The Econometrics Journal

Econometrics Journal (2023), volume **26**, pp. 124–146.
https://doi.org/10.1093/ectj/utac026

IV estimation of spatial dynamic panels with interactive effects: large sample theory and an application on bank attitude towards risk

GUOWEI CUI[†], VASILIS SARAFIDIS[‡] AND TAKASHI YAMAGATA[§]

[†]*School of Economics, Huazhong University of Science and Technology (HUST), Wuhan, 430074, China*

[‡]*Department of Economics, BI Norwegian Business School, 0484, Oslo, Norway*
E-mail: vasilis.sarafidis@bi.no

[§]*DERS, University of York, UK, and ISER, Osaka University, Japan*
E-mail: takashi.yamagata@york.ac.uk

First version received: 13 January 2022; final version accepted: 20 June 2022.

Summary: This paper develops a new instrumental variables estimator for spatial, dynamic panels with interactive effects under large N and T asymptotics. For this class of models, most approaches available in the literature are based on quasi-maximum likelihood estimation. The approach put forward here is appealing from both a theoretical and a practical point of view for a number of reasons. First, it is linear in the parameters of interest and computationally inexpensive. Second, the IV estimator is free from asymptotic bias. Third, the approach can accommodate endogenous regressors as long as external instruments are available. The IV estimator is consistent and asymptotically normal as $N, T \rightarrow \infty$, such that $N/T \rightarrow c$, where $0 < c < \infty$. We study the determinants of risk attitude of banking institutions. The results show that the capital regulation introduced by the Dodd–Frank Act has succeeded in influencing banks' behaviour.

Keywords: *Bank risk behaviour*, *capital regulation*, *common factors*, *instrumental variables*, *large N and T asymptotics*, *panel data*, *social interactions*, *state dependence*.

JEL codes: *C23*, *C26*, *C38*, *C51*, *G21*.

1. INTRODUCTION

Economic behaviour is intrinsically dynamic; that is, it is influenced by past own behaviour. This phenomenon, commonly described as 'state dependence', is due to habit formation, costs of adjustment and economic slack, among other factors.[1]

More recently, it has been emphatically pointed out that an economic agent's own behaviour is also influenced by the behaviour of other agents, typically their peers. This is due to network linkages, social interactions, and spillover effects—see, e.g., the pioneering work of Case (1991) and Manski (1993). At the same time, agents inhabit a common economic environment and so their behaviour is subject to aggregate (global) shocks, which may be due to shifts in technology

[1] See, e.g., the seminar papers by Balestra and Nerlove (1966), Anderson and Hsiao (1982) and Arellano and Bond (1991). A recent overview of this literature is provided by Bun and Sarafidis (2015).

and productivity, and changes in preferences, to mention only a few (Sarafidis and Wansbeek, 2021).

In panel data analysis, state dependence is characterised using dynamic models; peer effects can be modelled using spatial econometric techniques—see, e.g., Kelejian and Piras (2017) and Jing et al. (2018)—and aggregate shocks are typically captured by common factors, also known as 'interactive effects'.

The present paper develops a new instrumental variables (IV) estimator for spatial, dynamic panel data models with interactive effects under large N and T asymptotics, where N denotes the number of cross-sectional units and T denotes the number of time series observations. For this class of models, most approaches available in the literature are based on quasi-maximum likelihood estimation (QMLE)—see Shi and Lee (2017) and Bai and Li (2021).[2] The approach put forward in this paper is appealing both from a theoretical and from a practical point of view for a number of reasons.

First, the proposed IV estimator is linear in the parameters of interest and it is computationally inexpensive. In contrast, QML estimators are nonlinear and require estimation of the Jacobian matrix of the likelihood function, which may be subject to a high level of numerical complexity in spatial models with N large; see, e.g., Lee and Yu (2015, sec. 12.3.2). To provide some indication of the likely computational gains of our method, in the Monte Carlo section of this paper we found that the total length of time taken to perform 2,000 replications of the model when $N = T = 200$, was roughly 4.5 minutes for IV and 4.5 hours for QMLE. Hence, for every minute of running time using IV, it takes about an hour for QMLE.[3]

Second, the proposed IV approach is free from asymptotic bias. In contrast, existing QML estimators suffer from incidental parameter bias, depending on the sample size and the magnitude of unknown parameters of the data generating process (DGP). Unfortunately, approximate procedures aiming to recentre the limiting distribution of these estimators using first-order bias correction can fail to fully remove the bias in finite samples. This can lead to severe size distortions as confirmed in our Monte Carlo study. Moreover, bias expressions for QMLE are based on the true number of factors which is unknown in practice. Thus, if the number of factors is overestimated, bias correction becomes nontrivial.

Third, the proposed estimator retains the attractive feature of method of moments estimation in that it can potentially accommodate endogenous regressors, so long as external exogenous instruments are available.[4]

In a recent contribution, Chen et al. (2022) put forward an IV estimator for spatial static panels with heterogeneous coefficients and unobserved common factors under large N and T asymptotics. The common factors are controlled using a common correlated effects (CCE) type approach, as in Pesaran (2006).[5] In contrast, in this paper we consider estimation of models with homogeneous slopes and the common factors are estimated using the principal component method, following Bai (2009) and Norkute et al. (2021), among many others. Furthermore, we consider a spatial panel data model with a time lag, as well as a spatial-time lag. These specifications can capture much more complex cross-sectional and dynamic interdependencies

[2] The only exception that we are aware of is a recent paper by Chen et al. (2022), discussed further below.

[3] This ratio appears to decrease (increase) roughly exponentially with smaller (larger) values of N.

[4] Even in cases where such instruments are not easy to find, our approach provides a framework for testing for endogeneity, based on the overidentifying restrictions test statistic. In contrast, the exogeneity restriction is difficult to verify outside the IV framework and so it is typically taken for granted.

[5] As the factors are approximated using cross-sectional averages, this method crucially relies upon the rank condition, as it is the case with all CCE-type estimators.

than Chen et al. (2022). Due to these two features, the theoretical analysis differs substantially from that of Chen et al. (2022). On the contrary, the analysis of spatial dynamic panels with heterogeneous slopes is not a trivial extension of this paper and will not be considered.

There is substantial literature on dynamic panels under large N and T asymptotics—e.g., Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003), among others. More recently, several new methods have been developed to control for unobserved common shocks and strong cross-sectional dependence—see, e.g., Chudik and Pesaran (2015), Everaert and De Groote (2016), Moon and Weidner (2017), Juodis et al. (2021), Norkute et al. (2021), and Juodis and Sarafidis (2022). However, none of these papers considers spatial interactions and endogenous network effects.

There is also substantial literature on spatial panel data analysis and social interactions, which, however, mostly ignores the potential presence of common unobserved shocks. Some notable contributions include Yu et al. (2008), Korniotis (2010), Debarsy et al. (2012), and Lee and Yu (2014), among others.

The present paper sits on the intersection of the above two strands of literature. Despite the fact that such an intersection is highly relevant for the analysis of economic behaviour, the field is fairly new in the econometrics literature and, as such, it is sparse.

We put forward a two-step IV estimation approach that extends the methodology of Norkute et al. (2021) and Cui et al. (2022) to the case of panel data models with spatial interactions, in addition to state dependence and interactive effects. Unlike Norkute et al. (2021), where the moment conditions are independent and identically distributed (i.i.d.) conditional on the factors, in the present case the moment conditions are weakly correlated across i even conditional on the factors. Therefore, a central limit theorem for martingale differences is required, as in Kelejian and Prucha (2001). Our two-step procedure can be outlined as follows: in the first step, the common factors in the exogenous covariates are projected out using principal components analysis, as in Bai (2003). Next, the slope parameters are estimated using standard IV regression, which makes use of instruments constructed from defactored regressors. In the second step, the entire model is defactored based on factors extracted from the first step residuals. Subsequently, an IV regression is implemented again using the same instruments.

The proposed IV estimator is consistent and asymptotically normally distributed as $N, T \to \infty$ such that $N/T \to c$, where $0 < c < \infty$. Moreover, the estimator does not have asymptotic bias in either cross-sectional or time-series dimension. The main intuition of this result lies in that we extract factor estimates from two sets of information that are mutually independent, namely the exogenous covariates and the regression residuals. Therefore, there is no correlation between the regressors and the estimation error of the interactive fixed effects obtained in the second step. In addition, the proposed estimator is not subject to 'Nickell bias' that arises with QML-type estimators in dynamic panel data models.

We study the determinants of risk attitude of banking institutions, with emphasis on the impact of increased international capital regulation. The results bear important policy implications, and provide evidence that the more risk-sensitive capital regulation introduced by the Dodd–Frank Act in late 2010 has succeeded in influencing banks' behaviour in a substantial manner.

