IV Estimation of Heterogeneous Spatial Dynamic Panel Models with Interactive Effects

Chen, Jia and Cui, Guowei and Sarafidis, Vasilis and Yamagata, Takashi

University of Macau, HUST, Brunel University London, York University

30 January 2025

IV Estimation of Heterogeneous Spatial Dynamic Panel Models with Interactive Effects[*]

Jia Chen[†]
University of Macau

Guowei Cui[‡]
HUST

Vasilis Sarafidis[§]
Brunel University London

Takashi Yamagata[¶]
University of York and
Osaka University

Abstract

This paper develops a Mean Group Instrumental Variables (MGIV) estimator for spatial dynamic panel data models with interactive effects, under large N and T asymptotics. Unlike existing approaches that typically impose slope-parameter homogeneity, MGIV accommodates cross-sectional heterogeneity in slope coefficients. The proposed estimator is linear, making it computationally efficient and robust. Furthermore, it avoids the incidental parameters problem, enabling asymptotically valid inferences without requiring bias correction. The Monte Carlo experiments indicate strong finite-sample performance of the MGIV estimator across various sample sizes and parameter configurations. The practical utility of the estimator is illustrated through an application to regional economic growth in Europe. By explicitly incorporating heterogeneity, our approach provides fresh insights into the determinants of regional growth, underscoring the critical roles of spatial and temporal dependencies.

JEL classification: C3; C33; C55; O47.

Key Words: Dynamic panel data, spatial interactions, heterogeneous slopes, interactive effects, latent common factors, instrumental variables, large N and T asymptotics.

1 Introduction

Economic outcomes are shaped by complex dependencies that span both temporal dynamics and spatial interactions. Temporal dependencies arise from phenomena such as habit formation, adjustment costs, and economic slack, wherein past behavior influences current outcomes (e.g., Hamermesh 1995, Jappelli and Pistaferri 2017). Spatial interactions, on the other hand, reflect the influence of peers, spatial networks, and spillover mechanisms (e.g., Case 1991, Manski 1993, Bramoullé

[†]Department of Economics, University of Macau, Taipa, China. E-mail address: chenjia@um.edu.mo.

[‡]School of Economics, Huazhong University of Science and Technology, Wuhan 1037, China.

[§]Department of Economics and Finance, Brunel University London, Uxbridge UB8 2TL, UK. E-mail address: vasilis.sarafidis@brunel.ac.uk.

[¶]Corresponding author. Department of Economics and Related Studies, University of York, York YO10 5DD, UK. E-mail address: takashi.yamagata@york.ac.uk.

et al. 2009), where the behavior of one unit is affected by those of others. These dependencies are further complicated by the frequently pervasive impact of aggregate shocks, such as technological advancements, global market fluctuations, and economy-wide regulatory changes (e.g., Sarafidis and Wansbeek 2021). Together, these factors highlight the considerable challenges associated with modeling economic behavior.

Earlier contributions in the econometric panel data literature addressed temporal dynamics, spatial interactions and aggregate shocks largely in a fragmented manner. For instance, in the context of large-T panels, a considerable body of work focused on dynamic models with additive fixed effects, with limited consideration of spatial interactions or aggregate shocks (e.g., Hahn and Kuersteiner 2002, Alvarez and Arellano 2003, and Hayakawa 2015). Over the past decade, progress has been made with the development of dynamic panels incorporating interactive effects to account for aggregate shocks. Examples include Chudik and Pesaran (2015), Moon and Weidner (2017), Norkutė et al. (2021), De Vos and Everaert (2021) and Juodis and Sarafidis (2022). Parallel advancements have also been made in spatial dynamic panel data models with additive fixed effects, as explored by Yu et al. (2008), Korniotis (2010), Lee and Yu (2014) among others. In these models, spatial interdependence is captured through a pre-specified $N \times N$ adjacency matrix $\mathbf{W}$, which encodes the structure of the interactions among individual units.[1]

Advances in econometrics have since sought to bridge these three strands by developing spatial dynamic panel data models that integrate temporal dependencies, spatial interactions, and interactive error components, as in Shi and Lee (2017), Bai and Li (2021), Cui et. al. (2023), and Higgins and Martellosio (2023).[2] While these methodological advancements represent significant progress, a common assumption underlying much of this literature is slope-parameter homogeneity. That is, the magnitude of the relationships between dependent and independent variables is assumed to be identical across all cross-sectional units. Unfortunately, such assumption can be unduly restrictive in contexts characterized by substantial heterogeneity, which is a common feature of many economic systems. For instance, when cross-sectional heterogeneity in coefficients is captured by a random-coefficient model, it is well-established that dynamic pooled estimators fail to consistently estimate the population average, even for large T; see e.g., Robertson and Symons (1992) and Pesaran and Smith (1995).

To address this limitation, this paper develops a Mean Group Instrumental Variables (MGIV) estimator for spatial dynamic panel data models with interactive effects and heterogeneous slope coefficients, under large N and T asymptotics. Valid instruments are constructed by projecting out common factors from exogenous covariates using Principal Components Analysis (PCA), following the methodology of Bai (2003). Consequently, the individual-specific IV estimates are consistent. Our MGIV estimator then combines these IV estimates and averages them to obtain consistent estimates of population-level effects.

The present extension to incorporate heterogeneous slope coefficients represents a major step forward in the spatial econometrics literature. As noted by LeSage and Chih (2016), "space-time panel data samples covering longer time spans allow us to produce parameter estimates for all N spatial units, an exciting point of departure for future work. Allowing for heterogeneous coefficients for each spatial unit holds a natural appeal when contrasted with conventional static spatial panel

---

[1]Comprehensive overviews of spatial panel data models with additive effects can be found in Elhorst (2014) and Lee and Yu (2015).

[2]In Higgins and Martellosio (2023), it is assumed that $\mathbf{W}$ is observed only partially.

models."

In line with the aforementioned spatial literature, our analysis focuses on the estimation of heterogeneous slopes and spillover effects conditional on a pre-specified $\mathbf{W}$, which is treated as fixed and known. Consequently, the proposed method can be particularly appealing for panel datasets involving geographical entities, such as countries or administrative regions, where measures of spatial and economic proximity are readily available and may naturally inform the specification of $\mathbf{W}$. Furthermore, such datasets often include a substantial number of time-series observations per unit, as well as a large number of cross-sectional units, aligning well with the large N, T asymptotics considered in this paper.[3]

Our MGIV estimator is shown to be consistent and asymptotically normally distributed as $N, T \to \infty$. Importantly, MGIV is asymptotically unbiased. This property addresses the incidental parameters problem, and thereby standard inference procedures remain valid without the need for bias correction. Additionally, the estimator is linear and computationally efficient, making it particularly well-suited for empirical applications involving large datasets. Recently, our MGIV estimator was implemented in Stata by Kripfganz and Sarafidis (2025) via the `spxtivdfreg` command, providing researchers and practitioners with a readily available and user-friendly tool for empirical analysis.

The MGIV estimator developed in this paper extends the approach of Norkutė et al. (2021) to incorporate spatial interactions, addressing issues related to the identification of heterogeneous spatial parameters and the development of asymptotic theory for large N, T settings with spatially interdependent observations. Notably, our approach allows for more flexible expansion rates of N and T compared to Norkutė et al. (2021), permitting T to grow faster than N (but slower than N^2), proportionally to N, or slower than N. This broader flexibility in the relative growth of N and T enhances the applicability of the estimator in diverse settings where data dimensions vary significantly.

Recently, Chen et. al. (2022) proposed a related approach to our MGIV estimator, but their focus is limited to static panels without temporal dynamics.[4] Additionally, in their framework interactive effects are captured using cross-sectional averages à la Pesaran (2006), which relies on the so-called rank condition (e.g., non-zero mean factor loadings). In contrast, our method remains asymptotically valid even if the rank condition is violated.[5] Another related study is that of Aquaro et al. (2021), which assumes a purely idiosyncratic error structure without accounting for additive or interactive effects.

We illustrate the practical relevance of our method by estimating a regional spatial growth model, focusing on the magnitude of regional growth spillovers in Europe. Our approach explicitly

---

[3] A complementary strand of the economics literature examines settings where $\mathbf{W}$ is latent and estimated from the data using high-dimensional methods, as in De Paula (2017), Lam and Souza (2020), and De Paula et al. (2024). This approach offers appealing generality by avoiding the need for a priori specification of $\mathbf{W}$. However, to date, this literature also uniformly assumes slope parameter homogeneity and typically excludes temporal dynamics and aggregate shocks, both of which are explicitly addressed in the present paper. Extending our framework to accommodate the estimation of an unknown $\mathbf{W}$ remains an avenue for future research.

[4] Extending the approach of Chen et. al. (2022) to dynamic models is non-trivial, as it requires the construction of factor proxies from suitable cross-sectional averages of the dependent variable, the number of which grows with T. As noted on page 56 of their paper, the finite-sample bias arising from the correlation between cross-sectional averages and idiosyncratic errors is likely to be exacerbated in the presence of spatial lags. This issue becomes especially pronounced when the spatial weighting matrix remains relatively dense as the sample size increases, which often occurs when $\mathbf{W}$ is specified based on geographic proximity.

[5] Although De Vos et al. (2024) proposed a method for evaluating the rank condition for CCE estimators, it is currently applicable only to static panels and does not directly extend to settings with temporal dynamics.

accounts for heterogeneity across regions, reflecting differences in industrial structure, urbanization, and geographic characteristics. For instance, growth drivers in industrial Bavaria (Germany) likely differ from agricultural Thessaly (Greece), as do spillover effects between urban Greater London and rural Lapland (Finland). The results provide evidence of conditional convergence in regional growth dynamics. Spillovers play a dominant role, particularly for investment rates, where roughly four-fifths of the total impact on GDP per capita growth is attributable to neighboring regions, emphasizing the importance of inter-regional linkages. Human capital and R&D spillovers further reinforce the role of knowledge diffusion and innovation in fostering growth across regions.

Throughout, for an $m \times m$ matrix $\mathbf{A} = (a_{ij})_{1 \le i,j \le m}$, we denote its trace by $\text{tr}(\mathbf{A}) = \sum_{i=1}^{m} a_{ii}$. For an $m \times n$ matrix $\mathbf{B} = (b_{ij})_{1 \le i \le m, 1 \le j \le n}$, denote its column sum norm by $\|\mathbf{B}\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |b_{ij}|$, its Frobenius norm by $\|\mathbf{B}\| = \sqrt{\text{tr}(\mathbf{B}'\mathbf{B})}$, and its row sum norm by $\|\mathbf{B}\|_\infty = \max_{1 \le i \le m} \sum_{j=1}^{n} |b_{ij}|$. Also define $\mathbf{P_B} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$ and $\mathbf{M_B} = \mathbf{I}_m - \mathbf{P_B}$, where $\mathbf{I}_m$ is the $m \times m$ identity matrix. Denote by C a generic positive constant which need not be the same at each appearance, and $\delta_{NT}^2 = \min\{N, T\}$. We use $N, T \to \infty$ to denote that N and T pass to infinity jointly, and plim to denote the probability limit.

2 Model and Estimation Approach

We consider the following spatial dynamic panel data model with N cross-sectional units and T time periods:

$$
\begin{aligned}
y_{it} &= \psi_i \sum_{j=1}^{N} w_{ij} y_{jt} + \rho_i y_{i,t-1} + \mathbf{x}'_{it} \boldsymbol{\beta}_i + u_{it}; \\
u_{it} &= \boldsymbol{\lambda}'_i \mathbf{f}_t + \boldsymbol{\phi}'_i \boldsymbol{g}_t + \varepsilon_{it}, \quad i = 1, 2, \dots, N, \ t = 1, 2, \dots, T.
\end{aligned}
\tag{2.1}
$$

y_{it} denotes the outcome of interest for individual unit i at time t, while $\sum_{j=1}^{N} w_{ij} y_{jt}$ is the so-called "spatial-lag", a weighted sum of neighbor outcomes, where w_{ij} denotes the weight assigned to neighbor j in relation to i. These weights capture the connectedness structure among individuals and are specified within the $N \times N$ adjacency matrix $\mathbf{W} = [w_{ij}]$.[6] The vector $\mathbf{x}_{it} = (x_{1it}, \dots, x_{kit})'$ of dimension $k \times 1$, contains observed characteristics for individual i at time t. Additionally, in the composite error term $\mathbf{f}_t$ and $\boldsymbol{g}_t$ denote vectors of latent factors with dimensions $r_1 \times 1$ and $r_2 \times 1$, respectively, influencing y_{it}. These factors are associated with factor loadings $\boldsymbol{\lambda}_i$ and $\boldsymbol{\phi}_i$, which vary across individuals. The inclusion of latent factor components reflects the premise that individual agents inhabit a common economic environment and, as such, are subject to aggregate, economy-wide or "global" shocks that affect the entire population, albeit with different intensities. Examples of such shocks include technological disruptions, natural disasters, financial crises, pandemics, geopolitical conflicts, global market fluctuations and regulatory changes (e.g., Bai 2009 and Sarafidis and Wansbeek 2012). Note that the number of parameters associated with factors and factor loadings increases with T and N, respectively, as $N, T \to \infty$, leading to the presence of incidental parameters. Lastly, ε_{it} is a purely idiosyncratic error term.

---

[6]As discussed earlier, this paper aligns with standard practice in the spatial literature by assuming that $\mathbf{W}$ is fixed and known. Accordingly, the focus is placed on estimating the magnitude of heterogeneous spillover effects, conditional on a predetermined network structure.

There are three primary sources of endogeneity in model (2.1). First, $\sum_{j=1}^{N} w_{ij}y_{jt}$ is endogenous by construction. This term essentially represents the formal specification of an equilibrium outcome of a spatial interaction process, wherein the value of the dependent variable for one individual unit is simultaneously determined alongside that of its neighbours. This reciprocal interdependence underscores the networked nature of these relationships, as highlighted by Elhorst (2021).

Second, the lagged dependent variable, $y_{i,t-1}$ is also endogenous due to the presence of incidental parameters (e.g., Nickell 1981 and Phillips and Sul 2007).

Third, endogeneity also arises from the potential dependence of the covariates on the latent factors. To account for this, and consistent with the frameworks of Pesaran (2006) and Norkutė et al. (2021) among many others, we assume

$$\mathbf{x}_{it} = \boldsymbol{\Gamma}'_i \mathbf{f}_t + \mathbf{v}_{it}. \tag{2.2}$$

That is, $\mathbf{x}_{it}$ is influenced by $\mathbf{f}_t$, with the associated loadings represented by $\boldsymbol{\Gamma}_i$, a $k \times r_1$ matrix. Importantly, for the sake of generality, we allow the latent factors governing y_{it} and $\mathbf{x}_{it}$ to differ. This distinction justifies the inclusion of the additional term $\boldsymbol{\phi}'_i \boldsymbol{g}_t$ in model (2.1). Moreover, the loadings $\boldsymbol{\Gamma}_i$ are permitted to exhibit correlation with both $\boldsymbol{\lambda}_i$ and $\boldsymbol{\phi}_i$, further accommodating potential interdependencies that may arise in the model. Finally, $\mathbf{v}_{it}$ denotes the idiosyncratic error component for $\mathbf{x}_{it}$.

The individual-specific structural parameters reflect distinct mechanisms: ρ_i captures habit formation and adjustment costs, facilitating an important distinction between short- and long-run responses. $\boldsymbol{\beta}_i$ reflects the direct effects of an individual's own characteristics, and ψ_i encapsulates the influence of neighbours' outcomes, also known as spillover effects (e.g., Kelejian and Piras 2017 and Jing et al 2018).[7]

Stacking the equations in (2.1) over t yields

$$\mathbf{y}_i = \psi_i \sum_{j=1}^{N} w_{ij}\mathbf{y}_j + \rho_i \mathbf{y}_{i,-1} + \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{F}\boldsymbol{\lambda}_i + \mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i,$$
$$\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i, \quad i = 1, \ldots, N, \tag{2.3}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\mathbf{y}_j = (y_{j1}, \ldots, y_{jT})'$, $\mathbf{y}_{i,-1} = (y_{i0}, \ldots, y_{i,T-1})'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT})'$, $\mathbf{F} = (\mathbf{f}_1, \cdots, \mathbf{f}_T)'$, $\mathbf{G} = (\boldsymbol{g}_1, \cdots, \boldsymbol{g}_T)'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iT})'$ and $\mathbf{V}_i = (\mathbf{v}_{i1}, \ldots, \mathbf{v}_{iT})'$.

Define $\boldsymbol{\theta}_i = (\psi_i, \rho_i, \boldsymbol{\beta}'_i)'$, $\mathbf{C}_i = (\sum_{j=1}^{N} w_{ij}\mathbf{y}_j, \mathbf{y}_{i,-1}, \mathbf{X}_i)$, and $\mathbf{u}_i = \mathbf{F}\boldsymbol{\lambda}_i + \mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i$. Then, the first equation in (2.3) can be reformulated as

$$\mathbf{y}_i = \mathbf{C}_i \boldsymbol{\theta}_i + \mathbf{u}_i. \tag{2.4}$$

We use the method of Instrumental Variables to estimate $\boldsymbol{\theta}_i$. To this end, define the "defactoring" matrices that project out $\mathbf{F}$ and $\mathbf{F}_{-1} = (\mathbf{f}_0, \ldots, \mathbf{f}_{T-1})'$ as $\mathbf{M}_{\mathbf{F}} = \mathbf{I}_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$ and $\mathbf{M}_{\mathbf{F}_{-1}} = \mathbf{I}_T - \mathbf{F}_{-1}(\mathbf{F}'_{-1}\mathbf{F}_{-1})^{-1}\mathbf{F}'_{-1}$. Letting $\mathbf{X}_{i,-1} = (\mathbf{x}_{i0}, \ldots, \mathbf{x}_{i,T-1})'$, further define

$$\mathbf{Z}_i = \left(\sum_{j=1}^{N} w_{ij}\mathbf{M}_{\mathbf{F}}\mathbf{X}_j, \ \mathbf{M}_{\mathbf{F}}\mathbf{M}_{\mathbf{F}_{-1}}\mathbf{X}_{i,-1}, \ \mathbf{M}_{\mathbf{F}}\mathbf{X}_i \right), \tag{2.5}$$

---

[7]The main results of this paper naturally extend to models incorporating "contextual effects" (e.g., Manski 1993), also referred to as the spatial Durbin model (Elhorst 2014). This is further discussed in Remark 2.2.

whose elements are instruments for the elements of $\mathbf{C}_i$.[8] It's straightforward to see that, due to the de-factorisation, we have

$$\mathbf{Z}_i = \left(\sum_{j=1}^N w_{ij} \mathbf{M_F V}_j, \ \mathbf{M_F M_{F_{-1}} V}_{i,-1}, \ \mathbf{M_F V}_i \right) \quad \text{and} \quad \mathbf{Z}_i' \mathbf{u}_i = \mathbf{Z}_i' \big(\mathbf{G} \boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i \big), \qquad (2.6)$$

where $\mathbf{V}_{i,-1} = (\mathbf{v}_{i0}, \ldots, \mathbf{v}_{i,T-1})'$. Therefore, these instruments are exogenous.

Since $\mathbf{F}$ and $\mathbf{F}_{-1}$ are not observed, they are estimated using PCA on $\mathbf{X}$ and $\mathbf{X}_{-1}$. Assuming $T^{-1} \mathbf{F}' \mathbf{F} = \mathbf{I}_{r_1}$ and $T^{-1} \mathbf{F}'_{-1} \mathbf{F}_{-1} = \mathbf{I}_{r_1}$, the estimates $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{F}}_{-1}$ are obtained as $\sqrt{T}$ times the eigenvectors corresponding to the r_1 largest eigenvalues of the $T \times T$ matrices $(NT)^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{X}_i'$ and $(NT)^{-1} \sum_{i=1}^N \mathbf{X}_{i,-1} \mathbf{X}'_{i,-1}$, respectively.[9] Note that, since the factors are extracted from observed covariates, no estimation error arises in $\widehat{\mathbf{F}}$ that is associated with estimation of the slope coefficients. Furthermore, the factor loadings $\boldsymbol{\Gamma}_i$ can be estimated as $\widehat{\boldsymbol{\Gamma}}_i = T^{-1} \widehat{\mathbf{F}}' \mathbf{X}_i$.

Feasible instruments for $\boldsymbol{\theta}_i$ are constructed as

$$\widehat{\mathbf{Z}}_i = \left(\sum_{j=1}^N w_{ij} \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_j, \ \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{X}_{i,-1}, \ \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i \right), \qquad (2.7)$$

and the resulting IV estimator of $\boldsymbol{\theta}_i$ is given by

$$\widehat{\boldsymbol{\theta}}_i = \left(\widehat{\mathbf{A}}_i' \widehat{\mathbf{B}}_i^{-1} \widehat{\mathbf{A}}_i \right)^{-1} \widehat{\mathbf{A}}_i' \widehat{\mathbf{B}}_i^{-1} \widehat{\mathbf{c}}_{y,i}, \qquad (2.8)$$

where

$$\widehat{\mathbf{A}}_i = T^{-1} \widehat{\mathbf{Z}}_i' \mathbf{C}_i, \quad \widehat{\mathbf{B}}_i = T^{-1} \widehat{\mathbf{Z}}_i' \widehat{\mathbf{Z}}_i, \quad \widehat{\mathbf{c}}_{y,i} = T^{-1} \widehat{\mathbf{Z}}_i' \mathbf{y}_i. \qquad (2.9)$$

Remark 2.1 *Alternatively, a two-stage individual-specific IV estimator could be considered, which projects out the entire factor space in u_{it} using $\sqrt{T}$ times the eigenvectors corresponding to the $r_1 + r_2$ largest eigenvalues of the $T \times T$ matrix $(NT)^{-1} \sum_{i=1}^N \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i'$. However, due to the presence of heterogeneous slopes, $\widehat{\mathbf{u}}_i$ must be obtained from individual-specific time series IV regressions as $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{C}_i \widehat{\boldsymbol{\theta}}_i$. Since $\widehat{\boldsymbol{\theta}}_i$ is $\sqrt{T}$-consistent rather than $\sqrt{NT}$-consistent, the estimation of the full factor matrices $\mathbf{F}$ and $\mathbf{G}$ may become highly inefficient. Consequently, this estimator is not pursued further here. Note that the estimation of $\mathbf{F}$ for the IV estimator in Eq. (2.8) does not face this issue, as it can be directly estimated using the raw data $\{\mathbf{X}_i\}_{i=1}^N$.*

Remark 2.2 *Under a similar set of assumptions as in Section 3.1 below, the model in Eq. (2.3) can be straightforwardly extended to include either a spatial-time lag, $\sum_{j=1}^N w_{ij} \mathbf{y}_{j,-1}$, or spatial lags of the covariates, $\sum_{j=1}^N w_{ij} \mathbf{X}_j$, using instruments such as $\sum_{j=1}^N w_{ij} \mathbf{M_F M_{F_{-1}} X}_{j,-1}$. When both a spatial-time lag and spatial lags of covariates are included, additional instruments are required. These can be constructed, for example, as $\sum_{j=1}^N w_{ij} \mathbf{M_F M_{F_{-2}} X}_{j,-2}$ or $\sum_{j=1}^N w_{ij}^2 \mathbf{M_F M_{F_{-1}} X}_{j,-1}$, where the latter leverages information from the neighbors of an individual's neighbours. This setup is further explored in the empirical application in Section 5.*

---

[8]Loosely speaking, the term $\sum_{j=1}^N w_{ij} \mathbf{M_F X}_j$ instruments $\sum_{j=1}^N w_{ij} \mathbf{y}_j$, the term $\mathbf{M_F M_{F_{-1}} X}_{i,-1}$ instruments $\mathbf{y}_{i,-1}$, and $\mathbf{M_F X}_i$ instruments $\mathbf{X}_i$.

[9]For simplicity and without loss of generality, r_1 is treated as known. However, in practice it can be consistently estimated using established methods in the literature, such as the information criterion approach of Bai and Ng (2002) or the eigenvalue methods of Ahn and Horenstein (2013).

3 Assumptions and Asymptotic Theory

In this section, we examine the limiting properties of the individual-specific IV estimator, $\hat{\boldsymbol{\theta}}_i$, and the MGIV estimator. To facilitate the asymptotic analysis, we first introduce a set of assumptions.

3.1 Assumptions

Assumption A (idiosyncratic error in y) *The idiosyncratic error ε_{it} is independently and identically distributed across both i and t, and satisfies $\mathbb{E}(\varepsilon_{it}) = 0$, $\mathbb{E}(\varepsilon_{it}^2) = \sigma_\varepsilon^2 > 0$ and $\mathbb{E}|\varepsilon_{it}|^{8+\delta} \leq C$ for some $\delta > 0$.*

Assumption B (idiosyncratic error in x) *The idiosyncratic error $\mathbf{v}_{it}$ satisfies that*

1. *$\mathbf{v}_{it}$ is group-wise independent from ε_{it}, $\mathbb{E}(\mathbf{v}_{it}) = 0$, and $\mathbb{E}\|\mathbf{v}_{it}\|^{8+\delta} \leq C$;*

2. *Denoting $\boldsymbol{\Sigma}_{ij,st} = \mathbb{E}\left(\mathbf{v}_{is}\mathbf{v}_{jt}'\right)$, then $N^{-1}T^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{s=1}^T\sum_{t=1}^T \|\boldsymbol{\Sigma}_{ij,st}\| \leq C$; furthermore, there exist $\bar{\sigma}_{ij}$ and $\tilde{\sigma}_{st}$ such that $\|\boldsymbol{\Sigma}_{ij,st}\| \leq \bar{\sigma}_{ij}$ for all (s,t), $\|\boldsymbol{\Sigma}_{ij,st}\| \leq \tilde{\sigma}_{st}$ for all (i,j), and $T^{-1}\sum_{s=1}^T\sum_{t=1}^T \tilde{\sigma}_{st} \leq C$, $\sum_{j=1}^N \bar{\sigma}_{ij} \leq C$ for all i;*

3. *For all (s,t), $\mathbb{E}\left\|N^{-1/2}\sum_{i=1}^N(\mathbf{v}_{is}\mathbf{v}_{it}' - \boldsymbol{\Sigma}_{ii,st})\right\|^4 \leq C$, and for all (i,j),*
   $$\mathbb{E}\left\|T^{-1/2}\sum_{t=1}^T(\mathbf{v}_{it}\mathbf{v}_{jt}' - \boldsymbol{\Sigma}_{ij,tt})\right\|^4 \leq C;$$

4. *For all j and s, $\mathbb{E}\left\|N^{-1/2}T^{-1}\sum_{s_1=1}^T\sum_{i=1}^N\sum_{t=1}^T \mathbf{f}_{s_1}\mathbf{v}_{is_1}'\left[\mathbf{v}_{it}\mathbf{v}_{js}' - \mathbb{E}(\mathbf{v}_{it}\mathbf{v}_{js}')\right]\right\|^2 \leq C$ and $\mathbb{E}\left\|N^{-1/2}T^{-1/2}\sum_{i=1}^N\sum_{t=1}^T \boldsymbol{\Gamma}_i\left[\mathbf{v}_{it}\mathbf{v}_{js}' - \mathbb{E}(\mathbf{v}_{it}\mathbf{v}_{js}')\right]\right\|^2 \leq C$, and for all s,*
   $$\mathbb{E}\left\|N^{-1/2}T^{-1/2}\sum_{i=1}^N\sum_{t=1}^T \mathbf{h}_t[\mathbf{v}_{is}'\mathbf{v}_{it} - \mathbb{E}(\mathbf{v}_{is}'\mathbf{v}_{it})]\right\|^2 \leq C, \quad \text{where } \mathbf{h}_t = (\mathbf{f}_t', \mathbf{g}_t')';$$

5. *$N^{-1}T^{-2}\sum_{i=1}^N\sum_{j=1}^N\sum_{s_1=1}^T\sum_{s_2=1}^T\sum_{t_1=1}^T\sum_{t_2=1}^T |\text{cov}(\mathbf{v}_{is_1}'\mathbf{v}_{is_2}, \mathbf{v}_{jt_1}'\mathbf{v}_{jt_2})| \leq C$.*

Assumption C (factors) *The factors satisfy that*

1. *$\mathbf{f}_t$ and $\mathbf{g}_t$ are group-wise independent from $\mathbf{v}_{it}$ and ε_{it};*

2. *$\mathbb{E}\|\mathbf{f}_t\|^4 \leq C$ and $\mathbb{E}\|\mathbf{g}_t\|^4 \leq C$, and there exist non-random positive definite matrices $\boldsymbol{\Sigma}_F$ and $\boldsymbol{\Sigma}_G$ such that $\plim_{T\to\infty} T^{-1}\mathbf{F}'\mathbf{F} = \boldsymbol{\Sigma}_F$ and $\plim_{T\to\infty} T^{-1}\mathbf{G}'\mathbf{G} = \boldsymbol{\Sigma}_G$.*

Assumption D (loadings) *The factor loadings satisfy*

1. *$\boldsymbol{\Gamma}_i$, $\boldsymbol{\lambda}_i$ and $\boldsymbol{\phi}_i$ are group-wise independent from ε_{it}, $\mathbf{v}_{it}$, $\mathbf{f}_t$ and $\mathbf{g}_t$;*

2. *$\boldsymbol{\Gamma}_i \sim \text{i.i.d}(\mathbf{0}, \boldsymbol{\Sigma}_\Gamma)$, $\boldsymbol{\lambda}_i \sim \text{i.i.d}(\mathbf{0}, \boldsymbol{\Sigma}_\lambda)$, $\boldsymbol{\phi}_i \sim \text{i.i.d}(\mathbf{0}, \boldsymbol{\Sigma}_\phi)$, where $\boldsymbol{\Sigma}_\Gamma$ is positive definite and $\boldsymbol{\Sigma}_\lambda$ and $\boldsymbol{\Sigma}_\phi$ are positive semi-definite; furthermore, $\mathbb{E}\|\boldsymbol{\Gamma}_i\|^4 \leq C$, $\mathbb{E}\|\boldsymbol{\lambda}_i\|^4 \leq C$ and $\mathbb{E}\|\boldsymbol{\phi}_i\|^4 \leq C$.*

Assumption E (weight matrix) *The weight matrix, $\mathbf{W} = (w_{ij})_{N\times N} = (\mathbf{w}_1, \ldots, \mathbf{w}_N)'$, satisfies that*

1. *All diagonal elements of $\mathbf{W}$ are zero;*

2. $\mathbf{I}_N - \mathbf{\Psi W}$ *is invertible, where* $\mathbf{\Psi} = \mathrm{diag}(\psi_1, \dots, \psi_N)$;

3. $\mathbf{W}$ *has bounded row and column sum norms, i.e.,* $\|\mathbf{W}\|_\infty < C$ *and* $\|\mathbf{W}\|_1 < C$;

4. $\displaystyle\sup_{1 \le i \le N} |\psi_i| < \min\{1/\|\mathbf{W}\|_\infty, \ 1/\|\mathbf{W}\|_1\}$.

Assumption F (identification) *For each* i, *define* $\mathbf{A}_i = T^{-1}\mathbf{Z}'_i\mathbf{C}_i$ *and* $\mathbf{B}_i = T^{-1}\mathbf{Z}'_i\mathbf{Z}_i$. *It holds that*

1. *The matrices* $\mathbf{A}_i$ *and* $\mathbf{B}_i$ *have full column ranks for all* i *and* T;

2. $\mathbb{E}\|\mathbf{A}_i\|^{2+2\delta} \le C < \infty$ *and* $\mathbb{E}\|\mathbf{B}_i\|^{2+2\delta} \le C < \infty$ *for all* i *and* T;

3. *There exist two full-column-rank matrices,* $\mathbf{A}_{i,0}$ *and* $\mathbf{B}_{i,0}$, *such that* $\displaystyle\operatorname*{plim}_{T\to\infty}\mathbf{A}_i = \mathbf{A}_{i,0}$ *and* $\displaystyle\operatorname*{plim}_{T\to\infty}\mathbf{B}_i = \mathbf{B}_{i,0}$;

4. *Letting* $\mathbf{\Phi}_i = T^{-1}\mathbb{E}\big[\mathbf{Z}'_i\big(\mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i\big)\big(\mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i\big)'\mathbf{Z}_i\big]$, *then there exists a* $3k \times 3k$ *positive definite matrix* $\mathbf{\Phi}_{i,0}$ *such that* $\displaystyle\lim_{T\to\infty}\mathbf{\Phi}_i = \mathbf{\Phi}_{i,0}$.

Assumption G (random coefficients) *The heterogenous coefficients* $\boldsymbol{\theta}_i$ *are random and satisfy that*

1. $\boldsymbol{\theta}_i$ *follow the random-coefficient model* $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \mathbf{e}_i$, *where* $\boldsymbol{\theta} = (\psi, \rho, \boldsymbol{\beta}')'$ *and* $\mathbf{e}_i$ *is a random error that is independently and identically distributed with mean zero and variance* $\mathbf{\Sigma}_{\boldsymbol{\theta}}$; *furthermore,* $\mathbf{e}_i$ *is independent of* $\mathbf{\Gamma}_j$, $\boldsymbol{\lambda}_j$, $\boldsymbol{\phi}_j$, ε_{jt}, $\mathbf{v}_{jt}$, $\mathbf{f}_t$, *and* $\mathbf{g}_t$ *for all* i, j, t;

2. *Denoting* $\rho_w = \left(\dfrac{\displaystyle\sup_{1\le i\le N}|\rho_i|}{1 - \displaystyle\sup_{1\le i\le N}|\psi_i|\|\mathbf{W}'\|_1}\right)$, *then* $0 < \rho_w < 1$ *a.s. and* $\mathbb{E}\left[\left(\frac{1}{1-\rho_w}\right)\right]^2 \le C$.

