sectional independence among error terms. To relax the restrictions on error terms (say cross–sectional independence), we need to consistently estimate σ_{ij} in Assumption 3 of the main text. Then the discussions on "Estimation on Σ_e of Assumption 3" apply. More importantly, the above procedure assuming that $\{f_{0t}\}$ is observable. For the cases with unknown factors, these questions become more challenging than it looks like, so it deserves another research paper in view of the technical challenges involving unobservable factors and integrated time series, which essentially requires a combination of Su et al. (2015) and Sun et al. (2016).

For (D.4), model (1.3) of the main text reduces to

$$y_{it}=x_{it}'b_0+f_{0t}'\gamma_0(v_i)+e_{it}$$

under the null. Then one can adopt the PCA–based approach of Su et al. (2015) to conduct the hypothesis test. However, the nonstationarity of x_{it} needs to be taken into account. Sun et al. (2016) and Dong and Gao (2018) have clearly explained the difficulties of incorporating nonstationary variables in constancy test, so we refer interested readers to their works and the references therein.

To demonstrate the feasibility of (D.8), we implement some simple simulation studies below. The DGP is as follows. $x_{it}=x_{i,t-1}+$ i.i.d. $N(0,0.5I_{d_x})$, $f_{0t,\ell}\sim$ i.i.d. $U(1,2)$ for $\ell=1,\ldots,d_v$, $v_i\sim$ i.i.d. $N(0,4I_{d_v})$, $r_{it}\sim$ i.i.d. $U(-4,4)+v_{i,\ell}/4$, and $e_{it}\sim$ i.i.d. $N(0,1)$. We consider the following four cases for the coefficients.

- Size: $\beta_0=1_{d_x\times 1}$, and $\gamma_0=1_{d_v\times 1}$;
- Power:
    1. $\beta_0=\exp(-r_{it}^2/2)\cdot 1_{d_x\times 1}$, and $\gamma_0=1_{d_v\times 1}$;
    2. $\beta_0=1_{d_x\times 1}$, and $\gamma_0(v_i)=(\exp(-v_{i,1}),\ldots,\exp(-v_{i,d_v}))'$;
    3. $\beta_0=\exp(-r_{it}^2/2)\cdot 1_{d_x\times 1}$ and $\gamma_0(v_i)=(\exp(-v_{i,1}),\ldots,\exp(-v_{i,d_v}))'$;

After J replications, we report the following value for each case under different choices of N and T.

$$\overline{L} = \frac{1}{J}\sum_{j=1}^{J} 1(|L_{NT,j}| > 1.96),$$

where $L_{NT,j}$ stands for the value of L_{NT} at j^{th} replication.

We let $d_x = 1$, $d_v = 2$ and $J = 500$, and summarize the results in Table D.3 below. Overall, the size of Table D.3 converges to 5% as $(N, T) \to (\infty, \infty)$, which verifies (D.8) numerically. Also, the power of three cases converges to 1 sufficiently fast, which indicates good finite sample performance of our test.

Table D.3: Size and Power of (D.8)

	$N \setminus T$	80	160	240
Size	80	0.090	0.090	0.082
	160	0.070	0.058	0.054
	240	0.068	0.052	0.048
Power (1)	80	0.968	0.974	0.992
	160	0.984	0.992	0.99
	240	0.988	0.990	0.990
Power (2)	80	0.902	0.952	0.982
	160	0.946	0.976	0.984
	240	0.968	0.976	0.990
Power (3)	80	0.792	0.918	0.968
	160	0.872	0.958	0.958
	240	0.886	0.964	0.990

D.4 Cases with Mixed $I(1)/I(0)$ Regressors

Without too many difficulties, we can change the coefficient function of the main text from $\beta_0(r_{it})$ to $\beta_0(r_{it}, \tau_t)$, where $\tau_t = t/T$. We then consider a model with interactive fixed effects as follows.

$$y_{it} = x_{1it}'\beta_{10}(r_{it}, \tau_t) + x_{2it}'\beta_{20}(r_{it}, \tau_t) + f_{0t}'\gamma_i + e_{it}, \tag{D.9}$$

where x_{1it} and x_{2it} are $I(1)$ and $I(0)$ across t, respectively. As explained in Section 4 of Bai et al. (2009), the difficulty of considering such a model lies in the requirements of different normalizers, which further gives rise to a challenge of the degeneration of asymptotics since the covariance matrix would be singular. The detailed development of Appendix B of this study provides a clear solution to this type of challenge.