Throughout, we denote the largest and the smallest eigenvalues of the $N \times N$ matrix $\mathbf{A} = (a_{ij})$ by $\mu_{\max}(\mathbf{A})$ and $\mu_{\min}(\mathbf{A})$, respectively, its trace by $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{N} a_{ii}$, its column sum norm by $\|\mathbf{A}\|_1 = \max_{1\leq j\leq N} \sum_{i=1}^{N} |a_{ij}|$, its Frobenius norm by $\|\mathbf{A}\| = \sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})}$, and its row sum norm by $\|\mathbf{A}\|_\infty = \max_{1\leq i\leq N} \sum_{j=1}^{N} |a_{ij}|$. The projection matrix on $\mathbf{A}$ is $\mathbf{P}_\mathbf{A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ and $\mathbf{M}_\mathbf{A} =$

$\mathbf{I} - \mathbf{P}_{\mathbf{A}}$. C is a generic positive constant large enough, $\delta_{NT}^2 = \min\{N, T\}$. We use $N, T \to \infty$ to denote that N and T pass to infinity jointly.

2. MODEL AND TWO-STEP ESTIMATION APPROACH

In our baseline setup, we consider the following spatial dynamic panel data model with exogenous covariates:[6]

$$y_{it} = \psi \sum_{j=1}^{N} w_{ij} y_{jt} + \rho y_{it-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\varphi}_i^{0\prime}\mathbf{h}_t^0 + \varepsilon_{it}, \tag{2.1}$$

$i = 1, 2, \ldots, N, t = 1, 2, \ldots, T$, where y_{it} denotes the observation on the dependent variable for individual unit i at time period t, and $\mathbf{x}_{it}$ is a $k \times 1$ vector of regressors with slope coefficients $\boldsymbol{\beta}$. The spatial variable $\sum_{j=1}^{N} w_{ij} y_{jt}$ picks up endogenous network effects with corresponding parameter ψ. w_{ij} denotes the (i, j)th element of the $N \times N$ spatial weights matrix $\mathbf{W}_N$, which is assumed to be known. The lagged dependent variable captures dynamic or temporal effects. The error term of the model is composite: $\mathbf{h}_t^0$ and $\boldsymbol{\varphi}_i^0$ denote $r_y \times 1$ vectors of latent factors and factor loadings, respectively, and ε_{it} is an idiosyncratic error.

To ensure that the covariates are endogenous to the factor component, we assume that

$$\mathbf{x}_{it} = \boldsymbol{\Gamma}_i^{0\prime}\mathbf{f}_t^0 + \mathbf{v}_{it}, \tag{2.2}$$

where $\mathbf{f}_t^0$ denotes an $r_x \times 1$ vector of latent factors, $\boldsymbol{\Gamma}_i^0$ denotes an $r_x \times k$ factor loading matrix, while $\mathbf{v}_{it}$ is an idiosyncratic disturbance of dimension $k \times 1$. Note that $\mathbf{h}_t^0$ and $\mathbf{f}_t^0$ can be identical, share some common factors, or they can be completely different, but may be mutually correlated. Similarly, $\boldsymbol{\varphi}_i^0$ and $\boldsymbol{\Gamma}_i^0$ can be mutually correlated.[7] While the linear factor structure poses certain restrictions on the DGP, as argued by Freeman and Weidner (2021), it still provides a good approximation to more complex models in the sense that one can let the number of estimated factors grow asymptotically without consequences.

In the context of spatial panels, the above structure of the covariates has also been studied by Bai and Li (2013). The main difference between these two specifications is that the model in (2.1) allows for dynamics through the lagged dependent variable. Moreover, the covariates in (2.2) are not necessarily driven by the same factors as those entering into the error term of y. This has an appealing generality in that the common shocks that hit y and X may not be identical in practice.

Stacking the T observations for each i yields

$$\begin{aligned} \mathbf{y}_i &= \psi \mathbf{Y}\mathbf{w}_i + \rho \mathbf{y}_{i,-1} + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{H}^0\boldsymbol{\varphi}_i^0 + \boldsymbol{\varepsilon}_i; \\ \mathbf{X}_i &= \mathbf{F}^0\boldsymbol{\Gamma}_i^0 + \mathbf{V}_i, \end{aligned} \tag{2.3}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{it})'$, $\mathbf{y}_{i,-1} = (y_{i0}, \ldots, y_{i,T-1})'$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{it})'$ denote $T \times 1$ vectors, $\mathbf{X}_i = (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{it})'$ and $\mathbf{V}_i = (\mathbf{v}_{i1}, \ldots, \mathbf{v}_{it})'$ are matrices of order $T \times k$, while $\mathbf{H}^0 = (\mathbf{h}_1^0, \cdots, \mathbf{h}_T^0)'$ and $\mathbf{F}^0 = (\mathbf{f}_1^0, \cdots, \mathbf{f}_T^0)'$ are of dimensions $T \times r_y$ and $T \times r_x$, respectively.

[6] An extension of this model that allows for a spatial-time lag is analysed in Section 3.3. Exogenous network effects, e.g., through an additional term $\sum_{j=1}^{N} w_{ij}\mathbf{x}_{jt}'\boldsymbol{\delta}$, and lagged values of y_{it-1} can be accommodated in a straightforward manner without affecting the main derivations of the paper.

[7] Without loss of generality, r_y and r_x are treated as known. In practice, the number of factors can be estimated consistently using, e.g., the information criteria of Bai and Ng (2002), or the eigenvalue ratio test of Ahn and Horenstein (2013). The results of the Monte Carlo section indicate that these methods provide quite accurate estimates in our design.

Finally, $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_T)'$ denotes a $T \times N$, matrix and the $N \times 1$ vector $\mathbf{w}_i$ represents the ith row of $\mathbf{W}_N$.

The model in (2.3) can be written more succinctly as follows:

$$\mathbf{y}_i = \mathbf{C}_i \boldsymbol{\theta} + \mathbf{u}_i,$$

where $\mathbf{C}_i = (\mathbf{Y}\mathbf{w}_i, \mathbf{y}_{i,-1}, \mathbf{X}_i)$, $\boldsymbol{\theta} = \left(\psi, \rho, \boldsymbol{\beta}'\right)'$ and $\mathbf{u}_i = \mathbf{H}^0 \boldsymbol{\varphi}_i^0 + \boldsymbol{\varepsilon}_i$.

Let $\mathbf{X}_{i,-\tau} \equiv L^\tau \mathbf{X}_i$, where L^τ denotes the time-series lag operator of order τ. We make use of the convention $\mathbf{X}_{i,-0} = \mathbf{X}_i$. Our estimation approach involves two steps. In the first step, the common factors in $\mathbf{X}_{i,-\tau}$ are asymptotically eliminated using principal component analysis, as advanced by Bai (2003). Next, instruments are constructed using defactored covariates. The resulting first-step IV estimator of $\boldsymbol{\theta}$ is consistent. In the second step, the entire model is defactored based on estimated factors extracted from the first step IV residuals. Subsequently, a second IV regression is implemented, using the same instruments as in step one. That is, an IV regression is implemented in both of two stages.

In particular, define $\widehat{\mathbf{F}}_{-\tau}$ as $\sqrt{T}$ times the eigenvectors corresponding to the r_x largest eigenvalues of the $T \times T$ matrices $(NT)^{-1} \sum_{i=1}^N \mathbf{X}_{i,-\tau} \mathbf{X}'_{i,-\tau}$ for $\tau = 0, 1$.

The matrix of instruments is formulated as follows:

$$\widehat{\mathbf{Z}}_i = \left(\sum_{j=1}^N w_{ij} \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_j, \quad \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \mathbf{X}_{i,-1}, \quad \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i \right), \tag{2.4}$$

which is of dimension $T \times 3K$.[8]

The first-step IV estimator of $\boldsymbol{\theta}$ is defined as:

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

where

$$\widehat{\mathbf{A}} = \frac{1}{NT} \sum_{i=1}^N \widehat{\mathbf{Z}}'_i \mathbf{C}_i; \quad \widehat{\mathbf{B}} = \frac{1}{NT} \sum_{i=1}^N \widehat{\mathbf{Z}}'_i \widehat{\mathbf{Z}}_i; \quad \widehat{\mathbf{c}}_y = \frac{1}{NT} \sum_{i=1}^N \widehat{\mathbf{Z}}'_i \mathbf{y}_i.$$

Under certain regularity conditions, $\widehat{\boldsymbol{\theta}}$ is consistent (see Theorem 3.1 in Section 3), although asymptotically biased. Rather than bias-correcting this estimator, we put forward a second-step estimator, which is free from asymptotic bias and is potentially more efficient.

