Article

Common Correlated Effects Estimation for Dynamic Heterogeneous Panels with Non-Stationary Multi-Factor Error Structures

Shiyun Cao [1] and Qiankun Zhou [2,*]

1   School of Science, Guangxi University of Science and Technology, Liuzhou 545006, China
2   Department of Economics, Louisiana State University, Baton Rouge, LA 70803, USA
*   Correspondence: qzhou@lsu.edu

Abstract: In this paper, we consider the estimation of a dynamic panel data model with non-stationary multi-factor error structures. We adopted the common correlated effect (CCE) estimation and established the asymptotic properties of the CCE and common correlated effects mean group (CCEMG) estimators, as N and T tend to infinity. The results show that both the CCE and CCEMG estimators are consistent and the CCEMG estimator is asymptotically normally distributed. The theoretical findings were supported for small samples by an extensive simulation study, showing that the CCE estimators are robust to a wide variety of data generation processes. Empirical findings suggest that the CCE estimation is widely applicable to models with non-stationary factors. The proposed procedure is also illustrated by an empirical application to analyze the U.S. cigar dataset.

Keywords: dynamic panel models; cross-sectional dependence; non-stationary; common factors; common correlated effects

JEL Classification: C01; C13; C23

1. Introduction

Recently, there has been increased interest in the analysis of panel data models with cross-sectionally dependent errors (also known as unobserved common factors or multi-factor error structures), which are motivated by empirical applications in economics, such as common shocks and the global financial crisis, see Omay and Kan (2010); Bussiere et al. (2013); Eberhardt et al. (2013) and Chudik et al. (2017), etc. The dependencies across the units violate the traditional assumption of independent and identically distributed errors; conventional panel estimation methods (such as fixed effects estimation) could have serious consequences and lead to inconsistent estimations and misleading inferences. Therefore, in econometrics literature, much effort has been devoted to the estimations for panels with cross-sectional dependence, for example, Pesaran (2006); Bai (2009); Zaffaroni (2009); Greenaway-McGrevy et al. (2012); Kao et al. (2012); Chudik and Pesaran (2013); Moon and Weidner (2015, 2017), among others. See also Chudik and Pesaran (2015b) for a survey of recent developments in large panel models with cross-sectional dependence.

Among these studies, a predominant approach of dealing with cross-sectionally dependent errors in panel models is the so-called common correlated effect (CCE) method proposed by Pesaran (2006)[1]. The basic idea of CCE estimation is to proxy the unobserved common factors using the cross-sectional averages of the observables in the regression. Comparatively, it has several advantages. For instance, it can be computed by least squares to auxiliary regression, and it does not require the knowledge of the number of unobserved factors. The CCE method has been further developed and applies to different types of panel models. To name a few, Chudik and Pesaran (2015a) suggested the CCE approach to analyze dynamic heterogeneous panels with stationary unobserved common factors.

Kapetanios et al. (2011) extended the CCE method to static panel data models with non-stationary multi-factor error structures. Westerlund et al. (2019) considered the CCE for short panels, and Zhou and Zhang (2016) extended the CCE for unbalanced panels.

Among the aforementioned works, there is a gap in the CCE estimation for dynamic panels with non-stationary unobservable factors. To fill this gap, in this paper we consider a linear dynamic heterogeneous panel data model with non-stationary unobserved common factors when both the cross-sectional and time dimensions of the dataset grow to infinity. Under these settings, we find that the CCE estimator of the individual coefficient is consistent, and the CCE mean group (CCEMG) estimator is consistent and has a normal limit distribution. The practical implication of this finding is that for inferential purposes of the CCE estimation, one does not necessarily need to test the stationarity of the unobserved common factors in the model. The finite sample properties are examined through Monte Carlo simulations and the simulation results confirm our theoretical findings in the paper. Moreover, the proposed procedure is illustrated by an empirical application, which analyzes the U.S. cigar dataset.

The rest of the paper is organized as follows. Section 2 sets up the basic model and introduces the CCE estimation of the dynamic heterogeneous panel data model with common factors. The asymptotics of the CCE estimation with non-stationary unobserved common factors is provided in Section 3. Monte Carlo simulation results and an empirical application are reported in Sections 4 and 5, respectively. The concluding remarks are made in Section 6. Proof of the main results is provided in Appendix A.

Notation: The letter K stands for a finite positive constant. All vectors are column vectors represented by bold lower case letters, and matrices are represented by bold capital letters. Let $\|\mathbf{A}\| = \sqrt{tr(\mathbf{A}\mathbf{A}')}$ denote the Frobenius norm. $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq n} \Sigma_{i=1}^{n} |a_{ij}|$, and $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq n} \Sigma_{j=1}^{n} |a_{ij}|$ denote the maximum absolute column and row sum matrix norms, respectively. $\mathbf{A}^+$ denotes the Moore–Penrose inverse of $\mathbf{A}$, $rank(\mathbf{A})$ and $\varrho(\mathbf{A})$ denotes the rank and the spectral radius of $\mathbf{A}$, respectively.

2. Dynamic Panel Data Model with Non-Stationary Unobserved Common Factors

2.1. The Model

We assume the scalar dependent variable y_{it} and regressors $\mathbf{x}_{it}$ are generated as follows[2]

$$y_{it} = c_{yi} + \phi_i y_{i,t-1} + \beta_i' \mathbf{x}_{it} + \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \tag{1}$$

and

$$\mathbf{x}_{it} = \mathbf{c}_{xi} + \alpha_i y_{i,t-1} + \mathbf{\Gamma}_i' \mathbf{f}_t + \mathbf{u}_{it}, \tag{2}$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, where c_{yi} and $\mathbf{c}_{xi}$ are individual fixed effects for unit i, $\mathbf{x}_{it}$ is a $k \times 1$ vector of the regressors specific to cross-sectional unit i at time t, ε_{it} are the individual-specific (idiosyncratic) errors and $\mathbf{u}_{it}$ are the individual-specific components of $\mathbf{x}_{it}$, γ_i and $\mathbf{\Gamma}_i$ are $m \times 1$ and $m \times k$ factor loading matrices, and the $m \times 1$ vector $\mathbf{f}_t$ represents unobserved common factors. In what follows, we maintain the restriction that model (1) is stationary, such that $0 < |\phi_i| < 1$ for $i = 1, 2, \ldots, N$.

Models (1)–(2) have been widely studied in the literature; see, for instance, Pesaran (2006), Chudik and Pesaran (2015a), Westerlund et al. (2019), and the references therein. We follow these studies to consider the CCE estimation for ϕ_i and β_i, and reexamine the validity of the CCE estimation when $\mathbf{f}_t$ is non-stationary.

2.2. CCE Estimation

Following Chudik and Pesaran (2015a), let $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$, then (1) and (2) can be compactly written as

$$\mathbf{z}_{it} = \mathbf{c}_{zi} + \mathbf{A}_i \mathbf{z}_{i,t-1} + \mathbf{A}_{0i}^{-1} \mathbf{C}_i \mathbf{f}_t + \mathbf{e}_{zit}, \tag{3}$$

where $\mathbf{c}_{zi} = \mathbf{A}_{0i}^{-1}\mathbf{c}_i$, $\mathbf{A}_i = \mathbf{A}_{0i}^{-1}\mathbf{A}_{1i}$, $\mathbf{C}_i = (\gamma_i, \mathbf{\Gamma}_i)'$, and $\mathbf{e}_{zit} = \mathbf{A}_{0i}^{-1}\mathbf{e}_{it}$, with $\mathbf{c}_i = (c_{yi}, \mathbf{c}'_{xi})'$, $\mathbf{e}_{it} = (\varepsilon_{it}, \mathbf{u}'_{it})'$, and

$$\mathbf{A}_{0i} = \begin{pmatrix} 1 & -\beta'_i \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix}, \quad \mathbf{A}_{1i} = \begin{pmatrix} \phi_i & 0 \\ \alpha_i & \mathbf{0} \end{pmatrix}. \tag{4}$$

If the support of $\varrho(\mathbf{A}_i)$ lies strictly inside the unit circle, then (3) can be rewritten as the following distributed lag form

$$\mathbf{z}_{it} = \sum_{l=0}^{\infty} \mathbf{A}_i^l \left(\mathbf{c}_{zi} + \mathbf{A}_{0i}^{-1}\mathbf{C}_i\mathbf{f}_{t-l} + \mathbf{e}_{zi,t-l} \right), \tag{5}$$

for $i = 1, 2, \ldots, N$.

Taking the cross-sectional average of (5) yields

$$\bar{\mathbf{z}}_t = \bar{\mathbf{c}}_z + \mathbf{\Lambda}(L)\mathbf{C}\mathbf{f}_t + O_p(N^{-1/2}),$$

where $\bar{\mathbf{z}}_t = \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_{it}$ is a $k+1$ dimensional vector of the cross-section average, $\bar{\mathbf{c}}_z = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{I}_{k+1} - \mathbf{A}_i)^{-1}\mathbf{c}_{zi}$, and $\mathbf{C} = E(\mathbf{C}_i) = (\gamma, \mathbf{\Gamma})'$, $\mathbf{\Lambda}(L) = \sum_{l=0}^{\infty}\mathbf{\Lambda}_l L^l$ with $\mathbf{\Lambda}_l = E\left(\mathbf{A}_i^l \mathbf{A}_{0i}^{-1}\right)$, and L being the lag operator. Furthermore, if $\mathbf{\Lambda}(L)$ is invertible (see Assumption 4 below), then we have

$$\mathbf{C}\mathbf{f}_t = \mathbf{\Lambda}^{-1}(L)(\bar{\mathbf{z}}_t - \bar{\mathbf{c}}_z) + O_p(N^{-1/2}).$$

When the $(k+1) \times m$ matrix $\mathbf{C}$ has the full column rank, i.e., the rank condition

$$rank(\mathbf{C}) = m \leq k + 1, \tag{6}$$

holds, we have

$$\mathbf{f}_t = \mathbf{B}(L)(\bar{\mathbf{z}}_t - \bar{\mathbf{c}}_z) + O_p(N^{-1/2}), \tag{7}$$

where $\mathbf{B}(L) = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{\Lambda}^{-1}(L)$. This suggests that the contemporary and lagged value of $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{\mathbf{x}}'_t)'$ can be used as observable proxies for the unobserved common factors $\mathbf{f}_t$.

Substituting the observed proxies of the unobserved common factors (7) into (1) yields the following augmented regression

$$y_{it} = c^*_{yi} + \phi_i y_{i,t-1} + \beta'_i\mathbf{x}_{it} + \delta'_i(L)\bar{\mathbf{z}}_t + \varepsilon_{it} + O_p(N^{-1/2}),$$

or

$$y_{it} = c^*_{yi} + \phi_i y_{i,t-1} + \beta'_i\mathbf{x}_{it} + \sum_{l=0}^{p_T} \delta'_{il}\bar{\mathbf{z}}_{t-l} + w_{it}, \tag{8}$$

for $t = p_T + 1, p_T + 2, \ldots, T$, where $c^*_{yi} = c_{yi} - \delta'_i(1)\bar{\mathbf{c}}_z$, $\delta_i(L) = \mathbf{B}'(L)\gamma_i = \sum_{l=0}^{\infty}\delta_{il}L^l$, p_T is the number of lags used to truncate the infinite polynomial distributed lag function $\delta_i(L)$,[3] and the composite error w_{it} has the form of

$$w_{it} = \varepsilon_{it} + \sum_{l=p_T+1}^{\infty} \delta'_{il}\bar{\mathbf{z}}_{t-l} + O_p(N^{-1/2}).$$

For notational simplicity, let $\mathbf{y}_i = (y_{i,p_T+1}, y_{i,p_T+2}, \ldots, y_{iT})'$, $\mathbf{\Xi}_i = (\mathbf{y}_{i,-1}, \mathbf{X}_i)$, with $\mathbf{y}_{i,-1} = (y_{i,p_T}, y_{i,p_T+1}, \ldots, y_{i,T-1})'$ and $\mathbf{X}_i = (\mathbf{x}_{i,p_T+1}, \mathbf{x}_{i,p_T+2}, \ldots, \mathbf{x}_{iT})'$, and $\mathbf{w}_i = (w_{i,p_T+1}, w_{i,p_T+2}, \ldots, w_{iT})'$, then the augmented regression (8) can be expressed in vector form as

$$\mathbf{y}_i = \mathbf{\Xi}_i\pi_i + \bar{\mathbf{Q}}\mathbf{d}_i + \mathbf{w}_i, \tag{9}$$

where $\pi_i = \left(\phi_i, \beta_i'\right)'$ are the parameters of interest, $\mathbf{d}_i = (c_{yi}^*, \delta_{i0}', \delta_{i1}', \dots, \delta_{ip}')'$ are nuisance parameters, and[4]

$$\bar{\mathbf{Q}} = \begin{pmatrix} 1 & \bar{\mathbf{z}}_{p_T+1}' & \bar{\mathbf{z}}_{p_T}' & \cdots & \bar{\mathbf{z}}_1' \\ 1 & \bar{\mathbf{z}}_{p_T+2}' & \bar{\mathbf{z}}_{p_T+1}' & \cdots & \bar{\mathbf{z}}_2' \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \bar{\mathbf{z}}_T' & \bar{\mathbf{z}}_{T-1}' & \cdots & \bar{\mathbf{z}}_{T-p_T}' \end{pmatrix}. \tag{10}$$

Based on the cross-sectionally augmented regression model (9) and by the formula for partitioned regression, the CCE estimator of the individual coefficients π_i is given by

$$\hat{\pi}_i = \left(\boldsymbol{\Xi}_i' \mathbf{M}_q \boldsymbol{\Xi}_i\right)^{-1} \boldsymbol{\Xi}_i' \mathbf{M}_q \mathbf{y}_i, \tag{11}$$

which is an ordinary least squares estimate, where $\mathbf{M}_q = \mathbf{I}_{T-p_T} - \bar{\mathbf{Q}}(\bar{\mathbf{Q}}'\bar{\mathbf{Q}})^+\bar{\mathbf{Q}}'$ is an orthogonal projection matrix, with $\mathbf{I}_{T-p_T}$ a $(T - p_T)$-dimensional identity matrix. In panel models with N large, the primary parameters of interest are the means of the individual–specific coefficients, $E(\pi_i) = \pi$, which can be estimated by the common correlated effects mean group (CCEMG) estimator

$$\hat{\pi}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \hat{\pi}_i. \tag{12}$$

3. Asymptotics of CCE Estimators with Non-Stationary Factors

3.1. Assumptions

When the unobserved common factors $\mathbf{f}_t$ are stationary processes, Chudik and Pesaran (2015a) showed that the CCE estimator (11) of the individual coefficient is consistent, and the CCEMG estimator (12) is consistent and asymptotically normal. However, in practice, the common factors $\mathbf{f}_t$ may follow a non-stationary process (see Bai and Ng 2004, 2010; Pesaran 2007; Pesaran et al. 2013, among others). In this scenario, the validity of CCE estimators and their asymptotic properties need to be re-examined.

Following Kapetanios et al. (2011), we assume the unobserved common factors follow the multivariate unit root process

$$\mathbf{f}_t = \mathbf{f}_{t-1} + \varsigma_t. \tag{13}$$

To derive the asymptotic properties of the CCE type estimators (11) and (12) when $\mathbf{f}_t$ follows (13), we make the following assumptions.

Assumption 1. *(Individual-specific errors). (i) The individual–specific errors ε_{it} follow a linear stationary process with uniformly-bounded positive variance, $\sup_i \sigma_i^2 < K$, for some constant K, and uniformly-bounded fourth-order cumulants. $\mathbf{u}_{it}$ follows a linear stationary process with absolute summable auto-covariances (uniformly in i), with covariance matrices, $\boldsymbol{\Sigma}_{\mathbf{u}_i}$, which are non-singular and satisfy $\sup_i \|\boldsymbol{\Sigma}_{\mathbf{u}_i}\| < K$, and have uniformly-bounded fourth-order cumulants. (ii) ε_{it} are independently distributed of $\mathbf{u}_{jt'}$ for all $i, j, t,$ and t'. For each i, $e_{it} = (\varepsilon_{it}, \mathbf{u}_{it}')'$ is an $(k + 1) \times 1$ vector of $L_{2+\delta}, \delta > 0$, stationary near epoch dependent processes of size $2\delta/(2\delta + 4)$ on the α-mixing process of size $-(2 + \delta)/\delta$, and for $i = 1, 2, \dots, N$, $Var(e_{it}) = \boldsymbol{\Sigma}_{\mathbf{e}_i}$, which is a non-singular matrix and satisfies $\sup_i \|\boldsymbol{\Sigma}_{\mathbf{e}_i}\| < K$.*

Assumption 2. *(Factor loadings). The factor loadings γ_i and $\boldsymbol{\Gamma}_i$ are independently and identically distributed (IID) across i, and of the common factors $\mathbf{f}_t$, for all i and t, with means γ and $\boldsymbol{\Gamma}$, respectively, and the bounded second moments. In particular,*

$$\gamma_i = \gamma + \eta_{\gamma i}, \quad with \ \eta_{\gamma i} \sim IID(\mathbf{0}, \boldsymbol{\Sigma}_\gamma), for \ i = 1, 2, \dots, N,$$

and

$$vec(\mathbf{\Gamma}_i) = vec(\mathbf{\Gamma}) + \eta_{\Gamma i}, \text{ with } \eta_{\Gamma i} \sim IID(\mathbf{0}, \mathbf{\Sigma}_{\Gamma}), \text{for } i = 1, 2, \dots, N,$$

where $\mathbf{\Sigma}_{\gamma}$ *and* $\mathbf{\Sigma}_{\Gamma}$ *are* $m \times m$ *and* $mk \times mk$ *symmetric nonnegative definite matrices,* $\|\gamma\| < K, \|\mathbf{\Sigma}_{\gamma}\| < K, \|\mathbf{\Gamma}\| < K$ *and* $\|\mathbf{\Sigma}_{\Gamma}\| < K,$ *for some constant* K.