Assumption A is in line with existing spatial literature, see e.g. Lee and Yu (2014); Cui et. al. (2023). Cross-sectional and time-series homoskedasticity is imposed to simplify the asymptotic analysis of the variance-covariance estimator in panels where both N and T are large. In contrast, Norkutė et al. (2021) allow for cross-sectional/time-series heteroskedasticity by leveraging the results in Hansen (2007). However, Hansen (2007) assumes independence across cross-sectional units, a condition that is violated in the present setup. Although we do not formally derive theoretical results under heteroskedasticity, the finite-sample performance of a robust variance-covariance estimator is thoroughly investigated in Section 4.

Assumption B ensures that $\mathbf{x}_{it}$ is strictly exogenous with respect to ε_{it}, as e.g., in Pesaran (2006) and Bai (2009). Therefore, the defactored regressors are valid instruments. Moreover, this assumption accommodates cross-sectional and time series heteroskedasticity, as well as autocorrelation in $\mathbf{v}_{it}$. Unlike ε_{it}, here it is crucial to explicitly allow for this broader structure since, conditional on $\mathbf{F}$, the dynamics in $\mathbf{X}_i$ are driven by $\mathbf{V}_i$. Additionally, in contrast to Norkutė et al. (2021), $\mathbf{v}_{it}$ is allowed to exhibit weak cross-sectional correlation, aligning with the assumption of weak dependence in the process of y.

Assumptions C and D align with standard conditions in the PCA literature; see, for example Bai (2003). Assumption C permits both interdependence between $\mathbf{f}_t$ and $\mathbf{g}_t$, as well as within-group correlations in each term. Similarly, Assumption D allows for non-zero correlations not

only between $\boldsymbol{\Gamma}_i$, $\boldsymbol{\varphi}_i$ and $\boldsymbol{\lambda}_i$, but also within each of these components. These assumptions are particularly relevant in settings where y_{it} and $\mathbf{x}_{it}$ may be jointly influenced by common shocks.

Assumption E is standard in the spatial literature, as outlined in Kelejian and Prucha (2001). Specifically, Assumption E1 serves as a normalization condition, ensuring that no individual is treated as its own neighbor. Assumptions E2-E3 ensure the absence of a dominant unit, i.e., a unit that becomes asymptotically correlated with all others. Such scenarios are instead accommodated by the inclusion of latent factors. Assumption E4 pertains to the space of the autoregressive and spatial parameters, and are discussed in detail by Kelejian and Prucha (2010, Sec. 2.2). Importantly, these assumptions are invariant to the ordering of the data, which can be arbitrary as long as Assumption E is satisfied. Additionally, it is noteworthy that the spatial weighting matrix $\mathbf{W}$ is not required to be row-normalized.

Assumption F ensures IV-based identification, as e.g. in Wooldridge (2002, Ch. 5). For instance, Assumption F1 implies that the covariates and instruments are not perfectly collinear, and that their correlation is sufficiently strong to avoid degeneracy.

Finally, Assumption G is commonly employed in the random coefficients literature (e.g., Pesaran 2006). Assumption G2 imposes upper bounds on the random coefficients, ensuring the stability of the process.

3.2 Asymptotic theory

The following theorem demonstrates the asymptotic properties of the individual-specific IV estimator $\widehat{\boldsymbol{\theta}}_i$.

Theorem 3.1 *If Assumptions A–E and F1-F2 hold, then the individual-specific IV estimator, $\widehat{\boldsymbol{\theta}}_i$, defined in (2.8) is consistent as $N, T \to \infty$. If, in addition, Assumptions F3-F4 hold and $T/N^2 \to 0$ as $N, T \to \infty$, then $\widehat{\boldsymbol{\theta}}_i$ has the following asymptotic distribution*

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i) \xrightarrow{d} N\left(\mathbf{0}, \ \boldsymbol{\Sigma}_{i,0}\right), \quad \text{as } N, T \to \infty,$$

where $\boldsymbol{\Sigma}_{i,0} = \left(\mathbf{A}'_{i,0}\mathbf{B}^{-1}_{i,0}\mathbf{A}_{i,0}\right)^{-1}\mathbf{A}'_{i,0}\mathbf{B}^{-1}_{i,0}\boldsymbol{\Phi}_{i,0}\mathbf{B}^{-1}_{i,0}\mathbf{A}_{i,0}\left(\mathbf{A}'_{i,0}\mathbf{B}^{-1}_{i,0}\mathbf{A}_{i,0}\right)^{-1}$.

Note that the large-N requirement is indispensable because the validity of the instruments used by $\widehat{\boldsymbol{\theta}}_i$ necessitates consistent estimation (up to rotation) of the $T \times r_1$ matrix $\mathbf{F}$.

The relative expansion rate of N and T employed in Theorem 3.1 is more general than that in Norkutė et al. (2021), which imposes $T/N \to c$, where $0 < c < \infty$. Specifically, the relative expansion rate adopted here allows for greater flexibility, permitting T to grow faster than N (but slower than N^2), proportionally to N, or slower than N.[10]

Assuming the cross-sectionally heterogeneous coefficients $\boldsymbol{\theta}_i$ follow the random-coefficient model, as in Assumption G, it is known that the dynamic pooled estimator of the population average $\boldsymbol{\theta} = \mathbb{E}(\boldsymbol{\theta}_i)$ will be inconsistent (see Robertson and Symons 1992, Pesaran and Smith 1995 and Chen et. al. 2022). In the present case, the same holds true even in the absence of temporal dynamics in model (2.1) because the spatial lag variable, $\sum_{j=1}^{N} w_{ij}y_{jt}$, is endogenous by construction. For this reason, we develop a Mean Group IV estimator of $\boldsymbol{\theta}$, which combines the individual-specific IV estimates and averages them to obtain consistent estimates of population-level effects.

---

[10]It is straightforward to verify that any case where $T = cN$ also satisfies $T = o\left(N^2\right)$ but the converse does not necessarily hold.

Specifically, once $\widehat{\boldsymbol{\theta}}_i$ in (2.8) are obtained, the Mean Group IV (MGIV) estimator of $\boldsymbol{\theta}$ is obtained as

$$\widehat{\boldsymbol{\theta}}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\theta}}_i. \tag{3.1}$$

Theorem 3.2 below establishes the asymptotic properties of $\widehat{\boldsymbol{\theta}}_{MG}$.

Theorem 3.2 *If Assumptions A–E, F1-2, and G hold, and $N/T^2 \to 0$ as $N, T \to \infty$, then the mean-group estimator, $\widehat{\boldsymbol{\theta}}_{MG}$, is consistent for the population mean $\boldsymbol{\theta}$. If, it further holds that $N/T^{6/5} \to 0$, then the mean-group estimator has the following asymptotic distribution*

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}) \xrightarrow{d} N\left(\mathbf{0}, \ \boldsymbol{\Sigma_\theta}\right), \quad \text{as } N, T \to \infty,$$

where $\boldsymbol{\Sigma_\theta}$ was defined in Assumption G.

Remark 3.1 *As shown above, $\widehat{\boldsymbol{\theta}}_{MG}$ is correctly centered around the true parameter $\boldsymbol{\theta}$ without necessitating any bias correction. Intuitively, this property arises because the estimation error of $\widehat{\mathbf{F}}$ depends on $\mathbf{V}_i$, which is mean-independent of $\boldsymbol{\varepsilon}_i$. Consequently, the defactored regressors employed as instruments are asymptotically uncorrelated with the error term of the model. See Cui et. al. (2022) for further discussion.*

Remark 3.2 *Thus far, we have assumed that $\mathbf{X}_i$ is exogenous with respect to the purely idiosyncratic error, $\boldsymbol{\varepsilon}_i$. Such an assumption will be violated if e.g., $\mathbf{X}_i$ is subject to reverse causality or measurement error. The present method can accommodate such sources of endogeneity as well, provided that valid external instruments are available.[11] To formalise this, let $\mathbf{X}_i = \left(\mathbf{X}_i^{(exog)}, \mathbf{X}_i^{(endog)}\right)$, where $\mathbf{X}_i^{(exog)}$ and $\mathbf{X}_i^{(endog)}$ refer to those sets of regressors that are strictly exogenous and endogenous, respectively, with respect to $\boldsymbol{\varepsilon}_i$. These matrices have dimensions $T \times k^{(exog)}$ and $T \times k^{(endog)}$, respectively. Furthermore, define $\mathbf{X}_i^+ = (\mathbf{X}_i^{(exog)}, \mathbf{X}_i^{(ext)})$, a $T \times k^+$ matrix with $k^+ = k^{(exog)} + k^{(ext)}$, where $\mathbf{X}_i^{(ext)}$ represents a matrix of external exogenous covariates. $\mathbf{X}_i^{(ext)}$ can still be correlated with the factor component, i.e., it may be subject to a similar data generating process as in Equation (2.1), so long as it remains strictly exogenous with respect to $\boldsymbol{\varepsilon}_i$. Define $\widehat{\mathbf{F}}_x^+$ as $\sqrt{T}$ times the eigenvectors corresponding to the r_x^+ largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^{N} \mathbf{X}_i^+ \left(\mathbf{X}_i^+\right)^\top / NT$. The associated defactoring matrices are defined analogously to those in earlier sections, with appropriate adjustments. Under this framework, the matrix of instruments retains the same structure as in (2.5), with $\mathbf{X}_i$ replaced by $\mathbf{X}_i^+$.*

4 Monte Carlo Experiments

We investigate the finite sample behaviour of the proposed approach by means of Monte Carlo experiments. We shall focus on the mean, bias, RMSE, empirical size and power of the t-test.

4.1 Design

We consider the following heterogeneous, spatial dynamic panel data model:

$$y_{it} = \alpha_i + \rho_i y_{i,t-1} + \psi_i \sum_{j=1}^{N} w_{ij} y_{jt} + \sum_{\ell=1}^{k} \beta_{\ell,i} x_{\ell it} + u_{it}; \quad u_{it} = \sum_{s=1}^{r_y} \varphi_{si} f_{st} + \varepsilon_{it}, \tag{4.1}$$

---

[11]For a discussion of this issue in the context of CCE estimation, see Harding and Lamarche (2011).

$i = 1, ..., N$, $t = -49, ..., T$, where

$$f_{st} = \varrho_{fs}f_{s,t-1} + (1 - \varrho_{fs}^2)^{1/2}\zeta_{s,t}, \tag{4.2}$$

with $\zeta_{st} \sim i.i.d.N(0,1)$ for $s = 1, ..., r_y$. We set $\varrho_{fs} = 0.5 \ \forall s$, $k = 2$ and $r_y = 3$.

The spatial weighting matrix, $\mathbf{W} = [w_{ij}]$ is an invertible rook matrix of circular form (Kapoor Kelejian and Prucha 2007), such that its ith row, $1 < i < N$, has non-zero entries in positions $i - 1$ and $i + 1$, whereas the non-zero entries in rows 1 and N are in positions $(1, 2)$, $(1, N)$, and $(N, 1)$, $(N, N - 1)$, respectively. This matrix is row normalized so that all of its nonzero elements are equal to $1/2$.

The idiosyncratic error, ε_{it}, is non-normal and heteroskedastic across both i and t, such that $\varepsilon_{it} = \varsigma_\varepsilon \sigma_{it}(\epsilon_{it} - 1)/\sqrt{2}$, $\epsilon_{it} \sim i.i.d.\chi_1^2$, with $\sigma_{it}^2 = \eta_i\phi_t$, $\eta_i \sim i.i.d.\chi_2^2/2$, and $\phi_t = t/T$ for $t = 0, 1, ..., T$ and unity otherwise.

The stochastic process for the covariates is given by

$$x_{\ell it} = \mu_{\ell i} + \sum_{s=1}^{r_x} \gamma_{\ell si}f_{st} + v_{\ell it}; \ i = 1, 2, ..., N; \quad t = -49, -48, ..., T, \tag{4.3}$$

for $\ell = 1, 2$. We set $r_x = 2$. Thus, the first two factors in u_{it}, f_{1t}, f_{2t}, also drive the DGP for $x_{\ell it}$, $\ell = 1, 2$. However, f_{3t} does not enter into the DGP of the covariates directly.[12]

The idiosyncratic errors in the covariates are serially correlated, such that

$$v_{\ell it} = \varrho_{v,\ell}v_{\ell i,t-1} + (1 - \varrho_{v,\ell}^2)^{1/2}\varpi_{\ell it}; \ \varpi_{\ell it} \sim i.i.d.N(0, \varsigma_v^2), \tag{4.4}$$

for $\ell = 1, 2$. We set $\varrho_{v,\ell} = \varrho_v = 0.5$ for all ℓ.

All individual-specific effects and factor loadings are generated as correlated and mean-zero random variables. In particular, the individual-specific effects are drawn as

$$\alpha_i \sim i.i.d.N(0, (1 - \varrho)^2); \quad \mu_{\ell i} = \varrho_{\mu,\ell}\alpha_i + (1 - \varrho_{\mu,\ell}^2)^{1/2}\omega_{\ell i}, \tag{4.5}$$

where $\omega_{\ell i} \sim i.i.d.N(0, (1 - \varrho)^2)$, for $\ell = 1, 2$. We set $\varrho_{\mu,\ell} = 0.5$ for $\ell = 1, 2$.

The factor loadings in u_{it} are generated as $\varphi_{si} \sim i.i.d.N(0,1)$ for $s = 1, ..., r_y(= 3)$, and the factor loadings in x_{1it} and x_{2it} are drawn as

$$\gamma_{1si} = \varrho_{\gamma,1s}\varphi_{3i} + (1 - \varrho_{\gamma,1s}^2)^{1/2}\xi_{1si}; \ \xi_{1si} \sim i.i.d.N(0,1); \tag{4.6}$$

$$\gamma_{2si} = \varrho_{\gamma,2s}\gamma_{si} + (1 - \varrho_{\gamma,2s}^2)^{1/2}\xi_{2si}; \ \xi_{2si} \sim i.i.d.N(0,1); \tag{4.7}$$

respectively, for $s = 1, ..., r_x = 2$. The process in Eq. (4.6) allows the factor loadings to f_{1t} and f_{2t} in x_{1it} to be correlated with the factor loadings corresponding to the factor that does not enter into the DGP of the covariates, i.e., f_{3t}. On the other hand, Eq. (4.7) ensures that the factor loadings to f_{1t} and f_{2t} in x_{2it} are correlated with the factor loadings corresponding to the same factors in u_{it}, f_{1t} and f_{2t}. We consider $\varrho_{\gamma,11} = \varrho_{\gamma,12} \in \{0, 0.5\}$, whilst $\varrho_{\gamma,21} = \varrho_{\gamma,22} = 0.5$.

It is straightforward to see that the average variance of ε_{it} depends only on ς_ε^2. Let π_u denote the proportion of the average variance of u_{it} that is due to ε_{it}. That is, we define $\pi_u := \varsigma_\varepsilon^2/(r_y + \varsigma_\varepsilon^2)$. Thus, for example, $\pi_u = 3/4$ means that the variance of the idiosyncratic error accounts for 75% of the total variance in u. In this case most of the variation in the total error is due to the idiosyncratic component and the factor structure has relatively minor significance.

Solving in terms of ς_ε^2 yields

$$\varsigma_\varepsilon^2 = \frac{\pi_u}{(1 - \pi_u)}r_y. \tag{4.8}$$

We set ς_ε^2 such that $\pi_u \in \{1/4, 3/4\}$.[13]

The slope coefficients are generated as $\rho_i = \rho + \eta_{\rho,i}$, $\psi_{1i} = \beta_1 + \eta_{\psi,i}$, $\beta_{1i} = \beta_1 + \eta_{\beta_1,i}$ and

---

[12]Observe that, using notation of earlier sections, $f_{3t} = g_t$ in Eq. (2.1).

[13]These values of π_u are motivated by the results in Sargent and Smith (1977), in which they find that two common factors explain 86% of the variation in unemployment rate and 26% of the variation in residential construction.

$\beta_{2i} = \beta_2 + \eta_{\beta_2,i}$. We set $\rho = 0.4$, $\psi = 0.25$, and $\beta_1 = 3$ and $\beta_2 = 1$, following Bai (2009). In addition, we specify $\eta_{\rho,i} \sim i.i.d. \ U[-c_\rho, +c_\rho]$, $\eta_{\psi,i} \sim i.i.d. \ U[-c_\psi, +c_\psi]$ and

$$\eta_{\beta_\ell,i} = \left[(2c_\rho)^2/12\right]^{1/2} \rho_\beta \xi_{\beta_\ell,i} + \left(1 - \rho_\beta^2\right)^{1/2} \eta_{\rho,i},$$

where $\xi_{\beta_\ell,i}$ is the standardised squared idiosyncratic error in $x_{\ell i t}$, computed as

$$\xi_{\beta_\ell,i} = \frac{\overline{v_{\ell i}^2} - \overline{v_\ell^2}}{\left[N^{-1} \sum_{i=1}^N \left(\overline{v_{\ell i}^2} - \overline{v_\ell^2}\right)^2\right]^{1/2}},$$

with $\overline{v_{\ell i}^2} = T^{-1} \sum_{t=1}^T v_{\ell i t}^2$, $\overline{v_\ell^2} = N^{-1} \sum_{i=1}^N \overline{v_{\ell i}^2}$, for $\ell = 1, 2$. We set $c_\rho = 0.2$, $c_\psi = 0.15$, $\rho_\beta = 0.4$ for $\ell = 1, 2$.

We define the signal-to-noise ratio (SNR) conditional on the factor structure, the individual-specific effects and the spatial lag, as follows:

$$SNR := \frac{\text{var}\left[(y_{it} - \varepsilon_{it})\,|\mathcal{L}\right]}{\overline{\text{var}}\,(\varepsilon_{it})} = \frac{\left(\frac{\beta_1^2 + \beta_2^2}{1 - \rho^2}\right)\varsigma_v^2 + \frac{\varsigma_\varepsilon^2}{1 - \rho^2} - \varsigma_\varepsilon^2}{\varsigma_\varepsilon^2}, \tag{4.9}$$

where $\mathcal{L}$ is the information set that contains the factor structure, the individual-specific effects and the spatial lag[14], whereas $\overline{\text{var}}\,(\varepsilon_{it})$ is the overall average of $E\left(\varepsilon_{it}^2\right)$ over i and t. Solving for ς_v^2 yields

$$\varsigma_v^2 = \varsigma_\varepsilon^2 \left[SNR - \frac{\rho^2}{1 - \rho^2}\right] \left(\frac{\beta_1^2 + \beta_2^2}{1 - \rho^2}\right)^{-1}. \tag{4.10}$$

We set $SNR = 4$, following Juodis and Sarafidis (2018) and Cui et. al. (2023).

We consider two sets of instruments, namely

$$\widehat{\mathbf{Z}}_i^1 = \left(\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i, \quad \mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{i,-1}, \quad \mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-2}}\mathbf{X}_{i,-2}, \quad \sum_{j=1}^N w_{ij}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_j\right), \tag{4.11}$$

and

$$\widehat{\mathbf{Z}}_i^2 = \left(\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i, \quad \mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{i,-1}, \quad \sum_{j=1}^N w_{ij}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_j \quad \sum_{j=1}^N w_{ij}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{j,-1}\right). \tag{4.12}$$

Both matrices are of dimension $T \times 4K$. The difference between the two matrices is that $\widehat{\mathbf{Z}}_i^1$ uses an extra (second) lag of $\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i$, which helps identification of the autoregressive parameter, whereas $\widehat{\mathbf{Z}}_i^2$ replaces the second lag of $\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i$ with the first lag of $\sum_{j=1}^N w_{ij}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_j$.

The mean group estimator of $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}}_{MG} = \frac{1}{N}\sum_{i=1}^N \hat{\boldsymbol{\theta}}_i, \tag{4.13}$$

where $\hat{\boldsymbol{\theta}}_i$ is given by

$$\hat{\boldsymbol{\theta}}_i = \left(\hat{\bar{\mathbf{A}}}_{i,T}'\hat{\bar{\mathbf{B}}}_{i,T}^{-1}\hat{\bar{\mathbf{A}}}_{i,T}\right)^{-1}\hat{\bar{\mathbf{A}}}_{i,T}'\hat{\bar{\mathbf{B}}}_{i,T}^{-1}\hat{\bar{\mathbf{g}}}_{i,T}, \tag{4.14}$$

where

$$\hat{\bar{\mathbf{A}}}_{i,T} = \frac{1}{T}\left(\hat{\mathbf{Z}}_i^\ell\right)'\mathbf{C}_i; \ \hat{\bar{\mathbf{B}}}_{i,T} = \frac{1}{T}\left(\hat{\mathbf{Z}}_i^\ell\right)'\hat{\mathbf{Z}}_i^\ell; \ \hat{\bar{\mathbf{g}}}_{i,T} = \frac{1}{T}\left(\hat{\mathbf{Z}}_i^\ell\right)'\mathbf{y}_i, \tag{4.15}$$

with $\ell \in \{1, 2\}$ and $\mathbf{C}_i = (\mathbf{y}_{i,-1}, \mathbf{X}_i, \mathbf{Yw}_i)$.

The variance-covariance matrix of the MG estimator is given by

$$\widehat{\boldsymbol{\Sigma}}_\eta = \frac{1}{N(N-1)}\sum_{i=1}^N \left(\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{MG}\right)\left(\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{MG}\right)', \tag{4.16}$$

As a benchmark estimator, we consider the pooled two-stage IV (2SIV) estimator developed by Cui et. al. (2023), which is designed for spatial dynamic panel data models with homogeneous

---

[14]The reason for conditioning on these variables is that they influence both the composite error of y_{it}, as well as the covariates.

parameters. This estimator is defined as follows:

$$\widetilde{\boldsymbol{\theta}} = (\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}})^{-1}\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{c}}_y \tag{4.17}$$

where

$$\widetilde{\mathbf{A}} = \frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{C}_i \,, \widetilde{\mathbf{B}} = \frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i \,, \widetilde{\mathbf{c}}_y = \frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{y}_i \,,$$

and $\mathbf{M}_{\widehat{\mathbf{H}}} = \mathbf{I} - \widehat{\mathbf{H}}(\widehat{\mathbf{H}}'\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}'$, with $\widehat{\mathbf{H}}$ defined as $\sqrt{T}$ times the eigenvectors corresponding to the r_y largest eigenvalues of the $T \times T$ matrix $(NT)^{-1}\sum_{i=1}^{N}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'$, where $\widehat{\mathbf{u}}_i$ is the residual of the first-stage homogeneous IV estimator.

In terms of the sample size, we consider three cases. Case I specifies $N = 100\tau$ and $T = 25\tau$ for $\tau = 1, 2, 4$. This implies that while N and T increase by multiples of 2, the ratio N over T remains equal to 4 in all circumstances. Case II specifies $T = 100\tau$ with $N = 25\tau$ for $\tau = 1, 2, 4$. Therefore, $N/T = 0.25$, as both N and T grow. Finally, Case III sets $N = T = 50\tau$, $\tau = 1, 2, 4$. These choices allow us to consider different combinations of (N, T) in relatively small and large sample sizes. Note that Case I implies that the rate at which $N, T \to \infty$ violates the conditions outlined in Theorem 3.2 for MGIV. Studying the performance of the estimator under these circumstances is valuable, as in many applications N can be significantly larger than T. This is exemplified by our empirical application discussed in the following section.

All results are obtained based on 2,000 replications, and all tests are conducted at the 5% significance level. For the power of the "t-test", we specify $H_0 : \rho = \rho^0 + 0.1$ (or $H_0 : \psi = \psi^0 + 0.1$, and $H_0 : \beta_\ell = \beta_\ell^0 + 0.1$ for $\ell = 1, 2$) against two sided alternatives, where $\rho^0, \psi^0, \beta_1^0, \beta_2^0$ denote the true parameter values.

4.2 Results

Tables 4.1–4.3 report results for $\rho = 0.4$, $\psi = 0.25$ and $\beta_2 = 1$ in terms of the Mean, RMSE, ARB (Absolute Relative Bias), as well as empirical size (nominal level is 5%) and size-corrected power, which is computed based on the 2.5% and 97.5% quantiles of the empirical distribution of the t-ratio under the null hypothesis.[15] ARB is defined as $ARB \equiv \left(|\widehat{\theta}_\ell - \theta_\ell|/\theta_\ell\right)100$, where θ_ℓ denotes the ℓth entry of $\boldsymbol{\theta} = \left(\psi, \rho, \boldsymbol{\beta}'\right)'$. "IVMG" denotes the Mean Group IV estimator of $\boldsymbol{\theta}$ as defined in (4.13) with the matrix of instruments given by $\widehat{\mathbf{Z}}_i^1$.[16] In each of the tables, Panel A corresponds to $\pi_u = 3/4$ and Panel B to $\pi_u = 1/4$.

In regards to Table 4.1, IVMG appears to have very little bias under all circumstances. Furthermore, its ARB values decrease steadily with larger sample sizes. Empirical size is very close to nominal one; even in cases where T is rather small, there are only mild size distortions. Moreover, size approaches 5% quickly as both N and T increase.

On the other hand, 2SIV appears to be severely biased. This is not surprising given that the pooled IV estimator is inconsistent under slope parameter heterogeneity. In particular, ARB increases with larger sample sizes and empirical size is heavily distorted throughout. These properties hold true regardless of the value of π_u.

The overidentifying restrictions test statistic (J test) associated with the optimal 2SIV estimator has satisfactory power to reject the null hypothesis of slope parameter homogeneity, especially in relatively larger samples. It is worth noting that power increases dramatically when more instru-

---

[15]Results for $\beta_1 = 3$ are qualitatively similar to those for $\beta_2 = 1$ and therefore they are reported in the Appendix.

[16]The results for the IVMG and 2SIV estimators based on $\widehat{\mathbf{Z}}_i^2$ are reported in the Appendix.

ments with respect to $\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i$ are used. For example, when $\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-3}}\mathbf{X}_{i,-3}$ is added to the existing set of instruments in Eq. (4.11), then for $\pi_u = 3/4$ and $T = 100\tau$, $N = 25\tau$, $\tau = 1, 2, 4$, power increases to 33.25%, 51.45% and 94.7%, respectively, compared to 12.0%, 23.5% and 75.1% reported in Table 4.1 under Panel A.[17]

The results in Tables 4.2–4.3 are qualitatively similar to those in Table 4.1, albeit the 2SIV pooled estimator exhibits smaller ARB for ψ and β_2 compared to ρ. IVMG typically outperforms 2SIV in terms of ARB and has superior size properties.

In summary, MGIV performs well in all circumstances. In particular, the finite-sample bias of the estimator is negligible in almost all cases examined, and inferences are credible even with relatively small samples. Notably, the performance of the estimator appears to be robust to a wide range of values for π_u, which captures the proportion of variation in the total error is due to the idiosyncratic error. Thus, considering also the computational simplicity of our methodology, our IVMG estimator presents an attractive estimation approach in heterogeneous spatial dynamic panel data models with interactive effects when both N and T are large.

---

[17]Detailed results for the case of this expanded set of instruments are available from the authors upon request.

Table 4.1: **Simulation results for** $\rho = 0.4$.
Panel A $(\pi_u = 3/4)$

	IVMG					**2SIV**					
					Case I: $N = 25\tau$, $T = 100\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.4	0.027	0.09	0.058	0.937	0.414	0.032	3.425	0.15	0.931	0.12
2	0.401	0.018	0.13	0.058	1	0.416	0.024	4.025	0.188	1	0.235
3	0.401	0.012	0.153	0.05	1	0.418	0.021	4.375	0.327	1	0.751
					Case II: $N = 100\tau$, $T = 25\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.396	0.02	1.026	0.067	0.994	0.414	0.025	3.4	0.131	0.997	0.096
2	0.402	0.011	0.47	0.069	1	0.417	0.02	4.175	0.33	1	0.232
3	0.402	0.007	0.377	0.058	1	0.418	0.019	4.425	0.74	1	0.78
					Case III: $N = 50\tau$, $T = 50\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.401	0.022	0.331	0.063	0.993	0.416	0.028	3.875	0.149	0.989	0.107
2	0.401	0.013	0.293	0.054	1	0.417	0.022	4.225	0.261	1	0.231
3	0.401	0.009	0.14	0.053	1	0.418	0.02	4.375	0.523	1	0.768

Panel B $(\pi_u = 1/4)$

	IV					**2SIV**					
					Case I: $N = 25\tau$, $T = 100\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.403	0.026	0.707	0.057	0.969	0.414	0.03	3.406	0.139	0.951	0.139
2	0.402	0.018	0.495	0.059	1	0.416	0.024	4.028	0.191	1	0.284
3	0.401	0.012	0.352	0.055	1	0.418	0.021	4.381	0.321	1	0.868
					Case II: $N = 100\tau$, $T = 25\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.404	0.02	0.790	0.089	0.998	0.414	0.024	3.582	0.139	0.999	0.107
2	0.402	0.013	0.684	0.073	1	0.417	0.02	4.161	0.341	1	0.284
3	0.401	0.008	0.198	0.067	1	0.418	0.019	4.431	0.747	1	0.882
					Case III: $N = 50\tau$, $T = 50\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.406	0.022	1.496	0.084	0.998	0.415	0.027	3.807	0.143	0.994	0.1
2	0.404	0.014	0.99	0.069	1	0.417	0.022	4.258	0.261	1	0.287
3	0.402	0.009	0.536	0.064	1	0.418	0.02	4.402	0.528	1	0.864

Table 4.2: **Simulation results for** $\psi = 0.25$.

Panel A $(\pi_u = 3/4)$

			IVMG					2SIV			
					Case I: $N = 25\tau$, $T = 100\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.252	0.028	0.784	0.071	0.935	0.248	0.027	0.68	0.097	0.937	0.12
2	0.25	0.016	0.092	0.057	1	0.248	0.016	0.64	0.069	1	0.235
3	0.25	0.01	0.106	0.053	1	0.248	0.01	0.72	0.067	1	0.751
					Case II: $N = 100\tau$, $T = 25\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.255	0.03	2.076	0.051	0.942	0.249	0.025	0.44	0.058	0.971	0.096
2	0.25	0.013	0.062	0.056	1	0.248	0.013	0.72	0.063	1	0.232
3	0.25	0.007	0.087	0.055	1	0.249	0.007	0.56	0.064	1	0.78
					Case III: $N = 50\tau$, $T = 50\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.255	0.027	1.847	0.052	0.963	0.248	0.025	0.72	0.064	0.973	0.107
2	0.25	0.014	0.125	0.044	1	0.248	0.014	0.68	0.059	1	0.231
3	0.25	0.008	0.001	0.054	1	0.248	0.008	0.64	0.062	1	0.768

Panel B $(\pi_u = 1/4)$

			IV					2SIV			
					Case I: $N = 25\tau$, $T = 100\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.251	0.028	0.461	0.066	0.924	0.249	0.027	0.608	0.085	0.949	0.139
2	0.25	0.016	0.06	0.058	1	0.249	0.016	0.594	0.071	1	0.284
3	0.25	0.01	0.187	0.051	1	0.248	0.01	0.735	0.066	1	0.868
					Case II: $N = 100\tau$, $T = 25\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.253	0.026	1.276	0.052	0.957	0.248	0.023	0.675	0.058	0.985	0.107
2	0.25	0.013	0.219	0.05	1	0.248	0.012	0.648	0.067	1	0.284
3	0.25	0.007	0.091	0.053	1	0.249	0.007	0.525	0.055	1	0.882
					Case III: $N = 50\tau$, $T = 50\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.253	0.026	1.109	0.052	0.96	0.248	0.024	0.664	0.073	0.968	0.1
2	0.25	0.014	0.01	0.054	1	0.248	0.013	0.716	0.063	1	0.287
3	0.25	0.008	0.077	0.047	1	0.248	0.008	0.637	0.06	1	0.864

Table 4.3: **Simulation results for** $\beta_2 = 1$.
Panel A $(\pi_u = 3/4)$

	IVMG					**2SIV**					
					Case I: $N = 25\tau$, $T = 100\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.999	0.063	0.152	0.058	0.358	1.007	0.063	0.7	0.094	0.397	0.12
2	0.999	0.031	0.096	0.049	0.878	1.005	0.032	0.5	0.068	0.904	0.235
3	1	0.017	0.002	0.059	1	1.006	0.019	0.6	0.077	1	0.751
					Case II: $N = 100\tau$, $T = 25\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	1.005	0.08	0.512	0.07	0.248	1.026	0.074	2.6	0.116	0.338	0.096
2	1	0.033	0.006	0.055	0.857	1.013	0.033	1.3	0.082	0.946	0.232
3	1	0.015	0.017	0.048	1	1.01	0.019	1	0.11	1	0.78
					Case III: $N = 50\tau$, $T = 50\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	1	0.067	0.054	0.058	0.336	1.014	0.064	1.4	0.093	0.41	0.107
2	0.999	0.03	0.084	0.054	0.885	1.009	0.032	0.9	0.081	0.92	0.231
3	0.999	0.015	0.076	0.045	1	1.007	0.017	0.7	0.061	1	0.768

Panel B $(\pi_u = 1/4)$

	IV					**2SIV**					
					Case I: $N = 25\tau$, $T = 100\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.997	0.063	0.355	0.061	0.341	1.006	0.06	0.627	0.089	0.423	0.139
2	0.998	0.032	0.189	0.049	0.86	1.005	0.032	0.541	0.068	0.907	0.284
3	1	0.018	0.04	0.053	1	1.006	0.019	0.588	0.073	1	0.868
					Case II: $N = 100\tau$, $T = 25\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.998	0.07	0.161	0.052	0.291	1.018	0.059	1.806	0.074	0.506	0.107
2	0.998	0.032	0.236	0.057	0.845	1.012	0.031	1.247	0.083	0.951	0.284
3	0.998	0.016	0.156	0.057	1	1.01	0.018	1.006	0.113	1	0.882
					Case III: $N = 50\tau$, $T = 50\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.997	0.066	0.279	0.061	0.349	1.013	0.06	1.252	0.092	0.441	0.1
2	0.997	0.031	0.256	0.052	0.875	1.009	0.031	0.937	0.083	0.925	0.287
3	0.998	0.016	0.196	0.043	1	1.007	0.017	0.665	0.06	1	0.864

5 Illustration: Empirical Spatial Growth Model

This section demonstrates our methodology by estimating a heterogeneous spatial growth model across EU regions over the period 2001-2018. The analysis focuses on uncovering the magnitude and influence of regional growth spillovers, a topic of extensive investigation in the literature.