REMARK 2.1. Since our approach makes use of the defactored covariates as instruments, identification of the autoregressive and spatial parameters requires that *at least one* element of $\boldsymbol{\beta}$ is not equal to zero. Otherwise, it is easily seen that identification of ρ and ψ is not possible, since the lagged and spatial defactored covariates become irrelevant instruments. This requirement is mild and common in the estimation of spatial models using method of moments, see, e.g., Kelejian and Prucha (2007). Note that it is not necessary to know a priori which covariates have

[8] More instruments can be used with respect to further lags of $\mathbf{X}_i$ or spatial lags $\sum_{j=1}^N w_{ij} \mathbf{X}_{j,-\tau}$, for $\tau \geq 1$. Instruments constructed from powers of the spatial weights matrix can also be used, such as $\sum_{j=1}^N w_{ij}^{(\ell)} \mathbf{X}_j$, for $\ell = 2, 3, \ldots$, where $w_{ij}^{(\ell)}$ denotes the (i, j)th element of the $N \times N$ spatial weights matrix $\mathbf{W}_N^\ell$, which is defined as the product matrix taking $\mathbf{W}_N$ and multiplying it by itself ℓ-times. It is well documented in the literature that including a larger number of instruments may render the IV estimator more efficient, although such practice can also potentially magnify small sample bias. In principle, one could devise a lag selection procedure for optimising the bias-variance trade-off for the IV estimator, as per Okui (2009); however, we leave this avenue for future research. The present paper assumes that both $\tau \geq 1$ and $\ell \geq 1$ are small and do not depend on T.

nonzero coefficients, since by construction IV regression does not require all instruments to be relevant to all endogenous regressors.

To implement the second step, we estimate the space spanned by $\mathbf{H}^0$ from the first step IV residuals, i.e., $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{C}_i\hat{\boldsymbol{\theta}}$. To be specific, let $\widehat{\mathbf{H}}$ be defined as $\sqrt{T}$ times the eigenvectors corresponding to the r_y largest eigenvalues of the $T \times T$ matrix $(NT)^{-1}\sum_{i=1}^{N}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'$.

The proposed second-step IV estimator for $\boldsymbol{\theta}$ is defined as follows:

$$\widetilde{\boldsymbol{\theta}} = (\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}})^{-1}\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{c}}_y, \tag{2.5}$$

where

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

Section 3 shows that the second-step IV estimator is normally distributed and correctly centred around the true value.

A particularly useful diagnostic is the so-called overidentifying restrictions (J) test statistic, which is given by

$$J = \frac{1}{NT}\left(\sum_{i=1}^{N}\widetilde{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right)\widehat{\boldsymbol{\Omega}}^{-1}\left(\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widetilde{\mathbf{u}}_i\right), \tag{2.6}$$

where $\widetilde{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{C}_i\tilde{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Omega}} = \widetilde{\sigma}_\varepsilon^2\widetilde{\mathbf{B}}$ with $\widetilde{\sigma}_\varepsilon^2 = \sum_{i=1}^{N}\widetilde{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widetilde{\mathbf{u}}_i/NT$.

REMARK 2.2. The validity of our procedure crucially relies on the assumption that $\mathbf{X}_i$ is strictly exogenous with respect to $\boldsymbol{\varepsilon}_i$. Violations of such restriction are detectable using the J-test above. When strict exogeneity of $\mathbf{X}_i$ fails, identification of the model parameters requires the use of external instruments. These instruments can still be correlated with the common factor component, although they need to be exogenous with respect to $\boldsymbol{\varepsilon}_i$. The theoretical analysis of our approach based on external instruments remains exactly identical, with $\mathbf{X}_i$ in (2.4) replaced by the external instruments. As it is common practice in the literature—e.g., Robertson and Sarafidis (2015) and Kuersteiner and Prucha (2020)—in what follows, we do not explicitly account for this possibility in order to avoid the cost of additional notation to separate covariates that can be used as instruments from those that cannot. Finite sample results for a model with endogenous regressors are provided in Section S3 of the Online Appendix.

3. ASYMPTOTIC PROPERTIES

3.1. Assumptions

Before stating assumptions, define the population version of

$$\mathbf{Z}_i = \left(\sum_{j=1}^{N} w_{ij}\mathbf{M}_{\mathbf{F}^0}\mathbf{V}_j, \mathbf{M}_{\mathbf{F}^0_{-1}}\mathbf{V}_{i,-1}, \mathbf{M}_{\mathbf{F}^0}\mathbf{V}_i\right).$$

The following assumptions are employed throughout the paper.

ASSUMPTION 3.1 (IDIOSYNCRATIC ERROR IN y). *The disturbances ε_{it} are independently distributed across i and over t, with mean zero, $\mathbb{E}(\varepsilon_{it}^2) = \sigma_\varepsilon^2 > 0$ and $\mathbb{E}|\varepsilon_{it}|^{8+\delta} \leq C < \infty$ for some $\delta > 0$.*

ASSUMPTION 3.2 (IDIOSYNCRATIC ERROR IN x). The idiosyncratic error in the DGP for $\mathbf{x}_{it}$ *satisfies the following conditions:*

(1) $\mathbf{v}_{it}$ *is group-wise independent from ε_{it};*
(2) $\mathbb{E}(\mathbf{v}_{it}) = 0$ *and* $\mathbb{E}\|\mathbf{v}_{it}\|^{8+\delta} \leq C < \infty$;
(3) Let $\boldsymbol{\Sigma}_{ij,st} \equiv \mathbb{E}\left(\mathbf{v}_{is}\mathbf{v}_{jt}'\right)$. *We assume that there exist $\bar{\sigma}_{ij}$ and $\tilde{\sigma}_{st}$, $\|\boldsymbol{\Sigma}_{ij,st}\| \leq \bar{\sigma}_{ij}$ for all (s, t), and $\left\|\boldsymbol{\Sigma}_{ij,st}\right\| \leq \tilde{\sigma}_{st}$ for all (i, j), such that*

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\bar{\sigma}_{ij} \leq C < \infty\,,\ \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{\sigma}_{st} \leq C < \infty\,,\ \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\boldsymbol{\Sigma}_{ij,st}\| \leq C < \infty.$$

(4) For every (s, t), $\mathbb{E}\|N^{-1/2}\sum_{i=1}^{N}(\mathbf{v}_{is}\mathbf{v}_{it}' - \boldsymbol{\Sigma}_{ii,st})\|^4 \leq C < \infty$.
(5) The largest eigenvalue of $\mathbb{E}\left(\mathbf{V}_i\mathbf{V}_i'\right)$ *is bounded uniformly in i and T.*
(6) For any h, we have

$$\frac{1}{N}\sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_2=1}^{N}|w_{i_1j_1}||w_{i_2j_2}|\left\|\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{cov}(\mathbf{v}_{hs}\otimes\mathbf{v}_{j_2s},\,\mathbf{v}_{ht}\otimes\mathbf{v}_{j_1t})\right\| \leq C.$$

(7) For any s, we have

$$\mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{h=1}^{N}\sum_{t=1}^{T}\left[\mathbf{v}_{hs}'\mathbf{v}_{ht} - \mathbb{E}\left(\mathbf{v}_{hs}'\mathbf{v}_{ht}\right)\right]\mathbf{f}_t^0\right\|^2 \leq C.$$

(8)

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left\|\mathrm{cov}\left(\mathbf{v}_{is_1}'\mathbf{v}_{is_2},\,\mathbf{v}_{jt_1}'\mathbf{v}_{jt_2}\right)\right\| \leq C.$$

ASSUMPTION 3.3 (FACTORS). $\mathbb{E}\|\mathbf{f}_t^0\|^4 \leq C < \infty$, $T^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0 \xrightarrow{p} \boldsymbol{\Sigma}_F > 0$ *as* $T \to \infty$ *for some nonrandom positive definite matrix* $\boldsymbol{\Sigma}_F$. $\mathbb{E}\|\mathbf{h}_t^0\|^4 \leq C < \infty$, $T^{-1}\mathbf{H}^{0\prime}\mathbf{H}^0 \xrightarrow{p} \boldsymbol{\Sigma}_H > 0$ *as* $T \to \infty$ *for some nonrandom positive definite matrix* $\boldsymbol{\Sigma}_H$. $\mathbf{f}_t^0$ *and* $\mathbf{h}_t^0$ *are group-wise independent from* $\mathbf{v}_{it}$ *and* ε_{it}.

ASSUMPTION 3.4 (LOADINGS). $\boldsymbol{\Gamma}_i^0 \sim$ i.i.d.$(\mathbf{0}, \boldsymbol{\Sigma}_\Gamma)$, $\boldsymbol{\varphi}_i^0 \sim$ i.i.d.$(\mathbf{0}, \boldsymbol{\Sigma}_\varphi)$, *where* $\boldsymbol{\Sigma}_\Gamma$ *and* $\boldsymbol{\Sigma}_\varphi$ *are positive definite.* $\mathbb{E}\left\|\Gamma_i^0\right\|^4 \leq C < \infty$, $\mathbb{E}\left\|\boldsymbol{\varphi}_i^0\right\|^4 \leq C < \infty$. *In addition,* $\boldsymbol{\Gamma}_i^0$ *and* $\boldsymbol{\varphi}_i^0$ *are independent groups from* ε_{it}, $\mathbf{v}_{it}$, $\mathbf{f}_t^0$, *and* $\mathbf{h}_t^0$.