Assumption 3. *(Heterogeneous coefficients). The slope coefficients* $\pi_i = (\phi_i, \beta_i')'$ *follow the random coefficient model*

$$\pi_i = \pi + v_{\pi i}, \quad v_{\pi i} \sim IID(\mathbf{0}, \mathbf{\Sigma}_{\pi}), \quad for \ i = 1, 2, \dots, N,$$

where $\pi = E(\pi_i) = (\phi, \beta')'$, $\|\pi\| < K, \|\mathbf{\Sigma}_{\pi}\| < K$, $\mathbf{\Sigma}_{\pi}$ *is the* $(k+1) \times (k+1)$ *symmetric nonnegative definite matrix and the random deviations* $v_{\pi i}$ *are distributed independently of* $\gamma_j, \mathbf{\Gamma}_j, \varepsilon_{jt}, \mathbf{u}_{jt}$ *and* ς_t *for* $i, j,$ *and* t. *Furthermore, the support of* ϕ_i *lies strictly inside the unit circle, and* $E\|\mathbf{c}_i\| < K, E\|\alpha_i\| < K$ *for all* i, *where* $\mathbf{c}_i = (c_{yi}, \mathbf{c}_{xi}')'$.

Assumption 4. *(Exogenous regressors). Regressors* $\mathbf{x}_{it}$ *are either strictly exogenous and generated according to the canonical factor model (2) with* $\alpha_i = \mathbf{0}$, *or weakly exogenous and generated according to (2) with* α_i, *for* $i = 1, 2, \dots, N$, *IID across* i, *and independently distributed of* $v_{\pi j}$, $\gamma_j, \mathbf{\Gamma}_j, \varepsilon_{jt}, \mathbf{u}_{jt}$ *and* $\mathbf{f}_t$ *for all* $i, j,$ *and* t. *In the case where the regressors are weakly exogenous, we also assume:*

    (i) The support of $\varrho(\mathbf{A}_i)$ *lies strictly inside the unit circle, for* $i = 1, 2, \dots, N$, *where* $\mathbf{A}_i = \mathbf{A}_{0i}^{-1}\mathbf{A}_{1i}$ *with* $\mathbf{A}_{0i}$ *and* $\mathbf{A}_{1i}$ *are defined in (4).*

    (ii) The inverse of polynomial $\mathbf{\Lambda}(L) = \sum_{l=0}^{\infty} \mathbf{\Lambda}_l L_l$ *exists and has exponentially decaying coefficients, where* $\mathbf{\Lambda}_l = E(\mathbf{A}_i^l \mathbf{A}_{0i}^{-1})$.

Assumption 5. *(Rank condition). The* $(k+1) \times m$ *matrix* $\mathbf{C}$ *has a full column rank, such that*

$$rank(\mathbf{C}) = m \leq k + 1,$$

where $\mathbf{C} = E(\mathbf{C}_i) = E(\gamma_i, \mathbf{\Gamma}_i)'$.

Assumption 6. *(i) As* $(N, T) \to \infty$, *the* $(k+1) \times (k+1)$ *matrices* $\mathbf{\Psi}_{iT}^{-1} = (\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{\Xi}_i / T)^{-1}$, $\mathbf{\Psi}_{ih}^{-1} = (\mathbf{\Xi}_i' \mathbf{M}_h \mathbf{\Xi}_i / T)^{-1}$ *and* $\mathbf{\Psi}_{ig}^{-1} = (\mathbf{\Xi}_i' \mathbf{M}_g \mathbf{\Xi}_i / T)^{-1}$ *exist for all* i, *and* $\mathbf{\Psi}_{iT}^{-1}, \mathbf{\Psi}_{ih}^{-1}$ *and* $\mathbf{\Psi}_{ig}^{-1}$ *have finite second-order moments for all* i, *where* $\mathbf{M}_h = \mathbf{I}_{T-p_T} - \mathbf{H}(\mathbf{H}'\mathbf{H})^{+}\mathbf{H}'$ *and* $\mathbf{M}_g = \mathbf{I}_{T-p_T} - \mathbf{G}(\mathbf{G}'\mathbf{G})^{+}\mathbf{G}'$ *are projection matrices, where* $\mathbf{A}^{+}$ *denotes the Moore–Penrose generalized inverse of* $\mathbf{A}$, $\mathbf{H}$, *and* $\mathbf{G}$, *defined as*

$$\mathbf{H} = \begin{pmatrix} 1 & \mathbf{h}'_{p_T+1} & \cdots & \mathbf{h}'_1 \\ 1 & \mathbf{h}'_{p_T+2} & \cdots & \mathbf{h}'_2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{h}'_T & \cdots & \mathbf{h}'_{T-p_T} \end{pmatrix}, \mathbf{G} = (\tau, \tilde{\mathbf{F}}) = \begin{pmatrix} 1 & \mathbf{f}'_{p_T+1} & \cdots & \mathbf{f}'_1 \\ 1 & \mathbf{f}'_{p_T+2} & \cdots & \mathbf{f}'_2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{f}'_T & \cdots & \mathbf{f}'_{T-P_T} \end{pmatrix},$$

where $\mathbf{h}_t = \mathbf{\Psi}(L)\mathbf{f}_t + \bar{\mathbf{c}}_z$ *with* $\mathbf{\Psi}(L) = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1}\mathbf{A}_{0i}^{-1}\mathbf{C}_i$ *and* $\bar{\mathbf{c}}_z = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{I}_{k+1} - \mathbf{A}_i)^{-1}\mathbf{c}_{zi}$, $\tau = (1, 1, \dots 1)'$ *is a* $(T - p_T) \times 1$ *vector of ones.*

    (ii) The matrix $\mathbf{\Psi}^{*} = \lim_{N \to \infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{\Sigma}_{\Omega i}$ *is non-singular, where* $\mathbf{\Sigma}_{\Omega i} = \mathbf{\Psi}_{\xi i} \begin{pmatrix} \sigma_i^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\mathbf{u}_i} \end{pmatrix} \mathbf{\Psi}'_{\xi i}$ *with* $\mathbf{\Psi}_{\xi i} = \begin{pmatrix} \mathbf{S}'_y(\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1}L \\ \mathbf{S}'_x(\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} \end{pmatrix} \mathbf{A}_{0i}^{-1}$ *a* $(k+1) \times (k+1)$ *matrix,* $\mathbf{S}'_y = \begin{pmatrix} 1 & \mathbf{0} \end{pmatrix}_{1 \times (k+1)}$, *and* $\mathbf{S}'_x = \begin{pmatrix} \mathbf{0} & \mathbf{I}_k \end{pmatrix}_{k \times (k+1)}$.

Assumption 7. ς_t *in (13) is an* $m \times 1$ *vector of* $L_{2+\delta}, \delta > 0$, *stationary near epoch-dependent processes of size* $1/2$, *on an* α-*mixing process of size* $-(2+\delta)/\delta$, *and is distributed independently of the idiosyncratic errors* ε_{it} *and* $\mathbf{u}_{it}$ *for all* i *and* t.

Several remarks can be made for these assumptions. Assumptions 1–3 are quite standard in the literature for (dynamic) panel models with cross-sectional dependence, for example, see Pesaran (2006) and Kapetanios et al. (2011) and the references therein. Assumption 4 is also made on Chudik and Pesaran (2015a) for exogenous regressors and stationarity conditions for dynamic panels. Assumption 5 is a common condition for the implementation of the CCE estimation (e.g., Pesaran (2006) and Chudik and Pesaran (2015a), etc.), which implies that there are more included regressors than the unobserved factors in the model. See Juodis et al. (2021) for a detailed discussion of the validity of the rank condition and the resulting asymptotics for the CCE estimation. Assumption 6 is a common assumption for the CCE estimation and it is imposed for the partition regression in augmented regression for the dynamic panels (e.g., Chudik and Pesaran 2015a). Assumption 7 requires that the error structures in the unit root process $\mathbf{f}_t$ are stationary.

3.2. Asymptotics

Under these assumptions, we can establish the asymptotic properties of CCE estimators (11) and (12) when $\mathbf{f}_t$ is non-stationary. To begin with, we note that for the original model (1), it can be rewritten as in the vector form

$$\mathbf{y}_i = c_{yi}\tau + \phi_i \mathbf{y}_{i,-1} + \mathbf{X}_i \beta_i + \mathbf{F}\gamma_i + \varepsilon_i, \tag{14}$$

or more compactly as

$$\mathbf{y}_i = c_{yi}\tau + \Xi_i \pi_i + \mathbf{F}\gamma_i + \varepsilon_i, \tag{15}$$

for $i = 1, 2, \ldots, N$, where $\mathbf{y}_i = (y_{i,p_T+1}, y_{i,p_T+2}, \ldots, y_{iT})'$, $\mathbf{X}_i = (\mathbf{x}_{i,p_T+1}, \ldots, \mathbf{x}_{iT})'$, $\mathbf{y}_{i,-1} = (y_{i,p_T}, y_{i,p_T+1}, \ldots, y_{i,T-1})'$, $\Xi_i = (\mathbf{y}_{i,-1}, \mathbf{X}_i)$, $\mathbf{F} = (\mathbf{f}_{p_T+1}, \ldots, \mathbf{f}_T)'$, and $\varepsilon_i = (\varepsilon_{i,p_T+1}, \ldots, \varepsilon_{iT})'$.

Using the CCE estimator (11) into (15), we have

$$\hat{\pi}_i - \pi_i = \left(\frac{\Xi_i' \mathbf{M}_q \Xi_i}{T}\right)^{-1} \frac{\Xi_i' \mathbf{M}_q \mathbf{F}\gamma_i}{T} + \left(\frac{\Xi_i' \mathbf{M}_q \Xi_i}{T}\right)^{-1} \frac{\Xi_i' \mathbf{M}_q \varepsilon_i}{T}, \tag{16}$$

which shows that the asymptotics of $\hat{\pi}_i$ depends on the unobserved factors through $T^{-1}\Xi_i' \mathbf{M}_q \mathbf{F}$.

Using the results in Lemma A2, A5, and A6 in Appendix A, we obtain

$$\frac{\Xi_i' \mathbf{M}_q \mathbf{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \text{uniformly over } i, \tag{17}$$

and, thus,

$$\hat{\pi}_i - \pi_i = \left(\frac{\Xi_i' \mathbf{M}_q \Xi_i}{T}\right)^{-1} \frac{\Xi_i' \mathbf{M}_q \varepsilon_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{18}$$

when the rank condition (6) is satisfied. The above results are summarized in the following theorem, establishing the consistency of the CCE estimator of individual coefficients of interest.

Theorem 1. *Consider the panel models (1) and (2), suppose Assumption 1–7 hold, then, as $(N, T, p_T) \xrightarrow{j} \infty$, such that $p_T^3/T \to \lambda$, $0 < \lambda < \infty$, we have*

$$\hat{\pi}_i - \pi_i \xrightarrow{p} 0.$$

See the Appendix A for the proof.

Remark 1. *The above theorem suggests that the CCE estimator of the individual slope coefficient is consistent even if the common factors are non-stationary. When the rank condition (6) is not satisfied, the CCE estimator of the individual slope coefficients would be inconsistent due to the*

correlation of $\mathbf{x}_{it}$ and $\mathbf{f}_t$. See Juodis et al. (2021) for more discussions on the validity of the CCE estimator when the rank condition does not hold.

Next, we establish the asymptotic properties of the CCEMG estimator of the mean group coefficients, $\pi = E(\pi_i)$. We have

$$
\sqrt{N}(\hat{\pi}_{MG} - \pi) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_{\pi i} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{\Xi_i' \mathbf{M}_q \Xi_i}{T} \right)^{-1} \frac{\Xi_i' \mathbf{M}_q \mathbf{F}}{T} \gamma_i
$$
$$
+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{\Xi_i' \mathbf{M}_q \Xi_i}{T} \right)^{-1} \frac{\Xi_i' \mathbf{M}_q \varepsilon_i}{T}.
$$

When the rank condition is satisfied, by (17), we have

$$
\frac{\sqrt{N} \Xi_i' \mathbf{M}_q \mathbf{F}}{T} = O_p\left(\frac{1}{\sqrt{N}} \right) + O_p\left(\frac{1}{\sqrt{T}} \right),
$$

hence, we can obtain

$$
\sqrt{N}(\hat{\pi}_{MG} - \pi) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_{\pi i} + O_p\left(\frac{1}{\sqrt{N}} \right) + O_p\left(\frac{1}{\sqrt{T}} \right) \overset{d}{\sim} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_{\pi i},
$$

and, thus,

Theorem 2. *Consider the panel models (1) and (2), suppose Assumptions 1–7 hold, as $(N, T, p_T) \overset{j}{\to} \infty$, such that $p_T^3/T \to \lambda, 0 < \lambda < \infty$, then we have*

$$
\hat{\pi}_{MG} - \pi \overset{p}{\to} 0.
$$

If it is further assumed that $N/T \to \varphi, 0 < \varphi < \infty$, then

$$
\sqrt{N}(\hat{\pi}_{MG} - \pi) \overset{d}{\to} N(0, \Sigma_{MG}).
$$

The asymptotic variance of $\hat{\pi}_{MG}$ can be consistently estimated nonparametrically by

$$
\hat{\Sigma}_{MG} = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{\pi}_i - \hat{\pi}_{MG})(\hat{\pi}_i - \hat{\pi}_{MG})'.
$$

For the results in both Theorems 1 and 2, we find that, for models with non-stationary common factors, although the intermediate results needed for deriving the asymptotic properties of the common correlated effects estimators significantly differ from the stationary case, as in Chudik and Pesaran (2015a), the final results are surprisingly similar. This is in direct contrast to the usual phenomenon where distributional results of $I(1)$ processes are radically different from those of $I(0)$ processes.

Remark 2. *For the consistency of $\hat{\pi}_i$ and $\hat{\pi}_{MG}$, no restrictions on the relative expansion rates of N and T to infinity are required. However, they require $N/T \to \varphi, 0 < \varphi < \infty$ for the derivation of the asymptotic distribution of $\hat{\pi}_{MG}$ due to the time series bias, which arises from the presence of lagged values of the dependent variable; therefore, it is unsuitable for panels with T being small relative to N.*

Including a lagged dependent variable as the regressor in the model could induce the estimators with time series bias of order $O(T^{-1})$. When T is not large, the bias is non-negligible; hence, a certain bias correction approach should be considered. In the simulations

below, we consider the Jackknife bias-corrected method for bias reduction (e.g., see (21) below), which is used extensively in the relevant literature (e.g., Hahn and Newey 2004).

4. Monte Carlo Simulation

In this section, we investigate the finite sample properties of the CCEMG estimation for dynamic heterogeneous panels with non-stationary common factors. We consider the following data-generating processes[5]

$$y_{it} = c_{yi} + \phi_i y_{i,t-1} + \beta_{0i} x_{it} + \beta_{1i} x_{i,t-1} + \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \tag{19}$$

and

$$x_{it} = c_{xi} + \alpha_{xi} y_{i,t-1} + \gamma_{xi}' \mathbf{f}_t + u_{it}. \tag{20}$$

for $i = 1, 2, \ldots, N$, and $t = -99, \ldots, 0, \ldots, T$. Let $\phi_i \sim IIDU(0, 0.8)$, $\beta_{0i} \sim IIDU(0.5, 1)$, $\beta_{1i} = -0.5$ and $\alpha_{xi} \sim IIDU(0, 0.35)$, $c_{yi} \sim IIDN(1, 1)$, $c_{xi} = c_{yi} + \epsilon_{c_{xi}}$ with $\epsilon_{c_{xi}} \sim IIDN(0, 1)$. The main purpose of this paper is to illustrate the validity of the CCEMG estimator in the case of non-stationary unobserved common factors; hence, for the unobserved common factors $\mathbf{f}_t$, we consider the following three different non-stationary DGPs:

DGP 1. Two non-stationary unobserved common factors ($m = 2$), $f_{lt} = f_{l,t-1} + \varsigma_{flt}$, $\varsigma_{flt} \sim IIDN\left(0, \sigma_{fl}^2\right)$, where $\sigma_{fl} = 0.2$, for $l = 1, 2$, and $t = -99, \ldots, 0, \ldots, T$.