5.1 Context and Dataset

Estimating the magnitude of growth spillovers has been a focal point in regional economic studies. Ertur and Coch (2007) laid the groundwork with a spatially augmented neoclassical growth model, integrating productivity spillovers driven by capital investment. Their framework is supported by

extensive empirical evidence on the role of knowledge and technological diffusion (e.g., Audretsch and Feldman 2004, Autant-Bernard and LeSage 2011).

Further studies have refined methods to measure spillovers at the sub-national level, often focusing on urban regions. For example, Botazzi and Peri (2003) and Rodriguez-Rose and Crescenzi (2008) analysed the spatial extent of innovation spillovers, while Funke and Niebuhr (2005) assessed inter-regional economic interactions using spatial econometrics. Despite consensus on the existence of spillovers, their estimated magnitudes remain contentious. For instance, studies such as Ramajo et al. (2008), Benos et al. (2015), Ozyurt and Dees (2018) and Elhorst et al. (2024) report mixed findings, often influenced by methodological factors, including the choice of the weighting matrix and the temporal coverage of the dataset.[18]

Building on this literature, this section revisits the empirical specification proposed by Ertur and Coch (2007) and Elhorst et al. (2024), which include a spatial-time-lag, as well as spatial lags of the covariates. A key point of departure from these studies lies in allowing for a heterogeneous model that accommodates region-specific parameters. Accounting for slope parameter heterogeneity is crucial because European regions are inherently different. For example, industrially advanced regions such as Bavaria, may exhibit distinct economic dynamics compared to less industrialized or predominantly agricultural regions like Thessaly in Greece. Similarly, highly urbanized areas like Greater London may experience fundamentally different growth drivers and constraints compared to sparsely populated regions like Lapland in Finland.

Another significant departure from Ertur and Coch (2007) and Elhorst et al. (2024) is the inclusion of interactive fixed effects to capture unobserved common shocks and their heterogeneous impacts across regions. This is particularly relevant as the dataset spans the period 2001-2018, encompassing major economic events such as the 2008-2009 financial crisis, the Great Recession, and the European debt crisis (2010-2015).[19] These events likely induced common shocks that affected multiple regions simultaneously, albeit to varying degrees. Purging the effects of these exogenous shocks is therefore essential for obtaining reliable estimates of growth parameters.

5.2 Model Specification and Estimation

The empirical growth model is specified as:

$$\Delta ln\left(y_{it}\right) = \sum_{\tau=0}^{1} \psi_{\tau,i} \sum_{j=1}^{N} w_{ij} \Delta ln\left(y_{jt-\tau}\right) + \theta_i ln\left(y_{it-1}\right) + \sum_{\ell=1}^{k} \beta_{\ell,i} x_{\ell it} + \sum_{\ell=1}^{k} \gamma_{\ell,i} \sum_{j=1}^{N} w_{ij} x_{\ell jt} + u_{it}, \quad (5.1)$$

where the dependent variable, $\Delta ln\left(y_{it}\right)$, represents the change in the natural logarithm of real GDP per capita (adjusted for purchasing power parity) in region i at time t, with $i = 1, \ldots, 266$ and $t = 2001, \ldots, 2018$.[20]

Thus, the growth rate of real GDP per capita in region i at time t is influenced by the initial level of GDP per capita in region i, the contemporaneous growth rate of GDP per capita in neighboring regions and its lag,[21] and a set of $k = 4$ explanatory variables along with their spatial counterparts.

---

[18]Differences in empirical methods also contribute to varying results; while growth regressions typically suggest convergence, tests of distribution dynamics point toward divergence (see e.g., Di Vaio et al. 2014 for a high-level discussion).

[19]The dataset is publicly available in Paul Elhorst's website: https://spatial-panels.com/.

[20]The analysis focuses on EU NUTS-2 regions, which provide a harmonized dataset on GDP and other macroeconomic indicators, ensuring cross-regional comparability and offering a higher resolution than national aggregates. While data at the NUTS-3 level exist, they are often incomplete or inconsistent, making NUTS-2 the most suitable choice.

[21]The lagged own growth rate was found to be statistically insignificant throughout and was therefore excluded

These covariates are defined as follows:

$x_{1it} \equiv ln\,(inv_{it})$, where inv denotes the investment rate (investment as a share of GDP). Investment drives capital accumulation and serves as a proxy for the resources allocated to future production capacity;

$x_{2it} \equiv ln\,(n_{it}) = ln\,(p_{it} + g + q)$, where p_{it} represents the population growth rate for region i at time t, while g and q denote the rates of technological progress and capital depreciation, respectively.[22] Population growth influences capital dilution and, indirectly, the region's convergence speed toward its steady state;

$x_{3it} \equiv ln\,(educ_{it})$, where $educ$ denotes the share of the working-age population with tertiary education. This variable captures human capital, potentially a key driver of productivity and innovation in both neoclassical and endogenous growth frameworks; and

$x_{4it} \equiv ln\,(sci\&tech_{it})$, where $sci\&tech$ represents the share of employment in science and technology. This variable serves as a proxy for regional engagement in innovation-driven activities, reflecting the intensity of knowledge-based economic development. The inclusion of x_{3it} and x_{4it} aligns with endogenous growth theory, which emphasizes the role of human capital and technological innovation as engines of long-term growth. In contrast, Solow's neoclassical framework considers these factors as exogenous.

The error term is composite and given by:

$$u_{it} = \alpha_i + \boldsymbol{\lambda}'_i \mathbf{f}_t + \boldsymbol{\varphi}'_i \mathbf{g}_t + \varepsilon_{i,t}. \tag{5.2}$$

Conditional convergence occurs when $\theta_i < 0$, implying that regions with lower (higher) initial GDP per capita tend to experience faster (slower) growth, holding all else constant. In contrast, conditional divergence arises when $\theta_i > 0$, indicating that initial disparities in GDP per capita widen over time. A special case occurs when $\theta_i = 0$, which Yu et al. (2012) refer to as "spatial cointegration". This scenario suggests that while GDP per capita growth rates may fluctuate across economies over the business cycle, regional economies ultimately remain on distinct growth trajectories throughout the sample period.

The spatial weights matrix is specified as:

$$w_{ij} = \frac{exp\,(-\zeta d_{ij})}{\sum_{j=1}^{N} exp\,(-\zeta d_{ij})}, \tag{5.3}$$

where d_{ij} represents the great-circle distance between regions i and j.[23] Finally, ζ is the distance decay parameter. Following Ertur and Koch (2007), ζ is set to 0.02, which prioritises interactions among proximate regions. Alternative specifications for $\mathbf{W}$, such as inverse squared distances, are explored in Section 5.4.

The set of instruments is given by

$$\widehat{\mathbf{Z}}_i^1 = \left(\mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i, \ \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{X}_{i,-1}, \ \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{M}_{\widehat{\mathbf{F}}_{-2}} \mathbf{X}_{i,-2}, \ \mathbf{M}_{\widehat{\mathbf{F}}} \sum_{j=1}^{N} w_{ij} \mathbf{X}_j, \ \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \sum_{j=1}^{N} w_{ij} \mathbf{X}_{j,-1} \right), \tag{5.4}$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT})'$ with $\mathbf{x}_{it} = (x_{1it}, \ldots, x_{4it})'$. Thus, a total of 20 instruments is utilized.

---

from the model.

[22] In line with the standard assumptions of the neoclassical growth framework, the rates of technological progress (g) and depreciation (q) are not indexed, as they are assumed to remain uniform across all regions and time periods. Following Islam (1995), we set g + q = 0.05.

[23] This is measured in kilometers and constructed based on the latitude and longitude coordinates of the regions' centroids

5.3 Results

Table 5.1 below provides results for the model specified in Eq. (5.1), alongside several nested models that impose progressively stronger restrictions. Each column corresponds to a distinct specification, revealing the effects of these restrictions on the parameter estimates and their interpretations.

Column [6] represents the most general specification, allowing for spatial spillovers, interactive fixed effects, and slope-parameter heterogeneity. On the other hand, Column [1] represents the most restricted model. This specification imposes $u_{it} = \alpha_i + \tau_t + \varepsilon_{it}$, $\psi_{0,i} = \psi_{1,i} = \gamma_{\ell,i} = 0$ for all i and ℓ, $\theta_i = \theta$ and $\beta_{\ell,i} = \beta_\ell$ for all ℓ in Eq. (5.1).[24] In essence, this formulation rules out spatial spillover effects, enforces slope-parameter homogeneity, and assumes that the error structure adheres to a two-way error components model. The columns in-between present intermediate cases. Column [2] extends Column [1] by incorporating latent common factors in the error term, relaxing the restriction of an additive two-way error components structure. Column [3] relaxes slope-parameter homogeneity while maintaining no spatial spillovers or interactive effects. Column [4] introduces spatial spillovers, while retaining the restrictions $u_{it} = \alpha_i + \tau_t + \varepsilon_{it}$, $\theta_i = \theta$, $\psi_{\tau,i} = \psi_\tau$ for all τ and $\beta_{\ell,i} = \beta_\ell$ for all ℓ in Eq. (5.1). Finally, Column [5] builds on Column [4] by allowing for interactive fixed effects, although it still rules out slope-parameter heterogeneity.[25]

We start with the most restrictive specification first, Column [1]. The coefficient on $ln\,(y_{t-1})$ is negative and highly significant, supporting the hypothesis of conditional convergence among regions. This finding aligns with the Solow growth model, which predicts that regions with higher initial GDP per capita grow more slowly as they approach their steady-state income levels. The coefficient on $ln\,(inv_t)$ is positive and significant, indicating that higher investment rates contribute positively to regional growth. Conversely, the coefficient on $ln\,(n_t)$ is negative and significant, consistent with neoclassical growth theory's prediction that higher population growth reduces per capita output by diluting capital per worker. The coefficients on educational attainment and share of employment in science and technology are both close to zero and statistically insignificant. The J test strongly rejects the null hypothesis of valid instruments, indicating potential model misspecification.

Allowing for interactive effects in the error process, as in Column [2], or slope-parameter heterogeneity, as in Column [3], leads to a more negative coefficient on $ln\,(y_{t-1})$, which implies faster conditional convergence. This suggests that accounting for unobserved global shocks or shared regional trends and slope heterogeneity across regions is important.

Turning into the results incorporating spatial spillovers: The coefficients on $\mathbf{W}\Delta ln\,(y_t)$ and $\mathbf{W}\Delta ln\,(y_{t-1})$ in Columns [4] and [5] imply the presence of a spatial unit root. Taken literally, this suggests that shocks to neighboring regions' growth rates persist indefinitely and propagate throughout the spatial network without decay, which is counterintuitive. At the same time, the coefficient of $ln\,(y_{t-1})$ is close to zero, indicating the absence of conditional convergence under these specifications. This outcome, combined with the rejection of the null hypothesis of valid instruments using the J test, points to potential model mis-specification.

The most general specification presented in Column [6] yields a nuanced view of growth dynamics. Both $\mathbf{W}\Delta ln\,(y_t)$ and $\mathbf{W}\Delta ln\,(y_{t-1})$ remain positive and highly significant, with their combined

---

[24]The former restriction on the error term is equivalent to imposing one factor only in Eq. (5.2) with the corresponding loadings being constant across i.

[25]The results in Columns [1]-[3] are obtained using the Stata command `xtivdfreg`, developed by Kripfganz and Sarafidis (2021), while those in Columns [4]-[6] are based on the Stata wrapper command `spxtivdfreg`, developed by Kripfganz and Sarafidis (2025).

magnitude summing to 0.82, which is strictly less than one. This result indicates that growth spillovers, while highly influential, dissipate over larger regional distances and do not propagate indefinitely through the spatial network. The coefficient on the investment rate of a region's neighbours is positive and statistically significant, highlighting substantial spillovers of regional investment activity. Notably, a similar pattern emerges for the coefficient on tertiary educational attainment, suggesting that higher educational levels in neighboring regions contribute positively to a region's economic growth through knowledge diffusion and human capital externalities.

Table 5.1: **Results for empirical growth model.**

	[1]	[2]	[3]	[4]	[5]	[6]
$ln\,(y_{t-1})$	-0.137***	-0.267***	-0.216***	-0.011	-0.010	-0.473***
	(0.032)	(0.034)	(0.017)	(0.034)	(0.042)	(0.028)
$ln\,(inv_t)$	0.035***	0.013	0.016**	0.003	-0.004	0.027**
	(0.006)	(0.008)	(0.007)	(0.004)	(0.005)	(0.013)
$ln\,(n_t)$	-0.629***	-0.636***	-0.271	-0.617***	-0.679***	-0.582**
	(0.153)	(0.122)	(0.249)	(0.125)	(0.123)	(0.281)
$ln\,(educ_t)$	-0.006	0.013	0.002	-0.004	0.002	0.007
	(0.007)	(0.017)	(0.015)	(0.006)	(0.008)	(0.026)
$ln\,(sci\&tech_t)$	-0.006	-0.013	-0.034***	0.008	-0.009	-0.003
	(0.008)	(0.011)	(0.011)	(0.006)	(0.008)	(0.016)
W						
$\Delta ln\,(y_t)$				0.896***	0.997***	0.628***
				(0.145)	(0.048)	(0.033)
$\Delta ln\,(y_{t-1})$				0.075	0.039	0.192***
				(0.081)	(0.050)	(0.052)
$ln\,(inv_t)$				0.001	0.003	0.071***
				(0.012)	(0.011)	(0.021)
$ln\,(n_t)$				0.334*	0.568***	0.643
				(0.198)	(0.135)	(0.549)
$ln\,(educ_t)$				-0.000	0.007	0.150***
				(0.011)	(0.017)	(0.035)
$ln\,(sci\&tech_t)$				-0.017*	0.004	0.022
				(0.009)	(0.011)	(0.050)
N_g	266	266	266	266	266	266
χ^2_J	29.949	39.201		12.966	16.002	
p_J	0.000	0.000		0.024	0.067	
fact1	0	1	0	0	1	1
fact2	0	2	0	0	1	

Notes: Standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. [1] imposes $u_{it} = \alpha_i + \mu_t + \varepsilon_{it}$, $\psi_{\tau,i} = \gamma_{\ell,i} = 0$ for all i, τ and ℓ, $\theta_i = \theta$ and $\beta_{\ell,i} = \beta_\ell$ for all ℓ in Eq. (5.1). [2] is as in [1] except it relaxes the restriction $u_{it} = \alpha_i + \mu_t + \varepsilon_{it}$ in Eq. (5.1). [3] is as in [1] except it relaxes the slope-parameter homogeneity restriction. [4] imposes $u_{it} = \alpha_i + \mu_t + \varepsilon_{it}$, $\theta_i = \theta$, $\psi_{\tau,i} = \psi_\tau$ for all τ, and $\beta_{\ell,i} = \beta_\ell$ for all ℓ in Eq. (5.1). Column [5] is as in [4] except it further relaxes the restriction $u_{it} = \alpha_i + \varepsilon_{it}$. Finally, [6] relaxes the slope-parameter homogeneity restriction and presents the most general specification.

Figure 1 below depicts a map of European regions, shaded to reflect the magnitude and sign of θ_i, the coefficient of the initial level of GDP per capita. Regions shaded in dark blue indicate estimates close to -1, while those in dark red represent estimates closer to 1. Regions shaded in pale blue or pale red reflect estimates near 0, which are not statistically significant. As shown, most wealthy regions in are dark blue, indicating that higher initial GDP per capita tends to result

in slower growth, likely due to diminishing returns to capital. Conversely, dark blue regions in southern Europe, which are relatively poorer, suggest faster growth driven by low starting GDP levels and higher marginal returns to investment. In contrast, regions such as in central Spain and Portugal, in Wales and in northern Romania, exhibit θ_i values close to zero, indicating minimal evidence of conditional regional convergence. Note also that there is considerable heterogeneity in the magnitude of θ_i among blue-shaded regions, reflecting varying speeds of convergence.
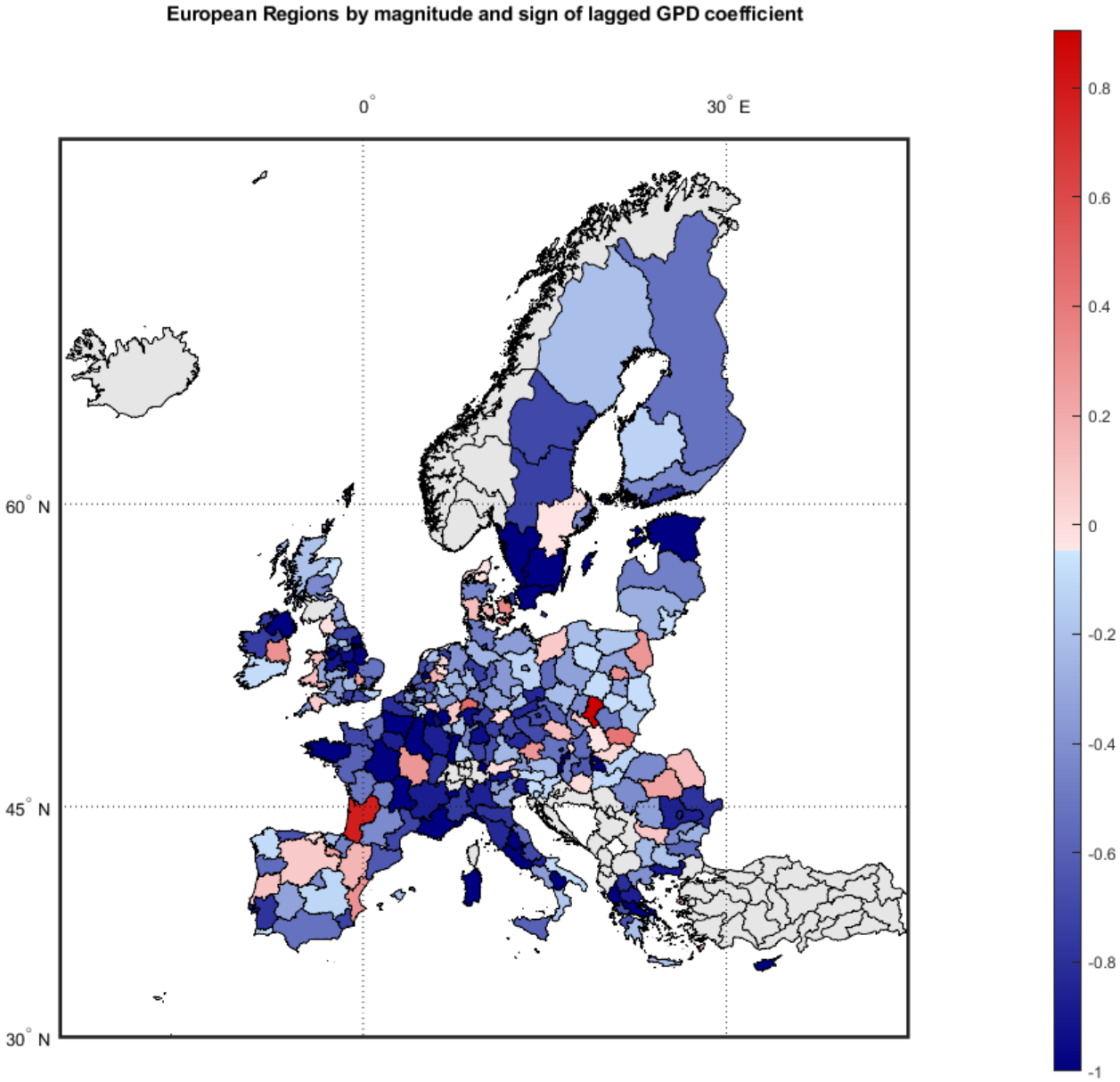

Figure 1: Convergence Patterns across Regional Europe

Table 5.2 below presents Mean Group direct, indirect (spillover), and total effects of $ln(inv_t)$, $ln(n_t)$, $ln(educ_t)$ and $ln(sci\&tech_t)$.

Specifically, by stacking the N observations for each t the model becomes:

$$\Delta\mathbf{y}_{(t)} = \mathbf{\Psi}_0\mathbf{W}\Delta\mathbf{y}_{(t)} + \mathbf{\Psi}_1\mathbf{W}\Delta\mathbf{y}_{(t-1)} + \mathbf{\Theta}\mathbf{y}_{(t-1)} + \sum_{\ell=1}^{k}\mathbf{B}_\ell\mathbf{x}_{\ell(t)} + \sum_{\ell=1}^{k}\mathbf{\Gamma}_\ell\mathbf{W}N\mathbf{x}_{\ell(t)} + \mathbf{u}_{(t)}, \qquad (5.5)$$

where $\Delta\mathbf{y}_{(t)} = (\Delta y_{1t},\ldots,\Delta y_{Nt})'$ is of dimension $N \times 1$, and similarly for the remaining variables, while $\mathbf{\Theta} \equiv diag\,(\theta_1,\ldots,\theta_N)$, $\mathbf{\Psi}_0 \equiv diag\,(\psi_{0,1},\ldots,\psi_{0,N})$ etc.

Solving the model for $\Delta\mathbf{y}_{(t)}$ yields:

$$\Delta\mathbf{y}_{(t)} = [\mathbf{I}_N - \mathbf{\Psi}_0\mathbf{W} - \mathbf{\Psi}_1\mathbf{W}L]^{-1}\left(\mathbf{\Theta}\mathbf{y}_{(t-1)} + \sum_{\ell=1}^{k}\mathbf{B}_\ell\mathbf{x}_{\ell(t)} + \sum_{\ell=1}^{k}\mathbf{\Gamma}_\ell\mathbf{W}\mathbf{x}_{\ell(t)}\right) + \mathbf{u}_{(t)}. \qquad (5.6)$$

where L denotes the lag operator The matrix of partial derivatives of the expected value of $\Delta\mathbf{y}_{(t)}$ with respect to the ℓth covariate is given by:

$$\left[\frac{\partial E\,(\mathbf{y})}{\partial x_{\ell 1}} \cdots \frac{\partial E\,(\mathbf{y})}{\partial x_{\ell N}}\right] = [\mathbf{I}_N - \mathbf{\Psi}_0\mathbf{W} - \mathbf{\Psi}_1\mathbf{W}L]^{-1}(\mathbf{B}_\ell\mathbf{I}_N + \mathbf{\Gamma}_\ell\mathbf{W}_N). \qquad (5.7)$$

Building on the framework established by LeSage and Pace (2009) and Debarsy et al. (2012), the Mean Group direct effect of a unit change in $\mathbf{x}_{\ell(t)}$ on $\Delta\mathbf{y}_{(t)}$ is calculated as the average of the diagonal elements of the matrix in Eq. (5.7). The Mean Group indirect effect, on the other hand, is defined as the average of the off-diagonal column sums, capturing the spillover effects of changes in one region's covariate on other regions. The total effect is obtained as the sum of the direct and indirect effects. To analyze heterogeneity, the direct effects for a specific region i correspond to the (i,i)th diagonal entry of the matrix in Eq. (5.7), while the indirect effects for region i are computed as the sum of the off-diagonal elements in the ith row of the same matrix.

The direct effect of $ln\,(inv_t)$ is small and statistically significant at the 10% level when employing a one-tailed test, indicating that a region's own investment rate has a modest but discernible impact on its economic growth. On the other hand, the indirect effect, capturing the influence of neighboring regions' investment rates, is much larger and also highly significant. This finding is consistent with the conclusions of Elhorst et al. (2024), who document that while the local impact of the investment rate tends to be small, its spillover effects are substantial. Overall, a 1% increase in the investment rate is associated with a 0.265% increase in the regional GDP per capita growth rate, with a remarkable 83.4% of this total effect attributable to spillovers from neighboring regions. This substantial contribution of spillovers highlights the pivotal role of cross-regional investment linkages, such as shared infrastructure projects, inter-regional supply chains, and trade networks, in driving economic growth.

For $ln\,(n_t)$ the direct and spillover effects exhibit a striking contrast in both magnitude and direction. Specifically, a 1% increase in a region's population growth rate corresponds to a 0.531% decrease in its GDP per capita growth rate, reflecting the neoclassical growth theory's prediction of capital dilution effects. Conversely, the same increase in neighboring regions' population growth rates leads to a 0.696% *increase* in the focal region's growth rate, likely due to enhanced labor market integration, demand spillovers, and knowledge diffusion across regional boundaries. The net result of these opposing forces is a near-cancellation of effects, which highlights an intricate balance of demographic dynamics, where the growth trajectory of a region is shaped not only by its internal factors but also by its interactions within the broader spatial network.

Turning to $ln\,(educ_t)$ and $ln\,(sci\&tech_t)$, the total effects for both variables are positive and

statistically significant at the 10% level under a one-tailed test.[26] Notably, these total effects are predominantly driven by their spillover components. For instance, the indirect effect of $ln\left(educ_t\right)$ is substantial, suggesting that a region benefits significantly from tertiary educational attainment of its neighbors. This outcome aligns with endogenous growth theories that emphasize the role of human capital spillovers in fostering innovation and productivity gains across regions. Similarly, the indirect effect of $ln\left(sci\&tech_t\right)$, though smaller in magnitude, indicates that regions appear to derive considerable benefits from innovation and knowledge diffusion originating in adjacent areas, reinforcing the argument for fostering cross-regional collaborations in research and development.

Table 5.2: **Direct, indirect and total effects.**

	Impact	Std. Err.	z	P>|z|	[95% Conf. Interval]
Direct					
$ln\left(inv_t\right)$	0.044	0.032	1.344	0.179	[-0.020, 0.107]
$ln\left(n_t\right)$	-0.531	0.313	-1.693	0.090	[-1.144, 0.083]
$ln\left(educ_t\right)$	0.036	0.314	0.114	0.909	[-0.580, 0.651]
$ln\left(sci\&tech_t\right)$	0.001	0.035	0.016	0.987	[-0.068, 0.069]
Indirect					
$ln\left(inv_t\right)$	0.220	0.032	6.779	0.000	[0.156, 0.284]
$ln\left(n_t\right)$	0.696	0.313	2.223	0.026	[0.082, 1.309]
$ln\left(educ_t\right)$	0.386	0.314	1.230	0.218	[-0.229, 1.002]
$ln\left(sci\&tech_t\right)$	0.050	0.035	1.432	0.152	[-0.018, 0.118]
Total					
$ln\left(inv_t\right)$	0.264	0.032	8.120	0.000	[0.200, 0.327]
$ln\left(n_t\right)$	0.165	0.313	0.527	0.598	[-0.448, 0.779]
$ln\left(educ_t\right)$	0.422	0.314	1.344	0.179	[-0.193, 1.038]
$ln\left(sci\&tech_t\right)$	0.050	0.035	1.446	0.148	[-0.018, 0.118]

Figure 2 below displays a map of all European regions, colored to represent the proportion of the total effect of the investment rate, $ln\left(inv_t\right)$, on growth attributable to the direct effect. Regions shaded in dark red indicate estimates near zero, suggesting that a substantial portion of the total effect of investment arises from spillover effects from neighboring regions. In contrast, regions shaded in dark blue correspond to estimates near unity, indicating that the majority of the total effect of investment on growth is driven by the region's own (direct) effect. As illustrated, many regions across Europe experience substantial spillover effects relative to the total. Notable exceptions include islands in the Mediterranean, such as Crete, Cyprus and Sicily, where the direct effect of a region's own investment is significantly more pronounced. Similar patterns are observed in northern Scotland, Northern Ireland, Thuringia in Germany (often referred to as "the green heart of Germany" due to its broad, dense forests), and a few capital regions, such as Lisbon in Portugal and Rome in Italy.

Taken together, these findings underscore the critical importance of incorporating both direct and spillover effects in growth models. Ignoring the spatial dimension of regional dynamics risks underestimating the true impact of key growth determinants and overlooking the complex interdependencies that shape economic growth. Moreover, the results highlight the necessity of accounting for slope parameter heterogeneity, as regions often exhibit distinct economic structures and re-

---

[26]Note here that the direct effect of $ln\left(sci\&tech_t\right)$ is (small and) positive even if the coefficient on $ln\left(sci\&tech_t\right)$ is (small and) negative. This phenomenon arises because a negative coefficient on a spatial lag of a covariate does not necessarily imply that the corresponding spillover effect is also negative. This issue is explained in LeSage and Pace (2009) (page 71).

Investment Rate: Proportion of Direct over Total Effect for European Regions
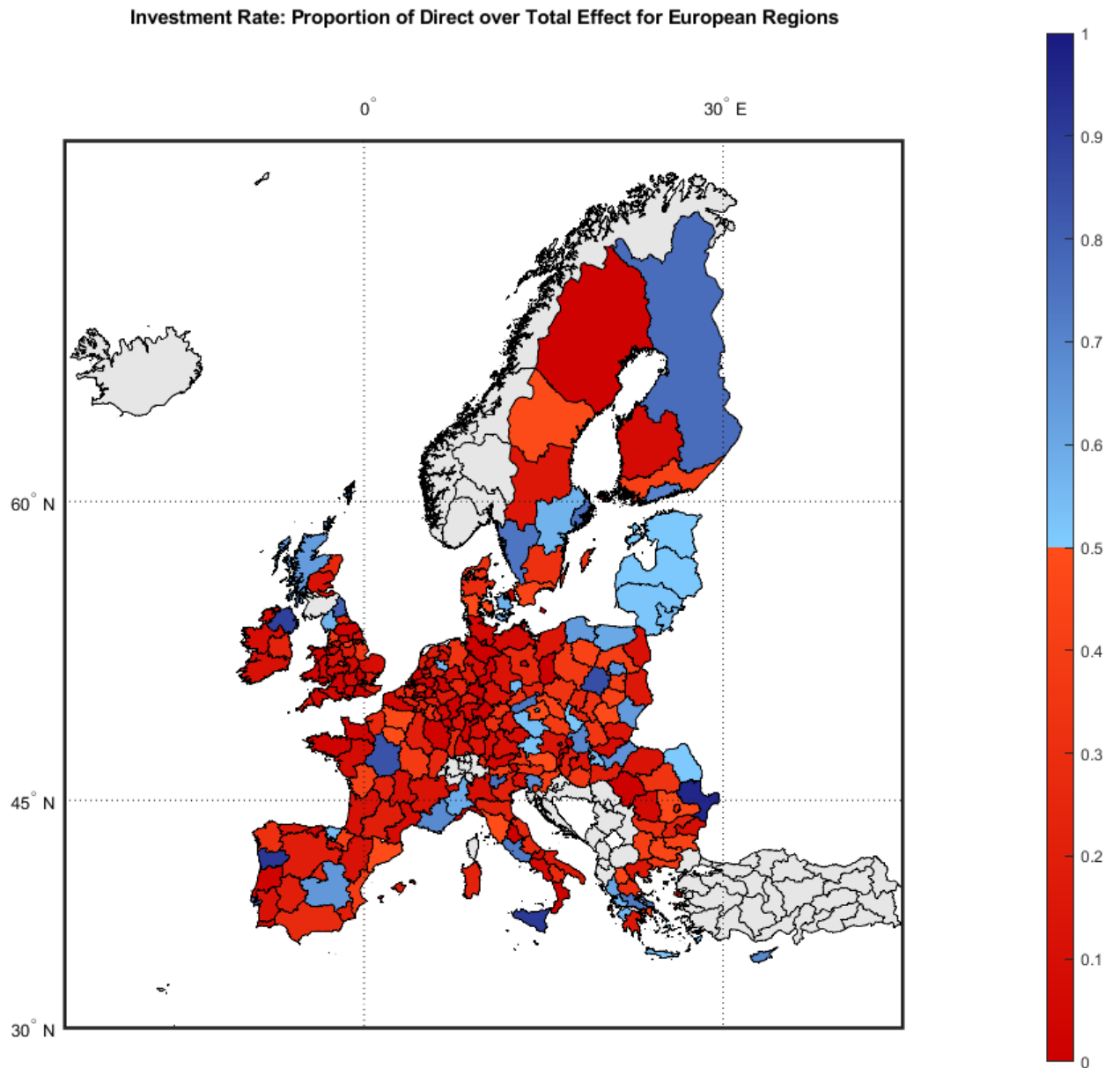
Figure 2: Investment Rate Effects on Regional Growth in Europe

source endowments. Allowing for such heterogeneity ensures that growth models capture these diverse regional dynamics meaningfully.