ASSUMPTION 3.5 (WEIGHTING MATRIX). *Denoting the true values of ρ and ψ as ρ^0 and ψ^0, respectively, the weights matrix* $\mathbf{W}_N$ *satisfies the following conditions:*

(1) All diagonal elements of $\mathbf{W}_N$ *are zeros;*
(2) The matrix $\mathbf{I}_N - \psi^0\mathbf{W}_N$ *is invertible;*
(3) The row and column sums of the matrices $\mathbf{W}_N$ *and* $\left(\mathbf{I}_N - \psi^0\mathbf{W}_N\right)^{-1}$ *are bounded uniformly in absolute value.*
(4) $\sum_{\ell=0}^{\infty} \|[\rho^0(\mathbf{I}_N - \psi^0\mathbf{W}_N)^{-1}]^{\ell}\|_{\infty} \leq C; \quad \sum_{\ell=0}^{\infty} \|[\rho^0(\mathbf{I}_N - \psi^0\mathbf{W}_N)^{-1}]^{\ell}\|_{1} \leq C.$

ASSUMPTION 3.6 (IDENTIFICATION). *We assume that*

(1) $\mathbf{A}_0 = plim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{C}_i$ *is fixed with full column rank, and* $\mathbf{B}_0 = plim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i$ *is fixed and positive definite.*
(2) $\mathbb{E}\left\|T^{-1}\mathbf{Z}_i'\mathbf{Z}_i\right\|^{2+2\delta} \leq C < \infty$ *and* $\mathbb{E}\left\|T^{-1}\mathbf{Z}_i'\mathbf{C}_i\right\|^{2+2\delta} \leq C < \infty$ *for all* i *and* T.

Assumption 3.1 is in line with existing spatial literature, see, e.g., Lee and Yu (2014). Cross-sectional/time-series homoscedasticity is imposed in order to simplify the asymptotic analysis for the variance–covariance estimator in panels with N and T both large. In contrast, Norkute et al. (2021) allow for cross-sectional/time-series heteroscedasticity as they invoke the results in Hansen (2007). However, Hansen (2007) assumes that the observations are independent across i, which is not the case here. While we do not formally derive theoretical results under heteroscedasticity, we study the finite sample properties of a robust variance–covariance estimator in Section 4.

Assumption 3.2 implies that $\mathbf{x}_{it}$ is strictly exogenous with respect to ε_{it}, i.e., the defactored regressors are valid instruments (see, e.g., Pesaran, 2006; Bai, 2009). In addition, Assumption 3.2 allows for cross-sectional and time-series heteroscedasticity as well as autocorrelation in $\mathbf{v}_{it}$. Note that, unlike with ε_{it}, here it is important to allow explicitly for this more general setup because, conditional on $\mathbf{F}^0$, the dynamics in $\mathbf{X}_i$ are solely driven by $\mathbf{V}_i$. Note also that in contrast with Norkute et al. (2021), $\mathbf{v}_{it}$ is permitted to be weakly correlated across i, which is in the same spirit as allowing for weak dependence in the process of y.

Assumptions 3.4 and 3.5 are standard in the principal components literature, see, e.g., Bai (2003), among others. Assumption 3.4 permits correlations between $\mathbf{f}_t^0$ and $\mathbf{h}_t^0$, and within each one of them. Assumption 3.5 allows for possible nonzero correlations between $\boldsymbol{\varphi}_i^0$ and $\boldsymbol{\Gamma}_i^0$, and within each one of them. Since for each i, y_{it} and $\mathbf{x}_{it}$ can be affected by common shocks in a related manner, it is potentially important to allow for this possibility in practice.

Assumption 3.5 is standard in the spatial literature, see, e.g., Kelejian and Prucha (2001). In particular, Assumption 3.5(1) is just a normalisation and implies that no individual is viewed as its own neighbour. Assumption 3.5(2) implies that there is no dominant unit, i.e., a unit that is asymptotically correlated with all remaining individuals. Assumptions 3.5(3)–(4) concern the space of the autoregressive and spatial parameters, and are discussed in detail in Kelejian and Prucha (2010, sec. 2.2). Notice that the assumptions above do not depend on a particular ordering of the data, which can be arbitrary so long as Assumption 3.5 holds true. Moreover, $\mathbf{W}_N$ is not required to be row-normalised.

Lastly, Assumption 3.6 ensures IV-based identification, see, e.g., Wooldridge (2002, ch. 5).

3.2. Asymptotic results

The asymptotic properties of the one-step estimator are determined primarily by those of $\widehat{\mathbf{Z}}_i'\mathbf{u}_i/\sqrt{NT}$. The following proposition provides an asymptotic expansion of this term.

Proposition 3.1. *Under Assumptions 3.1–3.6,*

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{Z}_i'\mathbf{u}_i + \sqrt{\frac{T}{N}}\mathbf{b}_1 + \sqrt{\frac{N}{T}}\mathbf{b}_2 + o_p(1),$$

where $\mathbb{Z}_i = \left(\sum_{j=1}^{N} w_{ij}\mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{V}}_j, \mathbf{M}_{\mathbf{F}^0_{-1}}\boldsymbol{\mathcal{V}}_{i,-1}, \mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{V}}_i\right)$ *with* $\boldsymbol{\mathcal{V}}_i = \mathbf{V}_i - \frac{1}{N}\sum_{\ell=1}^{N}\mathbf{V}_\ell\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_i^0$, $\boldsymbol{\mathcal{V}}_{i,-1} = \mathbf{V}_{i,-1} - \frac{1}{N}\sum_{\ell=1}^{N}\mathbf{V}_{\ell,-1}\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_i^0$, $\mathbf{V}_{i,-1} = L\mathbf{V}_i$, $\boldsymbol{\Upsilon}^0 = N^{-1}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i^0\boldsymbol{\Gamma}_i^{0\prime}$, *and* $\mathbf{b}_1$ *and* $\mathbf{b}_2$ *are* $O_p(1)$.[9]

Theorem 3.1 establishes convergence in probability of the one-step IV estimator, $\widehat{\boldsymbol{\theta}}$.

Theorem 3.1. *Under Assumptions 3.1–3.6, as* $N, T \to \infty$ *such that* $N/T \to c$, *where* $0 < c < \infty$, *we have*

$$\sqrt{NT}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = O_p(1).$$

The above theorem implies that $\widehat{\boldsymbol{\theta}}$ can be asymptotically biased. The bias term of the one-step IV estimator arises primarily due to the correlation between the factor loadings associated with $\mathbf{F}^0$ and those associated with $\mathbf{H}^0$.[10] Instead of correcting this bias, we proceed into the second stage of our approach and obtain $\widetilde{\boldsymbol{\theta}}$.

The following proposition provides an asymptotic expansion of $\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i/\sqrt{NT}$, the term that primarily determines the asymptotic properties of the second-step IV estimator.

Proposition 3.2. *Under Assumptions 3.1–3.6, we have*

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{Z}_i'\boldsymbol{\varepsilon}_i + O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{\sqrt{NT}}{\delta_{NT}^3}\right).$$

As we see from Proposition 3.2, the estimation effect in $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i$ can be ignored asymptotically. Since $\boldsymbol{\varepsilon}_i$ is independent of $\mathbf{Z}_i$ and $\mathbf{H}^0$ with zero mean, the limiting distribution of $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i$ is centred at zero. Hence, the asymptotic normality result can be readily obtained by applying the Central Limit Theorem (CLT) for martingale differences in Kelejian and Prucha (2001).

The following theorem formally establishes consistency and asymptotic normality for $\widetilde{\boldsymbol{\theta}}$.

Theorem 3.2. *Under Assumptions 3.1–3.6, as* $N, T \to \infty$ *such that* $N/T \to c$, *where* $0 < c < \infty$, *we have*

$$\sqrt{NT}\left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}),$$

where $\boldsymbol{\Psi} = \sigma_\varepsilon^2\left(\mathbf{A}_0'\mathbf{B}_0^{-1}\mathbf{A}_0\right)^{-1}$. *Moreover,* $\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi} \xrightarrow{p} \mathbf{0}$ *as* $N, T \to \infty$, *where* $\widetilde{\boldsymbol{\Psi}} = \tilde{\sigma}_\varepsilon^2\left(\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}}\right)^{-1}$.

[9] See the proof of Proposition 3.1 in Section S1 of the Online Appendix for explicit expressions of these bias terms. To save space, we do not report these expressions here, given also that we do not bias-correct the first-step estimator.