DGP 2. One non-stationary unobserved common factor and a stationary common factor ($m = 2$), $f_{1t} = f_{1,t-1} + \varsigma_{flt}$, $f_{2t} = 0.6 f_{2,t-1} + \varsigma_{flt}$, $\varsigma_{flt} \sim IIDN\left(0, \sigma_{fl}^2\right)$, where $\sigma_{fl} = 0.5$, for $l = 1, 2$, and $t = -99, \ldots, 0, \ldots, T$.

DGP 3. Cointegrated unobserved common factors ($m = 2$), $f_{1t} = f_{1,t-1} + \varsigma_{f1t}$, $f_{2t} = 0.5 f_{1t} + \varsigma_{f2t}$, $\varsigma_{flt} \sim IIDN\left(0, \sigma_{fl}^2\right)$, where $\sigma_{fl} = 1$, for $l = 1, 2$, and $t = -99, \ldots, 0, \ldots, T$.

For the above DGPs, the starting values are $f_{l,-100} = 0$, for $l = 1, 2$; the first 100 observations are discarded.

Correspondingly, the factor loadings are generated independently across replications as

$$\gamma_{il} = \gamma_l + \eta_{i,\gamma l}, \eta_{i,\gamma l} \sim IIDN\left(0, \sigma_{\gamma l}^2\right),$$

and

$$\gamma_{xil} = \gamma_{xl} + \eta_{i,\gamma xl}, \eta_{i,\gamma xl} \sim IIDN\left(0, \sigma_{\gamma xl}^2\right),$$

for $i = 1, 2, \ldots, N$ and $l = 1, 2$, where $\sigma_{\gamma l}^2 = 0.2^2$, $\sigma_{\gamma xl}^2 = 0.3^2$, and $\gamma_l = \sqrt{b_{\gamma l}}$, $\gamma_{xl} = \sqrt{lb_{xl}}$ for $l = 1, 2$, where $b_{\gamma l} = 1/2 - \sigma_{\gamma l}^2$ and $b_{xl} = 1/2 - \sigma_{\gamma xl}^2$ for $l = 1, 2$.

For the idiosyncratic errors, $\varepsilon_{it} \sim IIDN(0, 1)$ for all i and t, and the unit-specific components u_{it} are generated as independent stationary AR(1) processes:

$$u_{it} = \rho_{xi} u_{i,t-1} + \epsilon_{xit}, \rho_{xi} \sim IIDU(0, 0.95), \epsilon_{xit} \sim IIDN(0, 1),$$

for $i = 1, 2, \ldots, N$ and $t = -99, \ldots, 0, \ldots, T$ with the starting values $u_{i,-100} = 0$. The first 100 observations are discarded.

We consider the combination of $N = 50, 100, 200$, and $T = 50, 100, 150, 200$. The number of replications is set at 2000 times. In what follows, we focus on the lagged coefficient ϕ (the cross-section mean of ϕ_i), as well as β_0 (the cross-section mean of β_{0i}). To save space, we only report the results of β_0 since the results for β_1 are very similar to that of β_0 and they are available upon request.

Two estimators are considered in the simulation. The first is the main result of the CCEMG estimator $\hat{\pi}_{MG}$ given in (12), in which, the lag order p_T is selected to satisfy $p_T^3/T \to \lambda$, as $T \to \infty$, for some $0 < \lambda < \infty$; that is, $p_T = [T^{1/3}]$, which works well in our

Monte Carlo design[6]. The second is the Jackknife bias-corrected CCEMG estimator, which is constructed as

$$\hat{\pi}_{Jack-MG} = 2\hat{\pi}_{MG} - \frac{1}{2}\left(\hat{\pi}^a_{MG} + \hat{\pi}^b_{MG}\right), \tag{21}$$

where $\hat{\pi}^a_{MG}$ is the CCEMG estimator calculated using the first two-thirds of the available time period, namely over the period $t = 1, 2, \ldots, [2T/3]$, and $\hat{\pi}^b_{MG}$ denotes the CCEMG estimator computed using the observations over the period $t = [T/3], [T/3] + 1, \ldots, T$, where $[T/3]$ denotes the integer part of $T/3$. Note that a new strategy is applied to improve the performance of the Jackknife estimator, i.e., the whole time period is divided into three parts, the first two-thirds of the available period is applied to calculate the first estimator and another one is computed from the last two-thirds of the period. We find that, in our settings, this division strategy performs better than the half-panel Jackknife method discussed in Chudik and Pesaran (2015a).

We used the statistical software MATLAB to conduct the Monte Carlo experiments; the simulation results are summarized in Tables 1–3 for DGPs 1–3, respectively.

From Table 1, we note that for the estimation of ϕ, the CCEMG performs well in terms of bias and RMSE, with the bias diminishing as T is increased, and the associated RMSEs fall steadily when T increases, which implies that the CCEMG estimator is consistent. However, it still suffers from the time series bias when T is small. While the Jackknife bias-corrected CCEMG estimator is quite effective at reducing the time series bias of the CCEMG estimator, the bias has been significantly reduced compared with the original CCEMG estimator when T was not large, and the RMSE also decreased with the increase of either N or T. Similar findings can be observed for β_0.

In order to evaluate the robustness of various estimators, we considered additional results in Tables 2 and 3 for DGPs with both stationary and non-stationary factors or cointegrated factors. Similar to the case with non-stationary factors in DGP1, we find that the CCEMG estimator still performs well regardless of the number of common factors and the non-stationary type, and it can be improved by the Jackknife bias-corrected for the estimation of the autoregressive coefficient ϕ, the CCEMG estimator of the slope coefficient β_0 performs very well in almost all cases.

Table 1. Estimation results for DGP 1.

Parameter		Bias				RMSE			
	(N, T)	50	100	150	200	50	100	150	200
ϕ	CCEMG estimation								
	50	−0.1065	−0.0393	−0.0163	−0.0004	0.1131	0.0530	0.0392	0.0371
	100	−0.1085	−0.0392	−0.0156	−0.0004	0.1120	0.0476	0.0311	0.0284
	200	−0.1105	−0.0402	−0.0163	−0.0012	0.1116	0.0450	0.0265	0.0220
	Jackknife bias-corrected CCEMG estimation								
	50	−0.0508	−0.0126	−0.0071	−0.0043	0.0747	0.0417	0.0415	0.0440
	100	−0.0411	−0.0124	−0.0036	0.0034	0.0689	0.0324	0.0309	0.0365
	200	−0.0417	−0.0127	−0.0042	−0.0039	0.0664	0.0273	0.0253	0.0309
β_0	CCEMG estimation								
	50	0.0136	0.0071	0.0032	0.0012	0.0461	0.0332	0.0282	0.0275
	100	0.0129	0.0058	0.0029	0.0008	0.0341	0.0232	0.0200	0.0192
	200	0.0119	0.0049	0.0024	0.0003	0.0252	0.0169	0.0150	0.0139
	Jackknife bias-corrected CCEMG estimation								
	50	0.0112	0.0043	0.0011	−0.0007	0.0550	0.0361	0.0307	0.0289
	100	0.0098	0.0030	0.0003	−0.0015	0.0397	0.0251	0.0215	0.0206
	200	0.0091	0.0020	0.0000	0.0017	0.0281	0.0180	0.0160	0.0150

Table 2. Estimation results for DGP 2.

Parameter	(N, T)	Bias				RMSE			
		50	100	150	200	50	100	150	200
ϕ		CCEMG estimation							
	50	−0.0983	−0.0384	−0.0188	−0.0093	0.1053	0.0513	0.0386	0.0348
	100	−0.1004	−0.0389	−0.0200	−0.0104	0.1040	0.0461	0.0312	0.0259
	200	−0.1015	−0.0395	−0.0193	−0.0102	0.1036	0.0434	0.0271	0.0203
		Jackknife bias-corrected CCEMG estimation							
	50	−0.0506	−0.0179	−0.0069	−0.0014	0.0840	0.0410	0.0360	0.0346
	100	−0.0491	−0.0172	−0.0072	0.0015	0.0768	0.0316	0.0262	0.0245
	200	−0.0457	−0.0167	−0.0071	−0.0013	0.0704	0.0257	0.0194	0.0177
β_0		CCEMG estimation							
	50	0.0124	0.0077	0.0048	0.0036	0.0451	0.0334	0.0285	0.0275
	100	0.0122	0.0063	0.0042	0.0033	0.0335	0.0235	0.0202	0.0193
	200	0.0112	0.0056	0.0039	0.0028	0.0248	0.0171	0.0151	0.0138
		Jackknife bias-corrected CCEMG estimation							
	50	0.0109	0.0061	0.0037	0.0027	0.0535	0.0362	0.0306	0.0284
	100	0.0104	0.0045	0.0026	0.0020	0.0396	0.0252	0.0212	0.0200
	200	0.0090	0.0036	0.0024	0.0017	0.0279	0.0179	0.0156	0.0145

Table 3. Estimation results for DGP 3.

Parameter	(N, T)	Bias				RMSE			
		50	100	150	200	50	100	150	200
ϕ		CCEMG estimation							
	50	−0.0649	−0.0330	−0.0154	−0.0076	0.0733	0.0475	0.0363	0.0342
	100	−0.0760	−0.0370	−0.0159	−0.0119	0.0801	0.0440	0.0313	0.0259
	200	−0.0789	−0.0378	−0.0179	−0.0127	0.0814	0.0434	0.0284	0.0222
		Jackknife bias−corrected CCEMG estimation							
	50	−0.0195	−0.0073	0.0041	0.0010	0.0425	0.0368	0.0340	0.0335
	100	−0.0246	−0.0098	−0.0030	0.0001	0.0410	0.0272	0.0252	0.0235
	200	−0.0309	−0.0091	−0.0059	−0.0026	0.0395	0.0224	0.0182	0.0171
β_0		CCEMG estimation							
	50	0.0094	0.0063	0.0043	0.0045	0.0412	0.0329	0.0299	0.0276
	100	0.0092	0.0060	0.0045	0.0039	0.0312	0.0237	0.0205	0.0198
	200	0.0092	0.0062	0.0043	0.0038	0.0224	0.0174	0.0148	0.0141
		Jackknife bias-corrected CCEMG estimation							
	50	0.0069	0.0045	0.0032	0.0039	0.0441	0.0341	0.0307	0.0284
	100	0.0068	0.0041	0.0035	0.0027	0.0330	0.0243	0.0212	0.0202
	200	0.0058	0.0040	0.0029	0.0021	0.0230	0.0175	0.0148	0.0143

Overall, the findings of our Monte Carlo simulations show that, if the parameter of interest is the mean coefficient of the regressors, β_0, the CCEMG estimator performs well even if N and T are not large. For the mean coefficient of the lagged dependent, ϕ, the CCEMG estimator is still consistent, but it suffers from the time series bias unless T is sufficiently large and, thus, the Jackknife bias-corrected CCEMG estimator is proposed, it helps to mitigate the time series bias.

5. Empirical Study

In this section, we illustrate our method by considering the U.S. Cigar dataset, which is frequently used in the literature on panel models (e.g., Baltagi and Li 2004; Bada and

Liebl 2014). The panel contains the per capita cigarette consumption of $N = 46$ American states from 1963 to 1992 ($T = 30$) as well as data on the income per capita and cigarette prices; the dataset can be obtained from the R package *phtt*.

To test the cross-sectional dependence in the panel data, following Pesaran (2015) and Bailey et al. (2016), we compute the CD statistic and the α statistic for the variables of interest in Table 4. As can be seen from the table, the CD statistics turn out to be 101.519, 166.27, and 154.142 for consumption, income, and price, respectively; these are highly significant and reject the null hypothesis of weak cross-sectional dependence for all three variables. Additionally, the estimates of α together with their 95% confidence bands further confirm the above results. As a result, we can conclude that there is an obvious cross-sectional dependence for these three variables.

Table 4. Exponent of the cross-sectional dependence of variables.

Variable	CD	p-Value	α	$\alpha_{0.025}$	$\alpha_{0.975}$
consumption	101.519	0.000	0.975	0.887	1.064
income	166.270	0.000	1.004	-0.635	2.644
price	154.142	0.000	1.004	0.620	1.389

To investigate the relationship between the per capita cigarette consumption and the income per capita as well as cigarette prices, following Baltagi and Li (2004), we consider the panel model

$$y_{it} = c_i + \phi_i y_{i,t-1} + \beta_{1i} x_{1it} + \beta_{2i} x_{2it} + e_{it}, \tag{22}$$

where y_{it}, x_{1it}, and x_{2it} denote the per capita cigarette consumption, the income per capita, and cigarette price for the ith state at time t, respectively, and the idiosyncratic error has the multi-factor structure

$$e_{it} = \gamma_i' \mathbf{f}_t + \varepsilon_{it}. \tag{23}$$

The proposed dynamic CCE approach is applied to estimate the coefficients in model (22), and the augmented equation to be estimated can be written as

$$y_{it} = c_i + \phi_i y_{i,t-1} + \beta_{1i} x_{1it} + \beta_{2i} x_{2it} + \sum_{l=0}^{p_T} \delta_{il}' \bar{\mathbf{z}}_{t-l} + w_{it}, \tag{24}$$

where the number of lags $p_T = [\sqrt[3]{T}] = 3$, and $\bar{\mathbf{z}}_t = (\bar{y}_{t-1}, \bar{x}_{1t}, \bar{x}_{2t})'$. We focus on the CCEMG estimators and the results are presented in Table 5.

The following conclusions can be drawn from Table 5. On the one hand, the income per capita has a positive effect on the per capita cigarette consumption, while the increase in cigarette price will restrain cigarette consumption to a certain extent, and both are significant. These results are consistent with the conclusions of Bada and Liebl (2014). On the other hand, the lagged explained variable is highly significant, indicating that it is appropriate to use dynamic models for the per capita cigarette consumption.

Table 5. Estimation results (Jackknife bias-corrected CCEMG).

Variable	coef.	Std.Err.	p-Value	$CI_{0.025}$	$CI_{0.975}$
L. consumption	0.368	0.091	0.000	0.190	0.545
income	0.936	0.387	0.016	0.177	1.695
price	-0.629	0.115	0.000	-0.854	-0.404

Note: *L.* consumption denotes the first lag of cigarette consumption.

To illustrate the heterogeneous slopes across states, we display both the CCE and the CCEMG estimators in Figure 1, which clearly show that the estimates of coefficients vary from state to state, reflecting the heterogeneity among states. Moreover, to illustrate the

potential non-stationarity of unobservable common factors in (23), we consider the method proposed by Bada and Kneip (2014) to select the number of unobservable common factors and estimate the selected common factors. The results are given in Figure 2, where the top panel shows the estimated common factors and the bottom panel shows the estimated time-varying individual effects of $N = 46$ states. As can be seen from the figure, five common factors have been selected, among which the first and second common factors have obvious tendencies and violate the stationarity condition.
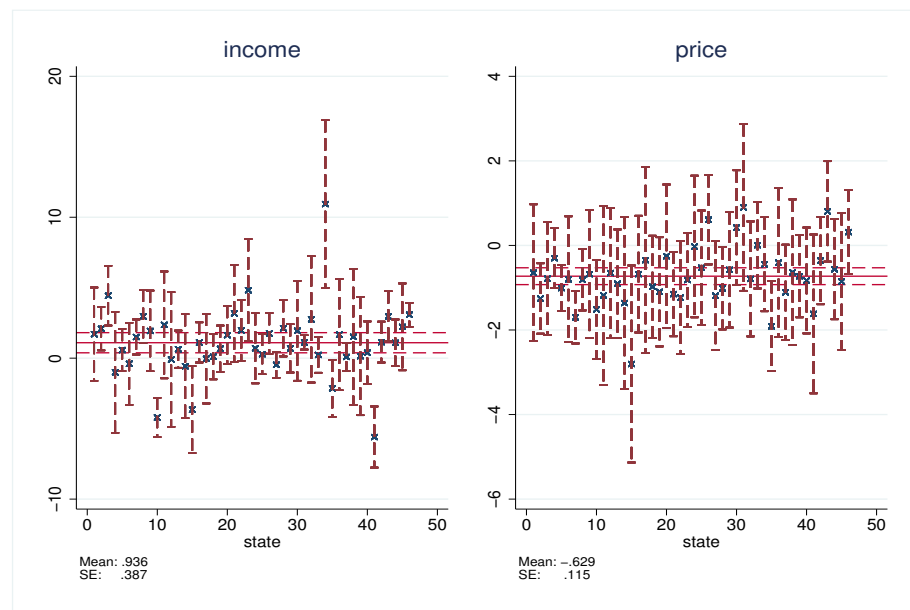
Figure 1. CCE and CCEMG estimations for income (**Left**) and price (**Right**), respectively (CCE estimates of individual coefficients are indicated by a cross, CCEMG estimates by the red line, and the 95% confidence interval by the upper and lower range and dashed red line).
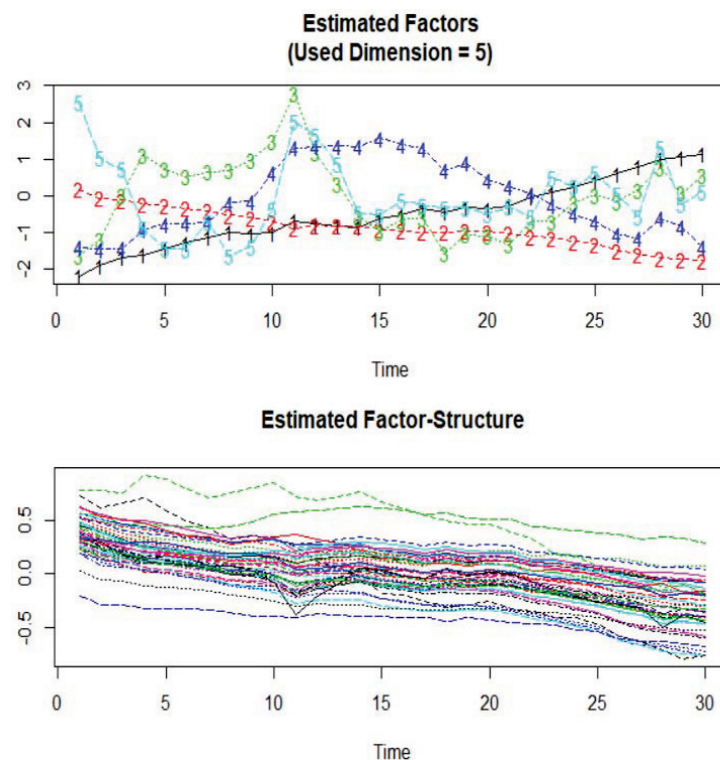

Figure 2. Estimated factors (**Top**) and the factor structure (**Bottom**).