From a policy perspective, the results emphasize the necessity of fostering inter-regional cooperation and investment. Policies that enhance connectivity –whether through infrastructure, trade facilitation, or collaborative innovation networks– are likely to yield substantial aggregate benefits. Moreover, the prominence of spillover effects in education and technology underscores the value of region-wide initiatives to elevate human capital and technological capabilities.

Finally, incorporating interactive fixed effects into growth models offers a robust framework

for addressing nonlinear unobserved heterogeneity, latent common shocks, and their region-specific impacts. This is particularly crucial in the context of global and economy-wide events, such as economic crises or technological revolutions, which may affect regions differently.

5.4 Robustness

Table 5.3 presents results from a series of robustness checks designed to test the stability of the estimated coefficients under varying assumptions and configurations.

Column [1] replicates the results from Column [6] of Table 5.1, serving as the benchmark specification. Columns [2] and [3] retain the same model structure but modify the distance decay parameter ζ in Eq. (5.3) to 0.015 and 0.01, respectively. Lowering ζ results in a slower exponential decay, assigning higher weights to more distant regions. Consequently, the influence of remote regions is amplified, leading to a weaker spatial localization effect. This adjustment reflects a scenario in which regional interactions are less constrained by geographical proximity and more broadly distributed across the spatial network.

In Column [4], the spatial weights matrix is specified as $w_{ij} = \frac{d_{ij}^{-2}}{\sum_{j=1}^{N} d_{ij}^{-2}}$, a specification also explored in Ertur and Koch (2007). This formulation imposes a quadratic decay in spatial interactions, whereby closer regions exert significantly larger weights relative to those farther away. Finally, Column [5] modifies the sample by excluding 10 island regions within the EU. These regions, due to their physical separation, may exhibit weaker integration into the mainland spatial network. This adjustment assesses whether the inclusion of such isolated regions disproportionately influences the estimated spatial dynamics.

Some key observations are as follows: Firstly, despite the variations in the weighting matrix and sample composition, the spatial spillover effects demonstrate remarkable consistency across specifications. In specific, the coefficients of $\mathbf{W}\Delta ln\,(y_t)$, $\mathbf{W}\Delta ln\,(y_{t-1})$, $\mathbf{W}ln\,(inv_t)$ and $\mathbf{W}ln\,(educ_t)$ are similar in magnitude and significance across all configurations. This stability affirms the robustness of the estimates to different choices of the weighting matrix, and reinforces the conclusion that spillovers from neighboring regions play a substantial and persistent role in shaping a region's economic growth trajectory.

In addition, the estimated rate of conditional convergence, captured by the coefficient of $ln\,(y_{t-1})$, also remains consistent across specifications. This invariance suggests that the process of regional income convergence is robust to changes in the spatial weighting structure and sample adjustments.

On the other hand, noticeable variations arise in the impact of a region's own investment rate and population growth: for $ln\,(inv_t)$, the direct effect appears diminished in some specifications, while its spatial counterpart increases in magnitude. This pattern highlights the dominant role of investment spillovers relative to localized investment impacts, particularly when greater weight is assigned to interactions with distant regions. The coefficient on $ln\,(n_t)$ also exhibits greater variability across specifications, reflecting the complex and context-dependent nature of demographic dynamics in influencing regional growth.

The robustness checks confirm the reliability of key findings, particularly the significance of spillover effects, while underscoring the need for carefully chosen spatial weighting matrices aligned with the model's theory and context. The consistent convergence rate further validates the framework's ability to capture regional growth dynamics.

Table 5.3: **Robustness results for empirical growth model.**

	[1]	[2]	[3]	[4]	[5]
$ln\,(y_{t-1})$	-0.473***	-0.481***	-0.497***	-0.517***	-0.477***
	(0.028)	(0.028)	(0.030)	(0.029)	(0.029)
$ln\,(inv_t)$	0.027**	0.023*	0.009	0.004	0.028**
	(0.013)	(0.012)	(0.013)	(0.012)	(0.014)
$ln\,(n_t)$	-0.582**	-0.379*	-0.297	-0.145	-0.684**
	(0.281)	(0.218)	(0.197)	(0.201)	(0.338)
$ln\,(educ_t)$	0.007	-0.006	-0.016	-0.024	0.008
	(0.026)	(0.024)	(0.028)	(0.025)	(0.027)
$ln\,(sci\&tech_t)$	-0.003	-0.008	-0.027	-0.019	-0.005
	(0.016)	(0.015)	(0.020)	(0.014)	(0.016)
W					
$\Delta ln\,(y_t)$	0.628***	0.644***	0.661***	0.636***	0.629***
	(0.033)	(0.035)	(0.040)	(0.038)	(0.033)
$\Delta ln\,(y_{t-1})$	0.192***	0.191***	0.188***	0.165***	0.203***
	(0.052)	(0.053)	(0.061)	(0.063)	(0.054)
$ln\,(inv_t)$	0.071***	0.083***	0.099***	0.131***	0.072***
	(0.021)	(0.021)	(0.030)	(0.034)	(0.021)
$ln\,(n_t)$	0.643	-0.242	-0.369	-0.696	1.102
	(0.549)	(0.668)	(0.549)	(0.521)	(0.744)
$ln\,(educ_t)$	0.150***	0.176***	0.197***	0.248***	0.160***
	(0.035)	(0.033)	(0.040)	(0.053)	(0.034)
$ln\,(sci\&tech_t)$	0.022	0.042	0.115**	0.108	0.009
	(0.050)	(0.041)	(0.050)	(0.081)	(0.050)
N_g	266	266	266	266	256
fact1	1	1	1	1	1

Notes: Standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Column [1] is the same specification as in Column [6] in Table 5.1, repeated here for benchmark. Columns [2] and [3] correspond to the same model specification except that ζ in Eq. (5.3) is set equal to 0.015 and 0.01, respectively. In Column [4] the weighting matrix is given by $w_{ij} = \frac{d_{ij}^{-2}}{\sum_{j=1}^{N} d_{ij}^{-2}}$. Finally, Column [5] is the same model as in [1] except that it excludes 10 regions in the sample that are islands.

6   Concluding Remarks

This paper develops a Mean Group Instrumental Variables (MGIV) estimator tailored for spatial dynamic panel data models with interactive fixed effects. In contrast to existing methods that assume slope-parameter homogeneity, the MGIV estimator explicitly accommodates heterogeneity across cross-sectional units. Theoretical results establish its consistency and asymptotic normality under large N and T asymptotics. Furthermore, the estimator is asymptotically unbiased, enabling valid inferences without requiring bias correction. Its linear structure also renders it computationally efficient, making it suitable for large-scale applications.

References

Ahn, S.C., Horenstein, A.R. (2013). Eigenvalue Ratio Test for the Number of Factors. *Econometrica*, 81(3), 1203–1227.

Alvarez, J., Arellano, M. (2003). The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators. *Econometrica*, 71(4), 1121–1159.

Aquaro, M., Bailey, N., Pesaran, M.H. (2015). Estimation and inference for spatial models with heterogeneous coefficients: An application to US houseprices. *Journal of Applied Econometrics*, 36(1), 18–44.

Audretsch, D.B., Feldman, M.P. (2004). Knowledge Spillovers and the Geography of Innovation. *Handbook of Regional and Urban Economics*, 4, Ch. 61, 2713–2739.

Autant-Bernard, C., LeSage, J.P. (2011). Quantifying Knowledge Spillovers Using Spatial Econometric Models. *Journal of Regional Science*, 51(3), 471–496.

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

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

Bai, J., Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1), 191–221.

Bai, J., Li, K.P. (2021). Spatial panel data models with common shocks. *Journal of Econometrics*, 224(1), 134–160.

Benos, N., Karagiannis, S., Karkalakos, S. (2015). Proximity and growth spillovers in European regions: The role of geographical, economic and technological linkages *Journal of Macroeconomics*, 43, 124–139.

Botazzi, L., Peri, G. (2003). Innovation and spillovers in regions: Evidence from European patent data. *European Economic Review*, 47(4), 687–710.

Bramoullé, Y., Djebbari, H., Fortin, B. (2009). Identification of Peer Effects Through Social Networks. *Journal of Econometrics*, 150(1), 41–55.

Chen, J., Shin, Y., Zheng, C. (2022). Estimation and inference in heterogenous spatial panel data models with a multifactor error structure. *Journal of Econometrics*, 229(1), 55-79.

Case, S. (1991). Spatial patterns in household demand. *Econometrica*, 59, 953–965.

Chudik, A., Pesaran, H. M. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188(2), 393–420.

Cui, G., Norkutė, M., Sarafidis, V., Yamagata, T. (2022). Two-Stage Instrumental Variable Estimation of Linear Panel Data Models with Interactive Effects. *Econometrics Journal*, 25(2), 340–361.

Cui, G., Sarafidis, V., Yamagata, T. (2023). IV Estimation of Spatial Dynamic Panels with Interactive Effects: Large Sample Theory and an Application on Bank Attitude Toward Risk. *Econometrics Journal*, 26(2), 124–146.

Debarsy, N., LeSage, J.P., Pace, R.K. (2012). The interpretation and presentation of results in spatial econometrics. *The Review of Regional Studies*, 42(2), 107–142.

De Paula, Á. (2017). Econometrics of Network Models. In B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson (Eds.), *Advances in Economics and Econometrics: Theory and Applications*, Cambridge University Press.

De Paula, Á., Rasul, I., Souza, P. C.L. (2024). Estimation and Selection of Spatial Weight Matrix in a Spatial Lag Model. *The Review of Economic Studies*, forthcoming.

De Vos, I., Everaert, G. (2021). Bias-Corrected Common Correlated Effects Pooled Estimation in Dynamic Panels. *Journal of Business & Economic Statistics*, 39(1), 294–306.

De Vos, I., Everaert, G., Sarafidis, V. (2024). A Method to Evaluate the Rank Condition for CCE Estimators. *Econometric Reviews*, 43(2-4), 123–155.

Di Vaio, G., Zeira, J., Battisti, M. (2014). A new look at global growth convergence and divergence. *VoxEU*, 9 Jan. 2014.

Higgins, H., Martellosio, F. (2023). Shrinkage estimation of network spillovers with factor structured errors. *Journal of Econometrics*, 233(1), 66–87.

Lam, C., Souza, P.C.L. (2020). Shrinkage estimation of network spillovers with factor structured errors. *Journal of Business & Economic Statistics*, 38(3), 693–710.

Lee, L-F., Yu, J. (2015). Spatial Panel Data Models. In B. Baltagi (Eds.), *The Oxford Handbook of Panel Data*, Ch. 12, 363–401.

Elhorst, P., (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, Springer.

Elhorst, P. (2021). Spatial panel models and common factors. In M. Fischer, P. Nijkamp (Eds.) *Handbook of Regional Science*, Springer, Berlin, Heidelberg. Ch. 12, 2141–2159.

Elhorst, P., Tziolas, I., Tan, C., Milionis, P. (2024). The distance decay effect and spatial reach of spillovers. *Journal of Geographical Systems*, 26, 265–289.

Ertur, C., Coch, W. (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics*, 22(6), 1033–1062.

Funke, M., Niebuhr, A. (2005). Regional Geographic Research and Development Spillovers and Economic Growth: Evidence from West Germany *Regional Studies*, 39(1), 143–153.

Hahn, J., Kuersteiner, G. (2002). Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both n and T Are Large *Econometrica*, 70(4), 1639–1657.

Harding, M., Lamarche, C. (2011). Least squares estimation of a panel data model with multifactor error structure and endogenous covariates. *Economics Letters*, 111(3), 197–199.

Hamermesh, D. S. (2015). Labor Demand and the Source of Adjustment Costs. *The Economic Journal*, 105(430), 620–634.

Hansen, C. (2007). Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics*, 141(2), 597–620.

Hayakawa, K. (2015). The Asymptotic Properties Of The System Gmm Estimator In Dynamic Panel Data Models When Both N And T Are Large. *Econometric Theory*, 31(3), 647–667.

Jappelli, T., Pistaferri, L. (2017). Time, Habits, and Consumer Durables. *The Economics of Consumption: Theory and Evidence*, Ch. 13, 234–256, Oxford University Press.

Jing, Z., Elhorst, J.P., Jakobs, J., de Haan J., (2018). The propagation of financial turbulence: Interdependence, spillovers, and direct and indirect effects. *Empirical Economics*, 55, 169–192.

Juodis, A., Sarafidis, V. (2018). Fixed T dynamic panel data estimators with multifactor errors. *Econometric Reviews*, 37(8), 893–929.

Juodis, A., Sarafidis, V. (2022). An Incidental Parameters Free Inference Approach for Panels with Common Factors. *Journal of Econometrics*, 229(1), 19–54.

Islam, N. (1995). Growth Empirics: A Panel Data Approach. *The Quarterly Journal of Economics*, 110(4), 1127–1170.

Kapoor, M., Kelejian, H.H., Prucha, I.R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1), 97–130.

Kelejian, H.H., Prucha, I.R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, 104(2), 219–257.

Kelejian, H.H., Prucha, I.R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157, 53–67.

Kelejian, H., Piras, G. (2017). *Spatial Econometrics*, 1st edn. London: Academic Press.

Korniotis, G.M. (2010). Estimating Panel Models With Internal and External Habit Formation *Journal of Business & Economic Statistics*, 28(1), 145–158.

Kripfganz, S., Sarafidis, V. (2021). Instrumental Variable Estimation of Large-T Panel Data Models with Common Factors. *Stata Journal*, 21(3), 1–28.

Kripfganz, S., Sarafidis, V. (2025). Estimating Spatial Dynamic Panel Data Models with Unobserved Common Factors in Stata *Journal of Statistical Software*, forthcoming.

Lee, L-F., Yu, J. (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects *Journal of Econometrics*, 180(2), 174–197.

Lee, L-F., Yu, J. (2015). Spatial Panel Data Models. In B. Baltagi (Eds.), *The Oxford Handbook of Panel Data*, Ch. 12, 363–401.

LeSage, J.P., Pace, R.K. (2009). *Introduction to Spatial Econometrics.* CRC Press.

LeSage, J.P., Chih, Y-Y. (2016). Interpreting Heterogeneous Coefficient Spatial Autoregressive Panel Models. *Economics Letters*, 142, 1–5.

LeSage, J.P., Vance, C., Chih, Y.Y. (2017). A Bayesian heterogeneous coefficients spatial autoregressive panel data model of retail fuel duopoly pricing. *Regional Science and Urban Economics*, 62, 46–55.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3), 531–542.

Moon, R.H., Weidner, M. (2017). Dynamic Linear Panel Regression Models With Interactive Fixed Effects. *Econometric Theory*, 33(1), 158–195.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6), 1417–1426.

Norkutė, M., Sarafidis, V., Yamagata, T., Cui, G. (2021). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. *Journal of Econometrics*, 220(2), 416–446.

Ozyurt, S., Dees, S. (2018). Regional Dynamics of Economic Performance in the EU: To What Extent Do Spatial Spillovers Matter? *REGION*, 5(3), 75–96.

Pesaran, M.H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4), 967–1012.

Pesaran, M.H., Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1), 79–113.

Phillips, P.C.B., Sul, D. (2007). Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence. *Journal of Econometrics*, 137(1), 162–188.

Ramajo, J., Márquez, M.A., Hewings, G.J.D., Salinas, M.M. (2008). Spatial heterogeneity and interregional spillovers in the European Union: Do cohesion policies encourage convergence across regions? *European Economic Review*, 52(3), 551–567.

Robertson, D., Symons, J. (1992). Some Strange Properties of Panel Data Estimators. *Journal of Applied Econometrics*, 7(2), 175–189.

Rodríguez-Pose, A., Crescenzi, R. (2008). Research and Development, Spillovers, Innovation Systems, and the Genesis of Regional Growth in Europe. *Regional Studies*, 42(1), 51–67.

Sargent, T., Sims, C. (1977). Business cycle modeling without pretending to have too much a priori economic theory. MIMEO.

Sarafidis, V., Wansbeek, T. (2012). Cross-Sectional Dependence in Panel Data Analysis. *Econometric Reviews*, 31(5), 483–531.

Sarafidis, V., Wansbeek, T. (2021). Celebrating 40 years of panel data analysis: Past, present and future. *Journal of Econometrics*, 220(2), 215–226.

Shi, W., Lee, L.F. (2017). Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 197(2), 323–347.

Yu, J., de Jong, R., Lee, L-F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146(1), 118–134.

Yu, J., de Jong, R., Lee, L-F. (2012). Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration. *Journal of Econometrics*, 167(1), 16–37.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, Massachusetts.

Appendix A: Proofs of the main theoretical results

Let $\boldsymbol{\Xi}$ be an $r_1 \times r_1$ diagonal matrix, which has the largest r_1 eigenvalues of the $T \times T$ matrix $(NT)^{-1} \sum_{i=1}^{N} \mathbf{X}_i \mathbf{X}_i'$ on its main diagonal. Similarly, denote $\boldsymbol{\Xi}_{-1}$ as the diagonal matrix whose main diagonal consists of the largest r_1 eigenvalues of $(NT)^{-1} \sum_{i=1}^{N} \mathbf{X}_{i,-1} \mathbf{X}_{i,-1}'$. Then by the definitions of $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{F}}_{-1}$, we have $\widehat{\mathbf{F}} \boldsymbol{\Xi} = (NT)^{-1} \sum_{i=1}^{N} \mathbf{X}_i \mathbf{X}_i' \widehat{\mathbf{F}}$ and $\widehat{\mathbf{F}}_{-1} \boldsymbol{\Xi}_{-1} = (NT)^{-1} \sum_{i=1}^{N} \mathbf{X}_{i,-1} \mathbf{X}_{i,-1}' \widehat{\mathbf{F}}_{-1}$. Following the argument in the proof of Lemma A.3 in Bai (2003), we can show that $\boldsymbol{\Xi}$ and $\boldsymbol{\Xi}_{-1}$ are invertible asymptotically. Then we obtain the following results for the estimation of factors:

$$\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R} = \frac{1}{NT} \sum_{i=1}^{N} \mathbf{F}\boldsymbol{\Gamma}_i \mathbf{V}_i' \widehat{\mathbf{F}} \boldsymbol{\Xi}^{-1} + \frac{1}{NT} \sum_{i=1}^{N} \mathbf{V}_i \boldsymbol{\Gamma}_i' \mathbf{F}' \widehat{\mathbf{F}} \boldsymbol{\Xi}^{-1} + \frac{1}{NT} \sum_{i=1}^{N} \mathbf{V}_i \mathbf{V}_i' \widehat{\mathbf{F}} \boldsymbol{\Xi}^{-1}, \qquad (A.1)$$

$$\widehat{\mathbf{F}}_{-1} - \mathbf{F}_{-1}\boldsymbol{\mathfrak{R}} = \frac{1}{NT} \sum_{i=1}^{N} \mathbf{F}_{-1}\boldsymbol{\Gamma}_i \mathbf{V}_{i,-1}' \widehat{\mathbf{F}}_{-1} \boldsymbol{\Xi}_{-1}^{-1} + \frac{1}{NT} \sum_{i=1}^{N} \mathbf{V}_{i,-1}\boldsymbol{\Gamma}_i' \mathbf{F}_{-1}' \widehat{\mathbf{F}}_{-1} \boldsymbol{\Xi}_{-1}^{-1}$$

$$+ \frac{1}{NT} \sum_{i=1}^{N} \mathbf{V}_{i,-1} \mathbf{V}_{i,-1}' \widehat{\mathbf{F}}_{-1} \boldsymbol{\Xi}_{-1}^{-1}, \qquad (A.2)$$

where $\mathbf{R} = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i' \mathbf{F}' \widehat{\mathbf{F}} \boldsymbol{\Xi}^{-1}$ and $\boldsymbol{\mathfrak{R}} = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i' \mathbf{F}_{-1}' \widehat{\mathbf{F}}_{-1} \boldsymbol{\Xi}_{-1}^{-1}$ are two rotation matrices. Again following the proof of Lemma A.3 in Bai (2003), we can show that $\mathbf{R}$ and $\boldsymbol{\mathfrak{R}}$ are asymptotically invertible.

We first present the following proposition, which plays an important role in the proof of Theorem 3.1. The proof of this proposition requires the lemmas given in Appendix B.

Proposition A.1 *Under Assumptions A–E, we have, for all i:*

$$(a) \quad T^{-1}\widehat{\mathbf{Z}}_i' \mathbf{u}_i = T^{-1}\mathbf{Z}_i' \mathbf{u}_i + O_p(\delta_{NT}^{-2}) = O_p(T^{-1/2}) + O_p(\delta_{NT}^{-2}),$$

$$(b) \quad T^{-1}\widehat{\mathbf{Z}}_i' \mathbf{C}_i = T^{-1}\mathbf{Z}_i' \mathbf{C}_i + O_p(\delta_{NT}^{-1}),$$

$$(c) \quad T^{-1}\widehat{\mathbf{Z}}_i' \widehat{\mathbf{Z}}_i = T^{-1}\mathbf{Z}_i' \mathbf{Z}_i + O_p(\delta_{NT}^{-1}),$$

as $N, T \to \infty$.

Proof of Proposition A.1. From Lemma B.3, we obtain

$$\frac{1}{T}\widehat{\mathbf{Z}}_i' \mathbf{u}_i = \frac{1}{T} \begin{pmatrix} \sum_{j=1}^{N} w_{ij} \mathbf{X}_j' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{u}_i \\ \mathbf{X}_{i,-1}' \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{u}_i \\ \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{u}_i \end{pmatrix} = \frac{1}{T} \begin{pmatrix} \sum_{j=1}^{N} w_{ij} \mathbf{X}_j' \mathbf{M}_{\mathbf{F}} \mathbf{u}_i \\ \mathbf{X}_{i,-1}' \mathbf{M}_{\mathbf{F}_{-1}} \mathbf{M}_{\mathbf{F}} \mathbf{u}_i \\ \mathbf{X}_i' \mathbf{M}_{\mathbf{F}} \mathbf{u}_i \end{pmatrix} + O_p(\delta_{NT}^{-2}) = \frac{1}{T}\mathbf{Z}_i' \mathbf{u}_i + O_p(\delta_{NT}^{-2}),$$

for all i. This proves the first result in (a). Next we show that

$$T^{-1}\mathbf{Z}_i' \mathbf{u}_i = O_p(T^{-1/2}). \qquad (A.3)$$

To this end, first note that (2.6) implies

$$\frac{1}{\sqrt{T}}\mathbf{Z}_i' \mathbf{u}_i = \frac{1}{\sqrt{T}} \begin{pmatrix} \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' \mathbf{M}_{\mathbf{F}} (\mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i) \\ \mathbf{V}_{i,-1}' \mathbf{M}_{\mathbf{F}_{-1}} \mathbf{M}_{\mathbf{F}} (\mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i) \\ \mathbf{V}_i' \mathbf{M}_{\mathbf{F}} (\mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i) \end{pmatrix} =: \begin{pmatrix} \boldsymbol{\xi}_{i,1} \\ \boldsymbol{\xi}_{i,2} \\ \boldsymbol{\xi}_{i,3} \end{pmatrix}. \qquad (A.4)$$

Under Assumptions A – D, it's easy to verify that

$$\mathbb{E}(\boldsymbol{\xi}_{i,1}) = \mathbf{0}, \quad \mathbb{E}(\boldsymbol{\xi}_{i,2}) = \mathbf{0}, \quad \mathbb{E}(\boldsymbol{\xi}_{i,3}) = \mathbf{0}. \qquad (A.5)$$

On the other hand,

$$\text{Var}(\boldsymbol{\xi}_{i,1}) = \frac{1}{T}\text{Var}\left(\sum_{j=1}^{N} w_{ij} \mathbf{V}_j' \mathbf{M}_{\mathbf{F}} \mathbf{G}\boldsymbol{\phi}_i\right) + \frac{1}{T}\text{Var}\left(\sum_{j=1}^{N} w_{ij} \mathbf{V}_j' \mathbf{M}_{\mathbf{F}} \boldsymbol{\varepsilon}_i\right)$$

$$+ \frac{1}{T}\mathbb{E}\left(\sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{V}_{j_1}' \mathbf{M}_{\mathbf{F}} \mathbf{G}\boldsymbol{\phi}_i \boldsymbol{\varepsilon}_i' \mathbf{M}_{\mathbf{F}} \mathbf{V}_{j_2}\right)$$

$$+ \frac{1}{T} \mathbb{E} \left(\sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{V}_{j_1}' \mathbf{M_F} \boldsymbol{\varepsilon}_i \boldsymbol{\phi}_i' \mathbf{G}' \mathbf{M_F} \mathbf{V}_{j_2} \right)$$

$$=: \boldsymbol{\Sigma}_{\xi,i,1,1} + \boldsymbol{\Sigma}_{\xi,i,1,2} + \boldsymbol{\Sigma}_{\xi,i,1,3} + \boldsymbol{\Sigma}_{\xi,i,1,4}. \tag{A.6}$$

For $\boldsymbol{\Sigma}_{\xi,i,1,1}$, by Assumptions B – E, we have

$$\boldsymbol{\Sigma}_{\xi,i,1,1} = \frac{1}{T} \mathbb{E} \left(\sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{V}_{j_1}' \mathbf{M_F} \mathbf{G} \boldsymbol{\phi}_i \boldsymbol{\phi}_i' \mathbf{G}' \mathbf{M_F} \mathbf{V}_{j_2} \right)$$

$$= \frac{1}{T} \mathbb{E} \left(\sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{V}_{j_1}' \mathbf{M_F} \mathbf{G} \boldsymbol{\Sigma}_\phi \mathbf{G}' \mathbf{M_F} \mathbf{V}_{j_2} \right)$$

$$\leq \frac{C}{T} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} |w_{ij_1}||w_{ij_2}| \left\| \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E} \left(\mathbf{v}_{j_1,s} \mathbf{v}_{j_2,t}' \right) \right\|$$

$$\leq \frac{C}{T} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} |w_{ij_1}||w_{ij_2}| \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{\sigma}_{st} \leq C. \tag{A.7}$$

In a similar manner, using Assumptions A – E we can prove

$$\boldsymbol{\Sigma}_{\xi,i,1,2} \leq C. \tag{A.8}$$

In addition, noticing that $\boldsymbol{\varepsilon}_i$ is independent of $\mathbf{F}$, $\mathbf{G}$, $\boldsymbol{\phi}_i$, and $\mathbf{V}_j$, $j = 1, \ldots, N$, and has zero mean, we have

$$\boldsymbol{\Sigma}_{\xi,i,1,3} = 0, \quad \boldsymbol{\Sigma}_{\xi,i,1,4} = 0. \tag{A.9}$$

Combing (A.6)–(A.9), we obtain

$$\mathrm{Var}\left(\boldsymbol{\xi}_{i,1} \right) \leq C,$$

which, given (A.5), in turn implies

$$\boldsymbol{\xi}_{i,1} = O_p(1). \tag{A.10}$$

By similar arguments, we can show that

$$\boldsymbol{\xi}_{i,2} = O_p(1), \quad \boldsymbol{\xi}_{i,3} = O_p(1). \tag{A.11}$$

In view of (A.4), (A.10), and (A.11), we have

$$\frac{1}{\sqrt{T}} \mathbf{Z}_i' \mathbf{u}_i = O_p(1), \tag{A.12}$$

which implies (A.3).

Next, we prove (c). Recall that $\widehat{\mathbf{Z}}_i = (\sum_{j=1}^{N} w_{ij} \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_j, \ \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{X}_{i,-1}, \ \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i)$. Hence,

$$\frac{1}{T} \widehat{\mathbf{Z}}_i' \widehat{\mathbf{Z}}_i = \frac{1}{T} \begin{pmatrix} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{X}_{j_1}' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_{j_2} & \sum_{j=1}^{N} w_{ij} \mathbf{X}_j' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{X}_{i,-1} & \sum_{j=1}^{N} w_{ij} \mathbf{X}_j' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i \\ \sum_{j=1}^{N} w_{ij} \mathbf{X}_{i,-1}' \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_j & \mathbf{X}_{i,-1}' \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{X}_{i,-1} & \mathbf{X}_{i,-1}' \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i \\ \sum_{j=1}^{N} w_{ij} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_j & \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{X}_{i,-1} & \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i \end{pmatrix}$$

First consider the $(1,1)$-th block in the above matrix, i.e., $\frac{1}{T} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{X}_{j_1}' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_{j_2}$. Note that

$$\frac{1}{T} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{X}_{j_1}' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_{j_2}$$

$$= \frac{1}{T} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{X}_{j_1}' \mathbf{M_F} \mathbf{X}_{j_2} + \frac{1}{T} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{X}_{j_1}' \left(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F} \right) \mathbf{X}_{j_2}. \tag{A.13}$$

By Lemma B.1(f) and Assumptions B – E, we have

$$\left\| \frac{1}{T} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{X}'_{j_1} \big(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}} \big) \mathbf{X}_{j_2} \right\|$$

$$\leq \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} |w_{ij_1}||w_{ij_2}| \left\| \frac{1}{\sqrt{T}} \mathbf{X}_{j_1} \right\| \left\| \frac{1}{\sqrt{T}} \mathbf{X}_{j_2} \right\| \| \mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}} \|$$

$$= O_p(\delta_{NT}^{-1}) \times \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} |w_{ij_1}||w_{ij_2}| \left\| \frac{1}{\sqrt{T}} \mathbf{X}_{j_1} \right\| \left\| \frac{1}{\sqrt{T}} \mathbf{X}_{j_2} \right\| = O_p(\delta_{NT}^{-1}), \qquad (A.14)$$

for all i. The last equality in (A.14) holds because

$$\mathbb{E} \left(\sum_{j_1=1}^{N} \sum_{j_2=1}^{N} |w_{ij_1}||w_{ij_2}| \left\| \frac{1}{\sqrt{T}} \mathbf{X}_{j_1} \right\| \left\| \frac{1}{\sqrt{T}} \mathbf{X}_{j_2} \right\| \right)$$

$$\leq \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} |w_{ij_1}||w_{ij_2}| \sqrt{\mathbb{E} \left(\left\| \frac{1}{\sqrt{T}} \mathbf{X}_{j_1} \right\|^2 \right)} \sqrt{\mathbb{E} \left(\left\| \frac{1}{\sqrt{T}} \mathbf{X}_{j_2} \right\|^2 \right)} \leq C,$$

by Assumption E and $\mathbb{E} \| \mathbf{x}_{it} \|^4 \leq C$. Combing (A.13) and (A.14), we have

$$\frac{1}{T} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{X}'_{j_1} \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_{j_2} = \frac{1}{T} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} w_{ij_1} w_{ij_2} \mathbf{X}'_{j_1} \mathbf{M}_{\mathbf{F}} \mathbf{X}_{j_2} + O_p(\delta_{NT}^{-1}).$$

Analogously, we can prove that each of the remaining blocks in $\frac{1}{T} \widehat{\mathbf{Z}}'_i \widehat{\mathbf{Z}}_i$ equals its counterpart with $\mathbf{M}_{\widehat{\mathbf{F}}}$ and/or $\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}$ replaced by $\mathbf{M}_{\mathbf{F}}$ and/or $\mathbf{M}_{\mathbf{F}_{-1}}$, plus a remainder term that is of order $O_p(\delta_{NT}^{-1})$. This completes the proof of (c).