[10] On the contrary, serial dependence and weak cross-sectional dependence in the idiosyncratic part of the x process, $\mathbf{v}_{it}$, does not result in bias because $\mathbf{v}_{it}$ is not correlated with the error term in the y equation, ε_{it}. Similarly, there is no 'Nickell bias', which typically occurs in the least squares estimation of dynamic panel models, because $\widehat{\boldsymbol{\theta}}$ is based on instrumental variables.

Note that $\tilde{\boldsymbol{\theta}}$ is asymptotically unbiased. This is in stark contrast with existing QML estimators available for spatial panels, which require bias correction.

The main reason for this result is that the estimation error of $\widehat{\mathbf{F}}$ depends on $\mathbf{V}_i$, which is independent from $\mathbf{u}_i$. Therefore, the defactored regressors are asymptotically uncorrelated with the error term of the model.[11] The limiting distribution of the overidentifying restrictions test statistic is established in the following theorem.

THEOREM 3.3. *Under Assumptions 3.1–3.6, as $N, T \to \infty$ such that $N/T \to c$, where $0 < c < \infty$, we have*

$$J \xrightarrow{d} \chi^2_\nu,$$

where $\nu = 3k - (k+2)$.

3.3. Extension to a model with a spatial-time lag

Our approach can be straightforwardly extended to a model with a spatial-time lag, which may capture more complex spatio-temporal dependence—see, e.g., Bai and Li (2021). In particular, we rewrite the model as follows:

$$\mathbf{y}_i = \psi \mathbf{Y}\mathbf{w}_i + \rho \mathbf{y}_{i,-1} + \varrho \mathbf{Y}_{-1}\mathbf{w}_i + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{H}^0\boldsymbol{\varphi}_i^0 + \boldsymbol{\varepsilon}_i, \tag{3.1}$$

where we introduce a spatial-time lag term, $\mathbf{Y}_{-1}\mathbf{w}_i$, with $\mathbf{Y}_{-1} = (\mathbf{y}_0, \cdots, \mathbf{y}_{T-1})'$, a $T \times N$ matrix. For simplicity, we assume that the spatial matrix for this term is $\mathbf{W}_N$, however, a different spatial matrix can be permitted provided it satisfies similar conditions as Assumption 3.5.

The matrix of instruments is given by

$$\widehat{\mathbf{Z}}_i = \left(\sum_{j=1}^N w_{ij}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_j, \quad \mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{i,-1}, \quad \mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i, \quad \sum_{j=1}^N w_{ij}\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{j,-1} \right), \tag{3.2}$$

which is of dimension $T \times 4K$.

For the main results to hold true, the only modification that needs to be made is the condition for model stationarity. Accordingly, the following assumption replaces Assumption 3.5(4):

Assumption 3.5(4′):

$$\sum_{\ell=0}^{\infty} \left\| \left[\rho^0 \left(\mathbf{I}_N - \psi^0\mathbf{W}_N\right)^{-1} + \varrho^0 \left(\mathbf{I}_N - \psi^0\mathbf{W}_N\right)^{-2} \right]^{\ell} \right\|_\infty \leq C;$$

$$\sum_{\ell=0}^{\infty} \left\| \left[\rho^0 \left(\mathbf{I}_N - \psi^0\mathbf{W}_N\right)^{-1} + \varrho^0 \left(\mathbf{I}_N - \psi^0\mathbf{W}_N\right)^{-2} \right]^{\ell} \right\|_1 \leq C,$$

where ϱ^0 denotes the true value of ϱ. We are ready to state the following result.

COROLLARY 3.1. *Consider model (3.1) instead of (2.3) and the matrix of instruments (3.2) replacing (2.4), where $\boldsymbol{\theta} = (\psi, \rho, \varrho, \boldsymbol{\beta}')'$ and the matrices $\mathbf{A}_0$ and $\mathbf{B}_0$ and their sample counterparts are redefined accordingly with $\mathbf{C}_i = (\mathbf{Y}\mathbf{w}_i, \mathbf{y}_{i,-1}, \mathbf{X}_i, \mathbf{Y}_{-1}\mathbf{w}_i)$ and $\mathbf{Z}_i = \left(\sum_{j=1}^N w_{ij}\mathbf{M}_{\mathbf{F}^0}\mathbf{V}_j, \; \mathbf{M}_{\mathbf{F}^0_{-1}}\mathbf{V}_{i,-1}, \mathbf{M}_{\mathbf{F}^0}\mathbf{V}_i, \sum_{j=1}^N w_{ij}\mathbf{M}_{\mathbf{F}^0_{-1}}\mathbf{V}_{j,-1}\right)$. Then, under Assumptions 3.1–3.6*

[11] For the case of a static panel without spatial lags, Cui et al. (2022) provide a detailed technical comparison between the present methodology and the one developed by Bai (2009).

with Assumption 3.5(4) replaced by 3.5(4′), Theorem 3.2 holds true, and Theorem 3.3 holds with $\nu = 4k - (k + 3)$.

4. MONTE CARLO EXPERIMENTS

We use Monte Carlo experiments to assess the finite sample behaviour of the two-stage IV estimator compared to the QML estimator of Shi and Lee (2017). Section S3 of the Online Appendix provides details of the design of our simulation study, as well as additional results for an augmented model with a spatial-time lag, and for a model with endogenous covariates.

We consider the following spatial dynamic panel data model:

$$y_{it} = \alpha_i + \rho y_{it-1} + \psi \sum_{j=1}^{N} w_{ij} y_{jt} + \sum_{\ell=1}^{2} \beta_\ell x_{\ell it} + u_{it}; \quad u_{it} = \sum_{s=1}^{3} \varphi_{si}^0 f_{s,t}^0 + \varepsilon_{it};$$

$$x_{\ell it} = \mu_{\ell i} + \sum_{s=1}^{2} \gamma_{\ell si}^0 f_{s,t}^0 + v_{\ell it}; \quad i = 1, ..., N, \quad t = -49, ..., T, \quad \text{for} \quad \ell = 1, 2.$$

All individual-specific effects and factor loadings are drawn as correlated mean-zero and unit-variance random variables. All individual-invariant, time-specific variables are drawn as autoregressive (AR) processes with mean-zero, unit variance and AR parameter equal to 0.5. The idiosyncratic error, ε_{it}, is non-normal and heteroscedastic 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 spatial weighting matrix, $\mathbf{W}_N = [w_{ij}]$, is an invertible rook matrix of circular form, such that its ith row, $1 < i < N$, has nonzero entries in positions $i - 1$ and $i + 1$, whereas the nonzero entries in rows 1 and N are in positions (1,2), (1, N), and (N, 1), (N, $N - 1$), respectively. This matrix is row-normalised so that all of its nonzero elements equal 1/2.

We set $\rho = 0.4$, $\psi = 0.25$, and $\beta_1 = 3$, $\beta_2 = 1$, following Bai (2009). The proportion of the (average) variance of u_{it} that is due to ε_{it}, denoted as π_u, is set equal to $\pi_u \in \{1/4, 3/4\}$. Thus, for example, $\pi_u = 3/4$ means that the variance of the idiosyncratic error accounts for 75% of the total variance in u_{it}. The signal-to-noise ratio of the model, defined in Section S3, is set equal to $SNR = 4$, as in Juodis and Sarafidis (2018).

We study the optimal two-step IV estimator, defined in (2.5), based on the same set of instruments as in (3.2) with $\mathbf{X}_i$ replaced by $\underline{\mathbf{X}}_i \equiv \mathbf{X}_i - \overline{\mathbf{X}}$, i.e., the covariates are cross-sectionally demeaned in order to control for individual-specific fixed effects.

To allow for heteroscedasticity, the variance estimator is given by

$$\widetilde{\Psi} = \left(\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}}\right)^{-1} \widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widehat{\Omega}\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}} \left(\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}}\right)^{-1}, \tag{4.1}$$

with

$$\widehat{\Omega} = \frac{1}{NT} \sum_{i=1}^{N} \hat{\mathbf{Z}}_i' \mathbf{M}_{\widehat{\mathbf{H}}} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{M}_{\widehat{\mathbf{H}}} \hat{\mathbf{Z}}_i, \tag{4.2}$$

and $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{C}_i \hat{\boldsymbol{\theta}}$.

Table 1. Baseline model with $\pi_u = 3/4$.