6. Conclusions

In this paper, we re-examined the CCE type estimator for dynamic heterogeneous panel regression models with non-stationary common factors. Asymptotic properties of CCE estimators are established when both N and T are large. It is shown that, under certain conditions, the main results of Pesaran (2006) and Chudik and Pesaran (2015a) hold for a dynamic panel with non-stationary factors. Monte Carlo simulations were conducted to investigate the finite sample properties of the CCE estimation for the panel with non-stationary factors. An empirical application to the U.S. cigarette consumption dataset shows that the real data may have cross-sectional dependence as well as dynamic and non-stationary common factors (at the same time). Based on the findings of this paper, together with the results by Pesaran (2006); Kapetanios et al. (2011), and Chudik and Pesaran (2015a), we can conclude that the CCE method can be widely used to deal with panel models with error cross-sectional dependence, regardless of whether the model is static or dynamic, and whether the unobservable common factors are stationary.

Appendix A. Useful Lemmas and Theoretical Derivations of Theorems

The appendix includes proofs of the theorems and lemmas used in the derivations of the main results in the paper.

Recall that

$$
\mathbf{H} = \begin{pmatrix} 1 & \mathbf{h}'_{p_T+1} & \cdots & \mathbf{h}'_1 \\ 1 & \mathbf{h}'_{p_T+2} & \cdots & \mathbf{h}'_2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{h}'_T & \cdots & \mathbf{h}'_{T-p_T} \end{pmatrix}, \mathbf{G} = (\tau, \tilde{\mathbf{F}}) = \begin{pmatrix} 1 & \mathbf{f}'_{p_T+1} & \cdots & \mathbf{f}'_1 \\ 1 & \mathbf{f}'_{p_T+2} & \cdots & \mathbf{f}'_2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{f}'_T & \cdots & \mathbf{f}'_{T-P_T} \end{pmatrix}, \quad (A1)
$$

$$
\bar{\mathbf{P}} = \begin{pmatrix} 1 & \bar{\mathbf{c}}'_z & \cdots & \bar{\mathbf{c}}'_z \\ \mathbf{0} & \Psi'(L) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Psi'(L) \end{pmatrix}, \bar{\mathbf{V}}^* = (\mathbf{0}, \tilde{\mathbf{V}}) = \begin{pmatrix} 0 & \bar{\mathbf{v}}'_{p_T+1} & \cdots & \bar{\mathbf{v}}'_1 \\ 0 & \bar{\mathbf{v}}'_{p_T+2} & \cdots & \bar{\mathbf{v}}'_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \bar{\mathbf{v}}'_T & \cdots & \bar{\mathbf{v}}'_{T-p_T} \end{pmatrix}, \quad (A2)
$$

where $\mathbf{h}_t = \Psi(L)\mathbf{f}_t + \bar{\mathbf{c}}_z$, $\Psi(L) = \frac{1}{N}\sum_{i=1}^N (\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1}\mathbf{A}_{0i}^{-1}\mathbf{C}_i$ and $\bar{\mathbf{v}}_t = \frac{1}{N}\sum_{i=1}^N (\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1}\mathbf{A}_{0i}^{-1}\mathbf{e}_{it}$. Then, the following equalities hold,

$$
\bar{\mathbf{Q}} = \mathbf{G}\bar{\mathbf{P}} + \bar{\mathbf{V}}^* \text{ and } \mathbf{H} = \mathbf{G}\bar{\mathbf{P}}, \quad (A3)
$$

where matrix $\bar{\mathbf{Q}}$ is given by (10).

Appendix A.1. Useful Lemmas

Now, let us turn to the lemmas, which are needed for the derivation of the results in the main paper.

Lemma A1. *(a) If* $\mathbf{A} \in \mathbb{R}_r^{m \times n}$, $r > 0$, *has a full-rank factorization* $\mathbf{A} = \mathbf{BC}$, *where* $\mathbf{B} \in \mathbb{R}_r^{m \times r}$, $\mathbf{C} \in \mathbb{R}_r^{r \times n}$, *then*

$$\mathbf{A}^+ = \mathbf{C}^+ \mathbf{B}^+.$$

(b) If $\mathbf{A} \in \mathbb{R}_m^{m \times n}$, *i.e.,* $\mathbf{A}$ *is full row rank, then* $\mathbf{A}^+ = \mathbf{A}'(\mathbf{AA}')^{-1}$.

Using the properties of Moore–Penrose inverse, Lemma A.1 can be easily established by the MacDuffe Theorem of Ben-Israe and Greville (2003).

Lemma A2. *If the rank condition (6) in the main text is satisfied, then*

$$\mathbf{M}_h = \mathbf{M}_g \tag{A4}$$

where $\mathbf{M}_h = \mathbf{I}_{T-p_T} - \mathbf{H}(\mathbf{H}'\mathbf{H})^+\mathbf{H}'$, $\mathbf{M}_g = \mathbf{I}_{T-p_T} - \mathbf{G}(\mathbf{G}'\mathbf{G})^+\mathbf{G}'$ *and* $\mathbf{H} = \mathbf{G}\bar{\mathbf{P}}$.

Lemma A3. $\Xi_i = (\mathbf{y}_{i,-1}, \mathbf{X}_i)$ *can be written as*

$$\Xi_i = \mathbf{G}_1 \mathbf{\Pi}_{i1} + \mathbf{\Omega}_i, \tag{A5}$$

or more concisely as

$$\Xi_i = \mathbf{G}_2 \mathbf{\Pi}_{i2} + \mathbf{\Omega}_i, \tag{A6}$$

where $\mathbf{G}_1 = \mathbf{G} = (\tau, \tilde{\mathbf{F}})$ *given by (A1),* $\mathbf{\Pi}_{i1} = (\mathbf{c}_{\xi i}, \mathbf{\Psi}_{\xi i}\mathbf{C}_i, \mathbf{0}, \dots, \mathbf{0})'$, $\mathbf{G}_2 = (\tau, \mathbf{F})$, $\mathbf{\Pi}_{i2} = (\mathbf{c}_{\xi i}, \mathbf{\Psi}_{\xi i}\mathbf{C}_i)'$ *and* $\mathbf{\Omega}_i = \mathbf{e}_i \mathbf{\Psi}'_{\xi i}$, *with* $\mathbf{e}_i = (\varepsilon_i, \mathbf{U}_i)$, $\mathbf{c}_{\xi i} = (\mathbf{S}_y, \mathbf{S}_x)'(\mathbf{I}_{k+1} - \mathbf{A}_i)^{-1}\mathbf{c}_{zi}$, $\mathbf{\Psi}_{\xi i} = \begin{pmatrix} \mathbf{S}'_y(\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} L \\ \mathbf{S}'_x(\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} \end{pmatrix} \mathbf{A}_{0i}^{-1}$ *is* $(k+1) \times (k+1)$ *matrix, and* $\mathbf{S}'_y = \begin{pmatrix} 1 & \mathbf{0} \\ & 1 \times k \end{pmatrix}$, $\mathbf{S}'_x = \begin{pmatrix} \mathbf{0} & \mathbf{I}_k \\ k \times 1 & \end{pmatrix}$.

Lemma A4. *Under Assumption 1–7 as well as restriction* $(N, T, p_T) \xrightarrow{j} \infty$, *such that* $p_T^2/T \to 0$ *and* $T/N \to \varphi \neq 0 < \infty$, *then the following holds,*

$$\left\| \frac{\bar{\mathbf{V}}^{*\prime}\bar{\mathbf{V}}^*}{T} \right\|_\infty = O_p\left(\frac{p_T}{N}\right) \tag{A7}$$

$$\left\| \frac{\varepsilon'_i\bar{\mathbf{V}}^*}{T} \right\|_\infty = O_p(\frac{p_T}{N}) + O_p(\frac{p_T}{\sqrt{NT}}), \quad \left\| \frac{\mathbf{U}'_i\bar{\mathbf{V}}^*}{T} \right\|_\infty = O_p(\frac{p_T}{N}) + O_p(\frac{p_T}{\sqrt{NT}}) \tag{A8}$$

$$\left\| \frac{\mathbf{G}'\bar{\mathbf{V}}^*}{T} \right\|_\infty = O_p(\frac{p_T}{\sqrt{N}}), \quad \left\| \frac{\mathbf{H}'\bar{\mathbf{V}}^*}{T} \right\|_\infty = O_p(\frac{p_T}{\sqrt{N}}), \quad \left\| \frac{\Xi'_i\bar{\mathbf{V}}^*}{T} \right\|_\infty = O_p(\frac{p_T}{\sqrt{N}}) \tag{A9}$$

$$\left\| \frac{\mathbf{G}'\varepsilon_i}{T} \right\|_\infty = O_p(1), \quad \left\| \frac{\mathbf{G}'\mathbf{U}_i}{T} \right\|_\infty = O_p(1), \quad \left\| \frac{\bar{\mathbf{Q}}'\varepsilon_i}{T} \right\|_\infty = O_p(1) \tag{A10}$$

$$\left\| \frac{\mathbf{G}'\mathbf{G}}{T^2} \right\|_\infty = O_p(p_T), \quad \left\| \frac{\mathbf{H}'\mathbf{H}}{T^2} \right\|_\infty = O_p(p_T), \quad \left\| \frac{\bar{\mathbf{Q}}'\bar{\mathbf{Q}}}{T^2} \right\|_\infty = O_p(p_T) \tag{A11}$$

$$\left\| \frac{\bar{\mathbf{Q}}'\mathbf{G}}{T^2} \right\|_\infty = O_p(p_T), \quad \left\| \frac{\mathbf{H}'\Xi_i}{T^2} \right\|_\infty = O_p(p_T), \quad \left\| \frac{\bar{\mathbf{Q}}'\Xi_i}{T^2} \right\|_\infty = O_p(p_T) \tag{A12}$$

Lemma A5. *Under Assumption 1–7 and* $(N, T, p_T) \xrightarrow{j} \infty$, *such that* $p_T^3/T \to \lambda$, $0 < \lambda < \infty$. *Then,*

$$\frac{\Xi'_i\mathbf{M}_g\Xi_i}{T} \xrightarrow{p} \mathbf{\Sigma}_{\Omega i}, \tag{A13}$$

where $\boldsymbol{\Sigma}_{\Omega i} = \boldsymbol{\Psi}_{\xi i} \begin{pmatrix} \sigma_i^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{u}_i} \end{pmatrix} \boldsymbol{\Psi}_{\xi i}'$ *is a positive definite matrix. Additionally, if the rank condition (6) is satisfied, then*

$$\frac{\boldsymbol{\Xi}_i' \mathbf{M}_q \mathbf{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \text{ uniformly over } i. \tag{A14}$$

Lemma A6. *If the rank condition (6) is satisfied, and* $(N, T, p_T) \overset{j}{\to} \infty$, *such that* $N/T \to \varphi$, $0 < \varphi < \infty$ *and* $p_T^2/T \to 0$, *it follows that*

$$\frac{\boldsymbol{\Xi}_i' \mathbf{M}_q \boldsymbol{\Xi}_i}{T} - \frac{\boldsymbol{\Xi}_i' \mathbf{M}_h \boldsymbol{\Xi}_i}{T} = O_p\left(\frac{p_T}{\sqrt{N}}\right), \text{ uniformly over } i, \tag{A15}$$

$$\frac{\boldsymbol{\Xi}_i' \mathbf{M}_q \mathbf{F}}{T} - \frac{\boldsymbol{\Xi}_i' \mathbf{M}_h \mathbf{F}}{T} = O_p\left(\frac{p_T}{\sqrt{N}}\right), \text{ uniformly over } i, \tag{A16}$$

$$\frac{\boldsymbol{\Xi}_i' \mathbf{M}_q \boldsymbol{\varepsilon}_i}{T} - \frac{\boldsymbol{\Xi}_i' \mathbf{M}_h \boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{p_T}{\sqrt{NT}}\right) + O_p\left(\frac{p_T}{N}\right), \text{ uniformly over } i. \tag{A17}$$

Appendix A.2. Theoretical Derivation of the Asymptotics of the CCE Estimators

Proof of Theorem 1. Since

$$\hat{\pi}_i - \pi_i = \left(\frac{\boldsymbol{\Xi}_i' \mathbf{M}_q \boldsymbol{\Xi}_i}{T}\right)^{-1} \frac{\boldsymbol{\Xi}_i' \mathbf{M}_q \mathbf{F} \boldsymbol{\gamma}_i}{T} + \left(\frac{\boldsymbol{\Xi}_i' \mathbf{M}_q \boldsymbol{\Xi}_i}{T}\right)^{-1} \frac{\boldsymbol{\Xi}_i' \mathbf{M}_g \boldsymbol{\varepsilon}_i}{T},$$

and using the results of Lemmas A5 and A6, we have

$$\hat{\pi}_i - \pi_i = \left(\frac{\boldsymbol{\Xi}_i' \mathbf{M}_q \boldsymbol{\Xi}_i}{T}\right)^{-1} \frac{\boldsymbol{\Xi}_i' \mathbf{M}_g \boldsymbol{\varepsilon}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Noting that under our assumptions, $T^{-1} \boldsymbol{\Xi}_i' \mathbf{M}_q \boldsymbol{\Xi}_i$ tends to a fixed positive definite matrix. Since $\boldsymbol{\Xi}_i = \mathbf{G} \boldsymbol{\Pi}_i + \boldsymbol{\Omega}_i$, then we have

$$\frac{\boldsymbol{\Xi}_i' \mathbf{M}_g \boldsymbol{\varepsilon}_i}{T} = -(\hat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i)' \frac{\mathbf{G}' \boldsymbol{\varepsilon}_i}{T} + \frac{\boldsymbol{\Omega}_i' \boldsymbol{\varepsilon}_i}{T},$$

where $\hat{\boldsymbol{\Pi}}_i$ is the OLS estimator of $\boldsymbol{\Xi}_i$ on $\mathbf{G}$. Since $(\hat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i) = O_p(T^{-1})$, the first part of (A10) in Lemma A4 implies that the first term is $O_p(T^{-1})$. Next, we establish $T^{-1} \boldsymbol{\Omega}_i' \boldsymbol{\varepsilon}_i \overset{p}{\to} 0$. Note that $\boldsymbol{\Omega}_i = \mathbf{e}_i \boldsymbol{\Psi}_{\xi i}'$ with $\mathbf{e}_i = (\boldsymbol{\varepsilon}_i, \mathbf{U}_i)$, i.e., $\boldsymbol{\Omega}_i$ contains the lags of ε_{it}, as well as the contemporary and lags of u_{it}, by Assumption 1, ε_{it} is the series uncorrelated and independent of $\mathbf{u}_{it}$, then we have $T^{-1} \boldsymbol{\Omega}_i' \boldsymbol{\varepsilon}_i \overset{p}{\to} 0$; consequently,

$$\frac{\boldsymbol{\Xi}_i' \mathbf{M}_g \boldsymbol{\varepsilon}_i}{T} \overset{p}{\to} \mathbf{0}, \text{ uniformly over } i. \tag{A18}$$

as $(N, T, p_T) \overset{j}{\to} \infty$ and $p_T^2/T \to 0$. Then it is followed by the consistency of $\hat{\pi}_i$. □

Proof of Theorem 2. Using the consistency of $\hat{\pi}_i$, and the definition of the mean group estimator $\hat{\pi}_{MG}$, we obtain

$$\hat{\pi}_{MG} - \frac{1}{N} \sum_{i=1}^{N} \pi_i \overset{p}{\to} \mathbf{0}, \tag{A19}$$

By the assumption of the random coefficient model, $\pi_i = \pi + v_{\pi i}$, it follows that

$$\frac{1}{N} \sum_{i=1}^{N} \pi_i = \pi + \frac{1}{N} \sum_{i=1}^{N} v_{\pi i}. \tag{A20}$$

Combining (A19) and (A20), we have

$$\hat{\pi}_{MG} \xrightarrow{p} \pi + \frac{1}{N} \sum_{i=1}^{N} v_{\pi i}, \tag{A21}$$

so we only need to show that $\frac{1}{N} \sum_{i=1}^{N} v_{\pi i} \xrightarrow{p} \mathbf{0}$. Since $v_{\pi i} \sim IID(\mathbf{0}, \mathbf{\Sigma}_\pi)$ by Assumption 3, we have $E\left(\frac{1}{N} \sum_{i=1}^{N} v_{\pi i}\right) = \mathbf{0}$ and $Var\left(\frac{1}{N} \sum_{i=1}^{N} v_{\pi i}\right) = \frac{1}{N^2} \sum_{i=1}^{N} Var(v_{\pi i}) = O(\frac{1}{N})$, which implies

$$\frac{1}{N} \sum_{i=1}^{N} v_{\pi i} \xrightarrow{p} \mathbf{0}. \tag{A22}$$

Using (A21) and (A22), we obtain $\hat{\pi}_{MG} \xrightarrow{p} \pi$ as desired.