Following the proof of Lemma B.6, we can show that

$$\mathbb{E} \left(\left\| \frac{1}{\sqrt{T}} \mathbf{Y} \mathbf{w}_i \right\|^2 \right) \leq C, \quad \mathbb{E} \left(\left\| \frac{1}{\sqrt{T}} \mathbf{y}_{i,-1} \right\|^2 \right) \leq C.$$

Then using the same argument as in the proof of (c), we can show that (b) holds. This completes the proof of Proposition A.1. $\qquad \square$

Proof of Theorem 3.1. By (2.8) and Proposition 1, we have

$$\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i = \left(\widehat{\mathbf{A}}'_i \widehat{\mathbf{B}}_i^{-1} \widehat{\mathbf{A}}_i \right)^{-1} \widehat{\mathbf{A}}'_i \widehat{\mathbf{B}}_i^{-1} \widehat{\mathbf{c}}_{y,i} - \boldsymbol{\theta}_i$$

$$= \left(\widehat{\mathbf{A}}'_i \widehat{\mathbf{B}}_i^{-1} \widehat{\mathbf{A}}_i \right)^{-1} \widehat{\mathbf{A}}'_i \widehat{\mathbf{B}}_i^{-1} \big[T^{-1} \widehat{\mathbf{Z}}'_i (\mathbf{y}_i - \mathbf{C}_i \boldsymbol{\theta}_i) \big]$$

$$= \left(\widehat{\mathbf{A}}'_i \widehat{\mathbf{B}}_i^{-1} \widehat{\mathbf{A}}_i \right)^{-1} \widehat{\mathbf{A}}'_i \widehat{\mathbf{B}}_i^{-1} \big(T^{-1} \widehat{\mathbf{Z}}'_i \mathbf{u}_i \big)$$

$$= \left(\mathbf{A}'_i \mathbf{B}_i^{-1} \mathbf{A}_i \right)^{-1} \mathbf{A}'_i \mathbf{B}_i^{-1} \big(T^{-1} \mathbf{Z}'_i \mathbf{u}_i \big) + O_p(T^{-1/2} \delta_{NT}^{-1} + \delta_{NT}^{-2}), \qquad (A.15)$$

as $N, T \to \infty$. By Assumptions F1-2 and the last result in Proposition 1 (a), we can see that the RHS of (A.15) tends to zero as $N, T \to \infty$. This completes the proof of the consistency of $\widehat{\boldsymbol{\theta}}_i$. Then, by (2.6) and Assumptions A, B.1, D and F.4, we can show that

$$\mathbb{E} \big(T^{-1/2} \mathbf{Z}'_i \mathbf{u}_i \big) = \mathbf{0}, \qquad \mathrm{Var} \big(T^{-1/2} \mathbf{Z}'_i \mathbf{u}_i \big) = \boldsymbol{\Phi}_i,$$

$$T^{-1/2} \mathbf{Z}'_i \mathbf{u}_i \xrightarrow{d} N(\mathbf{0}, \ \boldsymbol{\Phi}_{i,0}), \quad \text{as} \ T \to \infty. \qquad (A.16)$$

Combining (A.15)-(A.16) with Assumption F3, we can easily prove Theorem 3.1 when $T/N^2 \to 0$ as $N, T \to \infty$. $\qquad \square$

Similarly, we present Proposition A.2 before proving Theorem 3.2. The proof of this proposition requires the use of Lemma B.5, which is provided in Appendix B.

Proposition A.2 *Under Assumptions A–E, we have:*

(a) $\quad \dfrac{1}{NT}\sum_{i=1}^{N}\mathbf{X}'_i\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i = \dfrac{1}{NT}\sum_{i=1}^{N}\mathbf{X}'_i\mathbf{M}_{\mathbf{F}}\mathbf{u}_i + O_p(\delta_{NT}^{-2}) = O_p(N^{-1/2}T^{-1/2} + \delta_{NT}^{-2}),$

(b) $\quad \dfrac{1}{NT}\sum_{i=1}^{N}\mathbf{X}'_{i,-1}\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i = \dfrac{1}{NT}\sum_{i=1}^{N}\mathbf{X}'_{i,-1}\mathbf{M}_{\mathbf{F}_{-1}}\mathbf{M}_{\mathbf{F}}\mathbf{u}_i + O_p(\delta_{NT}^{-2}) = O_p(N^{-1/2}T^{-1/2} + \delta_{NT}^{-2}),$

(c) $\quad \dfrac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}\mathbf{X}'_j\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i = \dfrac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}\mathbf{X}'_j\mathbf{M}_{\mathbf{F}}\mathbf{u}_i + O_p(\delta_{NT}^{-2}) = O_p(N^{-1/2}T^{-1/2} + \delta_{NT}^{-2}),$

as $N, T \to \infty.$

Proof of Proposition A.2. We only provide the proof of (a), as the proofs of (b) and (c) are similar. By Lemma B.5 (a), we readily have the first result in (a). To prove the second result, we need only to show that

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}'_i\mathbf{M}_{\mathbf{F}}\mathbf{u}_i = O_p(N^{-1/2}T^{-1/2}). \tag{A.17}$$

First note that

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}'_i\mathbf{M}_{\mathbf{F}}\mathbf{u}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}'_i\mathbf{M}_{\mathbf{F}}\left(\mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i\right) \quad \text{and} \quad \mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}'_i\mathbf{M}_{\mathbf{F}}\left(\mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i\right)\right] = \mathbf{0}.$$

Furthermore, by Assumptions B − D, we have

$$\left\|\mathrm{Var}\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}'_i\mathbf{M}_{\mathbf{F}}\left(\mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i\right)\right)\right\|$$

$$=\frac{1}{N^2T^2}\left\|\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left[\mathbf{V}'_i\mathbf{M}_{\mathbf{F}}\left(\mathbf{G}\boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i\right)\left(\mathbf{G}\boldsymbol{\phi}_j + \boldsymbol{\varepsilon}_j\right)'\mathbf{M}_{\mathbf{F}}\mathbf{V}_j\right]\right\|$$

$$=\frac{1}{N^2T^2}\left\|\sum_{i=1}^{N}\left[\mathbb{E}\left(\mathbf{V}'_i\mathbf{M}_{\mathbf{F}}\mathbf{G}\boldsymbol{\Sigma}_\phi\mathbf{G}'\mathbf{M}_{\mathbf{F}}\mathbf{V}_i\right) + \sigma_\varepsilon^2\mathbb{E}\left(\mathbf{V}'_i\mathbf{M}_{\mathbf{F}}\mathbf{V}_i\right)\right]\right\|$$

$$\leq\frac{C}{N^2T^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{\sigma}_{st} \leq \frac{C}{NT}.$$

The above results implies (A.17). This completes the proof of (a). $\qquad\square$

Proof of Theorem 3.2. Note that

$$\left(\widehat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) + \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\theta}_i - \boldsymbol{\theta}\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) + \frac{1}{N}\sum_{i=1}^{N}\mathbf{e}_i. \tag{A.18}$$

By (A.15), we have

$$\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) = \frac{1}{NT}\sum_{i=1}^{N}\left(\widehat{\mathbf{A}}'_i\widehat{\mathbf{B}}_i^{-1}\widehat{\mathbf{A}}_i\right)^{-1}\widehat{\mathbf{A}}'_i\widehat{\mathbf{B}}_i^{-1}\left(\widehat{\mathbf{Z}}'_i\mathbf{u}_i\right)$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\left[(\widehat{\mathbf{A}}'_i\widehat{\mathbf{B}}_i^{-1}\widehat{\mathbf{A}}_i)^{-1}\widehat{\mathbf{A}}'_i\widehat{\mathbf{B}}_i^{-1} - (\mathbf{A}'_i\mathbf{B}_i^{-1}\mathbf{A}_i)^{-1}\mathbf{A}'_i\mathbf{B}_i^{-1}\right]\left[\widehat{\mathbf{Z}}'_i\boldsymbol{u}_i - \mathbf{Z}'_i\boldsymbol{u}_i\right]$$

$$\quad + \frac{1}{NT}\sum_{i=1}^{N}\left[(\widehat{\mathbf{A}}'_i\widehat{\mathbf{B}}_i^{-1}\widehat{\mathbf{A}}_i)^{-1}\widehat{\mathbf{A}}'_i\widehat{\mathbf{B}}_i^{-1} - (\mathbf{A}'_i\mathbf{B}_i^{-1}\mathbf{A}_i)^{-1}\mathbf{A}'_i\mathbf{B}_i^{-1}\right]\mathbf{Z}'_i\boldsymbol{u}_i$$

$$\quad + \frac{1}{NT}\sum_{i=1}^{N}(\mathbf{A}'_i\mathbf{B}_i^{-1}\mathbf{A}_i)^{-1}\mathbf{A}'_i\mathbf{B}_i^{-1}(\widehat{\mathbf{Z}}'_i\boldsymbol{u}_i - \mathbf{Z}'_i\boldsymbol{u}_i)$$

$$+ \frac{1}{NT} \sum_{i=1}^{N} (\mathbf{A}_i' \mathbf{B}_i^{-1} \mathbf{A}_i)^{-1} \mathbf{A}_i' \mathbf{B}_i^{-1} \mathbf{Z}_i' \boldsymbol{u}_i$$

$$= \Theta_{NT,1} + \Theta_{NT,2} + \Theta_{NT,3} + \Theta_{NT,4}. \tag{A.19}$$

On the other hand, Lemma B.5(d)–(i) and Lemma B.6 imply

$$\sup_{1 \le i \le N} \|\widehat{\mathbf{B}}_i - \mathbf{B}_i\| = O_p\big(N^{1/2} \delta_{NT}^{-2}\big), \tag{A.20}$$

$$\sup_{1 \le i \le N} \|\widehat{\mathbf{A}}_i - \mathbf{A}_i\| = O_p\big(\iota_{NT}\big), \tag{A.21}$$

where $\iota_{NT} = N^{1/2} \delta_{NT}^{-2} \log N + N^{1/4} \delta_{NT}^{-1} \log N + \big[N^{3/4} T^{-1/2} \delta_{NT}^{-1} + NT^{-1} \delta_{NT}^{-1}\big] \big[\delta_{NT}^{-1} \log N + 1\big]$. Then using (A.20)–(A.21) and following the proof of Lemma 18 in Norkutė et al. (2021), we can show that

$$\sup_{1 \le i \le N} \|(\widehat{\mathbf{A}}_i' \widehat{\mathbf{B}}_i^{-1} \widehat{\mathbf{A}}_i)^{-1} \widehat{\mathbf{A}}_i' \widehat{\mathbf{B}}_i^{-1} - (\mathbf{A}_i' \mathbf{B}_i^{-1} \mathbf{A}_i)^{-1} \mathbf{A}_i' \mathbf{B}_i^{-1}\| = O_p\big(\iota_{NT}\big). \tag{A.22}$$

Furthermore, by Lemma B.5(a)–(c) we have

$$\frac{1}{NT} \sum_{i=1}^{N} \|\widehat{\mathbf{Z}}_i' \boldsymbol{u}_i - \mathbf{Z}_i' \boldsymbol{u}_i\| = O_p(\delta_{NT}^{-2}). \tag{A.23}$$

Combining (A.22) and (A.23) we obtain

$$\|\Theta_{NT,1}\| \le \sup_{1 \le i \le N} \|(\widehat{\mathbf{A}}_i' \widehat{\mathbf{B}}_i^{-1} \widehat{\mathbf{A}}_i)^{-1} \widehat{\mathbf{A}}_i' \widehat{\mathbf{B}}_i^{-1} - (\mathbf{A}_i' \mathbf{B}_i^{-1} \mathbf{A}_i)^{-1} \mathbf{A}_i' \mathbf{B}_i^{-1}\| \cdot \frac{1}{NT} \sum_{i=1}^{N} \|\widehat{\mathbf{Z}}_i' \boldsymbol{u}_i - \mathbf{Z}_i' \boldsymbol{u}_i\|$$

$$= O_p\big(\delta_{NT}^{-2} \cdot \iota_{NT}\big). \tag{A.24}$$

Similarly, using Proposition A.1 (a) and (A.22), we obtain

$$\|\Theta_{NT,2}\| = O_p\big(T^{-1/2} \cdot \iota_{NT}\big). \tag{A.25}$$

And by Assumption F1-2 and (A.23), we have

$$\|\Theta_{NT,3}\| = O_p\big(\delta_{NT}^{-2}\big). \tag{A.26}$$

Finally, Proposition A.2 implies

$$\|\Theta_{NT,4}\| = O_p\big(N^{-1/2} T^{-1/2}\big). \tag{A.27}$$

If $N/T^2 \to 0$ as $N, T \to \infty$, it's easy to show that

$$\delta_{NT}^{-2} \cdot \iota_{NT} = o(1), \quad T^{-1/2} \cdot \iota_{NT} = o(1).$$

This together with (A.19) and (A.24)–(A.27) implies

$$\frac{1}{N} \sum_{i=1}^{N} \big(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\big) = o_p(1). \tag{A.28}$$

Meanwhile, by Assumption G and the Law of Large Numbers, we have

$$\frac{1}{N} \sum_{i=1}^{N} \mathbf{e}_i \xrightarrow{p} \mathbf{0}, \quad \text{as } N \to \infty.$$

In view of (A.18) and (A.28), the above proves the consistency of the mean-group estimator when $N, T \to \infty$.

To prove the asymptotic distribution for the mean-group estimator, note that under Assumption G, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{e}_i \xrightarrow{d} N\big(\mathbf{0}, \boldsymbol{\Sigma_\theta}\big),$$

as $N \to \infty$. In view of (A.18), it remains to show that

$$\sqrt{N} \cdot \frac{1}{N} \sum_{i=1}^{N} \big(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\big) = o_p(1), \quad \text{as } N, T \to \infty. \tag{A.29}$$

To prove this, we first note that when $N/T^{6/5} \to 0$, we have $(N/T^{c_1})^{c_2} \log N \to 0$ for any $c_1 > 6/5$

and $c_2 > 0$. Hence, from (A.24) and (A.25), we have
$$\sqrt{N}\cdot\Theta_{NT,1} = O_p(N^{1/2}\cdot\delta_{NT}^{-2}\cdot\iota_{NT}) = o_p(1), \quad \sqrt{N}\cdot\Theta_{NT,2} = O_p(N^{1/2}\cdot T^{-1/2}\cdot\iota_{NT}) = o_p(1). \quad (A.30)$$
Furthermore, (A.26) and (A.27) imply
$$\sqrt{N}\cdot\Theta_{NT,3} = O_p(N^{1/2}\cdot\delta_{NT}^{-2}) = o_p(1), \quad \sqrt{N}\cdot\Theta_{NT,4} = O_p(T^{-1/2}) = o_p(1). \quad (A.31)$$
In view of (A.19), (A.30), and (A.31), we complete the proof of (A.29). □

Appendix B: Lemmas and their proofs

In this appendix, we present the lemmas used in the proofs of the main theoretical results.

Lemma B.1 *Under Assumptions A – D, we have*

(a) $\dfrac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{FR}\|^2 = O_p(\delta_{NT}^{-2})$,

(b) $\dfrac{1}{T}(\widehat{\mathbf{F}} - \mathbf{FR})'\mathbf{F} = O_p(\delta_{NT}^{-2})$, $\dfrac{1}{T}(\widehat{\mathbf{F}} - \mathbf{FR})'\mathbf{F}_{-1} = O_p(\delta_{NT}^{-2})$, $\dfrac{1}{T}(\widehat{\mathbf{F}} - \mathbf{FR})'\mathbf{G} = O_p(\delta_{NT}^{-2})$,

(c) $\dfrac{1}{T}(\widehat{\mathbf{F}} - \mathbf{FR})'\widehat{\mathbf{F}} = O_p(\delta_{NT}^{-2})$, $\dfrac{1}{T}(\widehat{\mathbf{F}} - \mathbf{FR})'\widehat{\mathbf{F}}_{-1} = O_p\left(\delta_{NT}^{-2}\right)$,

(d) $\mathbf{\Xi} = O_p(1)$, $\mathbf{R} = O_p(1)$, $\mathbf{\Xi}^{-1} = O_p(1)$, $\mathbf{R}^{-1} = O_p(1)$,

(e) $\mathbf{RR}' - \left(\dfrac{1}{T}\mathbf{F}'\mathbf{F}\right)^{-1} = O_p(\delta_{NT}^{-2})$,

(f) $\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}} = O_p(\delta_{NT}^{-1})$, $\mathbf{M}_{\widehat{\mathbf{F}}_{-1}} - \mathbf{M}_{\mathbf{F}_{-1}} = O_p(\delta_{NT}^{-1})$,

(g) $\dfrac{1}{NT}\displaystyle\sum_{\ell=1}^{N}\mathbf{\Gamma}_\ell\mathbf{V}_\ell'(\widehat{\mathbf{F}} - \mathbf{FR}) = O_p(N^{-1}) + O_p(N^{-1/2}\delta_{NT}^{-2})$.

Proof of Lemma B.1. For the proofs of (a) – (e), and (g), see Proof of Lemma 4 in the Supplemental Material of Norkutė et al. (2021). For (f), we decompose $\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}$ as
$$\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}} = -T^{-1}\widehat{\mathbf{F}}\left(\widehat{\mathbf{F}} - \mathbf{FR}\right)' - T^{-1}\left(\widehat{\mathbf{F}} - \mathbf{FR}\right)\mathbf{R}'\mathbf{F}' - T^{-1}\mathbf{F}\left[\mathbf{RR}' - (T^{-1}\mathbf{F}'\mathbf{F})^{-1}\right]\mathbf{F}'$$
which is bounded in norm by
$$\left\|\frac{1}{\sqrt{T}}\widehat{\mathbf{F}}\right\|\cdot\left\|\frac{1}{\sqrt{T}}(\widehat{\mathbf{F}} - \mathbf{FR})\right\| + \|\mathbf{R}\|\cdot\left\|\frac{1}{\sqrt{T}}\mathbf{F}\right\|\cdot\left\|\frac{1}{\sqrt{T}}(\widehat{\mathbf{F}} - \mathbf{FR})\right\| + \left\|\frac{1}{\sqrt{T}}\mathbf{F}\right\|^2\cdot\left\|\mathbf{RR}' - \left(\frac{1}{T}\mathbf{F}'\mathbf{F}\right)^{-1}\right\|.$$
By Assumption C, $\mathbb{E}\|T^{-1/2}\mathbf{F}\|^2 \leq C$ and hence, $\|T^{-1/2}\mathbf{F}\| = O_p(1)$. Furthermore, by (a) and (d) $\|T^{-1/2}\widehat{\mathbf{F}}\| \leq \|T^{-1/2}\mathbf{FR}\| + \|T^{-1/2}\left(\widehat{\mathbf{F}} - \mathbf{FR}\right)\| \leq \|T^{-1/2}\mathbf{F}\|\cdot\|\mathbf{R}\| + \|T^{-1/2}\left(\widehat{\mathbf{F}} - \mathbf{FR}\right)\| = O_p(1)$. Using the above results and the results in (a), (d), and (e), we can easily see that the above is of order $O_p\left(\delta_{NT}^{-1}\right)$. This completes the proof of (f). □

Lemma B.2 *Under Assumptions A – E, we have*

(a) $\left\|\dfrac{1}{T}\boldsymbol{\varepsilon}_i'\left(\mathbf{F} - \widehat{\mathbf{F}}\mathbf{R}^{-1}\right)\right\| = O_p\left(\delta_{NT}^{-2}\right)$,

(b) $\displaystyle\sum_{j=1}^{N}|w_{ij}|\left\|\dfrac{1}{T}\mathbf{V}_j'\left(\widehat{\mathbf{F}} - \mathbf{FR}\right)\right\| = O_p\left(\delta_{NT}^{-2}\right)$.

Proof of Lemma B.2. First consider (a). By (A.1), we have
$$\left\|\frac{1}{T}\boldsymbol{\varepsilon}_i'\left(\mathbf{F} - \widehat{\mathbf{F}}\mathbf{R}^{-1}\right)\right\|$$
$$\leq\frac{1}{NT^2}\left\|\sum_{\ell=1}^{N}\boldsymbol{\varepsilon}_i'\mathbf{F}\mathbf{\Gamma}_\ell\mathbf{V}_\ell'\widehat{\mathbf{F}}\right\|\cdot\|\mathbf{\Xi}^{-1}\mathbf{R}^{-1}\| + \frac{1}{NT^2}\left\|\sum_{\ell=1}^{N}\boldsymbol{\varepsilon}_i'\mathbf{V}_\ell\mathbf{\Gamma}_\ell'\mathbf{F}'\widehat{\mathbf{F}}\right\|\cdot\|\mathbf{\Xi}^{-1}\mathbf{R}^{-1}\|$$

$$+ \frac{1}{NT^2} \left\| \sum_{\ell=1}^{N} \boldsymbol{\varepsilon}_i' \mathbf{V}_\ell \mathbf{V}_\ell' \widehat{\mathbf{F}} \right\| \cdot \| \boldsymbol{\Xi}^{-1} \mathbf{R}^{-1} \| =: \Psi_{i,1} + \Psi_{i,2} + \Psi_{i,3}. \tag{B.1}$$

By Lemma B.1 (d), $\| \boldsymbol{\Xi}^{-1} \mathbf{R}^{-1} \| = O_p(1)$. Hence, we ignore it in the following analysis. Then first note that

$$\Psi_{i,1} \leq C T^{-1/2} \cdot \left\| \left(T^{-1/2} \boldsymbol{\varepsilon}_i' \mathbf{F} \right) \left(N^{-1} T^{-1} \sum_{\ell=1}^{N} \boldsymbol{\Gamma}_\ell \mathbf{V}_\ell' \widehat{\mathbf{F}} \right) \right\|.$$

By Assumptions A and C, we have $\mathbb{E} \| T^{-1/2} \boldsymbol{\varepsilon}_i' \mathbf{F} \|^2 = T^{-1} \sum_{t=1}^{T} \sigma_\varepsilon^2 \mathbb{E} | \mathbf{f}_t |^2 \leq C$, which implies

$$\| T^{-1/2} \boldsymbol{\varepsilon}_i' \mathbf{F} \| = O_p(1). \tag{B.2}$$

By Assumptions B – D and Lemma B.1 (g), we have

$$\frac{1}{NT} \sum_{\ell=1}^{N} \boldsymbol{\Gamma}_\ell \mathbf{V}_\ell' \widehat{\mathbf{F}} = \frac{1}{NT} \sum_{\ell=1}^{N} \boldsymbol{\Gamma}_\ell \mathbf{V}_\ell' \mathbf{F} \mathbf{R} + \frac{1}{NT} \sum_{\ell=1}^{N} \boldsymbol{\Gamma}_\ell \mathbf{V}_\ell' (\widehat{\mathbf{F}} - \mathbf{F} \mathbf{R})$$

$$= O_p(N^{-1/2} T^{-1/2} + N^{-1} + N^{-1/2} \delta_{NT}^{-2}). \tag{B.3}$$

With (B.2) and (B.3), it's easy to see that

$$\Psi_{i,1} = O_p(N^{-1/2} T^{-1} + N^{-1} T^{-1/2} + N^{-1/2} T^{-1/2} \delta_{NT}^{-2}). \tag{B.4}$$

On the other hand,

$$\Psi_{i,2} \leq C \frac{1}{\sqrt{NT}} \left\| \frac{1}{\sqrt{NT}} \sum_{\ell=1}^{N} \boldsymbol{\varepsilon}_i' \mathbf{V}_\ell \boldsymbol{\Gamma}_\ell' \right\| \cdot \left\| \frac{1}{\sqrt{T}} \mathbf{F}' \right\| \cdot \left\| \frac{1}{\sqrt{T}} \widehat{\mathbf{F}} \right\| = O_p(N^{-1/2} T^{-1/2}), \tag{B.5}$$

where we have used the results $\| T^{-1/2} \widehat{\mathbf{F}} \| = O_p(1)$, $\| T^{-1/2} \mathbf{F} \| = O_p(1)$, and $\| N^{-1/2} T^{-1/2} \sum_{\ell=1}^{N} \boldsymbol{\varepsilon}_i' \mathbf{V}_\ell \boldsymbol{\Gamma}_\ell' \| = O_p(1)$, the last of which can be proved analogously to (B.2). Lastly, consider $\Psi_{i,3}$. We can easily prove $\mathbb{E} \| T^{-1/2} \boldsymbol{\varepsilon}_i \|^2 \leq C$. Then, by Assumptions A, B, C, and Cauchy-Schwardz inequality, we have

$$\mathbb{E} \left\| N^{-1} T^{-1} \sum_{\ell=1}^{N} \boldsymbol{\varepsilon}_i' \mathbb{E} \left(\mathbf{V}_\ell \mathbf{V}_\ell' \right) \mathbf{F} \right\|^2 = \mathbb{E} \left\| T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \left(N^{-1} \sum_{\ell=1}^{N} \mathbb{E} \left(\mathbf{v}_{\ell s}' \mathbf{v}_{\ell t} \right) \right) \varepsilon_{is} \mathbf{f}_t' \right\|^2$$

$$\leq T^{-2} \sum_{s_1=1}^{T} \sum_{t_1=1}^{T} \sum_{s_2=1}^{T} \sum_{t_2=1}^{T} \left| N^{-1} \sum_{\ell=1}^{N} \mathbb{E} \left(\mathbf{v}_{\ell s_1}' \mathbf{v}_{\ell t_1} \right) \right| \left| N^{-1} \sum_{\ell=1}^{N} \mathbb{E} \left(\mathbf{v}_{\ell s_2}' \mathbf{v}_{\ell t_2} \right) \right| \mathbb{E} \left(\| \varepsilon_{is_1} \mathbf{f}_{t_1} \| \| \varepsilon_{is_2} \mathbf{f}_{t_2} \| \right)$$

$$\leq T^{-2} \sum_{s_1=1}^{T} \sum_{t_1=1}^{T} \sum_{s_2=1}^{T} \sum_{t_2=1}^{T} \tilde{\sigma}_{s_1 t_1} \tilde{\sigma}_{s_2 t_2} \left(\mathbb{E} \varepsilon_{is_1}^4 \mathbb{E} \varepsilon_{is_2}^4 \mathbb{E} \| \mathbf{f}_{t_1} \|^4 \mathbb{E} \| \mathbf{f}_{t_2} \|^4 \right)^{1/4} \tag{B.6}$$

$$\leq C \left(T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{\sigma}_{st} \right)^2 \leq C.$$

With Assumption B5, we can follow the arguments in the proof of Lemma A.2 (i) in Bai (2003) to show that $\mathbb{E} \| N^{-1/2} T^{-1} \sum_{\ell=1}^{N} \boldsymbol{\varepsilon}_i' \left[\mathbf{V}_\ell \mathbf{V}_\ell' - \mathbb{E} \left(\mathbf{V}_\ell \mathbf{V}_\ell' \right) \right] \mathbf{F} \|^2 \leq C$. The above analysis implies

$$\| T^{-1/2} \boldsymbol{\varepsilon}_i \| = O_p(1), \tag{B.7}$$

$$\left\| N^{-1} T^{-1} \sum_{\ell=1}^{N} \boldsymbol{\varepsilon}_i' \mathbb{E} \left(\mathbf{V}_\ell \mathbf{V}_\ell' \right) \mathbf{F} \right\| = O_p(1), \tag{B.8}$$

$$\left\| N^{-1/2} T^{-1} \sum_{\ell=1}^{N} \boldsymbol{\varepsilon}_i' \left[\mathbf{V}_\ell \mathbf{V}_\ell' - \mathbb{E} \left(\mathbf{V}_\ell \mathbf{V}_\ell' \right) \right] \mathbf{F} \right\| = O_p(1). \tag{B.9}$$

In addition, note that by Assumption B, we have

$$\left\| N^{-1} T^{-1/2} \sum_{\ell=1}^{N} \mathbb{E} \left(\mathbf{V}_\ell \mathbf{V}_\ell' \right) \right\|^2 = T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \left| N^{-1} \sum_{i=1}^{N} \mathbb{E} \left(\mathbf{v}_{is}' \mathbf{v}_{it} \right) \right|^2$$

$$\leq C N^{-1} T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \left| \mathbb{E} \left(\mathbf{v}_{is}' \mathbf{v}_{it} \right) \right| \leq C T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{\sigma}_{st} \leq C,$$

where we have used the result
$$\left| N^{-1} \sum_{i=1}^{N} \mathbb{E}(\mathbf{v}'_{is}\mathbf{v}_{it}) \right| \leq N^{-1} \sum_{i=1}^{N} \left| \mathbb{E}(\mathbf{v}'_{is}\mathbf{v}_{it}) \right| \leq N^{-1} \sum_{i=1}^{N} \sqrt{\mathbb{E}\|\mathbf{v}_{is}\|^2 \mathbb{E}\|\mathbf{v}_{it})\|^2} \leq C,$$
and
$$\mathbb{E}\left\| N^{-1/2}T^{-1} \sum_{\ell=1}^{N} [\mathbf{V}_\ell \mathbf{V}'_\ell - \mathbb{E}(\mathbf{V}_\ell \mathbf{V}'_\ell)] \right\|^2 = \mathbb{E}\left(T^{-2} \sum_{s=1}^{T} \sum_{t=1}^{T} \left[N^{-1/2} \sum_{\ell=1}^{N} (\mathbf{v}'_{\ell s}\mathbf{v}_{\ell t} - \mathbb{E}(\mathbf{v}'_{\ell s}\mathbf{v}_{\ell t})) \right]^2 \right)$$
$$= T^{-2} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}\left[N^{-1/2} \sum_{\ell=1}^{N} (\mathbf{v}'_{\ell s}\mathbf{v}_{\ell t} - \mathbb{E}(\mathbf{v}'_{\ell s}\mathbf{v}_{\ell t})) \right]^2 \leq C.$$
Thus,
$$\left\| \frac{1}{NT} \sum_{\ell=1}^{N} \mathbf{V}_\ell \mathbf{V}'_\ell \right\| \leq N^{-1/2} \left\| N^{-1/2}T^{-1} \sum_{\ell=1}^{N} [\mathbf{V}_\ell \mathbf{V}'_\ell - \mathbb{E}(\mathbf{V}_\ell \mathbf{V}'_\ell)] \right\| + T^{-1/2} \left\| N^{-1}T^{-1/2} \sum_{\ell=1}^{N} \mathbb{E}(\mathbf{V}_\ell \mathbf{V}'_\ell) \right\|$$
$$= O_p(N^{-1/2} + T^{-1/2}) = O_p(\delta_{NT}^{-1}). \tag{B.10}$$
Using results in (B.7) – (B.10) and Lemma B.1 (a)&(d), we obtain
$$\Psi_{i,3} \leq \|T^{-1/2}\boldsymbol{\varepsilon}_i\| \left\| N^{-1}T^{-1} \sum_{\ell=1}^{N} \mathbf{V}_\ell \mathbf{V}'_\ell \right\| \left\| T^{-1/2} \left(\widehat{\mathbf{F}} - \mathbf{FR} \right) \right\| + T^{-1} \left\| N^{-1}T^{-1} \sum_{\ell=1}^{N} \boldsymbol{\varepsilon}'_i \mathbb{E}(\mathbf{V}_\ell \mathbf{V}'_\ell) \mathbf{F} \right\| \|\mathbf{R}\|$$
$$+ N^{-1/2}T^{-1} \left\| N^{-1/2}T^{-1} \sum_{\ell=1}^{N} \boldsymbol{\varepsilon}'_i [\mathbf{V}_\ell \mathbf{V}'_\ell - \mathbb{E}(\mathbf{V}_\ell \mathbf{V}'_\ell)] \mathbf{F} \right\| \|\mathbf{R}\| = O_p(\delta_{NT}^{-2}). \tag{B.11}$$
Collecting (B.1), (B.4), (B.5), and (B.11), we can prove the result in (a).