	IV					QMLE				
					Results for $\boldsymbol{\rho = 0.4}$. Case I: $N = 100\tau$, $T = 25\tau$					
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	0.399	0.016	0.363	0.059	1.000	0.394	0.013	1.611	0.309	1.000
2	0.400	0.007	0.131	0.056	1.000	0.396	0.006	0.951	0.283	1.000
4	0.400	0.003	0.041	0.048	1.000	0.398	0.003	0.515	0.249	1.000
					Results for $\rho = 0.4$. **Case II:** $N = 25\tau$, $T = 100\tau$					
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	0.399	0.014	0.267	0.084	1.000	0.404	0.010	0.939	0.186	1.000
2	0.400	0.007	0.042	0.062	1.000	0.401	0.005	0.315	0.171	1.000
4	0.400	0.003	0.003	0.045	1.000	0.400	0.002	0.067	0.079	1.000
					Results for $\boldsymbol{\psi = 0.25}$. Case I: $N = 100\tau$, $T = 25\tau$					
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	0.251	0.019	0.242	0.064	0.998	0.250	0.022	0.034	0.179	0.987
2	0.251	0.008	0.246	0.042	1.000	0.250	0.011	0.033	0.119	1.000
4	0.250	0.004	0.019	0.052	1.000	0.250	0.005	0.043	0.071	1.000
					Results for $\psi = 0.25$. Case II: $N = 25\tau$, $T = 100\tau$					
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	0.250	0.016	0.140	0.096	1.000	0.251	0.036	0.192	0.459	0.791
2	0.250	0.008	0.077	0.064	1.000	0.250	0.018	0.049	0.334	0.998
4	0.250	0.004	0.061	0.074	1.000	0.250	0.007	0.079	0.177	1.000
					Results for$\boldsymbol{\beta_2 = 1}$. Case I: $N = 100\tau$, $T = 25\tau$					
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	1.007	0.061	0.738	0.091	0.408	1.128	0.154	12.760	0.711	0.251
2	1.001	0.024	0.145	0.050	0.986	1.043	0.070	4.345	0.379	0.162
4	1.000	0.012	0.027	0.051	1.000	1.006	0.016	0.571	0.096	1.000
					Results for $\beta_2 = 1$. Case II: $N = 25\tau$, $T = 100\tau$					
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	1.001	0.049	0.137	0.097	0.554	1.130	0.170	12.950	0.713	0.167
2	0.999	0.024	0.069	0.065	0.979	1.051	0.088	5.120	0.503	0.149
4	1.000	0.011	0.048	0.045	1.000	1.008	0.024	0.764	0.205	1.000

The J-test statistic we use is as in (2.6) with $\widehat{\Omega}$ replaced by the expression in (4.2).

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 N and T increase by multiples of 2 and the ratio N/T remains equal to 4 in all circumstances. Case II specifies $T = 100\tau$ with $N = 25\tau$ for $\tau = 1, 2, 4$. Therefore, $N/T = 1/4$, as both N and T grow. Finally, Case III, for which results are reported in Section S3, 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.

We perform 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.

Table 1 reports results for the baseline model for $\pi_u = 3/4$. Results for $\pi_u = 1/4$ can be found in S3 of the Online Appendix. As a benchmark, we also consider the bias-corrected QML estimator proposed by Shi and Lee (2017).[12] To speed up the computation time of QMLE, we set $r_y = 3$, i.e., we assume knowledge of the true number of factors (this is only for QMLE). 'Mean' and 'RMSE' denotes the average value and average squared deviation of the estimated parameters from their true values across 2,000 replications, respectively. 'ARB' denotes absolute relative bias, defined as $ARB \equiv \left(|\widehat{\theta}_\ell - \theta_\ell|/\theta_\ell\right) 100$, where θ_ℓ denotes the ℓth entry of $\boldsymbol{\theta} = \left(\psi, \rho, \boldsymbol{\beta}'\right)'$. Size-corrected power is reported, based on the 2.5% and 97.5% quantiles of the empirical distribution of the t-ratio under the null hypothesis.

For both IV and QMLE the values obtained for the mean are close to the true parameters in most cases. Moreover, as predicted by theory, RMSE declines steadily with larger values of N and T, roughly at the rate of $\sqrt{NT}$. Therefore, in what follows we focus on *relative* RMSE performance, ARB, and size properties of the two estimators.

When it comes to the autoregressive parameter, ρ, QMLE outperforms IV in terms of RMSE. This reflects the higher efficiency of maximum likelihood/least squares compared to IV. However, QMLE exhibits substantial ARB and thereby it is severely size distorted. Both ARB and size distortions tend to become smaller as the sample size increases, albeit at a slow rate when $N/T = 4$. In contrast, IV has little ARB and good size properties in most cases, with some mild distortions observed only when N is small.

In regards to the spatial parameter, ψ, IV outperforms QMLE in terms of RMSE. As before, IV is subject to some small size distortion when N is small, which tends to be eliminated quickly with N. QMLE is severely size-distorted.

The results for β_2 are qualitatively no different from those for ψ, with one exception: when either N or T is small, size-adjusted power appears to be relatively lower. Moreover, IV often appears to have higher power than QMLE in moderate sample sizes.

Results for β_1 and for the performance of the J-test in terms of empirical size and power are reported in S3. In general, some mild distortions occur only for N small.

5. AN ANALYSIS OF BANK ATTITUDE TOWARDS RISK

We explore two simple and yet unresolved empirical questions. To what extent was the credit risk-taking behaviour of US banking institutions affected by the risk attitude of their peers during the period that culminated in the global financial crisis (GFC)? Has such relationship changed after the Dodd–Frank Wall Street Reform and Consumer Protection Act of 2010?

Peer influences on risk-taking behaviour may arise for several reasons. For instance, competition can lead institutions to pursue riskier policies in response to their peers' actions, in a bid to maintain the same level of profits—see, e.g., Keeley (1990), Hellmann et al. (2000) and Martynova et al. (2020). Peer influences can also manifest through spillover effects from one institution to another. In particular, while interbank financial networks normally offer the ability to externalise credit exposure by risk sharing, during periods of financial instability such interconnectedness may render banks highly vulnerable to the risk attitude of their peers.[13]

Our main objective is to analyse the effect of peer influences on bank credit risk-taking behaviour, and the impact of the Dodd–Frank Wall Act (hereafter DFA) on risk attitude. While there exist several insightful studies that model spatial interactions in the banking industry (e.g.,

[12] We are grateful to Wei Shi and Lung-fei Lee for providing us the algorithm for the QML estimator.

[13] The tension between these two forces has been explored in a variety of papers, including Allen and Gale (2000), Freixas et al. (2000), Vries (2005), and Gai et al. (2011).

Jing et al., 2018; Ding and Sickles, 2019), to the best of our knowledge, this is the first paper that estimates endogenous network effects and controls for state dependence, as well as for the impact of unobserved aggregate shocks.

5.1. Model specification

We estimate the same regression model as in (2.1) for $i = 1, \ldots, 350$, and $t = 1, \ldots, 56$, where $t = 1$ corresponds to 2006:Q1 and $t = 56$ corresponds to 2019:Q4. The following variables are employed:[14]

$y_{it} \equiv NPL_{it}$ denotes the ratio of nonperforming loans to total loans for bank i at time period t;

$x_{1it} \equiv INEFF_{it}$ denotes the time-varying operational inefficiency of bank i at period t, constructed using a cost frontier model with a translog functional form as in Altunbas et al. (2007);

$x_{2it} \equiv CAR_{it}$ stands for 'capital adequacy ratio', proxied by the ratio of core capital over risk-weighted assets;

$x_{3it} \equiv SIZE_{it}$ is proxied by the natural logarithm of banks' total assets;

$x_{4it} \equiv BUFFER_{it}$ denotes the amount of capital buffer, and it is computed by subtracting from the core capital (leverage) ratio the value of the minimum regulatory capital ratio (8%);

$x_{5it} \equiv PROFITABILITY_{it}$ is proxied by the return on equity (ROE), defined as annualised net income expressed as a percentage of average total equity on a consolidated basis;

$x_{6it} \equiv QUALITY_{it}$ is computed as the total amount of loan loss provisions (LLP) expressed as a percentage of assets.

$x_{7it} \equiv LIQUIDITY_{it}$ is proxied by the loan-to-deposit (LTD) ratio. When this ratio is too high, banks may not have enough liquidity to meet unforeseen funding requirements;

$x_{8it} \equiv PRESSURE_{it}$ takes the value of unity if a bank has a capital buffer that is less than or equal to the 10th percentile of the distribution of capital buffer in any given period, and zero otherwise.

The spatial weights matrix has been constructed following the methodology of Fernandez (2011). In particular, let

$$d_{ij} = \sqrt{2\left(1 - \rho_{ij}\right)},$$

where ρ_{ij} denotes Spearman's correlation coefficient between banks i and j, corresponding to a specific financial indicator observed over 56 time periods. Then, the (i, j)-element of the $N \times N$ spatial weights matrix, $\mathbf{W}_N$, is defined as $w_{ij} = exp(-d_{ij})$. Thus, more distant observations take a smaller weight. Each of the rows of $\mathbf{W}_N$ has been divided by the sum of its corresponding elements so that $\sum_j w_{ij} = 1$ for all j. Finally, the diagonal elements of $\mathbf{W}_N$ are set equal to zero in order to ensure that no individual is treated as its own neighbour.