Next, we establish the asymptotic distribution of $\hat{\pi}_{MG}$. We have

$$\sqrt{N}(\hat{\pi}_{MG} - \pi) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_{\pi i} + \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{\Xi}_i}{T}\right)^{-1} \frac{\sqrt{N} \mathbf{\Xi}_i' \mathbf{M}_q \mathbf{F}}{T} \gamma_i$$
$$+ \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{\Xi}_i}{T}\right)^{-1} \frac{\sqrt{N} \mathbf{\Xi}_i' \mathbf{M}_q \varepsilon_i}{T}. \tag{A23}$$

Using the result (A14) in Lemma A5, when the rank condition is satisfied, we have

$$\frac{\sqrt{N} \mathbf{\Xi}_i' \mathbf{M}_q \mathbf{F}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

which, together with the assumption of γ_i to be bounded, and the results of Lemma A2, A5 and Lemma A6, we obtain

$$\left(\frac{\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{\Xi}_i}{T}\right)^{-1} \frac{\sqrt{N} \mathbf{\Xi}_i' \mathbf{M}_q \mathbf{F}}{T} \gamma_i = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

uniformly over i, it follows that

$$\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{\Xi}_i}{T}\right)^{-1} \frac{\sqrt{N} \mathbf{\Xi}_i' \mathbf{M}_q \mathbf{F}}{T} \gamma_i = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A24}$$

For the third term, by Lemmas A2 and A6, we have

$$\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{\Xi}_i}{T}\right)^{-1} \frac{\sqrt{N} \mathbf{\Xi}_i' \mathbf{M}_q \varepsilon_i}{T}$$
$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{\mathbf{\Xi}_i' \mathbf{M}_g \mathbf{\Xi}_i}{T}\right)^{-1} \frac{\mathbf{\Xi}_i' \mathbf{M}_g \varepsilon_i}{T} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

Moreover, the result (A13) of Lemma A5 implies $T^{-1} \mathbf{\Xi}_i' \mathbf{M}_g \mathbf{\Xi}_i = O_p(1)$, and $T^{-1} \mathbf{\Xi}_i' \mathbf{M}_g \varepsilon_i = O_p(T^{-1/2})$ by (A18); hence, we have

$$\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{\Xi}_i}{T}\right)^{-1} \frac{\sqrt{N} \mathbf{\Xi}_i' \mathbf{M}_q \varepsilon_i}{T} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A25}$$

Using (A24) and (A25) in (A23), we can obtain

$$\sqrt{N}(\hat{\pi}_{MG} - \pi) \stackrel{d}{\sim} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_{\pi i}.$$

By the random coefficient assumption, it now follows that

$$\sqrt{N}(\hat{\pi}_{MG} - \pi) \xrightarrow{d} N(0, \Sigma_{MG}),$$

and Σ_{MG} can be consistently estimated nonparametrically by

$$\hat{\Sigma}_{MG} = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{\pi}_i - \hat{\pi}_{MG})(\hat{\pi}_i - \hat{\pi}_{MG})'.$$

$\square$

Appendix A.3. Proofs of Lemmas

Notation: All vectors are column vectors represented by bold lower case letters, and matrices are represented by bold capital letters. Let $\|\mathbf{A}\| = \sqrt{tr(\mathbf{A}\mathbf{A}')}$ denote the Frobenius norm. $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq n} \Sigma_{i=1}^{n} |a_{ij}|$ and $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq n} \Sigma_{j=1}^{n} |a_{ij}|$ denote the maximum absolute column and row sum matrix norms, respectively. $\lambda_{\min}(\mathbf{A})$ denotes the minimum eigenvalue of $\mathbf{A}$, and $\lambda_{\max}(\mathbf{A})$ denotes the maximum eigenvalue of $\mathbf{A}$. $\mathbf{A}^+$ denotes the Moore–Penrose inverse of $\mathbf{A}$, and $rk(\mathbf{A})$ denotes the rank of $\mathbf{A}$. We also let K denote a generic finite constant, which does not depend on N or T, and whose value may vary case by case.

Proof of Lemma A2. Since $\mathbf{H} = \mathbf{G}\bar{\mathbf{P}}$, where

$$\bar{\mathbf{P}} = \begin{pmatrix} 1 & \bar{\mathbf{c}}_z' & \bar{\mathbf{c}}_z' & \cdots & \bar{\mathbf{c}}_z' \\ 0 & \mathbf{\Psi}'(L) & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{\Psi}'(L) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{\Psi}'(L) \end{pmatrix},$$

with $\mathbf{\Psi}(L) = \mathbf{\Lambda}(L)\mathbf{C} + O_p(N^{-1/2})$, $\mathbf{\Lambda}(L)$ is invertible. If the rank condition (6) is satisfied, i.e., $\mathbf{C}$ is a full column rank matrix, then $\mathbf{\Lambda}(L)\mathbf{C}$ has a full column rank; hence, $\mathbf{\Psi}'(L)$ has full row rank asymptotically, which implies $\lim_{N \to \infty} \bar{\mathbf{P}}$ is a full row rank matrix. Moreover, noting that when the rank condition holds, matrix $\mathbf{G}'\mathbf{G}$ is full rank, so we have

$$\begin{aligned} \mathbf{M}_h &= \mathbf{I} - \mathbf{H}(\mathbf{H}'\mathbf{H})^+\mathbf{H}' \\ &= \mathbf{I} - \mathbf{G}\bar{\mathbf{P}}(\bar{\mathbf{P}}'\mathbf{G}'\mathbf{G}\bar{\mathbf{P}})^+\bar{\mathbf{P}}'\mathbf{G}' \\ &= \mathbf{I} - \mathbf{G}\bar{\mathbf{P}}\bar{\mathbf{P}}^+(\mathbf{G}'\mathbf{G})^+\bar{\mathbf{P}}'^+\bar{\mathbf{P}}'\mathbf{G}' \\ &= \mathbf{I} - \mathbf{G}\bar{\mathbf{P}}\bar{\mathbf{P}}'(\bar{\mathbf{P}}\bar{\mathbf{P}}')^{-1}(\mathbf{G}'\mathbf{G})^+(\bar{\mathbf{P}}\bar{\mathbf{P}}')^{-1}\bar{\mathbf{P}}\bar{\mathbf{P}}'\mathbf{G}' \\ &= \mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{G})^+\mathbf{G}' = \mathbf{M}_g. \end{aligned}$$

where the third equality follows from Lemma A1(a) since $\bar{\mathbf{P}}$ has the full row rank asymptotically, and the fourth equality is based on the result of Lemma A1(b). $\square$

Proof of Lemma A3. Denote $\Xi_i = (\xi_{i,p_T+1}, \xi_{i,p_T+2}, \ldots, \xi_{iT})'$, where $\xi_{it} = (y_{i,t-1}, \mathbf{x}_{i,t}')'$. Note that $\mathbf{z}_{it} = (y_{i,t}, \mathbf{x}_{i,t}')'$, so we can write $y_{i,t-1} = \mathbf{S}_y'\mathbf{z}_{i,t-1}$ and $\mathbf{x}_{i,t} = \mathbf{S}_x'\mathbf{z}_{it}$, where $\mathbf{S}_y' = \begin{pmatrix} 1 & \underset{1 \times k}{\mathbf{0}} \end{pmatrix}$, $\mathbf{S}_x' = \begin{pmatrix} \underset{k \times 1}{\mathbf{0}} & \mathbf{I}_k \end{pmatrix}$. Hence, we have

$$\xi_{it} = \begin{pmatrix} y_{i,t-1} \\ \mathbf{x}_{i,t} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_y'\mathbf{z}_{i,t-1} \\ \mathbf{S}_x'\mathbf{z}_{it} \end{pmatrix}. \tag{A26}$$

We also note that

$$
\begin{aligned}
\mathbf{z}_{it} &= \sum_{l=0}^{\infty} \mathbf{A}_i^l \left(\mathbf{c}_{zi} + \mathbf{A}_{0i}^{-1} C_i \mathbf{f}_{t-l} + \mathbf{e}_{zi,t-l} \right) \\
&= (\mathbf{I}_{k+1} - \mathbf{A}_i)^{-1} \mathbf{c}_{zi} + (\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} \mathbf{A}_{0i}^{-1} (\mathbf{C}_i \mathbf{f}_t + \mathbf{e}_{it}),
\end{aligned}
\tag{A27}
$$

using (A27) into (A26), we have

$$
\begin{aligned}
\xi_{it} &= \begin{pmatrix} \mathbf{S}_y'(\mathbf{I}_{k+1} - \mathbf{A}_i)^{-1} \\ \mathbf{S}_x'(\mathbf{I}_{k+1} - \mathbf{A}_i)^{-1} \end{pmatrix} \mathbf{c}_{zi} + \begin{pmatrix} \mathbf{S}_y'(\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} L \\ \mathbf{S}_x'(\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} \end{pmatrix} \mathbf{A}_{0i}^{-1}(\mathbf{C}_i \mathbf{f}_t + \mathbf{e}_{it}) \\
&= (\mathbf{S}_y, \mathbf{S}_x)'(\mathbf{I}_{k+1} - \mathbf{A}_i)^{-1} \mathbf{c}_{zi} + \begin{pmatrix} \mathbf{S}_y'(\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} L \\ \mathbf{S}_x'(\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} \end{pmatrix} \mathbf{A}_{0i}^{-1}(\mathbf{C}_i \mathbf{f}_t + \mathbf{e}_{it}) \\
&\equiv \mathbf{c}_{\xi i} + \mathbf{\Psi}_{\xi i}(\mathbf{C}_i \mathbf{f}_t + \mathbf{e}_{it}).
\end{aligned}
$$

Consequently, we have

$$
\begin{aligned}
\Xi_i &= \begin{pmatrix} \xi_{i,p_T+1}' \\ \xi_{i,p_T+2}' \\ \vdots \\ \xi_{iT}' \end{pmatrix} = \begin{pmatrix} \mathbf{c}_{\xi i}' + \mathbf{f}_{p_T+1}' \mathbf{C}_i' \mathbf{\Psi}_{\xi i}' \\ \mathbf{c}_{\xi i}' + \mathbf{f}_{p_T+2}' \mathbf{C}_i' \mathbf{\Psi}_{\xi i}' \\ \vdots \\ \mathbf{c}_{\xi i}' + \mathbf{f}_T' \mathbf{C}_i' \mathbf{\Psi}_{\xi i}' \end{pmatrix} + \begin{pmatrix} \mathbf{e}_{i,p_T+1}' \\ \mathbf{e}_{i,p_T+2}' \\ \vdots \\ \mathbf{e}_{iT}' \end{pmatrix} \mathbf{\Psi}_{\xi i}' \\
&= \begin{pmatrix} 1 & \mathbf{f}_{p_T+1}' & \mathbf{f}_{p_T}' & \cdots & \mathbf{f}_1' \\ 1 & \mathbf{f}_{p_T+2}' & \mathbf{f}_{p_T+1}' & \cdots & \mathbf{f}_2' \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{f}_T' & \mathbf{f}_{T-1}' & \cdots & \mathbf{f}_{T-P_T}' \end{pmatrix} \begin{pmatrix} \mathbf{c}_{\xi i}' \\ \mathbf{C}_i' \mathbf{\Psi}_{\xi i}' \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathbf{e}_i \mathbf{\Psi}_{\xi i}' \\
&= \mathbf{G}_1 \mathbf{\Pi}_{i1} + \mathbf{\Omega}_i,
\end{aligned}
$$

or more concisely as

$$
\Xi_i = \begin{pmatrix} 1 & \mathbf{f}_{p_T+1}' \\ 1 & \mathbf{f}_{p_T+2}' \\ \vdots & \vdots \\ 1 & \mathbf{f}_T' \end{pmatrix} \begin{pmatrix} \mathbf{c}_{\xi i}' \\ \mathbf{C}_i' \mathbf{\Psi}_{\xi i}' \end{pmatrix} + \mathbf{e}_i \mathbf{\Psi}_{\xi i}' = \mathbf{G}_2 \mathbf{\Pi}_{i2} + \mathbf{\Omega}_i.
$$

□

Proof of Lemma A4. We consider (A7) firstly. Note that

$$
\bar{\mathbf{V}}^* = (\mathbf{0}, \tilde{\mathbf{V}}) = \begin{bmatrix} 0 & \bar{\mathbf{v}}_{p_T+1}' & \bar{\mathbf{v}}_{p_T}' & \cdots & \bar{\mathbf{v}}_1' \\ 0 & \bar{\mathbf{v}}_{p_T+2}' & \bar{\mathbf{v}}_{p_T+1}' & \cdots & \bar{\mathbf{v}}_2' \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \bar{\mathbf{v}}_T' & \bar{\mathbf{v}}_{T-1}' & \cdots & \bar{\mathbf{v}}_{T-p_T}' \end{bmatrix},
$$

where

$$
\bar{\mathbf{v}}_t = \frac{1}{N} \sum_{i=1}^{N} (I_{k+1} - A_i L)^{-1} A_{0i}^{-1} \mathbf{e}_{it} = \frac{1}{N} \sum_{i=1}^{N} A_i^l \mathbf{e}_{zi,t-l}.
$$

So we only need to consider $T^{-1}(\tilde{\mathbf{V}}'\tilde{\mathbf{V}})$, which is a $(k+1)(p_T+1) \times (k+1)(p_T+1)$ matrix. Since the elements of $\mathbf{e}_{zit}$ are weakly cross-sectionally dependent, together with the random coefficient assumptions, we have $E\|\bar{\mathbf{v}}_t\| = O(N^{-\frac{1}{2}})$ and $E\|\bar{\mathbf{v}}_t\|^2 = O(N^{-1})$. Con-

sider the $(s,r)th$ block element of $T^{-1}(\tilde{\mathbf{V}}'\tilde{\mathbf{V}})$, which can be written as $T^{-1}(\sum_{t=p_T+1}^{T} \bar{\mathbf{v}}_{t-s}\bar{\mathbf{v}}'_{t-r})$, for $s,r \in \{0,1,\dots,p_T\}$. where the cross-product terms with finite means and variances. Hence,

$$E\left\| \frac{1}{T} \sum_{t=p_T+1}^{T} \bar{\mathbf{v}}_{t-s}\bar{\mathbf{v}}'_{t-r} \right\| \leq \frac{1}{T} \sum_{t=p_T+1}^{T} E\|\bar{\mathbf{v}}_t\|^2 = O(\frac{1}{N}), \tag{A28}$$

then we have

$$E\left\| \frac{\tilde{\mathbf{V}}'\tilde{\mathbf{V}}}{T} \right\|_{\infty} \leq O(\frac{p_T}{N}),$$

which establishes (A7).