Now consider (b). By (A.1), we have
$$\sum_{j=1}^{N} |w_{ij}| \left\| T^{-1}\mathbf{V}'_j \left(\widehat{\mathbf{F}} - \mathbf{FR} \right) \right\|$$
$$\leq N^{-1}T^{-2} \sum_{j=1}^{N} |w_{ij}| \left\| \sum_{\ell=1}^{N} \mathbf{V}'_j \mathbf{F} \boldsymbol{\Gamma}_\ell \mathbf{V}'_\ell \widehat{\mathbf{F}} \right\| \|\boldsymbol{\Xi}^{-1}\| + N^{-1}T^{-2} \sum_{j=1}^{N} |w_{ij}| \left\| \sum_{\ell=1}^{N} \mathbf{V}'_j \mathbf{V}_\ell \boldsymbol{\Gamma}'_\ell \mathbf{F}' \widehat{\mathbf{F}} \right\| \|\boldsymbol{\Xi}^{-1}\|$$
$$+ N^{-1}T^{-2} \sum_{j=1}^{N} |w_{ij}| \left\| \sum_{\ell=1}^{N} \mathbf{V}'_j \mathbf{V}_\ell \mathbf{V}'_\ell \widehat{\mathbf{F}} \right\| \|\boldsymbol{\Xi}^{-1}\|$$
$$=: \Pi_{i,1} + \Pi_{i,2} + \Pi_{i,3}. \tag{B.12}$$
By Lemma B.1 (d), $\|\boldsymbol{\Xi}^{-1}\| = O_p(1)$. So we ignore it in the following analysis. Then following the arguments in the proof of (B.4), we can show that
$$\Pi_{i,1} = O_p(N^{-1/2}T^{-1}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1/2}\delta_{NT}^{-2}). \tag{B.13}$$
On the other hand, note that
$$\Pi_{i,2} \leq \frac{1}{NT} \sum_{j=1}^{N} |w_{ij}| \cdot \left\| \sum_{\ell=1}^{N} \mathbf{V}'_j \mathbf{V}_\ell \boldsymbol{\Gamma}'_\ell \right\| \cdot \left\| \frac{1}{\sqrt{T}} \mathbf{F}' \right\| \cdot \left\| \frac{1}{\sqrt{T}} \widehat{\mathbf{F}} \right\|$$
$$= \frac{1}{NT} \sum_{j=1}^{N} |w_{ij}| \cdot \left\| \sum_{\ell=1}^{N} \mathbf{V}'_j \mathbf{V}_\ell \boldsymbol{\Gamma}'_\ell \right\| \times O_p(1), \tag{B.14}$$
and
$$\frac{1}{NT} \sum_{j=1}^{N} |w_{ij}| \cdot \left\| \sum_{\ell=1}^{N} \mathbf{V}'_j \mathbf{V}_\ell \boldsymbol{\Gamma}'_\ell \right\|$$
$$\leq \frac{1}{NT} \sum_{j=1}^{N} |w_{ij}| \cdot \left\| \sum_{\ell=1}^{N} \mathbb{E}(\mathbf{V}'_j \mathbf{V}_\ell) \boldsymbol{\Gamma}'_\ell \right\| + \frac{1}{NT} \sum_{j=1}^{N} |w_{ij}| \cdot \left\| \sum_{\ell=1}^{N} [\mathbf{V}'_j \mathbf{V}_\ell - \mathbb{E}(\mathbf{V}'_j \mathbf{V}_\ell)] \boldsymbol{\Gamma}'_\ell \right\|. \tag{B.15}$$

By Assumptions B2, D and E3, we have

$$\frac{1}{NT}\sum_{j=1}^{N}|w_{ij}|\left\|\sum_{\ell=1}^{N}\mathbb{E}(\mathbf{V}_j'\mathbf{V}_\ell)\mathbf{\Gamma}_\ell'\right\| = \frac{1}{NT}\sum_{j=1}^{N}|w_{ij}|\left\|\sum_{\ell=1}^{N}\left[\sum_{t=1}^{T}\mathbb{E}(\mathbf{v}_{jt}\mathbf{v}_{\ell t}')\right]\mathbf{\Gamma}_\ell'\right\|$$

$$\leq \frac{1}{NT}\sum_{j=1}^{N}|w_{ij}|\sum_{\ell=1}^{N}\sum_{t=1}^{T}\bar{\sigma}_{j\ell}\|\mathbf{\Gamma}_\ell\| = O_p(N^{-1}), \quad\quad\quad (\text{B.16})$$

and by Assumptions B4 and E3,

$$\frac{1}{NT}\sum_{j=1}^{N}|w_{ij}|\cdot\left\|\sum_{\ell=1}^{N}\left[\mathbf{V}_j'\mathbf{V}_\ell - \mathbb{E}(\mathbf{V}_j'\mathbf{V}_\ell)\right]\mathbf{\Gamma}_\ell'\right\|$$

$$= \frac{1}{NT}\sum_{j=1}^{N}|w_{ij}|\cdot\left\|\sum_{\ell=1}^{N}\sum_{t=1}^{T}\left[\mathbf{v}_{jt}\mathbf{v}_{\ell t}' - \mathbb{E}(\mathbf{v}_{jt}\mathbf{v}_{\ell t}')\right]\mathbf{\Gamma}_\ell'\right\| = O_p(N^{-1/2}T^{-1/2}). \quad (\text{B.17})$$

Combing (B.14) – (B.17), we obtain

$$\Pi_{i,2} = O_p(N^{-1} + N^{-1/2}T^{-1/2}) = O_p(\delta_{NT}^{-2}). \quad\quad\quad (\text{B.18})$$

Lastly, notice that by Assumptions B1 we have $\left\|\frac{1}{\sqrt{T}}\mathbf{V}_j\right\| = O_p(1)$ for all j, and analogous to (B.8), we can show that $\left\|\frac{1}{NT}\sum_{\ell=1}^{N}\mathbf{V}_j'\mathbb{E}(\mathbf{V}_\ell\mathbf{V}_\ell')\mathbf{F}\right\| = O_p(1)$. Furthermore, by Assumption B4, we have

$$\mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{\ell=1}^{N}\left[\mathbf{V}_\ell\mathbf{V}_\ell' - \mathbb{E}(\mathbf{V}_\ell\mathbf{V}_\ell')\right]\mathbf{F}\right\|^2 = \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{\ell=1}^{N}\sum_{t=1}^{T}\left[\mathbf{v}_{\ell s}'\mathbf{v}_{\ell t} - \mathbb{E}(\mathbf{v}_{\ell s}'\mathbf{v}_{\ell t})\right]\mathbf{f}_t'\right\|^2 \leq C.$$

Then using the above results as well as Assumption E3, (B.10), and Lemma B.1 (a) and (d), we obtain

$$\Pi_{i,3} \leq \sum_{j=1}^{N}|w_{ij}|\cdot\left\|\frac{1}{\sqrt{T}}\mathbf{V}_j\right\|\cdot\left\|\frac{1}{NT}\sum_{\ell=1}^{N}\mathbf{V}_\ell\mathbf{V}_\ell'\right\|\cdot\left\|\frac{1}{\sqrt{T}}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})\right\|$$

$$+ \frac{1}{T}\sum_{j=1}^{N}|w_{ij}|\cdot\left\|\frac{1}{NT}\sum_{\ell=1}^{N}\mathbf{V}_j'\mathbb{E}(\mathbf{V}_\ell\mathbf{V}_\ell')\mathbf{F}\right\|\cdot\|\mathbf{R}\|$$

$$+ \frac{1}{\sqrt{NT}}\sum_{j=1}^{N}|w_{ij}|\cdot\left\|\frac{1}{\sqrt{T}}\mathbf{V}_j\right\|\cdot\left\|\frac{1}{\sqrt{NT}}\sum_{\ell=1}^{N}\left[\mathbf{V}_\ell\mathbf{V}_\ell' - \mathbb{E}(\mathbf{V}_\ell\mathbf{V}_\ell')\right]\mathbf{F}\right\|\cdot\|\mathbf{R}\|$$

$$= O_p(\delta_{NT}^{-2} + T^{-1} + N^{-1/2}T^{-1/2}) = O_p(\delta_{NT}^{-2}). \quad\quad\quad (\text{B.19})$$

Combining (B.12) with (B.13), (B.18), and (B.19), we can see that (b) holds. This completes the proof of Lemma B.2. $\qquad\square$

Lemma B.3 *Under Assumptions A – E, we have, for all i,*

$$(a) \quad \frac{1}{T}\sum_{j=1}^{N}w_{ij}\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i = \frac{1}{T}\sum_{j=1}^{N}w_{ij}\mathbf{X}_j'\mathbf{M}_{\mathbf{F}}\mathbf{u}_i + O_p(\delta_{NT}^{-2}),$$

$$(b) \quad \frac{1}{T}\mathbf{X}_{i,-1}'\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i = \frac{1}{T}\mathbf{X}_{i,-1}'\mathbf{M}_{\mathbf{F}_{-1}}\mathbf{M}_{\mathbf{F}}\mathbf{u}_i + O_p(\delta_{NT}^{-2}),$$

$$(c) \quad \frac{1}{T}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i = \frac{1}{T}\mathbf{X}_i'\mathbf{M}_{\mathbf{F}}\mathbf{u}_i + O_p(\delta_{NT}^{-2}).$$

Proof of Lemma B.3. We only provide the proof for (a), as the proofs of (b) and (c) are similar.

Recall that $\mathbf{X}_j = \mathbf{F}\mathbf{\Gamma}_j + \mathbf{V}_j$. Hence,

$$\frac{1}{T}\sum_{j=1}^{N}w_{ij}\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i$$

$$= \frac{1}{T}\sum_{j=1}^{N}w_{ij}\mathbf{X}_j'\mathbf{M}_{\mathbf{F}}\mathbf{u}_i + \frac{1}{T}\sum_{j=1}^{N}w_{ij}\mathbf{X}_j'\left(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}\right)\mathbf{u}_i$$

$$= \frac{1}{T} \sum_{j=1}^{N} w_{ij} \mathbf{X}_j' \mathbf{M_F} \mathbf{u}_i + \frac{1}{T} \sum_{j=1}^{N} w_{ij} \mathbf{\Gamma}_j' \mathbf{F}' \big(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F} \big) \mathbf{u}_i + \frac{1}{T} \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' \big(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F} \big) \mathbf{u}_i \qquad \text{(B.20)}$$

First consider $\frac{1}{T} \sum_{j=1}^{N} w_{ij} \mathbf{\Gamma}_j' \mathbf{F}' \big(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F} \big) \mathbf{u}_i$. By Assumptions D2 and E3, $\mathbb{E} \| \sum_{j=1}^{N} w_{ij} \mathbf{\Gamma}_j \| \leq$ $\sum_{j=1}^{N} |w_{ij}| \mathbb{E} \| \mathbf{\Gamma}_j \| \leq C$ for all i. Hence, $\sum_{j=1}^{N} w_{ij} \mathbf{\Gamma}_j = O_p(1)$. Then,

$$\frac{1}{T} \sum_{j=1}^{N} w_{ij} \mathbf{\Gamma}_j' \mathbf{F}' \big(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F} \big) \mathbf{u}_i = O_p(1) \times \frac{1}{T} \mathbf{F}' \big(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F} \big) \mathbf{u}_i. \qquad \text{(B.21)}$$

Further notice that we have the following expansion

$$\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F} = -\frac{1}{T} (\widehat{\mathbf{F}} - \mathbf{FR}) \mathbf{R}' \mathbf{F}' - \frac{1}{T} \mathbf{FR} (\widehat{\mathbf{F}} - \mathbf{FR})' - \frac{1}{T} (\widehat{\mathbf{F}} - \mathbf{FR})(\widehat{\mathbf{F}} - \mathbf{FR})'$$
$$- \frac{1}{T} \mathbf{F} \left(\mathbf{RR}' - \big(T^{-1} \mathbf{F}' \mathbf{F} \big)^{-1} \right) \mathbf{F}'. \qquad \text{(B.22)}$$

As a result,

$$\frac{1}{T} \mathbf{F}' \big(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F} \big) \mathbf{u}_i = -\frac{1}{T^2} \mathbf{F}' (\widehat{\mathbf{F}} - \mathbf{FR}) \mathbf{R}' \mathbf{F}' \mathbf{u}_i - \frac{1}{T^2} \mathbf{F}' \mathbf{FR} (\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{u}_i$$
$$- \frac{1}{T^2} \mathbf{F}' (\widehat{\mathbf{F}} - \mathbf{FR})(\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{u}_i - \frac{1}{T^2} \mathbf{F}' \mathbf{F} \left(\mathbf{RR}' - \big(T^{-1} \mathbf{F}' \mathbf{F} \big)^{-1} \right) \mathbf{F}' \mathbf{u}_i$$
$$=: \mathbb{I}_{i,1} + \mathbb{I}_{i,2} + \mathbb{I}_{i,3} + \mathbb{I}_{i,4}. \qquad \text{(B.23)}$$

Consider $\mathbb{I}_{i,1}$. Given $\mathbf{u}_i = \mathbf{F} \boldsymbol{\lambda}_i + \mathbf{G} \boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i$, we have

$$\| T^{-1/2} \mathbf{u}_i \| \leq \| \boldsymbol{\lambda}_i \| \| T^{-1/2} \mathbf{F} \| + \| \boldsymbol{\phi}_i \| \| T^{-1/2} \mathbf{G} \| + \| T^{-1/2} \boldsymbol{\varepsilon}_i \|.$$

By Assumptions A, C, and D, we have $\| T^{-1/2} \mathbf{F} \| = O_p(1)$ and $\| T^{-1/2} \mathbf{u}_i \| = O_p(1)$ for all i. Then using Lemma B.1 (b) and (d), we obtain

$$\mathbb{I}_{i,1} = O_p(\delta_{NT}^{-2}). \qquad \text{(B.24)}$$

Similarly, using Lemma B.1 (a), (b) and (e), we have

$$\mathbb{I}_{i,3} = O_p(\delta_{NT}^{-3}), \qquad \mathbb{I}_{i,4} = O_p(\delta_{NT}^{-2}). \qquad \text{(B.25)}$$

Next, consider $\mathbb{I}_{i,2}$. By the fact that $\mathbf{u}_i = \mathbf{F} \boldsymbol{\lambda}_i + \mathbf{G} \boldsymbol{\phi}_i + \boldsymbol{\varepsilon}_i$ as well as Lemma B.1 (b)&(d) and Lemma B.2 (a), we have

$$\| \mathbb{I}_{i,2} \| \leq \left\| T^{-1} (\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{u}_i \right\| \cdot \left\| T^{-1/2} \mathbf{F} \right\|^2 \cdot \| \mathbf{R} \| = O_p \left(\left\| T^{-1} (\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{u}_i \right\| \right)$$
$$= O_p \left(\| \boldsymbol{\lambda}_i \| \left\| T^{-1} (\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{F} \right\| + \| \boldsymbol{\phi}_i \| \left\| T^{-1} (\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{G} \right\| + \| \mathbf{R} \| \left\| T^{-1} (\widehat{\mathbf{F}} \mathbf{R}^{-1} - \mathbf{F})' \boldsymbol{\varepsilon}_i \right\| \right)$$
$$= O_p(\delta_{NT}^{-2}). \qquad \text{(B.26)}$$

Combining (B.23) – (B.26), we have $\frac{1}{T} \mathbf{F}' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{u}_i = O_p(\delta_{NT}^{-2})$, which, in view of (B.21), implies

$$T^{-1} \sum_{j=1}^{N} w_{ij} \mathbf{\Gamma}_j' \mathbf{F}' \big(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F} \big) \mathbf{u}_i = O_p(\delta_{NT}^{-2}). \qquad \text{(B.27)}$$

Next, we prove that

$$T^{-1} \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' (\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F}) \mathbf{u}_i = O_p(\delta_{NT}^{-2}). \qquad \text{(B.28)}$$

By the expansion (B.22), we have

$$T^{-1} \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' (\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M_F}) \mathbf{u}_i$$
$$= -T^{-2} \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' (\widehat{\mathbf{F}} - \mathbf{FR}) \mathbf{R}' \mathbf{F}' \mathbf{u}_i - T^{-2} \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' \mathbf{FR} (\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{u}_i$$
$$- T^{-2} \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' (\widehat{\mathbf{F}} - \mathbf{FR})(\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{u}_i - T^{-2} \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' \mathbf{F} \left(\mathbf{RR}' - \big(T^{-1} \mathbf{F}' \mathbf{F} \big)^{-1} \right) \mathbf{F}' \mathbf{u}_i$$
$$= \mathbb{I}_{i,5} + \mathbb{I}_{i,6} + \mathbb{I}_{i,7} + \mathbb{I}_{i,8}. \qquad \text{(B.29)}$$

By Lemma B.1 (d), Lemma B.2(b), (B.2), and Assumptions C and D, we have $\sum_{j=1}^{N}|w_{ij}|\|\boldsymbol{\phi}_i\|\|T^{-1}\mathbf{V}_j'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})\| = O_p(\delta_{NT}^{-2})$ and $\sum_{j=1}^{N}|w_{ij}|\|T^{-1}\mathbf{V}_j'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})\|\|T^{-1/2}\mathbf{F}'\boldsymbol{\varepsilon}_i\| = O_p(\delta_{NT}^{-2})$.

$$
\begin{aligned}
\|\mathbb{I}_{i,5}\| \leq & \|\boldsymbol{\lambda}_i\| \cdot \sum_{j=1}^{N}|w_{ij}| \cdot \left\|T^{-1}\mathbf{V}_j'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})\right\| \cdot \left\|T^{-1/2}\mathbf{F}\right\|^2 \|\mathbf{R}\| \\
& + \|\boldsymbol{\phi}_i\| \sum_{j=1}^{N}|w_{ij}| \cdot \left\|T^{-1}\mathbf{V}_j'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})\right\| \left\|T^{-1/2}\mathbf{F}\right\| \cdot \left\|T^{-1/2}\mathbf{G}\right\| \cdot \|\mathbf{R}\| \\
& + T^{-1/2} \sum_{j=1}^{N}|w_{ij}| \cdot \left\|T^{-1}\mathbf{V}_j'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})\right\| \cdot \left\|T^{-1/2}\mathbf{F}'\boldsymbol{\varepsilon}_i\right\| \cdot \|\mathbf{R}\| \\
= & O_p\big(\delta_{NT}^{-2}\big).
\end{aligned}
\tag{B.30}
$$

For $\mathbb{I}_{i,6}$, first note that by Assumptions B2, C, and E3, we have

$$
\begin{aligned}
\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{j=1}^{N}w_{ij}\mathbf{V}_j'\mathbf{F}\right\|^2 \leq & \frac{1}{T}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}|w_{ij_1}||w_{ij_2}|\sum_{s=1}^{T}\sum_{t=1}^{T}\left|\mathbb{E}(\mathbf{f}_t'\mathbf{f}_s)\right|\left\|\mathbb{E}\left(\mathbf{v}_{j_1 t}\mathbf{v}_{j_2 s}'\right)\right\| \\
\leq & C\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}|w_{ij_1}||w_{ij_2}|\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{\sigma}_{st}\right) \leq C,
\end{aligned}
$$

which implies

$$
\frac{1}{\sqrt{T}}\sum_{j=1}^{N}w_{ij}\mathbf{V}_j'\mathbf{F} = O_p(1).
\tag{B.31}
$$

Then using Lemma B.1 (b)&(d) and Lemma B.2(a), we have

$$
\begin{aligned}
\mathbb{I}_{i,6} \leq & \frac{1}{\sqrt{T}}\left\|\frac{1}{\sqrt{T}}\sum_{j=1}^{N}w_{ij}\mathbf{V}_j'\mathbf{F}\right\|\|\mathbf{R}\|\left\|\frac{1}{T}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})'\mathbf{F}\right\|\|\boldsymbol{\lambda}_i\| \\
& + \frac{1}{\sqrt{T}}\left\|\frac{1}{\sqrt{T}}\sum_{j=1}^{N}w_{ij}\mathbf{V}_j'\mathbf{F}\right\|\|\mathbf{R}\|^2\left\|\frac{1}{T}(\widehat{\mathbf{F}}\mathbf{R}^{-1}-\mathbf{F})'\boldsymbol{\varepsilon}_i\right\| \\
& + \frac{1}{\sqrt{T}}\left\|\frac{1}{\sqrt{T}}\sum_{j=1}^{N}w_{ij}\mathbf{V}_j'\mathbf{F}\right\|\|\mathbf{R}\|\left\|\frac{1}{T}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})'\mathbf{G}\right\|\|\boldsymbol{\phi}_i\| \\
= & O_p(T^{-1/2}\delta_{NT}^{-2}).
\end{aligned}
\tag{B.32}
$$

On the other hand, by Lemmas B.1(b) and B.2(a)&(b), we have

$$
\begin{aligned}
\mathbb{I}_{i,7} \leq & \sum_{j=1}^{N}|w_{ij}|\left\|\frac{1}{T}\mathbf{V}_j'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})\right\|\left\|\frac{1}{T}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})'\mathbf{F}\right\|\|\boldsymbol{\lambda}_i\| \\
& + \sum_{j=1}^{N}|w_{ij}|\left\|\frac{1}{T}\mathbf{V}_j'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})\right\|\left\|\frac{1}{T}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})'\mathbf{G}\right\|\|\boldsymbol{\phi}_i\| \\
& + \sum_{j=1}^{N}|w_{ij}|\left\|\frac{1}{T}\mathbf{V}_j'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})\right\|\|\mathbf{R}\|\left\|\frac{1}{T}(\widehat{\mathbf{F}}\mathbf{R}^{-1}-\mathbf{F})'\boldsymbol{\varepsilon}_i\right\| \\
= & O_p(\delta_{NT}^{-4}).
\end{aligned}
\tag{B.33}
$$

Finally, by Lemma B.1(e) and (B.31), we have

$$
\mathbb{I}_{i,8} \leq \frac{1}{\sqrt{T}}\left\|\frac{1}{\sqrt{T}}\sum_{j=1}^{N}w_{ij}\mathbf{V}_j'\mathbf{F}\right\|\left\|\mathbf{R}\mathbf{R}' - \left(\frac{1}{T}\mathbf{F}'\mathbf{F}\right)^{-1}\right\|\left\|\frac{1}{T}\mathbf{F}'\mathbf{F}\right\|\|\boldsymbol{\lambda}_i\|
$$

$$+ \frac{1}{\sqrt{T}} \left\| \frac{1}{\sqrt{T}} \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' \mathbf{F} \right\| \left\| \mathbf{RR}' - \left(\frac{1}{T}\mathbf{F}'\mathbf{F}\right)^{-1} \right\| \left\| \frac{1}{\sqrt{T}}\mathbf{F} \right\| \left\| \frac{1}{\sqrt{T}}\mathbf{G} \right\| \|\boldsymbol{\phi}_i\|$$

$$+ \frac{1}{T} \left\| \frac{1}{\sqrt{T}} \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' \mathbf{F} \right\| \left\| \mathbf{RR}' - \left(\frac{1}{T}\mathbf{F}'\mathbf{F}\right)^{-1} \right\| \left\| \frac{1}{\sqrt{T}}\mathbf{F}'\boldsymbol{\varepsilon}_i \right\|$$

$$= O_p(T^{-1/2}\delta_{NT}^{-2}). \tag{B.34}$$

In view of (B.29) – (B.30), and (B.32) – (B.34), we readily have (B.28). With (B.20), (B.27), and (B.28), we can prove (a). $\qquad \square$

Lemma B.4 *Under Assumptions A – E, we have*

$$(a) \quad \sup_{1 \le i \le N} \left\| T^{-1}\mathbf{V}_i'(\widehat{\mathbf{F}} - \mathbf{FR}) \right\| = O_p(N^{1/4}\delta_{NT}^{-2} + N^{1/2}T^{-1/2}\delta_{NT}^{-2}),$$

$$(b) \quad \sup_{1 \le i \le N} \left\| T^{-1}\mathbf{V}_{i,-1}'(\widehat{\mathbf{F}}_{-1} - \mathbf{F}_{-1}\mathfrak{R}) \right\| = O_p(N^{1/4}\delta_{NT}^{-2} + N^{1/2}T^{-1/2}\delta_{NT}^{-2}),$$

$$(c) \quad \sup_{1 \le i \le N} \left\| T^{-1}\mathbf{V}_i'(\widehat{\mathbf{F}}_{-1} - \mathbf{F}_{-1}\mathfrak{R}) \right\| = O_p(N^{1/4}\delta_{NT}^{-2} + N^{1/2}T^{-1/2}\delta_{NT}^{-2}),$$

$$(d) \quad \sup_{1 \le i \le N} \left\| T^{-1}\mathbf{V}_{i,-1}'(\widehat{\mathbf{F}} - \mathbf{FR}) \right\| = O_p(N^{1/4}\delta_{NT}^{-2} + N^{1/2}T^{-1/2}\delta_{NT}^{-2}).$$

Proof of Lemma B.4. See the proof of Lemma 24 in Norkutė et al. (2021) with some notational modifications. $\qquad \square$

Lemma B.5 *Under Assumptions A – E, we have*

$$(a) \quad \frac{1}{NT} \sum_{i=1}^{N} \left\| \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i - \mathbf{X}_i'\mathbf{M}_{\mathbf{F}}\mathbf{u}_i \right\| = O_p(\delta_{NT}^{-2}),$$

$$(b) \quad \frac{1}{NT} \sum_{i=1}^{N} \left\| \mathbf{X}_{i,-1}'\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i - \mathbf{X}_{i,-1}'\mathbf{M}_{\mathbf{F}_{-1}}\mathbf{M}_{\mathbf{F}}\mathbf{u}_i \right\| = O_p(\delta_{NT}^{-2}),$$

$$(c) \quad \frac{1}{NT} \sum_{i=1}^{N} \left\| \sum_{j=1}^{N} w_{ij}\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i - \sum_{j=1}^{N} w_{ij}\mathbf{X}_j'\mathbf{M}_{\mathbf{F}}\mathbf{u}_i \right\| = O_p(\delta_{NT}^{-2}),$$

$$(d) \quad \sup_{1 \le i,j \le N} \frac{1}{T} \left\| \mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i - \mathbf{X}_j'\mathbf{M}_{\mathbf{F}}\mathbf{X}_i \right\| = O_p(N^{1/2}\delta_{NT}^{-2}),$$

$$(e) \quad \sup_{1 \le i,j \le N} \frac{1}{T} \left\| \mathbf{X}_{j,-1}'\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{i,-1} - \mathbf{X}_{j,-1}'\mathbf{M}_{\mathbf{F}_{-1}}\mathbf{M}_{\mathbf{F}}\mathbf{M}_{\mathbf{F}_{-1}}\mathbf{X}_{i,-1} \right\| = O_p(N^{1/2}\delta_{NT}^{-2}),$$

$$(f) \quad \sup_{1 \le i,j \le N} \frac{1}{T} \left\| \mathbf{X}_{j,-1}'\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i - \mathbf{X}_{j,-1}'\mathbf{M}_{\mathbf{F}_{-1}}\mathbf{M}_{\mathbf{F}}\mathbf{X}_i \right\| = O_p(N^{1/2}\delta_{NT}^{-2}),$$

$$(g) \quad \sup_{1 \le i \le N} \frac{1}{T} \left\| \sum_{j=1}^{N}\sum_{\ell=1}^{N} w_{ij}w_{i\ell}\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_\ell - \sum_{j=1}^{N}\sum_{\ell=1}^{N} w_{ij}w_{i\ell}\mathbf{X}_j'\mathbf{M}_{\mathbf{F}}\mathbf{X}_\ell \right\| = O_p(N^{1/2}\delta_{NT}^{-2}),$$

$$(h) \quad \sup_{1 \le i \le N} \frac{1}{T} \left\| \sum_{j=1}^{N} w_{ij}\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{i,-1} - \sum_{j=1}^{N} w_{ij}\mathbf{X}_j'\mathbf{M}_{\mathbf{F}}\mathbf{M}_{\mathbf{F}_{-1}}\mathbf{X}_{i,-1} \right\| = O_p(N^{1/2}\delta_{NT}^{-2}),$$

$$(i) \quad \sup_{1 \le i \le N} \frac{1}{T} \left\| \sum_{j=1}^{N} w_{ij}\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i - \sum_{j=1}^{N} w_{ij}\mathbf{X}_j'\mathbf{M}_{\mathbf{F}}\mathbf{X}_i \right\| = O_p(N^{1/2}\delta_{NT}^{-2}).$$

Proof of Lemma B.5. (a), (b), (d), (e) and (f) can be proved by following the proof of Lemma 25 in Norkutė et al. (2021) with some notational modifications. For the proof of (c), first note that

$$\frac{1}{NT} \sum_{i=1}^{N} \left\| \sum_{j=1}^{N} w_{ij} \mathbf{X}_j' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{u}_i - \sum_{j=1}^{N} w_{ij} \mathbf{X}_j' \mathbf{M}_{\mathbf{F}} \mathbf{u}_i \right\|$$

$$\leq \frac{1}{NT} \sum_{i=1}^{N} \left\| \sum_{j=1}^{N} w_{ij} \mathbf{\Gamma}_j' \mathbf{F}' (\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}) \mathbf{u}_i \right\| + \frac{1}{NT} \sum_{i=1}^{N} \left\| \sum_{j=1}^{N} w_{ij} \mathbf{V}_j' (\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}) \mathbf{u}_i \right\| =: \mathbb{J}_1 + \mathbb{J}_2. \quad \text{(B.35)}$$

By the expansion in (B.22), we have

$$\mathbb{J}_1 \leq \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |w_{ij}| \|\mathbf{\Gamma}_j\| \left\| \frac{1}{T^2} \mathbf{F}' \left[T (\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}) \right] \mathbf{u}_i \right\|$$

$$\leq \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |w_{ij}| \|\mathbf{\Gamma}_j\| \left\| \frac{1}{T^2} \mathbf{F}' (\widehat{\mathbf{F}} - \mathbf{FR}) \mathbf{R}' \mathbf{F}' \mathbf{u}_i \right\|$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |w_{ij}| \|\mathbf{\Gamma}_j\| \left\| \frac{1}{T^2} \mathbf{F}' \mathbf{FR} (\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{u}_i \right\|$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |w_{ij}| \|\mathbf{\Gamma}_j\| \left\| \frac{1}{T^2} \mathbf{F}' (\widehat{\mathbf{F}} - \mathbf{FR}) (\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{u}_i \right\|$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |w_{ij}| \|\mathbf{\Gamma}_j\| \left\| \frac{1}{T^2} \mathbf{F}' \mathbf{F} \left(\mathbf{RR}'^{-1} \mathbf{F}' \mathbf{F} \right)^{-1} \mathbf{F}' \mathbf{u}_i \right\|$$

$$=: \mathbb{J}_{1,1} + \mathbb{J}_{1,2} + \mathbb{J}_{1,3} + \mathbb{J}_{1,4}. \quad \text{(B.36)}$$

Following the proof of (B.24) and using Lemma B.1 (b)&(d), we obtain

$$\mathbb{J}_{1,1} \leq \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |w_{ij}| \|\mathbf{\Gamma}_j\| \left\| \frac{1}{T} \mathbf{F}' (\widehat{\mathbf{F}} - \mathbf{FR}) \right\| \|\mathbf{R}\| \left\| \frac{1}{T} \mathbf{F}' \mathbf{u}_i \right\| = O_p(\delta_{NT}^{-2}). \quad \text{(B.37)}$$

Similarly, using Lemma B.1 (a), (b) and (e), we can show that

$$\mathbb{J}_{1,3} = O_p(\delta_{NT}^{-3}), \quad \mathbb{J}_{1,4} = O_p(\delta_{NT}^{-2}). \quad \text{(B.38)}$$

For $\mathbb{J}_{1,2}$, by following the proof of (B.26), we can obtain

$$\mathbb{J}_{1,2} \leq \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |w_{ij}| \|\mathbf{\Gamma}_j\| \left\| \frac{1}{T} \mathbf{F}' \mathbf{F} \right\| \|\mathbf{R}\| \left\| \frac{1}{T} (\widehat{\mathbf{F}} - \mathbf{FR})' \mathbf{u}_i \right\| = O_p(\delta_{NT}^{-2}). \quad \text{(B.39)}$$

Collecting (B.36) – (B.39), we have

$$\mathbb{J}_1 = O_p(\delta_{NT}^{-2}). \quad \text{(B.40)}$$

Then, by following the proof of (B.28), we can show that

$$\mathbb{J}_2 = O_p(\delta_{NT}^{-2}). \quad \text{(B.41)}$$

Combining (B.35) with (B.40) and (B.41), we readily have (c).