We now briefly sketch how to estimate (D.9), and further implement a simple Monte Carlo to back up our arguments. We still need to restrict the set that the coefficient function of $I(1)$ regressors to a set like B_T of Appendix B. The objective function and the corresponding estimators are defined in the same fashion as Appendix B. Then the asymptotic properties can be derived with minor modification on the notations.

To support our arguments, we implement a simple Monte Carlo study here. Let x_{1it} and x_{2it} be scalars, and further let

$$x_{1it} = x_{1i,t-1} + \text{i.i.d. } N(0,1) \quad \text{and} \quad x_{2it} = 0.5\, x_{2i,t-1} + \text{i.i.d. } N(0,1).$$

The factors are generated by $f_{0t} \sim$ i.i.d. $N(0,1)$. For $\ell = 1, \ldots, d_v$, $\gamma_{i,\ell} \sim$ i.i.d. $\exp\left(-(v_{i,\ell} - \ell/4)^2\right)$ with $v_{i,\ell} \sim$ i.i.d. $U(0,1)$. The error terms are generated by $e_t = 0.4\, e_{t-1} + N(0, \Sigma_e)$ with $\Sigma_e = \{0.6^{|i-j|}\}_{N\times N}$. Similarly, generate $r_t^* = (r_{1t}^*, \ldots, r_{Nt}^*)'$, where $r_t^* = 0.8\, r_{t-1}^* + N(0, \Sigma_{r*})$ with $\Sigma_{r*} = \{0.4^{|i-j|}\}_{N\times N}$. Let $r_{it} = r_{it}^* + \|f_{0t}\|^2 + \sum_{\ell=1}^{d_v} |v_{i,\ell}|$, so that $\{r_{it}\}$ is correlated with the factor structure. For the coefficient functions, let $\beta_{01}(r) = \exp(-r^2/2)$ depend on r only, and let $\beta_{02}(\tau) = \tau^2$ depend on τ only. The supposition of this form may facilitate to plot the estimates of β_0 (see Figures D.1 and D.2 for details), because a three dimensional picture is not easy to draw for the purpose of comparison. We adopt Hermite functions to expand β_{01} with the truncation parameter $m_1 = \lfloor NT \rfloor^{\frac{1}{7}} + 1$, while we use the Fourier series (as used in Dong and Linton (2018)) to expand β_{02} with the truncation parameter $m_2 = \lfloor NT \rfloor^{\frac{1}{7}}$. Throughout the simulation studies, we choose $d_v = 3$.

In each replication, we estimate the coefficient functions using the PCA–based approach by assuming F_0 is known and unknown respectively, and record the estimated coefficient functions on some selected points over certain intervals (referred to as "M1" and "M2"). When F_0 is known, we just need to replace M_F with M_{F_0} in the objective function. After 1000 replications, we plot the lower and upper bounds of these values in Figures D.1–D.2 under a variety of choices of (N, T) (the true curve is referred to as "True"). For both figures, as the sample size goes up, the distance between lower and upper bounds becomes smaller, and all bounds move towards the real function. Moreover, the lower and upper bounds of M1 and M2 are almost identical, so it verifies the above arguments.