We make use of two financial indicators to construct weights: the debt ratio (total liabilities over total assets) and the dividend yield (dividends over market price per share).

5.2. Estimation

The model is estimated using the second-step IV estimator, combined with the robust variance–covariance estimator given by (4.1)–(4.2). $INEFF$ is treated as endogenous with respect to

[14] All data have been downloaded from Bank Data and Statistics (2022), a dataset maintained by the Federal Deposit Insurance Corporation.

Table 2. Results for different subperiods.

	Full	Basel I-II	DFA
$\widehat{\rho}$ (AR parameter)	0.492*** (0.061)	0.369*** (0.081)	0.383*** (0.131)
$\widehat{\psi}$ (spatial parameter)	0.376*** (0.091)	0.224** (0.115)	0.583** (0.230)
$\widehat{\beta}_1$ (inefficiency)	0.329*** (0.081)	0.565** (0.146)	0.242** (0.104)
$\widehat{\beta}_2$ (CAR)	0.013*** (0.005)	0.031*** (0.012)	0.007 (0.005)
$\widehat{\beta}_3$ (size)	0.115* (0.022)	0.940*** (0.344)	0.275 (0.265)
$\widehat{\beta}_4$ (buffer)	−0.033** (0.014)	−0.027 (0.026)	−0.004 (0.015)
$\widehat{\beta}_5$ (profitability)	−0.003 (0.002)	−0.002 (0.004)	−0.010** (0.004)
$\widehat{\beta}_6$ (quality)	0.221*** (0.038)	0.243*** (0.045)	0.035 (0.092)
$\widehat{\beta}_7$ (liquidity)	1.301*** (0.178)	2.717*** (0.563)	1.585*** (0.435)
$\widehat{\beta}_8$ (inst. pressure)	0.045 (0.047)	0.010 (0.067)	0.015 (0.074)
$\widehat{r}_y$	1	1	1
$\widehat{r}_x$	1	1	2
J-test	36.019 [0.030]	27.119 [0.207]	24.662 [0.313]

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

ε_{it} due to reverse causality. This arises because higher levels of risk imply additional costs and managerial efforts incurred by banks, in order to improve existing loan underwriting and monitoring procedures. $INEFF$ is instrumented by the ratio of interest expenses paid on deposits over the value of total deposits.[15]

The remaining covariates are treated as exogenous with respect to ε_{it}. However, these covariates can be endogenous with the common factor component $\boldsymbol{\varphi}_i'\mathbf{h}_t^0$. The matrix of instruments is the same as in (3.2) with $\mathbf{X}_i$ replaced by $\widetilde{\underline{\mathbf{X}}}_i = \left(\widetilde{\underline{\mathbf{x}}}_{1i}, \underline{\mathbf{x}}_{2i}, \ldots, \underline{\mathbf{x}}_{8i}\right)$, a matrix of order $T \times 8$, where $\underline{\mathbf{x}}_{\ell i} = \mathbf{x}_{\ell i} - \overline{\mathbf{x}}_{\ell}$ and $\mathbf{x}_{\ell i}$ is a $T \times 1$ vector that denotes the ℓth covariate corresponding to β_ℓ, for $\ell = 2, \ldots, k$, whereas $\widetilde{\underline{\mathbf{x}}}_{1i}$ is the external instrument. Thus, we make use of 32 moment conditions in total, and with 10 parameters the number of degrees of freedom equals 22.

The projection matrix $\mathbf{M}_{\widehat{\mathbf{F}}}$ is computed based on $\hat{r}_x$ factors estimated from $(NT)^{-1}\sum_{i=1}^{N}\widetilde{\underline{\mathbf{X}}}_i\widetilde{\underline{\mathbf{X}}}_i'$ using the eigenvalue ratio test of Ahn and Horenstein (2013). $\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}$ and $\mathbf{M}_{\widehat{\mathbf{H}}}$ are computed in a similar manner.

5.3. Results

Column 'Full' in Table 2 reports results for the entire period of the sample.[16] Columns 'Basel I-II' and 'DFA' present results for two different subperiods, namely 2006:Q1–2010:Q4 and 2011:Q1–2019:Q4, respectively. The first subsample corresponds to the Basel I-II regulatory framework and includes the GFC. The second subsample corresponds to the Dodd–Frank Act.

In regards to column 'Full', we see that both $\widehat{\rho}$ and $\widehat{\psi}$ are statistically significant, providing evidence for state dependence and endogenous peer effects.

The coefficient of operational inefficiency is positive and statistically significant, an outcome consistent with Williams (2004). This provides support for the so-called bad management hypothesis (see, e.g., Fiordelisi et al., 2011). The coefficient of capital adequacy ratio on bank risk is positive and statistically significant at the 5% level. However, bank size appears to exert

[15] The correlation between these two variables in the sample equals 0.22.

[16] Tables S4.2–S4.3 in the Online Appendix report robustness results in terms of different specifications and/or different estimation approaches.

a small impact on risk. While this finding is in contrast with the 'too-big-to-fail hypothesis', the conclusions change when the model is re-estimated during 2006:Q1–2010:Q4 only. We shall discuss this in detail shortly.

Capital buffer has a negative and significant effect on risk attitude, consistent with capital buffer theory.

In line with the findings of Aggarwal and Jacques (2001), asset quality (or lack of thereof) has a strong positive effect on risk attitude, i.e., banks with higher levels of loan loss provision also have a larger proportion of risky assets in their portfolios. Similarly, liquidity (or lack of thereof) exerts a strong positive effect on risk.

Finally, the number of estimated factors in y equals 1 across all samples, i.e., 'Full', 'Basel I-II', and 'DFA'. As it turns out, the estimated factor is highly correlated with the US gross private domestic investment (PDI), and thus it appears to capture one of the major indices of economic activity that influence aggregate demand.[17]

Regarding columns 'Basel I-II' and 'DFA', the following major differences are worth noting. First, the size effect is much larger in magnitude during the period under Basel I-II. However, following the introduction of the DFA, the effect of size falls and is no longer statistically significant.

Second, the effect of operational inefficiency appears to be much larger under the Basel I-II than that under the DFA. A similar result applies to the coefficients of quality and liquidity. Finally, it appears that more profitable banks are less willing to take on more risk during the DFA, whereas there seems to be no effect during Basel I-II.

These results bear important policy implications and provide evidence that the more risk-sensitive capital regulation introduced by the DFA framework has succeeded in influencing banks' behaviour in a substantial manner.

Section S4 in the Online Appendix provides additional results in terms of direct, indirect, and total effects, computed as in Debarsy et al. (2012), and LeSage and Pace (2009). Under Basel I-II the direct effects appear to be larger than the indirect ones, contributing roughly three-quarters of the total effect. In contrast, under the DFA period direct effects contribute about 42% of the total effect.

6. CONCLUDING REMARKS

This paper develops a new IV estimator for spatial, dynamic panel data models with interactive effects under large N and T asymptotics. The proposed estimator is computationally inexpensive and free from asymptotic bias in either cross-sectional or time-series dimension. Last, the proposed estimator retains the attractive feature of method of moments estimation in that it can potentially accommodate endogenous regressors, so long as external exogenous instruments are available.

ACKNOWLEDGEMENTS

We are grateful to the Editor, Dennis Kristensen, for a highly efficient reviewing process, as well as two anonymous referees for very useful comments and suggestions. We are also grateful to Arturas Juodis, and seminar participants at Michigan State University, BI Norwegian Business School, University of Birmingham, the 2021 Dutch Workshop on Panel

[17] More details are provided in Section S4 of the Online Appendix.

Data Analysis, the Norwegian Bureau of Statistics, the 2022 RCEA conference, and the 2022 AMEF conference for valuable feedback and discussion. This work has been supported by JSPS KAKENHI (grant numbers 20H05631, 21H00700, and 21H04397).

REFERENCES

Aggarwal, R. and K. T. Jacques (2001). The impact of FDICIA and prompt corrective action on bank capital and risk: Estimates using a simultaneous equations model. *Journal of Banking and Finance 25*(6), 1139–60.

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–27.

Allen, F. and D. Gale (2000). Financial contagion. *Journal of Political Economy 108*(1), 1–33.

Altunbas, Y., S. Carbo, E. P. Gardener and P. Molyneux (2007). Examining the relationships between capital, risk and efficiency in European banking. *European Financial Management 13*(1), 49–70.

Alvarez, J. and M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica 71*(4), 1121–59.

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*, 47–82.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277–97.

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

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–79.

Bai, J. and K. Li (2013). Spatial panel data models with common shocks, Ms., Mimeo.

Bai, J. and K. Li (2021). Dynamic spatial panel data models with common shocks. *Journal of Econometrics 224*, 134–60.

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

Balestra, P. and M. Nerlove (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica 34*(3), 585–612.