For (g), we use Assumption E3 and the result in (d) to obtain

$$\sup_{1 \leq i \leq N} \frac{1}{T} \left\| \sum_{j=1}^{N} \sum_{\ell=1}^{N} w_{ij} w_{i\ell} \mathbf{X}_j' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_\ell - \sum_{j=1}^{N} \sum_{\ell=1}^{N} w_{ij} w_{i\ell} \mathbf{X}_j' \mathbf{M}_{\mathbf{F}} \mathbf{X}_\ell \right\|$$

$$\leq \sup_{1 \leq i \leq N} \sum_{j=1}^{N} \sum_{\ell=1}^{N} |w_{ij}| |w_{i\ell}| \cdot \sup_{1 \leq j, \ell \leq N} \frac{1}{T} \left\| \mathbf{X}_j' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_\ell - \mathbf{X}_j' \mathbf{M}_{\mathbf{F}} \mathbf{X}_\ell \right\| = O_p(N^{1/2} \delta_{NT}^{-2}).$$

Analogously, we can prove (h) and (i). This completes the proof of Lemma B.5. $\qquad \square$

Lemma B.6 *Denoting*

$$\iota_{NT} = N^{1/2} \delta_{NT}^{-2} \log N + N^{1/4} \delta_{NT}^{-1} \log N + \left[N^{3/4} T^{-1/2} \delta_{NT}^{-1} + NT^{-1} \delta_{NT}^{-1} \right] \left[\delta_{NT}^{-1} \log N + 1 \right],$$

then under Assumptions A – E and G, we have

$$(a) \quad \sup_{1 \le i,j \le N} \frac{1}{T} \left\| \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{y}_j - \mathbf{X}_i' \mathbf{M}_{\mathbf{F}} \mathbf{y}_j \right\| = O_p(\iota_{NT}),$$

$$(b) \quad \sup_{1 \le i,j \le N} \frac{1}{T} \left\| \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{y}_{j,-1} - \mathbf{X}_i' \mathbf{M}_{\mathbf{F}} \mathbf{y}_{j,-1} \right\| = O_p(\iota_{NT}),$$

$$(c) \quad \sup_{1 \le i,j \le N} \frac{1}{T} \left\| \mathbf{X}_{i,-1}' \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{y}_j - \mathbf{X}_{i,-1}' \mathbf{M}_{\mathbf{F}_{-1}} \mathbf{M}_{\mathbf{F}} \mathbf{y}_j \right\| = O_p(\iota_{NT}),$$

$$(d) \quad \sup_{1 \le i \le N} \frac{1}{T} \left\| \mathbf{X}_{i,-1}' \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{y}_{i,-1} - \mathbf{X}_{i,-1}' \mathbf{M}_{\mathbf{F}_{-1}} \mathbf{M}_{\mathbf{F}} \mathbf{y}_{i,-1} \right\| = O_p(\iota_{NT}).$$

Proof of Lemma B.6. We first prove (a). Let $x_{\ell,it}$ be the ℓ-th element of $\mathbf{x}_{it}$, $\beta_{\ell i}$ be its corresponding coefficient, $\boldsymbol{\gamma}_{\ell i}$ $(r_1 \times 1)$ be the ℓ-th column of $\boldsymbol{\Gamma}_i$ and $v_{\ell,it}$ be the ℓ-th element of $\mathbf{v}_{it}$ for $1 \le \ell \le k$. Define $\mathbf{X}^{(\ell)} = (x_{\ell,it})_{T \times N}$, $\boldsymbol{\beta}^{(\ell)} = \mathrm{diag}(\beta_{\ell 1}, \cdots, \beta_{\ell N})'$, $\boldsymbol{\Gamma}^{(\ell)} = (\gamma_{\ell 1}, \cdots, \gamma_{\ell N})'$, and $\mathbf{V}^{(\ell)} = (v_{\ell,it})_{T \times N}$. Then $\mathbf{X}^{(\ell)} = \mathbf{F} \boldsymbol{\Gamma}^{(\ell)'} + \mathbf{V}^{(\ell)}$. Thus,

$$\mathbf{Y} = \mathbf{Y} \mathbf{W}' \boldsymbol{\Psi} + \mathbf{Y}_{-1} \boldsymbol{\rho} + \sum_{\ell=1}^{k} \mathbf{X}^{(\ell)} \boldsymbol{\beta}^{(\ell)} + \mathbf{F} \boldsymbol{\Lambda}' + \mathbf{G} \boldsymbol{\Phi}' + \boldsymbol{\varepsilon}$$

where $\mathbf{W} = (w_{ij})_{N \times N} = (\mathbf{w}_1, \cdots, \mathbf{w}_N)'$, $\boldsymbol{\Psi} = \mathrm{diag}(\psi_1, \cdots, \psi_N)$, $\boldsymbol{\rho} = \mathrm{diag}(\rho_1, \cdots, \rho_N)$, $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)'$, and $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N)'$. In addition, we define $\mathbf{Y}_{-s} = (y_{i,t-s})_{T \times N}$, $\boldsymbol{\varepsilon}_{-s} = (\varepsilon_{i,t-s})_{T \times N}$, $\mathbf{F}_{-s} = (\mathbf{f}_{1-s}, \cdots, \mathbf{f}_{T-s})'$, $\mathbf{G}_{-s} = (\mathbf{g}_{1-s}, \cdots, \mathbf{g}_{T-s})'$, $\mathbf{X}_{-s}^{(\ell)} = (x_{\ell,i,t-s})_{T \times N}$ and $\mathbf{V}_{-s}^{(\ell)} = (v_{\ell,i,t-s})_{T \times N}$. Then, we have

$$\mathbf{Y} = \mathbf{Y}_{-1} \boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} + \sum_{\ell=1}^{k} \mathbf{X}^{(\ell)} \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}$$

$$+ \mathbf{F} \boldsymbol{\Lambda}' (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} + \mathbf{G} \boldsymbol{\Phi}' (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} + \boldsymbol{\varepsilon} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}$$

and

$$\mathbf{Y} = \sum_{s=0}^{\infty} \sum_{\ell=1}^{k} \mathbf{X}_{-s}^{(\ell)} \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} [\boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}]^s + \sum_{s=0}^{\infty} \mathbf{F}_{-s} \boldsymbol{\Lambda}' (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} [\boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}]^s$$

$$+ \sum_{s=0}^{\infty} \mathbf{G}_{-s} \boldsymbol{\Phi}' (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} [\boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}]^s + \sum_{s=0}^{\infty} \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} [\boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}]^s.$$

Hence,

$$T^{-1} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{y}_j = T^{-1} \sum_{s=0}^{\infty} \sum_{\ell=1}^{k} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_{-s}^{(\ell)} \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} [\boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j$$

$$+ T^{-1} \sum_{s=0}^{\infty} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{F}_{-s} \boldsymbol{\Lambda}' (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} [\boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j$$

$$+ T^{-1} \sum_{s=0}^{\infty} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{G}_{-s} \boldsymbol{\Phi}' (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} [\boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j$$

$$+ T^{-1} \sum_{s=0}^{\infty} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} [\boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j,$$

where $\boldsymbol{\iota}_j$ is the j-th column vector in the identity matrix $\mathbf{I}_N$. As a result,

$$\sup_{1 \le i,j \le N} T^{-1} \left\| \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{y}_j - \mathbf{X}_i' \mathbf{M}_{\mathbf{F}} \mathbf{y}_j \right\|$$

$$\le \sup_{1 \le i,j \le N} T^{-1} \left\| \sum_{s=0}^{\infty} \sum_{\ell=1}^{k} \mathbf{X}_i' (\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}) \mathbf{X}_{-s}^{(\ell)} \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} [\boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \le i,j \le N} T^{-1} \left\| \sum_{s=0}^{\infty} \mathbf{X}_i' (\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}) \mathbf{F}_{-s} \boldsymbol{\Lambda}' (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1} [\boldsymbol{\rho} (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \leq i,j \leq N} T^{-1} \left\| \sum_{s=0}^{\infty} \mathbf{X}_i' (\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}) \boldsymbol{G}_{-s} \boldsymbol{\Phi}' (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \leq i,j \leq N} T^{-1} \left\| \sum_{s=0}^{\infty} \mathbf{X}_i' (\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}) \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$=: \mathbb{F}_1 + \mathbb{F}_2 + \mathbb{F}_3 + \mathbb{F}_4. \tag{B.42}$$

First consider $\mathbb{F}_2$. By (B.22), we have

$$\mathbb{F}_2 \leq \sup_{1 \leq i,j \leq N} T^{-2} \left\| \sum_{s=0}^{\infty} \mathbf{X}_i' (\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R}) \mathbf{R}' \mathbf{F}' \mathbf{F}_{-s} \boldsymbol{\Lambda}' (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \leq i,j \leq N} T^{-2} \left\| \sum_{s=0}^{\infty} \mathbf{X}_i' \mathbf{F}\mathbf{R} (\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})' \mathbf{F}_{-s} \boldsymbol{\Lambda}' (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \leq i,j \leq N} T^{-2} \left\| \sum_{s=0}^{\infty} \mathbf{X}_i' (\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R}) (\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})' \mathbf{F}_{-s} \boldsymbol{\Lambda}' (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \leq i,j \leq N} T^{-2} \left\| \sum_{s=0}^{\infty} \mathbf{X}_i' \mathbf{F} \left(\mathbf{R}\mathbf{R}' - (T^{-1}\mathbf{F}'\mathbf{F})^{-1} \right) \mathbf{F}' \mathbf{F}_{-s} \boldsymbol{\Lambda}' (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$=: \mathbb{F}_{2,1} + \mathbb{F}_{2,2} + \mathbb{F}_{2,3} + \mathbb{F}_{2,4}. \tag{B.43}$$

Since $\sup_{1 \leq i \leq N} |\psi_i| < \min\{1/\|\mathbf{W}\|_1, 1/\|\mathbf{W}\|_\infty\}$, we have the expansion

$$(\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} = \mathbf{I}_N + \mathbf{W}'\boldsymbol{\Psi} + (\mathbf{W}'\boldsymbol{\Psi})^2 + (\mathbf{W}'\boldsymbol{\Psi})^3 + \cdots,$$

which implies

$$\left\| (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} \right\|_1 \leq \|\mathbf{I}_N\|_1 + \|\mathbf{W}'\boldsymbol{\Psi}\|_1 + \|(\mathbf{W}'\boldsymbol{\Psi})^2\|_1 + \|(\mathbf{W}'\boldsymbol{\Psi})^3\|_1 + \cdots$$

$$\leq 1 + \sup_{1 \leq i \leq N} |\psi_i| \|\mathbf{W}'\|_1 + \left(\sup_{1 \leq i \leq N} |\psi_i| \|\mathbf{W}'\|_1 \right)^2 + \left(\sup_{1 \leq i \leq N} |\psi_i| \|\mathbf{W}'\|_1 \right)^3 + \cdots$$

$$= \frac{1}{1 - \sup_{1 \leq i \leq N} |\psi_i| \|\mathbf{W}'\|_1} \leq C.$$

Denoting

$$\rho_w = \left(\frac{\sup_{1 \leq i \leq N} |\rho_i|}{1 - \sup_{1 \leq i \leq N} |\psi_i| \|\mathbf{W}'\|_1} \right),$$

then for any $s \geq 0$, we have

$$\left\| (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1}]^s \right\|_1 \leq \|\boldsymbol{\rho}\|_1^s \cdot \left\| (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} \right\|_1^{s+1}$$

$$\leq C \left(\frac{\sup_{1 \leq i \leq N} |\rho_i|}{1 - \sup_{1 \leq i \leq N} |\psi_i| \|\mathbf{W}'\|_1} \right)^s = C \rho_w^s.$$

By Assumption G2, we have $0 < \rho_w < 1$ a.s.. Further define the $N \times 1$ vector, $\boldsymbol{a}_j^{(s)} = (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j$, $s \geq 0$, $1 \leq j \leq N$, and denote its k-th element by $a_{kj}^{(s)}$, $k = 1, \ldots, N$. Then,

$$\sup_{1 \leq j \leq N} \sum_{k=1}^{N} |a_{kj}^{(s)}| = \left\| (\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}'\boldsymbol{\Psi})^{-1}]^s \right\|_1 \leq C \rho_w^s. \tag{B.44}$$

Using the above results and Lemma B.1(b)&(d) and Lemma B.4(a), we have

$$\mathbb{F}_{2,1} \leq \sup_{1 \leq i \leq N} \left\| \frac{1}{T} \mathbf{X}_i' (\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R}) \right\| \|\mathbf{R}\| \sum_{s=0}^{\infty} \left\| \frac{1}{T} \mathbf{F}'\mathbf{F}_{-s} \right\| \sup_{1 \leq j \leq N} \left\| \boldsymbol{\Lambda}' (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$= \sup_{1 \leq i \leq N} \left\| \frac{1}{T} \mathbf{X}_i' (\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R}) \right\| \|\mathbf{R}\| \sum_{s=0}^{\infty} \left\| \frac{1}{T} \mathbf{F}'\mathbf{F}_{-s} \right\| \sup_{1 \leq j \leq N} \left\| \sum_{k=1}^{N} \boldsymbol{\lambda}_k a_{kj}^{(s)} \right\|$$

47

$$\leq \sup_{1\leq i\leq N}\left\|\frac{1}{T}\mathbf{X}_i'(\widehat{\mathbf{F}}-\mathbf{FR})\right\|\|\mathbf{R}\|\sum_{s=0}^{\infty}\left\|\frac{1}{T}\mathbf{F}'\mathbf{F}_{-s}\right\|\sup_{1\leq k\leq N}\|\boldsymbol{\lambda}_k\|\left(\sup_{1\leq j\leq N}\sum_{k=1}^{N}|a_{kj}^{(s)}|\right)$$

$$\leq \sup_{1\leq i\leq N}\left\|\frac{1}{T}\mathbf{X}_i'(\widehat{\mathbf{F}}-\mathbf{FR})\right\|\cdot\|\mathbf{R}\|\cdot\sup_{1\leq k\leq N}\|\boldsymbol{\lambda}_k\|\cdot C\sum_{s=0}^{\infty}\left\|\frac{1}{T}\mathbf{F}'\mathbf{F}_{-s}\right\|\rho_w^s$$

$$\leq \sup_{1\leq i\leq N}\|\boldsymbol{\Gamma}_i\|\left\|\frac{1}{T}\mathbf{F}'(\widehat{\mathbf{F}}-\mathbf{FR})\right\|\|\mathbf{R}\|\sup_{1\leq k\leq N}\|\boldsymbol{\lambda}_k\|\cdot C\sum_{s=0}^{\infty}\left\|\frac{1}{T}\mathbf{F}'\mathbf{F}_{-s}\right\|\rho_w^s$$

$$+\sup_{1\leq i\leq N}\left\|\frac{1}{T}\mathbf{V}_i'(\widehat{\mathbf{F}}-\mathbf{FR})\right\|\|\mathbf{R}\|\sup_{1\leq k\leq N}\|\boldsymbol{\lambda}_k\|\cdot C\sum_{s=0}^{\infty}\left\|\frac{1}{T}\mathbf{F}'\mathbf{F}_{-s}\right\|\rho_w^s$$

$$=O_p(N^{1/4}\delta_{NT}^{-2}+N^{1/2}T^{-1/2}\delta_{NT}^{-2})\cdot O_p(N^{1/4})$$

$$=O_p(N^{1/2}\delta_{NT}^{-2}+N^{3/4}T^{-1/2}\delta_{NT}^{-2}),\tag{B.45}$$

where we have used the results

$$\sup_{1\leq i\leq N}\|\boldsymbol{\Gamma}_i\|=O_p(N^{1/4}),\qquad\sup_{1\leq i\leq N}\|\boldsymbol{\lambda}_i\|=O_p(N^{1/4}),\qquad\sum_{s=0}^{\infty}\left\|\frac{1}{T}\mathbf{F}'\mathbf{F}_{-s}\right\|\rho_w^s=O_p(1),$$

the last of which is derived from

$$\mathbb{E}\left(\sum_{s=0}^{\infty}\left\|\frac{1}{T}\mathbf{F}'\mathbf{F}_{-s}\right\|\rho_w^s\right)=\sum_{s=0}^{\infty}\mathbb{E}\left\|\frac{1}{T}\mathbf{F}'\mathbf{F}_{-s}\right\|\mathbb{E}(\rho_w^s)\leq\sum_{s=0}^{\infty}\sqrt{\mathbb{E}\left\|\frac{1}{\sqrt{T}}\mathbf{F}\right\|^2\mathbb{E}\left\|\frac{1}{\sqrt{T}}\mathbf{F}_{-s}\right\|^2}\mathbb{E}(\rho_w^s)$$

$$\leq C\mathbb{E}\left(\sum_{s=0}^{\infty}\rho_w^s\right)=C\mathbb{E}\left(\frac{1}{1-\rho_w}\right)\leq C.\tag{B.46}$$

Similarly, we can show that

$$\mathbb{F}_{2,4}=O_p(N^{1/2}\delta_{NT}^{-2}).\tag{B.47}$$

For $\mathbb{F}_{2,2}$, note that

$$\mathbb{F}_{2,2}\leq\sup_{1\leq i\leq N}\left\|\frac{1}{T}\mathbf{X}_i'\mathbf{F}\right\|\|\mathbf{R}\|\sum_{s=0}^{\infty}\left\|\frac{1}{T}(\widehat{\mathbf{F}}-\mathbf{FR})'\mathbf{F}_{-s}\right\|\sup_{1\leq j\leq N}\left\|\boldsymbol{\Lambda}'(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$\leq\sup_{1\leq i\leq N}\left\|\frac{1}{T}\mathbf{X}_i'\mathbf{F}\right\|\cdot\|\mathbf{R}\|\cdot\sup_{1\leq k\leq N}\|\boldsymbol{\lambda}_k\|\cdot C\sum_{s=0}^{\infty}\left\|\frac{1}{T}(\widehat{\mathbf{F}}-\mathbf{FR})'\mathbf{F}_{-s}\right\|\rho_w^s.\tag{B.48}$$

Note that

$$\mathbb{E}\left(\sup_{1\leq i\leq N}\left\|\frac{1}{T}\mathbf{V}_i'\mathbf{F}\right\|^2\right)\leq\sum_{i=1}^{N}\mathbb{E}\left(\left\|\frac{1}{T}\mathbf{V}_i'\mathbf{F}\right\|^2\right)\leq C\frac{1}{T^2}\sum_{i=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sqrt{\mathbb{E}\|\mathbf{f}_{t_1}\|^2\mathbb{E}\|\mathbf{f}_{t_2}\|^2}tr(\boldsymbol{\Sigma}_{ii,t_1t_2})$$

$$\leq C\frac{N}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\tilde{\sigma}_{t_1t_2}\leq C\frac{N}{T}.$$

Then,

$$\sup_{1\leq i\leq N}\left\|\frac{1}{T}\mathbf{X}_i'\mathbf{F}\right\|\leq\sup_{1\leq i\leq N}\|\boldsymbol{\Gamma}_i'\|\left\|\frac{1}{T}\mathbf{F}'\mathbf{F}\right\|+\sup_{1\leq i\leq N}\left\|\frac{1}{T}\mathbf{V}_i'\mathbf{F}\right\|=O_p(N^{1/4}+N^{1/2}T^{-1/2}).\tag{B.49}$$

Furthermore, for any s, by Assumptions B – D, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i\mathbf{V}_i'\mathbf{F}_{-s}=O_p(1),$$

and by following the proof of (B.19), we can show that

$$\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{F}}'\mathbf{V}_i\mathbf{V}_i'\mathbf{F}_{-s}=O_p(\delta_{NT}^{-2}).$$

Then in the same spirit of (B.46), we can show that

$$\sum_{s=0}^{\infty}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i\mathbf{V}_i'\mathbf{F}_{-s}\right\|\rho_w^s=O_p(1),\qquad\sum_{s=0}^{\infty}\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{F}}'\mathbf{V}_i\mathbf{V}_i'\mathbf{F}_{-s}\right\|\rho_w^s=O_p(\delta_{NT}^{-2}).$$

Using the above results as well as (A.1), Lemma B.1 (d), and (B.3), we have

$$\sum_{s=0}^{\infty}\left\|\frac{1}{T}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})'\mathbf{F}_{-s}\right\|\rho_w^s$$

$$\leq\|\boldsymbol{\Xi}^{-1}\|\left\|\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{F}}'\mathbf{V}_i\boldsymbol{\Gamma}_i'\right\|\sum_{s=0}^{\infty}\left\|\frac{1}{T}\mathbf{F}'\mathbf{F}_{-s}\right\|\rho_w^s+\|\boldsymbol{\Xi}^{-1}\|\cdot\sum_{s=0}^{\infty}\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{F}}'\mathbf{V}_i\mathbf{V}_i'\mathbf{F}_{-s}\right\|\rho_w^s$$

$$+\frac{1}{\sqrt{NT}}\|\boldsymbol{\Xi}^{-1}\|\left\|\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{F}\right\|\cdot\sum_{s=0}^{\infty}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i\mathbf{V}_i'\mathbf{F}_{-s}\right\|\rho_w^s$$

$$=O_p(N^{-1/2}T^{-1/2}+N^{-1}+\delta_{NT}^{-2})=O_p(\delta_{NT}^{-2}). \tag{B.50}$$

From (B.48) – (B.50), we can see that

$$\mathbb{F}_{2,2}=O_p(N^{1/2}\delta_{NT}^{-2}+N^{3/4}T^{-1/2}\delta_{NT}^{-2}).$$

Similarly, we have $\mathbb{F}_{2,3}=O_p(N^{1/2}\delta_{NT}^{-4}+N^{3/4}T^{-1/2}\delta_{NT}^{-4})$. Combing these with (B.43), (B.45), and (B.47), we have

$$\mathbb{F}_2=O_p(N^{1/2}\delta_{NT}^{-2}+N^{3/4}T^{-1/2}\delta_{NT}^{-2}). \tag{B.51}$$

Analogously, we can show that

$$\mathbb{F}_3=O_p(N^{1/2}\delta_{NT}^{-2}+N^{3/4}T^{-1/2}\delta_{NT}^{-2}). \tag{B.52}$$

Next, consider $\mathbb{F}_4$. First note that by (B.22), we have

$$\mathbb{F}_4\leq\sup_{1\leq i,j\leq N}\left\|\frac{1}{T^2}\sum_{s=0}^{\infty}\mathbf{X}_i'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})\mathbf{R}'\mathbf{F}'\boldsymbol{\varepsilon}_{-s}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$+\sup_{1\leq i,j\leq N}\left\|\frac{1}{T^2}\sum_{s=0}^{\infty}\mathbf{X}_i'\mathbf{F}\mathbf{R}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})'\boldsymbol{\varepsilon}_{-s}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$+\sup_{1\leq i,j\leq N}\left\|\frac{1}{T^2}\sum_{s=0}^{\infty}\mathbf{X}_i'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})'\boldsymbol{\varepsilon}_{-s}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$+\sup_{1\leq i,j\leq N}\left\|\frac{1}{T^2}\sum_{s=0}^{\infty}\mathbf{X}_i'\mathbf{F}\left(\mathbf{R}\mathbf{R}'-\left(T^{-1}\mathbf{F}'\mathbf{F}\right)^{-1}\right)\mathbf{F}'\boldsymbol{\varepsilon}_{-s}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$=:\mathbb{F}_{4.1}+\mathbb{F}_{4.2}+\mathbb{F}_{4.3}+\mathbb{F}_{4.4}. \tag{B.53}$$

Consider the term $\mathbb{F}_{4.1}$. It's easy to show that $\mathbb{E}\|T^{-1/2}\sum_{t=1}^{T}\mathbf{f}_t\varepsilon_{k,t-s}\|^2\leq C$ for any k and s. Then by (B.44), we have

$$\mathbb{E}\left(\sup_{1\leq j\leq N}\sum_{s=0}^{\infty}\sum_{k=1}^{N}|a_{kj}^{(s)}|\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{f}_t\varepsilon_{k,t-s}\right\|\right)^2$$

$$\leq\sum_{j=1}^{N}\sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty}\sum_{k_1=1}^{N}\sum_{k_2=1}^{N}\mathbb{E}\left(|a_{k_1j}^{(s_1)}||a_{k_2j}^{(s_2)}|\right)\cdot\mathbb{E}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t_1=1}^{T}\mathbf{f}_{t_1}\varepsilon_{k_1,t_1-s_1}\right\|\left\|\frac{1}{\sqrt{T}}\sum_{t_2=1}^{T}\mathbf{f}_{t_2}\varepsilon_{k_2,t_2-s_2}\right\|\right)$$

$$\leq\sum_{j=1}^{N}\sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty}\sum_{k_1=1}^{N}\sum_{k_2=1}^{N}\mathbb{E}\left(|a_{k_1j}^{(s_1)}||a_{k_2j}^{(s_2)}|\right)\cdot\sqrt{\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t_1=1}^{T}\mathbf{f}_{t_1}\varepsilon_{k_1,t_1-s_1}\right\|^2\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t_2=1}^{T}\mathbf{f}_{t_2}\varepsilon_{k_2,t_2-s_2}\right\|^2}$$

$$\leq C\sum_{j=1}^{N}\sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty}\mathbb{E}\left[\left(\sum_{k_1=1}^{N}|a_{k_1j}^{(s_1)}|\right)\left(\sum_{k_2=1}^{N}|a_{k_2j}^{(s_2)}|\right)\right]$$

$$\leq C\sum_{j=1}^{N}\sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty}\mathbb{E}\left(\rho_w^{s_1}\rho_w^{s_2}\right)=CN\cdot\mathbb{E}\left[\left(\frac{1}{1-\rho_w}\right)^2\right]\leq CN, \tag{B.54}$$

which in turn implies

$$\sup_{1\leq j\leq N}\sum_{s=0}^{\infty}\left\|\frac{1}{T}\mathbf{F}'\boldsymbol{\varepsilon}_{-s}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

49

$$= \sup_{1 \le j \le N} \sum_{s=0}^{\infty} \left\| \frac{1}{T} \sum_{k=1}^{N} \sum_{t=1}^{T} \mathbf{f}_t \varepsilon_{k,t-s} a_{kj}^{(s)} \right\|$$

$$\le \frac{1}{\sqrt{T}} \sup_{1 \le j \le N} \left(\sum_{s=0}^{\infty} \sum_{k=1}^{N} |a_{kj}^{(s)}| \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{f}_t \varepsilon_{k,t-s} \right\| \right) = O_p(N^{1/2} T^{-1/2}). \tag{B.55}$$

With the above result and (B.45), we have

$$\mathbb{F}_{4,1} \le \sup_{1 \le i \le N} \left\| \frac{1}{T} \mathbf{X}_i'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R}) \right\| \|\mathbf{R}\| \sup_{1 \le j \le N} \sum_{s=0}^{\infty} \left\| \frac{1}{T} \mathbf{F}' \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$= O_p(N^{3/4} T^{-1/2} \delta_{NT}^{-2} + N T^{-1} \delta_{NT}^{-2}). \tag{B.56}$$

Analogously, we can show that

$$\mathbb{F}_{4,4} = O_p(N^{3/4} T^{-1/2} \delta_{NT}^{-2} + N T^{-1} \delta_{NT}^{-2}). \tag{B.57}$$

Next we consider $\mathbb{F}_{4,2}$. By (A.1), we have

$$\sup_{1 \le j \le N} \left\| \frac{1}{T} \sum_{s=0}^{\infty} (\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})' \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$\le \sup_{1 \le j \le N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Xi}^{-1} \widehat{\mathbf{F}}' \mathbf{V}_k \boldsymbol{\Gamma}_k' \mathbf{F}' \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \le j \le N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Xi}^{-1} \widehat{\mathbf{F}}' \mathbf{F} \boldsymbol{\Gamma}_k \mathbf{V}_k' \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \le j \le N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Xi}^{-1} \widehat{\mathbf{F}}' \mathbf{V}_k \mathbf{V}_k' \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$\le \|\boldsymbol{\Xi}^{-1}\| \left\| \frac{1}{NT} \sum_{k=1}^{N} \widehat{\mathbf{F}}' \mathbf{V}_k \boldsymbol{\Gamma}_k' \right\| \sup_{1 \le j \le N} \left\| \frac{1}{T} \sum_{s=0}^{\infty} \mathbf{F}' \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \left\| \frac{1}{T} \widehat{\mathbf{F}}' \mathbf{F} \right\| \sup_{1 \le j \le N} \left\| \frac{1}{NT} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Gamma}_k \mathbf{V}_k' \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \|\mathbf{R}\| \sup_{1 \le j \le N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \mathbf{F}' \mathbb{E}(\mathbf{V}_k \mathbf{V}_k') \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \|\mathbf{R}\| \sup_{1 \le j \le N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \mathbf{F}' [\mathbf{V}_k \mathbf{V}_k' - \mathbb{E}(\mathbf{V}_k \mathbf{V}_k')] \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R}\| \sup_{1 \le j \le N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \mathbb{E}(\mathbf{V}_k \mathbf{V}_k') \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R}\| \sup_{1 \le j \le N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} [\mathbf{V}_k \mathbf{V}_k' - \mathbb{E}(\mathbf{V}_k \mathbf{V}_k')] \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$=: \mathbb{L}_1 + \mathbb{L}_2 + \mathbb{L}_3 + \mathbb{L}_4 + \mathbb{L}_5 + \mathbb{L}_6. \tag{B.58}$$

From (B.3) and (B.55), we obtain

$$\mathbb{L}_1 = O_p(T^{-1} + N^{-1/2} T^{-1/2} + T^{-1/2} \delta_{NT}^{-2}). \tag{B.59}$$

Following the argument in the proof of (B.55), we can show that

$$\sup_{1 \le j \le N} \left\| \frac{1}{NT} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Gamma}_k \mathbf{V}_k' \boldsymbol{\varepsilon}_{-s} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\| = O_p(T^{-1/2}),$$

which implies

$$\mathbb{L}_2 = O_p(1) \cdot O_p(T^{-1/2}) = O_p(T^{-1/2}). \tag{B.60}$$

On the other hand, following the argument in the proof of (B.6), we can show that

$$\mathbb{E}\left\|\frac{1}{NT}\sum_{k=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathbf{f}_{t_1}\boldsymbol{\varepsilon}_{i,t_2-s}tr(\boldsymbol{\Sigma}_{kk,t_2t_1})\right\|^2 \leq C$$

and following the argument in the proof of (B.55), we have

$$\sup_{1\leq j\leq N}\left\|\frac{1}{NT^2}\sum_{s=0}^{\infty}\sum_{k=1}^{N}\mathbf{F}'\mathbb{E}(\mathbf{V}_k\mathbf{V}_k')\boldsymbol{\varepsilon}_{-s}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$\leq \sup_{1\leq j\leq N}\frac{1}{T}\sum_{s=0}^{\infty}\sum_{i=1}^{N}\left\|\frac{1}{NT}\sum_{k=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathbf{f}_{t_1}\boldsymbol{\varepsilon}_{i,t_2-s}tr(\boldsymbol{\Sigma}_{kk,t_2t_1})\right\||a_{ij}^{(s)}| = O_p(N^{1/2}T^{-1}).$$

The above implies

$$\mathbb{L}_3 = O_p(1)\cdot O_p(N^{1/2}T^{-1}) = O_p(N^{1/2}T^{-1}). \tag{B.61}$$

Analogously, we can show that

$$\mathbb{L}_4 = O_p(T^{-1}). \tag{B.62}$$

Furthermore, by Assumptions A and B, we have, for any i, s, and t_1,

$$\mathbb{E}\left(\frac{1}{N\sqrt{T}}\sum_{k=1}^{N}\sum_{t_2=1}^{T}\boldsymbol{\varepsilon}_{i,t_2-s}tr(\boldsymbol{\Sigma}_{kk,t_2t_1})\right)^2 \leq C.$$

Then with similar arguments to (B.54), we obtain

$$\mathbb{E}\left(\sup_{1\leq j\leq N}\left\|\frac{1}{NT^2}\sum_{s=0}^{\infty}\sum_{k=1}^{N}\mathbb{E}(\mathbf{V}_k\mathbf{V}_k')\boldsymbol{\varepsilon}_{-s}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|\right)^2$$

$$\leq \frac{1}{T^3}\sum_{j=1}^{N}\mathbb{E}\left[\sum_{t_1=1}^{T}\left(\sum_{s=0}^{\infty}\sum_{i=1}^{N}\left|\frac{1}{N\sqrt{T}}\sum_{k=1}^{N}\sum_{t_2=1}^{T}\boldsymbol{\varepsilon}_{i,t_2-s}tr(\boldsymbol{\Sigma}_{kk,t_2t_1})\right|\cdot|a_{ij}^{(s)}|\right)^2\right]$$

$$\leq C\frac{1}{T^2}\sum_{j=1}^{N}\sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty}\mathbb{E}\left[\left(\sum_{i_1=1}^{N}|a_{i_1j}^{(s_1)}|\right)\left(\sum_{i_2=1}^{N}|a_{i_2j}^{(s_2)}|\right)\right] \leq C\frac{N}{T^2},$$

which implies

$$\mathbb{L}_5 = O_p(1)\cdot O_p(T^{1/2}\delta_{NT}^{-1})\cdot O_p(N^{1/2}T^{-1}) = O_p(N^{1/2}T^{-1/2}\delta_{NT}^{-1}). \tag{B.63}$$

Analogously, we can prove that

$$\mathbb{L}_6 = O_p(T^{-1/2}\delta_{NT}^{-1}). \tag{B.64}$$

Collecting (B.58)–(B.64), we obtain

$$\sup_{1\leq j\leq N}\left\|\frac{1}{T}\sum_{s=0}^{\infty}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})'\boldsymbol{\varepsilon}_{-s}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\| = O_p(N^{1/2}T^{-1/2}\delta_{NT}^{-1}). \tag{B.65}$$

In view of (B.49) and (B.65), we have

$$\mathbb{F}_{4.2} \leq \sup_{1\leq i\leq N}\left\|\frac{1}{T}\mathbf{X}_i'\mathbf{F}\right\|\cdot\|\mathbf{R}\|\cdot\sup_{1\leq j\leq N}\left\|\frac{1}{T}\sum_{s=0}^{\infty}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})'\boldsymbol{\varepsilon}_{-s}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$= O_p(N^{3/4}T^{-1/2}\delta_{NT}^{-1} + NT^{-1}\delta_{NT}^{-1}). \tag{B.66}$$

For $\mathbb{F}_{4.3}$, using Lemma B.1(b), Lemma B.4(a), and (B.65), we obtain

$$\mathbb{F}_{4.3} \leq \sup_{1\leq i\leq N}\left\|\frac{1}{T}\mathbf{X}_i'(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})\right\|\sup_{1\leq j\leq N}\left\|\frac{1}{T}\sum_{s=0}^{\infty}(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{R})'\boldsymbol{\varepsilon}_{-s}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$= O_p(N^{3/4}T^{-1/2}\delta_{NT}^{-3} + NT^{-1}\delta_{NT}^{-3}). \tag{B.67}$$

Combing (B.53), (B.56), (B.57), (B.66), and (B.67), we have

$$\mathbb{F}_4 = O_p(N^{3/4}T^{-1/2}\delta_{NT}^{-1} + NT^{-1}\delta_{NT}^{-1}). \tag{B.68}$$

Lastly, consider $\mathbb{F}_1$. Given the fact that $\mathbf{X}_{-s}^{(\ell)} = \mathbf{F}_{-s}\boldsymbol{\Gamma}^{(\ell)'} + \mathbf{V}_{-s}^{(\ell)}$, $1 \leq \ell \leq k$, we have

$$\mathbb{F}_1 \leq \sum_{\ell=1}^{k}\sup_{1\leq i,j\leq N}\left\|\frac{1}{T}\sum_{s=0}^{\infty}\mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{F}}}-\mathbf{M}_{\mathbf{F}})\mathbf{F}_{-s}\boldsymbol{\Gamma}^{(\ell)'}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N-\mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$+ \sum_{\ell=1}^{k} \sup_{1 \leq i,j \leq N} \left\| \frac{1}{T} \sum_{s=0}^{\infty} \mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}})\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$=: \mathbb{F}_{1.1} + \mathbb{F}_{1.2}. \tag{B.69}$$