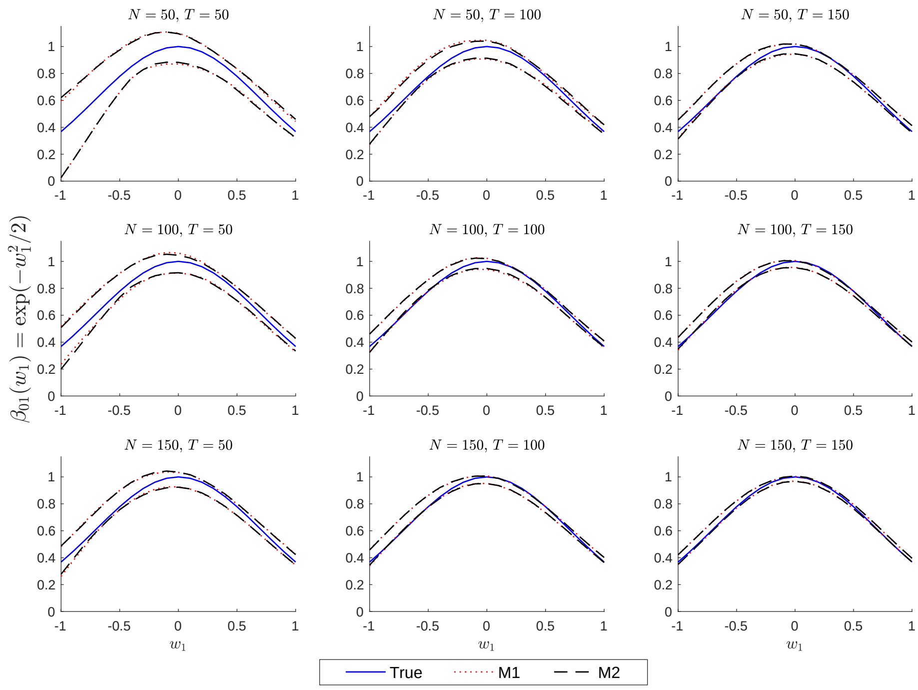

Figure D.1: $\beta_{01}(w_1) = \exp(-w_1^2/2)$ and its estimates

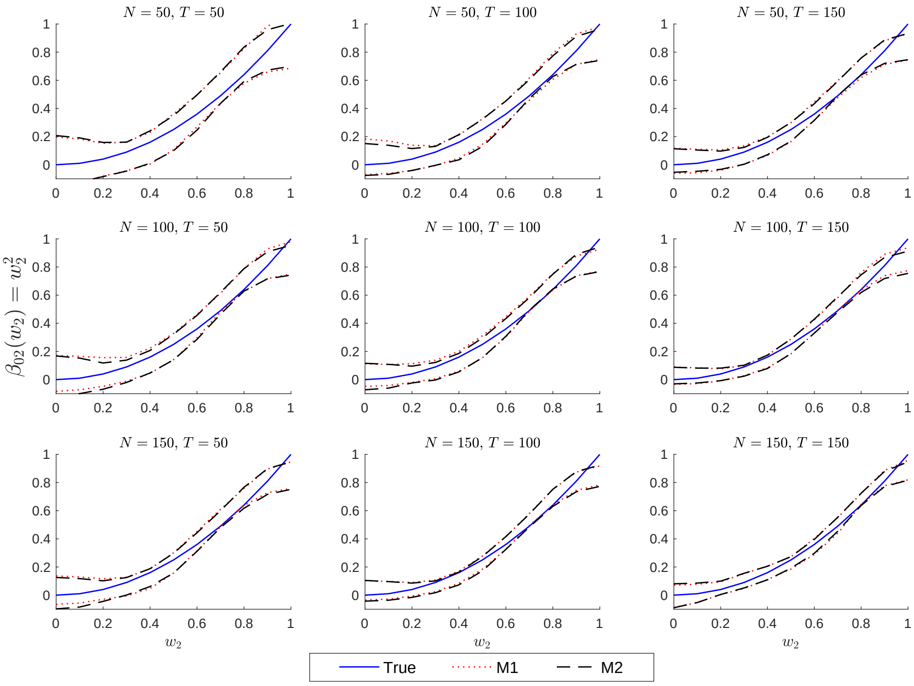

Figure D.2: $\beta_{02}(w_2) = w_2^2$ and its estimates

D.5 Assumption on r_{it}

Recall that as mentioned in the main text, the independence assumption between x_{it} and $\{r_{it}, v_i, f_{0t}\}$ can be further relaxed by following Assumption B.1 of Dong and Linton (2018).

Moreover, the assumptions on r_{it} in fact can be further relaxed to the locally stationary process by following Vogt (2012) and Dong and Linton (2018) in order to account for more cases.