Now, we establish (A8), as before, we consider $T^{-1}\varepsilon_i'\tilde{\mathbf{V}}$ here, and note that the lth column block of $T^{-1}\varepsilon_i'\tilde{\mathbf{V}}$ is $T^{-1}\left(\sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\mathbf{v}}'_{t-l}\right)$, for $l = 0,1,\dots,p_T$, which can be partitioned as

$$\frac{1}{T} \sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\mathbf{v}}'_{t-l} = \left(\begin{array}{cc} \frac{1}{T}\sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\varepsilon}_{t-l}, & \frac{1}{T}\sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\mathbf{u}}'_{t-l} \end{array} \right). \tag{A29}$$

We consider the first term and note that

$$\begin{aligned} \bar{\varepsilon}_t &= \frac{1}{N} \sum_{j=1}^{N} (\mathbf{I}_{k+1} - \mathbf{A}_j L)^{-1} \mathbf{A}_{0j}^{-1} \varepsilon_{jt} = \frac{1}{N} \sum_{j=1}^{N} \sum_{l=0}^{\infty} \mathbf{A}_j^l \mathbf{A}_{0j}^{-1} \varepsilon_{j,t-l} \\ &= \frac{1}{N} \sum_{j=1}^{N} \left(\mathbf{A}_{0j}^{-1}\varepsilon_{jt} + \mathbf{A}_j \mathbf{A}_{0j}^{-1}\varepsilon_{j,t-1} + \mathbf{A}_j^2 \mathbf{A}_{0j}^{-1}\varepsilon_{j,t-2} + \cdots \right), \end{aligned}$$

which implies that

$$\frac{1}{T} \sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\varepsilon}_{t-l} = \frac{1}{NT} \sum_{t=p_T+1}^{T} \sum_{j=1}^{N} \varepsilon_{it} \left(\mathbf{A}_{0j}^{-1}\varepsilon_{j,t-l} + \mathbf{A}_j \mathbf{A}_{0j}^{-1}\varepsilon_{j,t-l-1} + \cdots \right).$$

Under the assumption of the individual–specific error, we have $cov(\varepsilon_{it}, \varepsilon_{js}) = 0, i \neq j$; hence,

$$\frac{1}{T} \sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\varepsilon}_{t-l} = \frac{1}{NT} \sum_{t=p_T+1}^{T} \varepsilon_{it} \left(\mathbf{A}_{0i}^{-1}\varepsilon_{i,t-l} + \mathbf{A}_i \mathbf{A}_{0i}^{-1}\varepsilon_{i,t-l-1} + \cdots \right),$$

when $l = 0$, $\frac{1}{T}\sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\varepsilon}_{t-l} = \frac{\mathbf{A}_{0i}^{-1}}{NT} \sum_{t=p_T+1}^{T} \varepsilon_{it}^2$, since by Assumption 1, $E(\varepsilon_{it}^2) = O(1)$, and $\left\| \mathbf{A}_{0i}^{-1} \right\| \leq K$, then it easily follows that $\frac{1}{T}\sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\varepsilon}_{t-l} = O_p(\frac{1}{N})$. When $l = 1,2,\cdots,p_T$, we have the result $\frac{1}{T}\sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\varepsilon}_{t-l} = O_p(\frac{1}{N})$ since ε_{it} is a serial uncorrelated covariance stationary process under Assumption 2. Combining these results yields

$$\frac{1}{T} \sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\varepsilon}_{t-l} = O_p(\frac{1}{N}), \text{ uniformly over } i. \tag{A30}$$

Now, we consider the second term $\frac{1}{T}\sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\mathbf{u}}'_{t-l}$, noting that ε_{it} and $\mathbf{u}_{it}$ are independently distributed stationary processes with zero means, it follows that

$$\sup_i Var\left(\frac{1}{T} \sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\mathbf{u}}'_{t-l} \right) = O(\frac{1}{N}) \left[\frac{1}{T^2} \sum_{t=p_T+1}^{T} \sum_{t'=p_T+1}^{T} E(\varepsilon_{it}\varepsilon_{it'}) = O(\frac{1}{NT}) \right],$$

which follows that

$$\frac{1}{T} \sum_{t=p_T+1}^{T} \varepsilon_{it}\bar{\mathbf{u}}'_{t-l} = O_p(\frac{1}{\sqrt{NT}}), \text{ uniformly over } i. \tag{A31}$$

Using (A30) and (A31) in (A29), we have

$$\frac{1}{T}\sum_{t=p_T+1}^{T}\varepsilon_{it}\bar{\mathbf{v}}'_{t-l} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}}), \text{uniformly over } i.$$

Consequently, we have

$$\left\|\frac{\varepsilon'_i\tilde{\mathbf{V}}}{T}\right\|_{\infty} = O_p(\frac{p_T}{N}) + O_p(\frac{p_T}{\sqrt{NT}}), \text{uniformly over } i,$$

hence, the first part of (A8) is established. Similarly, the result for $T^{-1}\mathbf{U}'_i\bar{\mathbf{V}}^*$ of the second part of (A8) is established.

For the first part of (A9), since $\mathbf{G} = (\tau, \tilde{\mathbf{F}})$ and $\bar{\mathbf{V}}^* = (\mathbf{0}, \tilde{\mathbf{V}})$, we consider

$$
\frac{\tilde{\mathbf{F}}'\tilde{\mathbf{V}}}{T} = \frac{1}{T}
\begin{pmatrix}
\mathbf{f}_{p_T+1} & \mathbf{f}_{p_T+2} & \cdots & \mathbf{f}_T \\
\mathbf{f}_{p_T} & \mathbf{f}_{p_T+1} & \cdots & \mathbf{f}_{T-1} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{f}_1 & \mathbf{f}_2 & \cdots & \mathbf{f}_{T-p_T}
\end{pmatrix}
\begin{pmatrix}
\bar{\mathbf{v}}'_{p_T+1} & \bar{\mathbf{v}}'_{p_T} & \cdots & \bar{\mathbf{v}}'_1 \\
\bar{\mathbf{v}}'_{p_T+2} & \bar{\mathbf{v}}'_{p_T+1} & \cdots & \bar{\mathbf{v}}'_2 \\
\vdots & \vdots & \ddots & \vdots \\
\bar{\mathbf{v}}'_T & \bar{\mathbf{v}}'_{T-1} & \cdots & \bar{\mathbf{v}}'_{T-p_T}
\end{pmatrix}
$$

$$
= \frac{1}{T}
\begin{pmatrix}
\sum_{t=p_T+1}^{T}\mathbf{f}_t\bar{\mathbf{v}}'_t & \sum_{t=p_T+1}^{T}\mathbf{f}_t\bar{\mathbf{v}}'_{t-1} & \cdots & \sum_{t=p_T+1}^{T}\mathbf{f}_t\bar{\mathbf{v}}'_{t-p_T} \\
\sum_{t=p_T+1}^{T}\mathbf{f}_{t-1}\bar{\mathbf{v}}'_t & \sum_{t=p_T+1}^{T}\mathbf{f}_{t-1}\bar{\mathbf{v}}'_{t-1} & \cdots & \sum_{t=p_T+1}^{T}\mathbf{f}_{t-1}\bar{\mathbf{v}}'_{t-p_T} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{t=p_T+1}^{T}\mathbf{f}_{t-p_T}\bar{\mathbf{v}}'_t & \sum_{t=p_T+1}^{T}\mathbf{f}_{t-p_T}\bar{\mathbf{v}}'_{t-1} & \cdots & \sum_{t=p_T+1}^{T}\mathbf{f}_{t-p_T}\bar{\mathbf{v}}'_{t-p_T}
\end{pmatrix},
$$

which is a $m(p_T+1) \times (k+1)(p_T+1)$ matrix. Without loss of generality, we consider the first block element, $T^{-1}\sum_{t=p_T+1}^{T}\mathbf{f}_t\bar{\mathbf{v}}'_t$, and note that the lth row of that can be written as $T^{-1}\sum_{t=p_T+1}^{T}f_{tl}\bar{\mathbf{v}}'_t, l = 1, 2, \ldots, m$. According to the assumption of f_{tl} and $\bar{\mathbf{v}}_t$ (independently distributed processes), it easily follows that

$$\left\|E\frac{1}{T}\sum_{t=p_T+1}^{T}f_{tl}\bar{\mathbf{v}}'_t\right\| = \mathbf{0},$$

and

$$Var\left(\frac{1}{T}\sum_{t=p_T+1}^{T}f_{tl}\bar{\mathbf{v}}'_t\right) = O(\frac{1}{N})\left[\frac{\sum_t\sum_{t'}E(f_{tl}f_{t'l})}{T^2}\right] = O(\frac{1}{N}).$$

by the standard unit root asymptotic analysis result $T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}E(f_{tl}f_{t'l}) = O(1)$, which establishes that $T^{-1}\sum_{t=p_T+1}^{T}f_{tl}\bar{\mathbf{v}}'_t$ converges to its limit at the desired rate of $O_p(\frac{1}{\sqrt{N}})$. Consequently, we have

$$\left\|\frac{\tilde{\mathbf{F}}'\tilde{\mathbf{V}}}{T}\right\|_{\infty} = O_p(\frac{p_T}{\sqrt{N}}),$$

then the first part of (A9) is proven.

To establish the second part of (A9), recalling that $\mathbf{H} = \mathbf{G}\bar{\mathbf{P}}$, we have $\left\|\frac{\mathbf{H}'\bar{\mathbf{V}}^*}{T}\right\|_{\infty} \leq \|\bar{\mathbf{P}}'\|_{\infty}\left\|\frac{\mathbf{G}'\bar{\mathbf{V}}^*}{T}\right\|_{\infty} = O_p(\frac{p_T}{\sqrt{N}})$, since the norm of $\bar{\mathbf{P}}$ is assumed to be bounded.

To establish the third part of (A9), noting that $\Xi_i = G\Pi_i + \Omega_i$, and using triangle inequality and the submultiplicative property of matrix norm $\|\cdot\|_\infty$, we have

$$
\begin{aligned}
\left\| \frac{\Xi_i' \bar{V}^*}{T} \right\|_\infty &= \left\| \Pi_i' \frac{G' \bar{V}^*}{T} + \frac{\Omega_i' \bar{V}^*}{T} \right\|_\infty \\
&\leq \left\| \Pi_i' \frac{G' \bar{V}^*}{T} \right\|_\infty + \left\| \Psi_{\xi i}(L) \frac{(\varepsilon_i, U_i)' \bar{V}^*}{T} \right\|_\infty \\
&\leq \left\| \Pi_i' \right\|_\infty \left\| \frac{G' \bar{V}^*}{T} \right\|_\infty + \left\| \Psi_{\xi i}(L) \right\|_\infty \left\| \frac{(\varepsilon_i, U_i)' \bar{V}^*}{T} \right\|_\infty \\
&= O_p \left(\frac{p_T}{\sqrt{N}} \right),
\end{aligned}
$$

by (A8), the first and second parts of (A9), as well as the norm of Π_i and $\Psi_{\xi i}(L)$ are assumed to be bounded in probability uniformly over i.

To establish the first part of (A10), recalling $G = (\tau, \tilde{F})$, consider the $m(p_T + 1) \times 1$ vector $T^{-1} \tilde{F}' \varepsilon_i$, the element of $T^{-1} \tilde{F}' \varepsilon_i$ can be written as $T^{-1} \sum_{t=p_T+1}^{T} f_{t-l,s} \varepsilon_{it}$, $l = 0, 1, \ldots, p_T$, $s = 1, 2, \ldots, m$. Since by the assumption of $f_{t-l,s}$ and ε_{it} (independently distributed processes), it easily follows that

$$
\left\| E \frac{1}{T} \sum_{t=p_T+1}^{T} f_{t-l,s} \varepsilon_{it} \right\| = 0,
$$

and

$$
\sup_i Var \left(\frac{1}{T} \sum_{t=p_T+1}^{T} f_{t-l,s} \varepsilon_{it} \right) = O(1) \left[\frac{\sum_t \sum_{t'} E(f_{t-l,s} f_{t'-l,s})}{T^2} \right] = O(1).
$$

by the standard unit root asymptotic analysis result $T^{-2} \sum_t \sum_{t'} E(f_{t-l,s} f_{t'-l,s}) = O(1)$, which establishes that $\frac{1}{T} \sum_{t=p_T+1}^{T} f_{t-l,s} \varepsilon_{it}$ converges to its limit at the desired rate of $O_p(1)$. It follows that $\left\| T^{-1} \tilde{F}' \varepsilon_i \right\|_\infty = O_p(1)$; hence, the first part of (A10) is established. Moreover, the second part of (A10) can be proven similarly.

Recalling that $\bar{Q} = G\bar{P} + \bar{V}^*$, the third part of (A10) is established because

$$
\begin{aligned}
\left\| \frac{\bar{Q}' \varepsilon_i}{T} \right\|_\infty &= \left\| \bar{P}' \frac{G' \varepsilon_i}{T} + \frac{\bar{V}^{*\prime} \varepsilon_i}{T} \right\|_\infty \\
&\leq \left\| \bar{P}' \right\|_\infty \left\| \frac{G' \varepsilon_i}{T} \right\|_\infty + \left\| \frac{\bar{V}^{*\prime} \varepsilon_i}{T} \right\|_\infty = O_p(1),
\end{aligned}
$$

by (A8) and the first part of (A10).

For the first part of (A11), note that $G = (\tau, \tilde{F})$, we only need to consider the $T^{-2} \tilde{F}' \tilde{F}$, a $m(p_T + 1) \times m(p_T + 1)$ matrix. We consider the $(s, r)th$ block element of $T^{-2} \tilde{F}' \tilde{F}$, which can be written as $T^{-2} (\sum_{t=p_T+1}^{T} f_{t-s} f_{t-r}')$, for $s, r \in \{0, 1, \ldots, p_T\}$. Without loss of generality, we consider the first block, $T^{-2} \sum_{t=p_T+1}^{T} f_t f_t'$, and the element of $T^{-2} \sum_{t=p_T+1}^{T} f_t f_t'$ can be written as $T^{-2} \sum_{t=p_T+1}^{T} f_{tl} f_{tl'}$, $l, l' \in \{1, 2, \ldots, m\}$. By the standard unit root asymptotic analysis, we have

$$
\frac{1}{T^2} \sum_{t=p_T+1}^{T} E(f_{tl} f_{tl'}) = O(1),
$$

which implies that $T^{-2} \sum_{t=p_T+1}^{T} \mathbf{f}_t \mathbf{f}_t' = O_p(1)$, then we have $\left\| T^{-2} \tilde{\mathbf{F}}' \tilde{\mathbf{F}} \right\|_\infty = O_p(p_T)$, which establishes the first part of (A11). The second part of (A11) is established by

$$
\begin{aligned}
\left\| \frac{\mathbf{H}'\mathbf{H}}{T^2} \right\|_\infty &= \left\| \bar{\mathbf{P}}' \frac{\mathbf{G}'\mathbf{G}}{T^2} \bar{\mathbf{P}} \right\|_\infty \\
&\leq \|\bar{\mathbf{P}}'\|_\infty \left\| \frac{\mathbf{G}'\mathbf{G}}{T^2} \right\|_\infty \|\bar{\mathbf{P}}\|_\infty = O_p(p_T),
\end{aligned}
$$

since the norm of $\bar{\mathbf{P}}$ is assumed bounded (and the above result).

To prove the third part of (A11), note that $\bar{\mathbf{Q}} = \mathbf{H} + \bar{\mathbf{V}}^*$, by (A7), the second part of (A9), and the previous result in (A11), we have

$$
\begin{aligned}
\left\| \frac{\bar{\mathbf{Q}}'\bar{\mathbf{Q}}}{T^2} \right\|_\infty &= \frac{(\mathbf{H} + \bar{\mathbf{V}}^*)'(\mathbf{H} + \bar{\mathbf{V}}^*)}{T^2} \\
&\leq \left\| \frac{\mathbf{H}'\mathbf{H}}{T^2} \right\|_\infty + \left\| \frac{\mathbf{H}'\bar{\mathbf{V}}^*}{T^2} \right\|_\infty + \left\| \frac{\bar{\mathbf{V}}^{*\prime}\mathbf{H}}{T^2} \right\|_\infty + \left\| \frac{\bar{\mathbf{V}}^{*\prime}\bar{\mathbf{V}}^*}{T^2} \right\|_\infty \\
&= O_p(p_T).
\end{aligned}
$$

Noting that $\bar{\mathbf{Q}} = \mathbf{G}\bar{\mathbf{P}} + \bar{\mathbf{V}}^*$, by the first part of (A9) and (A11), the first part of (A12) can be established.