First consider $\mathbb{F}_{1.2}$. Note that by (B.22), we have, for any $1 \leq \ell \leq k$,

$$\sup_{1 \leq i,j \leq N} \left\| \frac{1}{T} \sum_{s=0}^{\infty} \mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}})\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$\leq \sup_{1 \leq i,j \leq N} \left\| \frac{1}{T^2} \sum_{s=0}^{\infty} \mathbf{X}_i'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})\mathbf{R}'\mathbf{F}'\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \leq i,j \leq N} \left\| \frac{1}{T^2} \sum_{s=0}^{\infty} \mathbf{X}_i'\mathbf{F}\mathbf{R}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})'\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \leq i,j \leq N} \left\| \frac{1}{T^2} \sum_{s=0}^{\infty} \mathbf{X}_i'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})'\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \leq i,j \leq N} \left\| \frac{1}{T^2} \sum_{s=0}^{\infty} \mathbf{X}_i'\mathbf{F} \left(\mathbf{R}\mathbf{R}' - (T^{-1}\mathbf{F}'\mathbf{F})^{-1} \right) \mathbf{F}'\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$=: \mathbb{F}_{1.2.1} + \mathbb{F}_{1.2.2} + \mathbb{F}_{1.2.3} + \mathbb{F}_{1.2.4}. \tag{B.70}$$

From the proof of (B.45), we can see that

$$\sup_{1 \leq i \leq N} \left\| \frac{1}{T}\mathbf{X}_i'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R}) \right\| = O_p(N^{1/4}\delta_{NT}^{-2} + N^{1/2}T^{-1/2}\delta_{NT}^{-2}). \tag{B.71}$$

On the other hand, using similar arguments to the proof of (B.55), we obtain

$$\sup_{1 \leq j \leq N} \left\| \frac{1}{T} \sum_{s=0}^{\infty} \mathbf{F}'\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$= \sup_{1 \leq j \leq N} \left\| \frac{1}{\sqrt{T}} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{f}_t v_{l,k,t-s} \right) \beta_{lk} a_{kj}^{(s)} \right\|$$

$$\leq \sup_{1 \leq j \leq N} \frac{1}{\sqrt{T}} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{f}_t v_{l,k,t-s} \right\| |a_{kj}^{(s)}| \sup_{1 \leq k \leq N} |\beta_{lk}|$$

$$= O_p(N^{1/2}T^{-1/2}) \cdot O_p(\log N) = O_p(N^{1/2}T^{-1/2} \log N). \tag{B.72}$$

With (B.71), (B.72), and Lemma B.1(d), we have

$$\mathbb{F}_{1.2.1} \leq \sup_{1 \leq i \leq N} \left\| \frac{1}{T}\mathbf{X}_i'(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R}) \right\| \cdot \|\mathbf{R}\| \cdot \sup_{1 \leq j \leq N} \left\| \frac{1}{T} \sum_{s=0}^{\infty} \mathbf{F}'\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$= O_p\left(N^{3/4}T^{-1/2}\delta_{NT}^{-2} \log N + NT^{-1}\delta_{NT}^{-2} \log N \right). \tag{B.73}$$

Given (B.49), (B.72), and Lemma B.1(e), we can analogously show that

$$\mathbb{F}_{1.2.4} = O_p\left(N^{3/4}T^{-1/2}\delta_{NT}^{-2} \log N + NT^{-1}\delta_{NT}^{-2} \log N \right). \tag{B.74}$$

For $\mathbb{F}_{1.2.2}$ and $\mathbb{F}_{1.2.3}$, first note that by (A.1), we have

$$\sup_{1 \leq j \leq N} \left\| \frac{1}{T} \sum_{s=0}^{\infty} (\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})'\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$\leq \sup_{1 \leq j \leq N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Xi}^{-1}\widehat{\mathbf{F}}'\mathbf{V}_k\boldsymbol{\Gamma}_k'\mathbf{F}'\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \leq j \leq N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Xi}^{-1}\widehat{\mathbf{F}}'\mathbf{F}\boldsymbol{\Gamma}_k\mathbf{V}_k'\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \sup_{1 \leq j \leq N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Xi}^{-1} \widehat{\mathbf{F}}' \mathbf{V}_k \mathbf{V}_k' \mathbf{V}_{-s}^{(\ell)} \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$\leq \|\boldsymbol{\Xi}^{-1}\| \left\| \frac{1}{NT} \sum_{k=1}^{N} \widehat{\mathbf{F}}' \mathbf{V}_k \boldsymbol{\Gamma}_k' \right\| \sup_{1 \leq j \leq N} \left\| \frac{1}{T} \sum_{s=0}^{\infty} \mathbf{F}' \mathbf{V}_{-s}^{(\ell)} \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \left\| \frac{1}{T} \widehat{\mathbf{F}}' \mathbf{F} \right\| \sup_{1 \leq j \leq N} \left\| \frac{1}{NT} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Gamma}_k \mathbb{E}(\mathbf{V}_k' \mathbf{V}_{-s}^{(\ell)}) \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \left\| \frac{1}{T} \widehat{\mathbf{F}}' \mathbf{F} \right\| \sup_{1 \leq j \leq N} \left\| \frac{1}{NT} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Gamma}_k \left[\mathbf{V}_k' \mathbf{V}_{-s}^{(\ell)} - \mathbb{E}(\mathbf{V}_k' \mathbf{V}_{-s}^{(\ell)}) \right] \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \|\mathbf{R}\| \sup_{1 \leq j \leq N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \mathbf{F}' \mathbf{V}_k \mathbb{E}(\mathbf{V}_k' \mathbf{V}_{-s}^{(\ell)}) \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \|\mathbf{R}\| \sup_{1 \leq j \leq N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \mathbf{F}' \mathbf{V}_k \left[\mathbf{V}_k' \mathbf{V}_{-s}^{(\ell)} - \mathbb{E}(\mathbf{V}_k' \mathbf{V}_{-s}^{(\ell)}) \right] \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \sup_{1 \leq j \leq N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} (\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})' \mathbf{V}_k \mathbb{E}(\mathbf{V}_k' \mathbf{V}_{-s}^{(\ell)}) \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$+ \|\boldsymbol{\Xi}^{-1}\| \|\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R}\| \sup_{1 \leq j \leq N} \left\| \frac{1}{NT^2} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \mathbf{V}_k \left[\mathbf{V}_k' \mathbf{V}_{-s}^{(\ell)} - \mathbb{E}(\mathbf{V}_k \mathbf{V}_{-s}^{(\ell)}) \right] \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$=: \mathbb{M}_1 + \mathbb{M}_2 + \mathbb{M}_3 + \mathbb{M}_4 + \mathbb{M}_5 + \mathbb{M}_6 + \mathbb{M}_7. \tag{B.75}$$

First note that by Assumptions B, D, and G, (B.44) and similar arguments in (B.54), we have

$$\mathbb{E} \left(\sup_{1 \leq j \leq N} \left[\frac{1}{NT} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \|\boldsymbol{\Gamma}_k\| \|\mathbb{E}(\mathbf{v}_{kt} v_{\ell,i,t-s})\| |a_{ij}^{(s)}| \right]^2 \right)$$

$$\leq \sum_{j=1}^{N} \mathbb{E} \left(\left[\frac{1}{NT} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \|\boldsymbol{\Gamma}_k\| \|\mathbb{E}(\mathbf{v}_{kt} v_{\ell,i,t-s})\| |a_{ij}^{(s)}| \right]^2 \right)$$

$$= \frac{1}{N^2 T^2} \sum_{j=1}^{N} \sum_{s_1=0}^{\infty} \sum_{s_2=0}^{\infty} \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \|\mathbb{E}(\mathbf{v}_{k_1 t_1} v_{\ell,i_1,t_1-s_1})\| \|\mathbb{E}(\mathbf{v}_{k_2 t_2} v_{\ell,i_2,t_2-s_2})\|$$

$$\times \mathbb{E}(\|\boldsymbol{\Gamma}_{k_1}\| \|\boldsymbol{\Gamma}_{k_2}\|) \mathbb{E}(|a_{i_1 j}^{(s_1)}| |a_{i_2 j}^{(s_2)}|)$$

$$\leq \frac{C}{N^2} \sum_{j=1}^{N} \sum_{s_1=0}^{\infty} \sum_{s_2=0}^{\infty} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \mathbb{E}(|a_{i_1 j}^{(s_1)}| |a_{i_2 j}^{(s_2)}|) \left(\sum_{k_1=1}^{N} \bar{\sigma}_{k_1, i_1} \right) \left(\sum_{k_2=1}^{N} \bar{\sigma}_{k_2, i_2} \right)$$

$$\leq \frac{C}{N} \sum_{s_1=0}^{\infty} \sum_{s_2=0}^{\infty} \mathbb{E}(\rho_w^{s_1} \rho_w^{s_2}) \leq \frac{C}{N} \mathbb{E} \left[\left(\frac{1}{1 - \rho_w} \right)^2 \right] \leq \frac{C}{N}.$$

Thus, we have

$$\sup_{1 \leq j \leq N} \left\| \frac{1}{NT} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \boldsymbol{\Gamma}_k \mathbb{E}(\mathbf{V}_k' \mathbf{V}_{-s}^{(\ell)}) \boldsymbol{\beta}^{(\ell)} (\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1} [\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s \boldsymbol{\iota}_j \right\|$$

$$= \sup_{1 \leq j \leq N} \left\| \frac{1}{NT} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{\Gamma}_k \mathbb{E}(\mathbf{v}_{kt} v_{\ell,i,t-s}) \beta_{li} a_{ij}^{(s)} \right\|$$

$$\leq \sup_{1 \leq i \leq N} |\beta_{li}| \cdot \sup_{1 \leq j \leq N} \frac{1}{NT} \sum_{s=0}^{\infty} \sum_{k=1}^{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \|\boldsymbol{\Gamma}_k\| \|\mathbb{E}(\mathbf{v}_{kt} v_{\ell,i,t-s})\| |a_{ij}^{(s)}|$$

$$= O_p(\log N) \cdot O_p(N^{-1/2}) = O_p(N^{-1/2} \log N).$$

By Lemma B.1(d) and the fact that $\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{F} = \frac{1}{T}\mathbf{R}'\mathbf{F}'\mathbf{F} + \frac{1}{T}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})'\mathbf{F} = O_p(1)$, we obtain

$$\mathbb{M}_2 = O_p(N^{-1/2}\log N). \tag{B.76}$$

Analogously, using the result, $\mathbb{E}\left\|\frac{1}{\sqrt{T}}\mathbf{F}'\mathbf{V}_k\right\|^2 \leq C$, for any k, we have

$$\mathbb{M}_4 = O_p(T^{-1/2}N^{-1/2}\log N), \tag{B.77}$$

and using Lemma B.4(a), we have

$$\mathbb{M}_6 = O_p(N^{1/4}\delta_{NT}^{-2} + N^{1/2}T^{-1/2}\delta_{NT}^{-2}) \cdot O_p(N^{-1/2}\log N)$$

$$= O_p(N^{-1/4}\delta_{NT}^{-2}\log N + T^{-1/2}\delta_{NT}^{-2}\log N). \tag{B.78}$$

On the other hand, by (B.3) and (B.72), we have

$$\mathbb{M}_1 = O_p(N^{-1/2}T^{-1/2} + N^{-1} + N^{-1/2}\delta_{NT}^{-2}) \cdot O_p(N^{1/2}T^{-1/2}\log N)$$

$$= O_p(T^{-1}\log N + N^{-1/2}T^{-1/2}\log N + T^{-1/2}\delta_{NT}^{-2}\log N). \tag{B.79}$$

For $\mathbb{M}_3$, first note that by Assumption B4, we have

$$\mathbb{E}\left[\sup_{1\leq j\leq N}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=0}^{\infty}|a_{ij}^{(s)}|\left\|\sum_{k=1}^{N}\sum_{t=1}^{T}\mathbf{\Gamma}_k\left[\mathbf{v}_{kt}v_{l,i,t-s} - \mathbb{E}(\mathbf{v}_{kt}v_{l,i,t-s})\right]\right\|\right)^2\right]$$

$$\leq \frac{1}{N^2T^2}\sum_{j=1}^{N}\sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\mathbb{E}(|a_{i_1j}^{(s_1)}||a_{i_2j}^{(s_2)}|)\mathbb{E}\left(\left\|\sum_{k_1=1}^{N}\sum_{t_1=1}^{T}\mathbf{\Gamma}_{k_1}\left[\mathbf{v}_{k_1t_1}v_{l,i_1,t_1-s_1} - \mathbb{E}(\mathbf{v}_{k_1t_1}v_{l,i_1,t_1-s_1})\right]\right\|\right.$$

$$\left.\times\left\|\sum_{k_2=1}^{N}\sum_{t_2=1}^{T}\mathbf{\Gamma}_{k_2}\left[\mathbf{v}_{k_2t_2}v_{l,i_2,t_2-s_2} - \mathbb{E}(\mathbf{v}_{k_2t_2}v_{l,i_2,t_2-s_2})\right]\right\|\right)$$

$$\leq \frac{C}{N^2T^2}\cdot N\cdot NT\cdot\sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\mathbb{E}(|a_{i_1j}^{(s_1)}||a_{i_2j}^{(s_2)}|) \leq \frac{C}{T},$$

which implies

$$\sup_{1\leq j\leq N}\left\|\frac{1}{NT}\sum_{s=0}^{\infty}\sum_{k=1}^{N}\mathbf{\Gamma}_k\left[\mathbf{V}_k'\mathbf{V}_{-s}^{(\ell)} - \mathbb{E}(\mathbf{V}_k'\mathbf{V}_{-s}^{(\ell)})\right]\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$= \sup_{1\leq j\leq N}\left\|\frac{1}{NT}\sum_{s=0}^{\infty}\sum_{k=1}^{N}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{\Gamma}_k\left[\mathbf{v}_{kt}v_{l,i,t-s} - \mathbb{E}(\mathbf{v}_{kt}v_{l,i,t-s})\right]\beta_{li}a_{ij}^{(s)}\right\|$$

$$\leq \sup_{1\leq i\leq N}|\beta_{li}|\sup_{1\leq j\leq N}\frac{1}{NT}\sum_{i=1}^{N}\left\|\sum_{s=0}^{\infty}\sum_{k=1}^{N}\sum_{t=1}^{T}\mathbf{\Gamma}_k\left[\mathbf{v}_{kt}v_{l,i,t-s} - \mathbb{E}(\mathbf{v}_{kt}v_{l,i,t-s})\right]a_{ij}^{(s)}\right\|$$

$$\leq \sup_{1\leq i\leq N}|\beta_{li}|\sup_{1\leq j\leq N}\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=0}^{\infty}|a_{ij}^{(s)}|\left\|\sum_{k=1}^{N}\sum_{t=1}^{T}\mathbf{\Gamma}_k\left[\mathbf{v}_{kt}v_{l,i,t-s} - \mathbb{E}(\mathbf{v}_{kt}v_{l,i,t-s})\right]\right\|$$

$$= O_p(T^{-1/2}\log N).$$

Hence,

$$\mathbb{M}_3 = O_p(1)\times O_p(T^{-1/2}\log N) = O_p(T^{-1/2}\log N). \tag{B.80}$$

Analogously, using Assumption B4 and Lemma B.1(a), we can show that

$$\mathbb{M}_5 = O_p(T^{-1}\log N), \qquad \mathbb{M}_7 = O_p(T^{1/2}\delta_{NT}^{-1})\cdot O_p(T^{-1}\log N) = O_p(T^{-1/2}\delta_{NT}^{-1}\log N). \tag{B.81}$$

Collecting (B.75)–(B.81), we obtain

$$\sup_{1\leq j\leq N}\left\|\frac{1}{T}\sum_{s=0}^{\infty}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{R})'\mathbf{V}_{-s}^{(\ell)}\boldsymbol{\beta}^{(\ell)}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}[\boldsymbol{\rho}(\mathbf{I}_N - \mathbf{W}\boldsymbol{\Psi})^{-1}]^s\boldsymbol{\iota}_j\right\|$$

$$= O_p((N^{-1/2} + T^{-1/2})\log N) = O_p(\delta_{NT}^{-1}\log N).$$

Combining this result with (B.49) and (B.71), we obtain

$$\mathbb{F}_{1.2.2} = O_p(N^{1/4}\delta_{NT}^{-1}\log N + N^{1/2}T^{-1/2}\delta_{NT}^{-1}\log N) \tag{B.82}$$

and

$$\mathbb{F}_{1.2.3} = O_p(N^{1/4}\delta_{NT}^{-3}\log N + N^{1/2}T^{-1/2}\delta_{NT}^{-3}\log N), \tag{B.83}$$

respectively. In view of (B.70), (B.73), (B.74), (B.82), and (B.83), we have

$$\mathbb{F}_{1.2} = O_p\big(N^{1/4}\delta_{NT}^{-1}\log N + N^{1/2}T^{-1/2}\delta_{NT}^{-1}\log N + N^{3/4}T^{-1/2}\delta_{NT}^{-2}\log N + NT^{-1}\delta_{NT}^{-2}\log N\big). \tag{B.84}$$

Following the proof of (B.51) for $\mathbb{F}_2$ and the proof of (B.84) for $\mathbb{F}_{1.2}$, we can show that

$$\mathbb{F}_{1.1} = O_p(N^{1/2}\delta_{NT}^{-2}\log N + N^{3/4}T^{-1/2}\delta_{NT}^{-2}\log N). \tag{B.85}$$

Combining (B.69) with (B.84) and (B.85), we obtain

$$\mathbb{F}_1 = O_p\big(N^{1/2}\delta_{NT}^{-2}\log N + N^{3/4}T^{-1/2}\delta_{NT}^{-2}\log N + N^{1/4}\delta_{NT}^{-1}\log N + NT^{-1}\delta_{NT}^{-2}\log N\big), \tag{B.86}$$

where we have used the result that $\delta_{NT}^{-1} \geq T^{-1/2}$ and hence, $N^{1/2}\delta_{NT}^{-2}\log N \geq N^{1/2}T^{-1/2}\delta_{NT}^{-1}\log N$. Collecting (B.42), (B.51), (B.52), (B.68), and (B.86), we have

$$\sup_{1\leq i,j\leq N} T^{-1} \left\| \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{y}_j - \mathbf{X}_i'\mathbf{M}_{\mathbf{F}}\mathbf{y}_j \right\|$$
$$=O_p\big(N^{1/2}\delta_{NT}^{-2}\log N + N^{3/4}T^{-1/2}\delta_{NT}^{-2}\log N + N^{1/4}\delta_{NT}^{-1}\log N$$
$$+ NT^{-1}\delta_{NT}^{-2}\log N + N^{3/4}T^{-1/2}\delta_{NT} + NT^{-1}\log N\big),$$

which proves Lemma B.6(a). With similar arguments, we can prove B.6(b).

For Lemma B.6(c), first note that

$$\sup_{1\leq i,j\leq N} \frac{1}{T} \left\| \mathbf{X}_{i,-1}'\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{y}_j - \mathbf{X}_{i,-1}'\mathbf{M}_{\mathbf{F}_{-1}}\mathbf{M}_{\mathbf{F}}\mathbf{y}_j \right\|$$
$$\leq \sup_{1\leq i,j\leq N} \frac{1}{T} \left\| \mathbf{X}_{i,-1}'\big(\mathbf{M}_{\widehat{\mathbf{F}}_{-1}} - \mathbf{M}_{\mathbf{F}_{-1}}\big)\mathbf{M}_{\mathbf{F}}\mathbf{y}_j \right\| + \sup_{1\leq i,j\leq N} \frac{1}{T} \left\| \mathbf{X}_{i,-1}'\mathbf{M}_{\mathbf{F}_{-1}}\big(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}\big)\mathbf{y}_j \right\|$$
$$+ \sup_{1\leq i,j\leq N} \frac{1}{T} \left\| \mathbf{X}_{i,-1}'\big(\mathbf{M}_{\widehat{\mathbf{F}}_{-1}} - \mathbf{M}_{\mathbf{F}_{-1}}\big)\big(\mathbf{M}_{\widehat{\mathbf{F}}} - \mathbf{M}_{\mathbf{F}}\big)\mathbf{y}_j \right\|$$
$$=: \mathbb{H}_1 + \mathbb{H}_2 + \mathbb{H}_3. \tag{B.87}$$

Following similar arguments in the proof of Lemma B.6(a), we can show that

$$\mathbb{H}_1 = O_p(\iota_{NT}), \qquad \mathbb{H}_2 = O_p(\iota_{NT}). \tag{B.88}$$

Furthermore, by Lemma B.1(f), we can prove that

$$\mathbb{H}_3 = O_p(\delta_{NT}^{-1}\iota_{NT}). \tag{B.89}$$

With (B.87)–(B.89), we readily have the result in Lemma B.6(c). Similarly, we can prove Lemma B.6(d). This completes the proof of Lemma B.6. $\qquad\square$

Appendix C: Additional Simulation Results

In this appendix, we present additional simulation results. Table C.1 below presents results for $\beta_1 = 3$ for the same design as in Section 4, with the matrix of instruments given by Eq. (4.11).

Table C.1: Simulation results for $\beta_2 = 3$.

Panel A ($\pi_u = 3/4$)

	IVMG					2SIV					

Case I: $N = 25\tau$, $T = 100\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	2.996	0.063	0.12	0.048	0.366	2.99	0.062	0.333	0.085	0.289	0.12
2	2.998	0.032	0.065	0.044	0.862	2.991	0.033	0.3	0.068	0.782	0.235
3	2.999	0.017	0.021	0.045	1	2.992	0.02	0.267	0.081	0.998	0.751

Case II: $N = 100\tau$, $T = 25\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	3.004	0.08	0.147	0.057	0.279	3.001	0.063	0.033	0.063	0.368	0.096
2	3	0.032	0.005	0.042	0.878	2.999	0.03	0.033	0.057	0.899	0.232
3	2.998	0.016	0.072	0.047	1	2.996	0.016	0.133	0.054	1	0.78

Case III: $N = 50\tau$, $T = 50\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	2.992	0.066	0.254	0.045	0.319	2.997	0.06	0.1	0.063	0.366	0.107
2	2.998	0.031	0.082	0.046	0.877	2.994	0.032	0.2	0.063	0.836	0.231
3	2.998	0.016	0.063	0.051	1	2.993	0.018	0.233	0.077	1	0.768

Panel B ($\pi_u = 1/4$)

	IV					2SIV					

Case I: $N = 25\tau$, $T = 100\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	2.991	0.061	0.294	0.044	0.353	2.989	0.059	0.367	0.086	0.311	0.139
2	2.995	0.031	0.176	0.04	0.865	2.992	0.031	0.279	0.065	0.824	0.284
3	2.997	0.017	0.086	0.04	1	2.992	0.019	0.275	0.075	0.999	0.868

Case II: $N = 100\tau$, $T = 25\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	3.002	0.07	0.056	0.05	0.328	3	0.057	0.012	0.064	0.436	0.107
2	2.994	0.031	0.199	0.048	0.846	2.999	0.028	0.019	0.053	0.931	0.284
3	2.993	0.017	0.232	0.067	1	2.996	0.015	0.139	0.052	1	0.882

Case III: $N = 50\tau$, $T = 50\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	2.989	0.063	0.368	0.037	0.352	2.996	0.055	0.132	0.059	0.416	0.1
2	2.992	0.031	0.251	0.05	0.846	2.994	0.03	0.215	0.061	0.858	0.287
3	2.995	0.016	0.173	0.054	1	2.993	0.017	0.22	0.07	1	0.864

Tables C.2–C.5 below present results for the same design as in Section 4, except that the matrix of instruments is given by Eq. (4.12). The results are qualitatively similar when it comes to the performance of IVMG. The same holds true for 2SIV except that the overidentifying restrictions test statistic has very little power to detect violation of the null of slope parameter homogeneity. This implies that in order to achieve satisfactory power for the J test, one needs to include a sufficient number of lags of $\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i$ as instruments.

Table C.2: Simulation results for $\rho = 0.4$.

Panel A ($\pi_u = 3/4$)

	IVMG					2SIV					
					Case I: $N = 25\tau$, $T = 100\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.401	0.027	0.219	0.061	0.94	0.409	0.03	2.222	0.129	0.926	0.093
2	0.402	0.018	0.363	0.063	1	0.411	0.022	2.848	0.132	1	0.069
3	0.401	0.012	0.268	0.054	1	0.413	0.017	3.17	0.193	1	0.092
					Case II: $N = 100\tau$, $T = 25\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.397	0.02	0.853	0.061	0.993	0.409	0.023	2.248	0.092	0.998	0.063
2	0.404	0.012	0.91	0.077	1	0.412	0.017	2.953	0.197	1	0.061
3	0.403	0.007	0.725	0.076	1	0.413	0.015	3.209	0.46	1	0.074
					Case III: $N = 50\tau$, $T = 50\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.403	0.022	0.804	0.058	0.994	0.411	0.026	2.756	0.111	0.991	0.059
2	0.403	0.014	0.68	0.058	1	0.412	0.018	3.04	0.158	1	0.055
3	0.402	0.009	0.385	0.062	1	0.413	0.016	3.197	0.313	1	0.087

Panel B ($\pi_u = 1/4$)

	IV					2SIV					
					Case I: $N = 25\tau$, $T = 100\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.404	0.026	0.987	0.063	0.967	0.409	0.029	2.213	0.113	0.962	0.093
2	0.403	0.018	0.722	0.063	1	0.412	0.021	2.866	0.141	1	0.076
3	0.402	0.012	0.476	0.058	1	0.413	0.017	3.18	0.195	1	0.093
					Case II: $N = 100\tau$, $T = 25\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.405	0.02	0.602	0.082	0.999	0.41	0.021	2.432	0.096	0.999	0.058
2	0.403	0.014	0.545	0.069	1	0.412	0.016	2.952	0.203	1	0.066
3	0.401	0.009	0.41	0.061	1	0.413	0.015	3.223	0.484	1	0.095
					Case III: $N = 50\tau$, $T = 50\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.407	0.022	1.845	0.078	0.998	0.411	0.025	2.689	0.122	0.994	0.079
2	0.405	0.014	1.315	0.077	1	0.412	0.018	3.064	0.169	1	0.067
3	0.403	0.009	0.765	0.069	1	0.413	0.016	3.212	0.325	1	0.096

Table C.3: Simulation results for $\psi = 0.25$.

			IVMG					**2SIV**			
					Panel A ($\pi_u = 3/4$)						
					Case I: $N = 25\tau$, $T = 100\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.251	0.028	0.532	0.067	0.945	0.248	0.027	0.637	0.101	0.943	0.093
2	0.25	0.016	0.038	0.061	1	0.249	0.016	0.482	0.074	1	0.069
3	0.25	0.01	0.141	0.052	1	0.249	0.01	0.595	0.064	1	0.092
					Case II: $N = 100\tau$, $T = 25\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.255	0.03	2.118	0.054	0.939	0.249	0.024	0.467	0.063	0.974	0.063
2	0.25	0.013	0.012	0.057	1	0.248	0.012	0.708	0.062	1	0.061
3	0.25	0.007	0.023	0.053	1	0.249	0.007	0.433	0.056	1	0.074
					Case III: $N = 50\tau$, $T = 50\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.255	0.027	1.891	0.055	0.959	0.248	0.024	0.673	0.072	0.975	0.059
2	0.25	0.013	0.068	0.052	1	0.249	0.013	0.613	0.068	1	0.055
3	0.25	0.008	0.031	0.045	1	0.249	0.008	0.509	0.058	1	0.087

			IV					**2SIV**			
					Panel B ($\pi_u = 1/4$)						
					Case I: $N = 25\tau$, $T = 100\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.25	0.028	0.122	0.066	0.91	0.249	0.026	0.59	0.088	0.952	0.093
2	0.25	0.016	0.083	0.056	1	0.249	0.016	0.473	0.069	1	0.076
3	0.249	0.01	0.247	0.053	1	0.249	0.01	0.599	0.066	1	0.093
					Case II: $N = 100\tau$, $T = 25\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.253	0.027	1.155	0.055	0.945	0.248	0.023	0.717	0.059	0.983	0.058
2	0.249	0.013	0.375	0.048	1	0.248	0.012	0.624	0.066	1	0.066
3	0.249	0.007	0.245	0.054	1	0.249	0.007	0.417	0.055	1	0.095
					Case III: $N = 50\tau$, $T = 50\tau$						
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.253	0.027	1.025	0.062	0.95	0.248	0.024	0.622	0.075	0.97	0.079
2	0.25	0.014	0.114	0.052	1	0.248	0.014	0.635	0.067	1	0.067
3	0.25	0.008	0.17	0.043	1	0.249	0.008	0.505	0.053	1	0.096

Table C.4: Simulation results for $\beta_1 = 3$.

Panel A ($\pi_u = 3/4$)

	IVMG					2SIV					

Case I: $N = 25\tau$, $T = 100\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	2.994	0.063	0.207	0.048	0.332	2.998	0.062	0.053	0.086	0.348	0.093
2	2.996	0.032	0.12	0.049	0.836	3	0.033	0.006	0.057	0.856	0.069
3	2.999	0.018	0.048	0.047	1	3.001	0.019	0.023	0.055	1	0.092

Case II: $N = 100\tau$, $T = 25\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	3.002	0.079	0.077	0.053	0.274	3.006	0.064	0.184	0.065	0.382	0.063
2	2.997	0.033	0.095	0.047	0.839	3.006	0.031	0.206	0.058	0.921	0.061
3	2.996	0.016	0.149	0.052	1	3.004	0.016	0.136	0.055	1	0.074

Case III: $N = 50\tau$, $T = 50\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	2.99	0.066	0.318	0.043	0.313	3.003	0.061	0.109	0.065	0.397	0.059
2	2.995	0.032	0.171	0.048	0.845	3.002	0.032	0.072	0.062	0.886	0.055
3	2.996	0.017	0.12	0.055	1	3.002	0.017	0.068	0.052	1	0.087

Panel B ($\pi_u = 1/4$)

	IV					2SIV					

Case I: $N = 25\tau$, $T = 100\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	2.989	0.062	0.38	0.042	0.333	2.997	0.059	0.105	0.074	0.382	0.093
2	2.993	0.031	0.232	0.047	0.834	3	0.031	0.004	0.062	0.891	0.076
3	2.997	0.017	0.115	0.041	1	3.001	0.018	0.019	0.052	1	0.093

Case II: $N = 100\tau$, $T = 25\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	2.999	0.07	0.025	0.046	0.315	3.005	0.058	0.181	0.065	0.449	0.058
2	2.991	0.033	0.286	0.06	0.789	3.007	0.029	0.223	0.051	0.959	0.066
3	2.991	0.018	0.305	0.088	0.999	3.004	0.015	0.134	0.051	1	0.095

Case III: $N = 50\tau$, $T = 50\tau$

τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	2.987	0.063	0.422	0.04	0.31	3.003	0.056	0.085	0.058	0.441	0.079
2	2.99	0.031	0.327	0.052	0.827	3.002	0.029	0.054	0.054	0.924	0.067
3	2.993	0.017	0.226	0.06	1	3.002	0.016	0.065	0.046	1	0.096

Table C.5: Simulation results for $\beta_2 = 1$.

Panel A ($\pi_u = 3/4$)

	IVMG					2SIV					
				Case I: $N = 25\tau,\ T = 100\tau$							
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.998	0.063	0.254	0.057	0.366	1.01	0.063	0.986	0.097	0.413	0.093
2	0.999	0.031	0.147	0.054	0.869	1.008	0.033	0.811	0.073	0.903	0.069
3	1	0.017	0.026	0.06	1	1.009	0.021	0.879	0.089	1	0.092
				Case II: $N = 100\tau,\ T = 25\tau$							
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	1.005	0.082	0.478	0.07	0.258	1.027	0.075	2.75	0.118	0.339	0.063
2	0.999	0.033	0.096	0.059	0.834	1.015	0.034	1.542	0.094	0.948	0.061
3	0.999	0.015	0.059	0.049	1	1.013	0.02	1.312	0.141	1	0.074
				Case III: $N = 50\tau,\ T = 50\tau$							
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.999	0.068	0.096	0.058	0.34	1.016	0.065	1.642	0.097	0.418	0.059
2	0.999	0.03	0.152	0.052	0.889	1.012	0.033	1.213	0.083	0.92	0.055
3	0.999	0.015	0.136	0.044	1	1.01	0.018	0.964	0.086	1	0.087

Panel B ($\pi_u = 1/4$)

	IV					2SIV					
				Case I: $N = 25\tau,\ T = 100\tau$							
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.995	0.063	0.462	0.062	0.341	1.009	0.061	0.895	0.091	0.439	0.093
2	0.998	0.032	0.25	0.052	0.851	1.008	0.033	0.821	0.073	0.911	0.076
3	0.999	0.018	0.067	0.054	1	1.009	0.021	0.888	0.096	1	0.093
				Case II: $N = 100\tau,\ T = 25\tau$							
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.998	0.072	0.176	0.057	0.284	1.02	0.06	2.008	0.077	0.513	0.058
2	0.997	0.033	0.34	0.056	0.828	1.015	0.032	1.494	0.092	0.95	0.066
3	0.998	0.015	0.237	0.053	1	1.013	0.02	1.287	0.148	1	0.095
				Case III: $N = 50\tau,\ T = 50\tau$							
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power	Size J
1	0.997	0.067	0.274	0.064	0.334	1.015	0.061	1.5	0.092	0.45	0.079
2	0.997	0.031	0.321	0.054	0.865	1.012	0.032	1.217	0.096	0.932	0.067
3	0.997	0.016	0.259	0.045	1	1.01	0.018	0.953	0.082	1	0.096