Definition D.1. *The $d \times 1$ dimensional process $\{r_t \mid t = 1, \ldots, T\}$ is locally stationary if for each rescaled time point $u \in [0, 1]$ there exists an associated process $\{r_t[u] \mid t = 1, \ldots, T\}$ with the following two properties:*

1. *$\{r_t[u] \mid t = 1, \ldots, T\}$ is strictly stationary with density $f_u(r)$;*
2. *It holds that $\|r_t - r_t[u]\|_\nu \leq \left(|\tau_t - u| + T^{-1}\right) R_t(u)$ a.s., where $\tau_t = t/T$, $\{R_t(u)\}$ is a process of positive variables satisfying $E|R_t(u)|^\rho < C$ for some $\rho > 0$ and $C < \infty$ independent of u, t, and T. Moreover, $\|\cdot\|_\nu$ denotes an arbitrary norm on $\mathbb{R}^d$.*

Some detailed development has been given in an earlier version of this paper. As it is not a main concern for the partially observed factor structure, we remove this setting from the main text in the revised version.

References

Ahn, S. C. and Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica* **81**(3), 1203–1227.

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.

Bai, J. and Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Bai, J. and Ng, S. (2013), 'Principal components estimation and identification of static factors', *Journal of Econometrics* **176**(1), 18–99.

Bernstein, D. S. (2005), *Matrix Mathematics: Theory, Facts, and Formulas*, Princeton University Press.

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), 1144–1163.

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, Z. and Leng, C. (2016), 'Dynamic covariance models', *Journal of the American Statistical Association* **111**(515), 1196–1208.

Connor, G., Hagmann, M. and Linton, O. (2012), 'Efficient semiparametric estimation of the fama-french model and extensions', *Econometrica* **80**(2), 713–754.

Dong, C. and Gao, J. (2018), 'Specification testing driven by orthogonal series for nonlinear cointegration with endogeneity', *Econometric Theory* **34**(4), 754789.

Dong, C., Gao, J. and Peng, B. (2019), 'Estimation in a semiparametric panel data model with nonstationarity', *Econometric Reviews* **38**, 961–977.

Dong, C. and Linton, O. (2018), 'Additive nonparametric models with time variable and both stationary and nonstationary regressors', *Journal of Econometrics* **207**(1), 212–236.

Fan, J., Liao, Y. and Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *Journal of the Royal Statistical Society: Series B* **75**(4), 603–680.

Fan, J., Liao, Y. and Wang, W. (2016), 'Projected principal component analysis in factor models', *Annals of Statistics* **44**(1), 219–254.

Fan, Y. and Li, Q. (1996), 'Consistent model specification tests: omitted variables, parametric and semiparametric functional forms', *Econometrica* **64**(4), 865–890.

Jiang, B., Yang, Y., Gao, J. and Hsiao, C. (2017), Recursive estimation in large panel data models: Theory and practice. Working paper available at https://ssrn.com/abstract=2915749.

Lam, C. and Yao, Q. (2012), 'Factor modeling for high-dimensional time series: Inference for the number of factors', *Annals of Statistics* **40**(2), 694–726.

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

Moon, H. R. and Weidner, M. (2015), 'Linear regression for panel with unknown number of factors as interactive fixed effects', *Econometrica* **83**(4), 1543–1579.

Su, L. and Chen, Q. (2013), 'Testing homogeneity in panel data models with interactive fixed effects', *Econometric Theory* **29**(6), 1079–1135.

Su, L., Jin, S. and Zhang, Y. (2015), 'Specification test for panel data models with interactive fixed effects', *Journal of Econometrics* **186**(1), 222–244.

Sun, Y., Cai, Z. and Li, Q. (2016), 'A consistent nonparametric test on semiparametric smooth coefficient models with integrated time series', *Econometric Theory* **32**(4), 988–1022.

Vogt, M. (2012), 'Nonparametric regression for locally stationary time series', *Annals of Statistics* **40**(5), 2601–2633.

Wang, H. and Xia, Y. (2009), 'Shrinkage estimation of the varying coefficient', *Journal of the American Statistical Association* **104**(486), 747–757.