To establish the second part of (A12), note that $\mathbf{H} = \mathbf{G}\bar{\mathbf{P}}$ and $\Xi_i = \mathbf{G}\Pi_i + \Omega_i$, and recalling that $\Omega_i = \mathbf{e}_i \Psi'_{\xi i}(L)$ with $\mathbf{e}_i = (\varepsilon_i, \mathbf{U}_i)$, we have

$$
\begin{aligned}
\left\| \frac{\mathbf{H}'\Xi_i}{T^2} \right\|_\infty &= \left\| \frac{\bar{\mathbf{P}}'\mathbf{G}'(\mathbf{G}\Pi_i + \Omega_i)}{T^2} \right\|_\infty \leq \left\| \bar{\mathbf{P}}' \frac{\mathbf{G}'\mathbf{G}}{T^2} \Pi_i \right\|_\infty + \left\| \bar{\mathbf{P}}' \frac{\mathbf{G}'\Omega_i}{T^2} \right\|_\infty \\
&\leq \|\bar{\mathbf{P}}'\|_\infty \left\| \frac{\mathbf{G}'\mathbf{G}}{T^2} \right\|_\infty \|\Pi_i\|_\infty + \|\bar{\mathbf{P}}'\|_\infty \left\| \frac{\mathbf{G}'\Omega_i}{T^2} \right\|_\infty = O_p(p_T).
\end{aligned}
$$

by (A10) and the first part of (A11), as well as the assumption that the norm of $\bar{\mathbf{P}}$, Π_i, and $\Psi_{\xi i}(L)$ is assumed bounded in probability uniformly over i. The third part of (A12) is proven straightforwardly since $\bar{\mathbf{Q}} = \mathbf{H} + \bar{\mathbf{V}}^*$, using (A8) and the second part of (A12). □

Proof of Lemma A5. To proof (A13), we note that

$$
\Xi_i = \mathbf{G}\Pi_i + \Omega_i, \tag{A32}
$$

where $\mathbf{G} = (\tau, \mathbf{F}, \mathbf{F}_{-1}, \ldots, \mathbf{F}_{-p_T})$ is a matrix of $I(1)$ factors, $\Omega_i = \mathbf{e}_i \Psi'_{\xi i}$ with $\mathbf{e}_i = (\varepsilon_i, \mathbf{U}_i)$. Denote the OLS estimator of the multiple regression (A32) as $\hat{\Pi}_i = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\Xi_i$. Since that $\Xi_i'\mathbf{M}_g\Xi_i = \Omega_i'\mathbf{M}_g\Omega_i = \hat{\Omega}_i'\hat{\Omega}_i$, where $\hat{\Omega}_i$ is the OLS residuals, i.e., $\hat{\Omega}_i = \Xi_i - \mathbf{G}\hat{\Pi}_i$, and in the light of Assumption, $T^{-1}(\Omega_i'\Omega_i) \to \Sigma_{\Omega_i}$, we only need to show that $T^{-1}(\hat{\Omega}_i'\hat{\Omega}_i) - T^{-1}(\Omega_i'\Omega_i) \to 0$. In fact, we can write

$$
\begin{aligned}
T^{-1}(\hat{\Omega}_i'\hat{\Omega}_i) - T^{-1}(\Omega_i'\Omega_i) &= T^{-1}\hat{\Omega}_i'(\hat{\Omega}_i - \Omega_i) + T^{-1}(\hat{\Omega}_i - \Omega_i)\Omega_i \\
&= -T^{-1}\Xi_i'\mathbf{M}_g\mathbf{G}(\hat{\Pi}_i - \Pi_i) - T^{-1}(\hat{\Pi}_i - \Pi_i)\mathbf{G}'\Omega_i \\
&= -(\hat{\Pi}_i - \Pi_i)(T^{-1}\mathbf{G}'\Omega_i),
\end{aligned}
$$

because $\mathbf{M}_g\mathbf{G} = \mathbf{0}$. However, since $\left\| T^{-1}\mathbf{G}'\Omega_i \right\|_\infty = O_p(1)$ by (A10) of Lemma A4, $\hat{\Pi}_i - \Pi_i = O_p(T^{-1})$, it follows that $T^{-1}(\hat{\Omega}_i'\hat{\Omega}_i) - T^{-1}(\Omega_i'\Omega_i) = O_p(T^{-1})$. Hence, $\frac{\Xi_i'\mathbf{M}_g\Xi_i}{T} \xrightarrow{p} \Sigma_{\Omega_i}$, (A13) is established.

To prove (A14), we follow the same spirit of Lemma A.4 in Kapetanios et al. (2011), but need more attention because of the lags. Specifically, note that

$$
\mathbf{M}_q\bar{\mathbf{Q}} = \mathbf{M}_q(\mathbf{G}\bar{\mathbf{P}} + \bar{\mathbf{V}}^*),
$$

since $\bar{\mathbf{Q}} = \mathbf{G}\bar{\mathbf{P}} + \bar{\mathbf{V}}^*$, where $\bar{\mathbf{Q}} = (\tau, \bar{\mathbf{Z}}, \bar{\mathbf{Z}}_{-1}, \ldots, \bar{\mathbf{Z}}_{-p_T})$, $\mathbf{G} = (\tau, \tilde{\mathbf{F}}) = (\tau, \mathbf{F}, \mathbf{F}_{-1}, \ldots, \mathbf{F}_{-p_T})$ and $\bar{\mathbf{V}}^* = (\mathbf{0}, \bar{\mathbf{V}}, \bar{\mathbf{V}}_{-1}, \ldots, \bar{\mathbf{V}}_{-p_T})$. However, $\mathbf{M}_q \bar{\mathbf{Q}} = \mathbf{0}$ and $\mathbf{M}_q \tau = \mathbf{0}$ since $\tau \in \bar{\mathbf{Q}}$. Then

$$(\mathbf{0}, \mathbf{M}_q \tilde{\mathbf{F}}) \begin{pmatrix} 1 & \tilde{\mathbf{c}}_z \\ \mathbf{0} & \tilde{\mathbf{\Psi}} \end{pmatrix} + (\mathbf{0}, \mathbf{M}_q \bar{\mathbf{V}}^*) = \mathbf{0},$$

or $\mathbf{M}_q \tilde{\mathbf{F}} \tilde{\mathbf{\Psi}} = -\mathbf{M}_q \bar{\mathbf{V}}^*$.

For the second column block of the above equation, we have $\mathbf{M}_q \mathbf{F} \mathbf{\Psi}'(L) = -\mathbf{M}_q \bar{\mathbf{V}}$ or $\mathbf{M}_q \mathbf{F} \mathbf{C}' \mathbf{\Lambda}'(L) = -\mathbf{M}_q \bar{\mathbf{V}}$ as $N \xrightarrow{P} \infty$, since $\tilde{\mathbf{\Psi}} = diag(\mathbf{\Psi}'(L))$ and $\mathbf{\Psi}(L) = \mathbf{\Lambda}(L)\mathbf{C} + O_p(N^{-1/2})$, Hence

$$\bar{\mathbf{V}}' \mathbf{M}_q \mathbf{F} \mathbf{C}' \mathbf{\Lambda}'(L) = -\bar{\mathbf{V}}' \mathbf{M}_q \bar{\mathbf{V}}, \tag{A33}$$

and

$$\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{F} \mathbf{C}' \mathbf{\Lambda}'(L) = -\mathbf{\Xi}_i' \mathbf{M}_q \bar{\mathbf{V}}. \tag{A34}$$

Since $\mathbf{\Lambda}(L)$ is invertible under the assumption, then (A34) can be rewritten as

$$\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{F} \mathbf{C}' = -\mathbf{\Xi}_i' \mathbf{M}_q \bar{\mathbf{V}} \mathbf{\Lambda}^{-1}(L).$$

When the rank condition is satisfied, we have

$$\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{F} = -\mathbf{\Xi}_i' \mathbf{M}_q \bar{\mathbf{V}} \mathbf{\Lambda}^{-1}(L) \mathbf{C} (\mathbf{C}' \mathbf{C})^{-1}. \tag{A35}$$

Note that $\mathbf{\Xi}_i$ can be written as $\mathbf{\Xi}_i = \mathbf{G}_{2i} \mathbf{\Pi}_{2i} + \mathbf{\Omega}_i$, where $\mathbf{G}_{2i} = (\tau, \mathbf{F})$ and $\mathbf{\Pi}_{2i} = (\mathbf{c}_{\xi i}, \mathbf{\Psi}_{\xi i} \mathbf{C}_i)'$, then

$$\begin{aligned} \mathbf{\Xi}_i' \mathbf{M}_q \bar{\mathbf{V}} &= (\mathbf{G}_{2i} \mathbf{\Pi}_{2i} + \mathbf{\Omega}_i)' \mathbf{M}_q \bar{\mathbf{V}} = \mathbf{\Pi}_{2i}' \mathbf{G}_{2i}' \mathbf{M}_q \bar{\mathbf{V}} + \mathbf{\Omega}_i' \mathbf{M}_q \bar{\mathbf{V}} \\ &= (\mathbf{c}_{\xi i}', \mathbf{C}_i' \mathbf{\Psi}_{\xi i}') \begin{pmatrix} \tau' \\ \mathbf{F}' \end{pmatrix} \mathbf{M}_q \bar{\mathbf{V}} + \mathbf{\Omega}_i' \mathbf{M}_q \bar{\mathbf{V}} \\ &= (\mathbf{c}_{\xi i}', \mathbf{C}_i' \mathbf{\Psi}_{\xi i}') \begin{pmatrix} \mathbf{0} \\ \mathbf{F}' \mathbf{M}_q \bar{\mathbf{V}} \end{pmatrix} + \mathbf{\Omega}_i' \mathbf{M}_q \bar{\mathbf{V}} \\ &= \mathbf{C}_i' \mathbf{\Psi}_{\xi i}' \mathbf{F}' \mathbf{M}_q \bar{\mathbf{V}} + \mathbf{\Psi}_{\xi i} \mathbf{e}_i' \mathbf{M}_q \bar{\mathbf{V}}. \end{aligned} \tag{A36}$$

Substituting (A36) into (A35), we obtain

$$\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{F} = -\mathbf{C}_i' \mathbf{\Psi}_{\xi i}' \mathbf{F}' \mathbf{M}_q \bar{\mathbf{V}} \mathbf{\Lambda}^{-1}(L) \mathbf{C} (\mathbf{C}' \mathbf{C})^{-1} - \mathbf{\Psi}_{\xi i} \mathbf{e}_i' \mathbf{M}_q \bar{\mathbf{V}} \mathbf{\Lambda}^{-1}(L) \mathbf{C} (\mathbf{C}' \mathbf{C})^{-1}. \tag{A37}$$

Moreover, from (A33), $\mathbf{\Lambda}(L) \mathbf{C} \mathbf{F}' \mathbf{M}_q \bar{\mathbf{V}} = -\bar{\mathbf{V}}' \mathbf{M}_q \bar{\mathbf{V}}$, which directly follows

$$\mathbf{F}' \mathbf{M}_q \bar{\mathbf{V}} = -(\mathbf{C}' \mathbf{C})^{-1} \mathbf{C}' \mathbf{\Lambda}^{-1}(L) \bar{\mathbf{V}}' \mathbf{M}_q \bar{\mathbf{V}}, \tag{A38}$$

under the assumption of $\mathbf{\Lambda}(L)$ is invertible and the rank condition is satisfied. Then, using this result in (A37), we have

$$\begin{aligned} \left\| \frac{\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{F}}{T} \right\| &= \left\| \mathbf{C}_i' \right\| \left\| \mathbf{\Psi}_{\xi i}' \right\| \left\| (\mathbf{C}' \mathbf{C})^{-1} \mathbf{C}' \right\|^2 \left\| \mathbf{\Lambda}^{-1}(L) \right\|^2 \left\| \frac{\bar{\mathbf{V}}' \mathbf{M}_q \bar{\mathbf{V}}}{T} \right\| \\ &\quad + \left\| \mathbf{\Psi}_{\xi i} \right\| \left\| \frac{\mathbf{e}_i' \mathbf{M}_q \bar{\mathbf{V}}}{T} \right\| \left\| \mathbf{\Lambda}^{-1}(L) \right\| \left\| \mathbf{C} (\mathbf{C}' \mathbf{C})^{-1} \right\|. \end{aligned} \tag{A39}$$

Since the norms of $\mathbf{C}_i$, $\mathbf{\Lambda}^{-1}(L)$ and $\mathbf{\Psi}_{\xi i}$ are assumed to be bounded, we need to establish the probability orders of $\left\| \bar{\mathbf{V}}' \mathbf{M}_q \bar{\mathbf{V}} / T \right\|$ and $\left\| \mathbf{e}_i' \mathbf{M}_q \bar{\mathbf{V}} / T \right\|$. For $\bar{\mathbf{V}}' \mathbf{M}_q \bar{\mathbf{V}} / T$, since

$\bar{\mathbf{V}}$ is a $(T - p_T) \times (k+1)$ submatrix of $\bar{\mathbf{V}}^*$, (A7) and (A9) imply $\bar{\mathbf{V}}'\bar{\mathbf{V}}/T = O_p(N^{-1})$ and $\left\|\bar{\mathbf{Q}}'\bar{\mathbf{V}}/T\right\|_\infty = O_p(N^{-1/2})$, which together with (A7), we obtain

$$\frac{\bar{\mathbf{V}}'\mathbf{M}_q\bar{\mathbf{V}}}{T} = O_p(\frac{1}{N}).$$

Similarly, by (A7)–(A9),

$$\frac{\mathbf{e}_i'\mathbf{M}_q\bar{\mathbf{V}}}{T} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}}), \quad \text{uniformly over } i,$$

Substituting the above two results into (A39) establishes the result. □

Proof of Lemma A6. To prove (A15), we need to determine the order of probability of $\left\|\frac{\Xi_i'\mathbf{M}_q\Xi_i}{T} - \frac{\Xi_i'\mathbf{M}_h\Xi_i}{T}\right\|_\infty$, by the triangle inequality of the matrix norm $\|\cdot\|_\infty$, which equals

$$\left\|\frac{\Xi_i'\bar{\mathbf{Q}}(\bar{\mathbf{Q}}'\bar{\mathbf{Q}})^{-1}\bar{\mathbf{Q}}'\Xi_i}{T} - \frac{\Xi_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\Xi_i}{T}\right\|_\infty$$

$$\leq \left\|\frac{1}{T}(\Xi_i'\bar{\mathbf{Q}} - \Xi_i'\mathbf{H})(\bar{\mathbf{Q}}'\bar{\mathbf{Q}})^{-1}\bar{\mathbf{Q}}'\Xi_i\right\|_\infty$$

$$+ \left\|\frac{1}{T}\Xi_i'\mathbf{H}\left((\bar{\mathbf{Q}}'\bar{\mathbf{Q}})^{-1} - (\mathbf{H}'\mathbf{H})^{-1}\right)\bar{\mathbf{Q}}'\Xi_i\right\|_\infty$$

$$+ \left\|\frac{1}{T}\Xi_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}(\bar{\mathbf{Q}}'\Xi_i - \mathbf{H}'\Xi_i)\right\|_\infty. \tag{A40}$$

Using the results of Lemma A.3 and the submultiplicative property of the matrix norm, and noting that $\bar{\mathbf{Q}} = \mathbf{H} + \bar{\mathbf{V}}^*$, we focus on the individual elements on the right side of (A40).

For the first term, we have

$$\left\|\frac{1}{T}(\Xi_i'\bar{\mathbf{Q}} - \Xi_i'\mathbf{H})(\bar{\mathbf{Q}}'\bar{\mathbf{Q}})^{-1}\bar{\mathbf{Q}}'\Xi_i\right\|_\infty \leq \left\|\frac{\Xi_i'\bar{\mathbf{V}}^*}{T}\right\|_\infty \left\|\left(\frac{\bar{\mathbf{Q}}'\bar{\mathbf{Q}}}{T^2}\right)^{-1}\frac{\bar{\mathbf{Q}}'\Xi_i}{T^2}\right\|_\infty$$

$$= O_p(\frac{p_T}{\sqrt{N}}), \quad \text{uniformly over } i, \tag{A41}$$

by the third parts of (A9), (A11), and (A12).

For the second term, we also have

$$\left\|\frac{1}{T}\Xi_i'\mathbf{H}\left((\bar{\mathbf{Q}}'\bar{\mathbf{Q}})^{-1} - (\mathbf{H}'\mathbf{H})^{-1}\right)\bar{\mathbf{Q}}'\Xi_i\right\|_\infty$$

$$\leq \left\|\frac{\bar{\mathbf{V}}^{*\prime}\bar{\mathbf{V}}^*}{T} + \frac{\mathbf{H}'\bar{\mathbf{V}}^*}{T} + \frac{\bar{\mathbf{V}}^{*\prime}\mathbf{H}}{T}\right\|_\infty \left\|\frac{\Xi_i'\mathbf{H}}{T^2}\left(\frac{\bar{\mathbf{Q}}'\bar{\mathbf{Q}}}{T^2}\right)^{-1}\right\|_\infty \left\|\left(\frac{\mathbf{H}'\mathbf{H}}{T^2}\right)^{-1}\frac{\bar{\mathbf{Q}}'\Xi_i}{T^2}\right\|_\infty$$

$$= O_p(\frac{p_T}{\sqrt{N}}), \quad \text{uniformly over } i, \tag{A42}$$

by (A7), as well as some results in (A9), (A11), and (A12).

Finally, we have

$$\left\|\frac{1}{T}\Xi_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}(\bar{\mathbf{Q}}'\Xi_i - \mathbf{H}'\Xi_i)\right\|_\infty \leq \left\|\frac{\Xi_i'\mathbf{H}}{T^2}\left(\frac{\mathbf{H}'\mathbf{H}}{T^2}\right)^{-1}\right\|_\infty \left\|\frac{\Xi_i'\bar{\mathbf{V}}^*}{T}\right\|_\infty$$

$$= O_p(\frac{p_T}{\sqrt{N}}), \quad \text{uniformly over } i, \tag{A43}$$

by (A9), (A11), and (A12). Substituting (A41)–(A43) into (A40), we have

$$\left\| \frac{\mathbf{\Xi}_i' \mathbf{M}_q \mathbf{\Xi}_i}{T} - \frac{\mathbf{\Xi}_i' \mathbf{M}_h \mathbf{\Xi}_i}{T} \right\|_\infty = O_p(\frac{p_T}{\sqrt{N}}), \quad \text{uniformly over } i,$$

as required.

To establish result (A16), similar to the proof of (A15), we have

$$\left\| \frac{\mathbf{\Xi}_i' \bar{\mathbf{Q}} (\bar{\mathbf{Q}}' \bar{\mathbf{Q}})^{-1} \bar{\mathbf{Q}}' \mathbf{F}}{T} - \frac{\mathbf{\Xi}_i' \mathbf{H} (\mathbf{H}' \mathbf{H})^{-1} \mathbf{H}' \mathbf{F}}{T} \right\|_\infty$$

$$\leq \left\| \frac{1}{T} (\mathbf{\Xi}_i' \bar{\mathbf{Q}} - \mathbf{\Xi}_i' \mathbf{H}) (\bar{\mathbf{Q}}' \bar{\mathbf{Q}})^{-1} \bar{\mathbf{Q}}' \mathbf{F} \right\|_\infty$$

$$+ \left\| \frac{1}{T} \mathbf{\Xi}_i' \mathbf{H} \left((\bar{\mathbf{Q}}' \bar{\mathbf{Q}})^{-1} - (\mathbf{H}' \mathbf{H})^{-1} \right) \bar{\mathbf{Q}}' \mathbf{F} \right\|_\infty$$

$$+ \left\| \frac{1}{T} \mathbf{\Xi}_i' \mathbf{H} (\mathbf{H}' \mathbf{H})^{-1} (\bar{\mathbf{Q}}' \mathbf{F} - \mathbf{H}' \mathbf{F}) \right\|_\infty, \tag{A44}$$

then, examine each term of (A44), and note that $\mathbf{F} \in \mathbf{G}$.

For the first term, the third part of (A9) and (A12) imply

$$\left\| \frac{1}{T} (\mathbf{\Xi}_i' \bar{\mathbf{Q}} - \mathbf{\Xi}_i' \mathbf{H}) (\bar{\mathbf{Q}}' \bar{\mathbf{Q}})^{-1} \bar{\mathbf{Q}}' \mathbf{F} \right\|_\infty \leq \left\| \frac{\mathbf{\Xi}_i' \bar{\mathbf{V}}^*}{T} \right\|_\infty \left\| \left(\frac{\bar{\mathbf{Q}}' \bar{\mathbf{Q}}}{T^2} \right)^{-1} \frac{\bar{\mathbf{Q}}' \mathbf{F}}{T^2} \right\|_\infty$$

$$= O_p(\frac{p_T}{\sqrt{N}}), \quad \text{uniformly over } i. \tag{A45}$$

Using some results in Lemma A4, we have

$$\left\| \frac{1}{T} \mathbf{\Xi}_i' \mathbf{H} \left((\bar{\mathbf{Q}}' \bar{\mathbf{Q}})^{-1} - (\mathbf{H}' \mathbf{H})^{-1} \right) \bar{\mathbf{Q}}' \mathbf{F} \right\|_\infty$$

$$\leq \left\| \frac{\bar{\mathbf{V}}^{*\prime} \bar{\mathbf{V}}^*}{T} + \frac{\mathbf{H}' \bar{\mathbf{V}}^*}{T} + \frac{\bar{\mathbf{V}}^{*\prime} \mathbf{H}}{T} \right\|_\infty \left\| \frac{\mathbf{\Xi}_i' \mathbf{H}}{T^2} \left(\frac{\bar{\mathbf{Q}}' \bar{\mathbf{Q}}}{T^2} \right)^{-1} \right\|_\infty \left\| \left(\frac{\mathbf{H}' \mathbf{H}}{T^2} \right)^{-1} \frac{\bar{\mathbf{Q}}' \mathbf{F}}{T^2} \right\|_\infty$$

$$= O_p(\frac{p_T}{\sqrt{N}}), \quad \text{uniformly over } i. \tag{A46}$$

Finally, by the first part of (A9), the second part of (A11) and (A12), we have

$$\left\| \frac{1}{T} \mathbf{\Xi}_i' \mathbf{H} (\mathbf{H}' \mathbf{H})^{-1} (\bar{\mathbf{Q}}' \mathbf{F} - \mathbf{H}' \mathbf{F}) \right\|_\infty \leq \left\| \frac{\mathbf{\Xi}_i' \mathbf{H}}{T^2} \left(\frac{\mathbf{H}' \mathbf{H}}{T^2} \right)^{-1} \right\|_\infty \left\| \frac{\bar{\mathbf{V}}^{*\prime} \mathbf{F}}{T} \right\|_\infty$$

$$= O_p(\frac{p_T}{\sqrt{N}}), \quad \text{uniformly over } i. \tag{A47}$$

Substituting (A45)–(A47) into (A44), we have

$$\left\| \frac{\mathbf{\Xi}_i' \bar{\mathbf{M}}_q \mathbf{F}}{T} - \frac{\mathbf{\Xi}_i' \mathbf{M}_h \mathbf{F}}{T} \right\|_\infty = O_p(\frac{p_T}{\sqrt{N}}), \quad \text{uniformly over } i,$$

which completes the proof of (A16).

Result (A17) can also be established in a similar way, we have

$$\left\| \frac{\boldsymbol{\Xi}_i' \bar{\mathbf{M}}_q \varepsilon_i}{T} - \frac{\boldsymbol{\Xi}_i' \mathbf{M}_h \varepsilon_i}{T} \right\|_\infty = \left\| \frac{\boldsymbol{\Xi}_i' \bar{\mathbf{Q}} (\bar{\mathbf{Q}}' \bar{\mathbf{Q}})^{-1} \bar{\mathbf{Q}}' \varepsilon_i}{T} - \frac{\boldsymbol{\Xi}_i' \mathbf{H} (\mathbf{H}' \mathbf{H})^{-1} \mathbf{H}' \varepsilon_i}{T} \right\|_\infty$$

$$\leq \left\| \frac{1}{T} (\boldsymbol{\Xi}_i' \bar{\mathbf{Q}} - \boldsymbol{\Xi}_i' \mathbf{H}) (\bar{\mathbf{Q}}' \bar{\mathbf{Q}})^{-1} \bar{\mathbf{Q}}' \varepsilon_i \right\|_\infty$$

$$+ \left\| \frac{1}{T} \boldsymbol{\Xi}_i' \mathbf{H} \left((\bar{\mathbf{Q}}' \bar{\mathbf{Q}})^{-1} - (\mathbf{H}' \mathbf{H})^{-1} \right) \bar{\mathbf{Q}}' \varepsilon_i \right\|_\infty$$

$$+ \left\| \frac{1}{T} \boldsymbol{\Xi}_i' \mathbf{H} (\mathbf{H}' \mathbf{H})^{-1} (\bar{\mathbf{Q}}' \varepsilon_i - \mathbf{H}' \varepsilon_i) \right\|_\infty, \qquad (A48)$$

then we examine each of the above terms. The first term equals

$$\left\| \frac{1}{T} (\boldsymbol{\Xi}_i' \bar{\mathbf{Q}} - \boldsymbol{\Xi}_i' \mathbf{H}) (\bar{\mathbf{Q}}' \bar{\mathbf{Q}})^{-1} \bar{\mathbf{Q}}' \varepsilon_i \right\|_\infty \leq \frac{1}{T} \left\| \frac{\boldsymbol{\Xi}_i' \bar{\mathbf{V}}^*}{T} \right\|_\infty \left\| \left(\frac{\bar{\mathbf{Q}}' \bar{\mathbf{Q}}}{T^2} \right)^{-1} \frac{\bar{\mathbf{Q}}' \varepsilon_i}{T} \right\|_\infty$$

$$= O_p\left(\frac{1}{\sqrt{NT}} \right), \quad \text{uniformly over } i, \qquad (A49)$$

by the third part of (A9)–(A11). Next, we have

$$\left\| \frac{1}{T} \boldsymbol{\Xi}_i' \mathbf{H} \left((\bar{\mathbf{Q}}' \bar{\mathbf{Q}})^{-1} - (\mathbf{H}' \mathbf{H})^{-1} \right) \bar{\mathbf{Q}}' \varepsilon_i \right\|_\infty$$

$$\leq \frac{1}{T} \left\| \frac{\bar{\mathbf{V}}^{*\prime} \bar{\mathbf{V}}^*}{T} + \frac{\mathbf{H}' \bar{\mathbf{V}}^*}{T} + \frac{\bar{\mathbf{V}}^{*\prime} \mathbf{H}}{T} \right\|_\infty \left\| \frac{\boldsymbol{\Xi}_i' \mathbf{H}}{T^2} \left(\frac{\bar{\mathbf{Q}}' \bar{\mathbf{Q}}}{T^2} \right)^{-1} \right\|_\infty \left\| \left(\frac{\mathbf{H}' \mathbf{H}}{T^2} \right)^{-1} \frac{\bar{\mathbf{Q}}' \varepsilon_i}{T} \right\|_\infty$$

$$= O_p\left(\frac{1}{\sqrt{NT}} \right), \quad \text{uniformly over } i, \qquad (A50)$$

by (A7), and the results in (A9)–(A12). Finally,

$$\left\| \frac{1}{T} \boldsymbol{\Xi}_i' \mathbf{H} (\mathbf{H}' \mathbf{H})^{-1} (\bar{\mathbf{Q}}' \varepsilon_i - \mathbf{H}' \varepsilon_i) \right\|_\infty$$

$$\leq \left\| \frac{\boldsymbol{\Xi}_i' \mathbf{H}}{T^2} \left(\frac{\mathbf{H}' \mathbf{H}}{T^2} \right)^{-1} \right\|_\infty \left\| \frac{\bar{\mathbf{V}}^{*\prime} \varepsilon_i}{T} \right\|_\infty = O_p\left(\frac{p_T}{\sqrt{NT}} \right) + O_p\left(\frac{p_T}{N} \right), \qquad (A51)$$

by (A8), the second part of (A11) and (A12). Using (A49)–(A51) into (A48), (A17) is proven. □

Notes

[1] An alternative approach to deal with cross-sectional dependence is the principle component analysis proposed by Bai (2009).

[2] As in Pesaran (2006) and Kapetanios et al. (2011), observed factors, such as time effects, can also be included in model (1). For notational simplicity and illustration purpose, we do not include such factors in the model (1).

[3] As Chudik and Pesaran (2015a) point out, the number of lags p_T needs to be restricted. Letting $p_T^3/T \to \lambda, 0 < \lambda < \infty$ can ensures that, on the one hand, the number of lags is not too large, so that there are sufficient degrees of freedom for the consistent estimator, and on the other hand, the number of lags is not too small, so that the bias due to the truncation of infinite lag polynomials is sufficiently small

[4] We note that $\bar{\mathbf{Q}}$ can be denoted as $\bar{\mathbf{Q}} \overset{\triangle}{=} (\tau, \tilde{\mathbf{Z}})$, where $\tau = (1, 1, \ldots 1)'$ is a $(T - p_T) \times 1$ vector of ones, $\tilde{\mathbf{Z}}$ is the $(T - p_T) \times (k+1)p_T$ matrices of observations on $\bar{z}_t$ for $t = p_T + 1, p_T + 2, \ldots, T$.

[5] To illustrate the validity and robustness of the CCE estimator in the case of non-stationary common factors, the data-generating process and parameter settings are similar to the settings in Chudik and Pesaran (2015a), except for unobserved common factors.

[6] We also conducted additional Monte Carlo simulations for other settings, such as $p_T = [0.75T^{1/3}]$ and $p_T = [1.25T^{1/3}]$; the corresponding results are slightly worse than that of $p_T = [T^{1/3}]$, these results are not reported to save space.

References

Bada, Oualid, and Alois Kneip. 2014. Parameter cascading for panel models with unknown number of unobserved factors: An application to the credit spread puzzle. *Computational Statistics and Data Analysis* 76: 95–115. [CrossRef]

Bada, Oualid, and Dominik Liebl. 2014. The R package phtt: Panel data analysis with heterogeneous time trends. *Journal of Statistical Software* 59: 1–34. [CrossRef]

Bai, Jushan. 2009. Panel data models with interactive fixed effects. *Econometrica* 77: 1229–79.

Bai, Jushan, and Serena Ng. 2004. A panic on unit root tests and cointegration. *Econometrica* 72: 1127–77. [CrossRef]

Bai, Jushan, and Serena Ng. 2010. Panel unit root tests with cross section dependence: A further investigation. *Econometric Theory* 26: 1088–114. [CrossRef]

Bailey, Natalia, George Kapetanios, and M. Hashem Pesaran. 2016. Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics* 31: 929–1196. [CrossRef]

Baltagi, Badi H., and Dong Li. 2004. Prediction in the panel data model with spatial correlation. In *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Edited by Luc Anselin, Raymond J. G. M. Florax and Sergio J. Rey. Berlin/Heidelberg: Springer, pp. 283–295.

Ben-Israel, Adi, and Thomas N. E. Greville. 2003. *Generalized Inverses: Theory and Applications*, 2nd ed. New York: Springer.

Bussiere, Matthieu, Alexander Chudik, and Arnaud Mehl. 2013. How have global shocks impacted the real effective exchange rates of individual euro area countries since the euro's creation? *The B.E. Journal of Macroeconomics* 13: 1–48. [CrossRef]

Chudik, Alexander, and M. Hashem Pesaran. 2013. Econometric analysis of high dimensional VARs featuring a dominant unit. *Econometric Reviews* 32: 592-649. [CrossRef]

Chudik, Alexander, and M. Hashem Pesaran. 2015a. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics* 188: 393–420. [CrossRef]

Chudik, Alexander, and M. Hashem Pesaran. 2015b. Large panel data models with cross-sectional dependence: A survey. In *The Oxford Handbook of Panel Data*. Edited by Badi H. Baltagi. Oxford: Oxford University Press, pp. 2–45.

Chudik, Alexander, Kamiar Mohaddes, M. Hashem Pesaran, and Mehdi Raissi. 2017. Is there a debt-threshold effect on output growth? *Review of Economics and Statistics* 99: 135–50. [CrossRef]

Eberhardt, Markus, Christian Helmers, and Hubert Strauss. 2013. Do spillovers matter when estimating private returns to R&D? *Review of Economics and Statistics* 95: 436–48.

Greenaway-McGrevy, Ryan, Chirok Han, and Donggyu Sul. 2012. Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics* 169: 48–53. [CrossRef]

Hahn, Jinyong, and Whitney Newey. 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72: 1295–319. [CrossRef]

Juodis, Artūras, Hande Karabiyik, and Joakim Westerlund. 2021. On the robustness of the pooled CCE estimator. *Journal of Econometrics* 220: 325–48. [CrossRef]

Kao, Chihwa, Lorenzo Trapani, and Giovanni Urga. 2012. Asymptotics for panel models with common shocks. *Econometric Reviews* 31: 390–439. [CrossRef]

Kapetanios, George, M. Hashem Pesaran, and Takashi Yamagata. 2011. Panels with non-stationary multifactor error structures. *Journal of Econometrics* 160: 326–48. [CrossRef]

Moon, Hyungsik Roger, and Martin Weidner. 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83: 1543–79. [CrossRef]

Moon, Hyungsik Roger, and Martin Weidner. 2017. Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory* 33: 158–95. [CrossRef]

Omay, Tolga, and Elif Oznur Kan. 2010. Re-examing the threshold effects in the inflation-growth nexus with cross-sectionally dependent non-linear panel: Evidence from six industrialized economies. *Economic Modelling* 27: 996–1005. [CrossRef]

Pesaran, M. Hashem. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74: 967–1012. [CrossRef]

Pesaran, M. Hashem. 2007. A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics* 22: 265–312. [CrossRef]

Pesaran, M. Hashem. 2015. Testing weak cross-sectional dependence in large panels. *Econometric Reviews* 34: 1089–117. [CrossRef]

Pesaran, M. Hashem, Ron Smith, and Takashi Yamagata. 2013. Panel unit root tests in the presence of multifactor error structure. *Journal of Econometrics* 175: 94–115. [CrossRef]

Westerlund, Joakim, Petrova Yana, and Norkute Milda. 2019. CCE in fixed-T panels. *Journal of Applied Econometrics* 34: 746–761. [CrossRef]

Zaffaroni, Paolo. 2009. Generalized least estimation of panel with common shocks. *Unpublished Manuscript*.

Zhou, Qiankun, and Yonghui Zhang. 2016. Common correlated effects estimation of unbalanced panel data models with cross-sectional dependence. *Journal of Economic Theory and Econometrics* 27: 25–45.