Bank Data and Statistics,. (2022). *Statistics on Depository Institutions (SDI) [database]*, Federal Deposit Insurance Corporation, Washington, DC. https://www7.fdic.gov/sdi/main.asp?formname=customddownload (accessed 5 July 2022).

Bun, M. J. G. and V. Sarafidis (2015). Dynamic panel data models. In (B. H. Baltagi (ed.)), *The Oxford Handbook of Panel Data*, pp. 76–110. Oxford: Oxford University Press.

Case, A. (1991). Spatial patterns in household demand. *Econometrica 59*, 953–65.

Chen, J., Y. Shin and C. Zheng (2022). Estimation and inference in heterogeneous spatial panels with a multifactor error structure. *Journal of Econometrics 229*, 55–79.

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

Cui, G., M. Norkutuė, V. Sarafidis and T. Yamagata (2022). Two-stage instrumental variable estimation of linear panel data models with interactive effects. *Econometrics Journal 25*(2), 340–61.

Debarsy, N., C. Ertur and J. P. LeSage (2012). Interpreting dynamic space time panel data models. *Statistical Methodology 9*(1), 158–71.

Ding, D. and R. Sickles (2019). Capital regulation, efficiency, and risk taking: A spatial panel analysis of US banks. In (M. Tsionas (ed.)), *Panel Data Econometrics: Empirical Applications*, pp. 405–66. London: Elsevier.

Everaert, G. and T. De Groote (2016). Common correlated effects estimation of dynamic panels with cross-sectional dependence. *Econometric Reviews 35*, 428–63.

Fernandez, V. (2011). Spatial linkages in international financial markets. *Quantitative Finance 11*(2), 237–45.

Fiordelisi, F., D. Marques-Ibanez and P. Molyneux (2011). Efficiency and risk in European banking. *Journal of Banking and Finance 35*(5), 1315–26.

Freeman, H. and M. Weidner (2021). Linear panel regressions with two-way unobserved heterogeneity, Ms., Mimeo.

Freixas, X., B. M. Parigi and J.-C. Rochet (2000). Systemic risk, interbank relations, and liquidity provision by the central bank. *Journal of Money, Credit and Banking 32*(3), 611–38.

Gai, P., A. Haldane and S. Kapadia (2011). Complexity, concentration and contagion. *Journal of Monetary Economics 58*, 453–70.

Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica 70*(4), 1639–57.

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

Hellmann, T. F., K. C. Murdock and J. E. Stiglitz (2000). Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review 90*(1), 147–65.

Jing, Z., J. P. Elhorst, J. P. A. M. Jakobs and J. de Haan (2018). The propagation of financial turbulence: Interdependence, spillovers, and direct and indirect effects. *Empirical Economics 55*, 169–92.

Juodis, A., H. Karabıyık and J. Westerlund (2021). On the robustness of the pooled CCE estimator. *Journal of Econometrics 220*(2), 325–48.

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

Juodis, A. and V. Sarafidis (2022). An incidental parameters free inference approach for panels with common shocks. *Journal of Econometrics 229*, 19–54.

Keeley, M. C. (1990). Deposit insurance, risk, and market power in banking. *American Economic Review 80*(5), 1183–200.

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

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics 140*, 219–57.

Kelejian, H. H. and I. R. Prucha (2007). HAC estimation in a spatial framework. *Journal of Econometrics 140*, 131–54.

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

Korniotis, G. M. (2010). Estimating panel models with internal and external habit formation. *Journal of Business and Economic Statistics 28*(1), 145–58.

Kuersteiner, G. and I. R. Prucha (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica 88*(5), 2109–46.

Lee, L. F. and J. Yu (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics 180*, 174–97.

Lee, L. F. and J. Yu (2015). Spatial panel data models. In (B. Baltagi (ed.)), *The Oxford Handbook of Panel Data*, pp. 363–401. Oxford: Oxford University Press.

LeSage, J. P. and R. Pace (2009). *Introduction to Spatial Econometrics*, 1st edn. Boca Raton, FL: Chapman and Hall.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–42.

Martynova, N., L. Ratnovski and R. Vlahu (2020). Bank profitability, leverage constraints, and risk-taking. *Journal of Financial Intermediation 44*, 100821.

Moon, H. R. and M. Weidner (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory 33*(1), 158–95.

Norkute, M., V. Sarafidis, T. Yamagata and G. Cui (2021). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. *Journal of Econometrics 220*(2), 416–46.

Okui, R. (2009). The optimal choice of moments in dynamic panel data models. *Journal of Econometrics 151*, 1–16.

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

Robertson, D. and V. Sarafidis (2015). IV estimation of panels with factor residuals. *Journal of Econometrics 185*(2), 526–41.

Sarafidis, V. and T. J. Wansbeek (2021). Celebrating 40 years of panel data analysis: Past, present and future. *Journal of Econometrics 220*(2), 215–26.

Shi, W. and L. Lee (2017). Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics 197*(2), 323–47.

Vries, C. G. D. (2005). The simple economics of bank fragility. *Journal of Banking and Finance 29*, 803–25.

Williams, J. (2004). Determining management behaviour in European banking. *Journal of Banking and Finance 28*(10), 2427–60.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*, 1st edn. Cambridge, MA: MIT Press.

Yu, J., R. de Jong and L.-F. Lee (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), 118–34.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendix
Replication Package

Co-editor Dennis Kristensen handled this manuscript.

APPENDIX A: PROOFS OF RESULTS

Proof of Proposition 3.1. With the definition of $\widehat{\mathbf{Z}}_i$, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{u}_i = \begin{pmatrix} \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i \\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i,-1}'\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{u}_i \\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i \end{pmatrix}. \tag{A.1}$$

By Norkute et al. (2021), we have

$$\begin{pmatrix} \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\mathcal{X}}_{i,-1}'\mathbf{M}_{\mathbf{F}_{-1}^0}\mathbf{u}_i \\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\mathcal{X}}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i \end{pmatrix} + \sqrt{\frac{T}{N}}\begin{pmatrix}\mathbf{b}_{12}\\ \mathbf{b}_{13}\end{pmatrix} + \sqrt{\frac{N}{T}}\begin{pmatrix}\mathbf{b}_{22}\\ \mathbf{b}_{23}\end{pmatrix} + o_p(1), \tag{A.2}$$

where

$$\begin{aligned}
\mathbf{b}_{12} &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\boldsymbol{\mathcal{V}}_{i,-1}'\mathbf{V}_{j,-1}}{T}\boldsymbol{\Gamma}_j'(\boldsymbol{\Upsilon}^0)^{-1}\left(\frac{\mathbf{F}_{-1}'\mathbf{F}_{-1}}{T}\right)^{-1}\frac{\mathbf{F}_{-1}'\mathbf{u}_i}{T}\\
\mathbf{b}_{13} &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\boldsymbol{\mathcal{V}}_i'\mathbf{V}_j}{T}\boldsymbol{\Gamma}_j'(\boldsymbol{\Upsilon}^0)^{-1}\left(\frac{\mathbf{F}^{0\prime}\mathbf{F}^0}{T}\right)^{-1}\frac{\mathbf{F}^{0\prime}\mathbf{u}_i}{T}\\
\mathbf{b}_{22} &= -\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\left(\frac{\mathbf{F}_{-1}'\mathbf{F}_{-1}}{T}\right)^{-1}\mathbf{F}_{-1}'\boldsymbol{\Sigma}_{-1}\mathbf{M}_{\mathbf{F}_{-1}^0}\mathbf{u}_i\\
\mathbf{b}_{23} &= -\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\left(\frac{\mathbf{F}^{0\prime}\mathbf{F}^0}{T}\right)^{-1}\mathbf{F}^{0\prime}\boldsymbol{\Sigma}\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i
\end{aligned} \tag{A.3}$$

with $\boldsymbol{\mathcal{X}}_i = \mathbf{X}_i - \frac{1}{N}\sum_{\ell=1}^{N}\mathbf{X}_\ell\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_i$, $\boldsymbol{\mathcal{X}}_{i,-1} = \mathbf{X}_{i,-1} - \frac{1}{N}\sum_{\ell=1}^{N}\mathbf{X}_{\ell,-1}\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_i$, $\boldsymbol{\mathcal{V}}_i = \mathbf{V}_i - \frac{1}{N}\sum_{\ell=1}^{N}\mathbf{V}_\ell\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_i$, $\boldsymbol{\mathcal{V}}_{i,-1} = \mathbf{V}_{i,-1} - \frac{1}{N}\sum_{\ell=1}^{N}\mathbf{V}_{\ell,-1}\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_i$, $\boldsymbol{\Upsilon} = N^{-1}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i'$. In addition, $\boldsymbol{\Sigma} = N^{-1}\sum_{\ell=1}^{N}\mathbb{E}\left(\mathbf{V}_\ell\mathbf{V}_\ell'\right)$ and $\boldsymbol{\Sigma}_{-1} = N^{-1}\sum_{l=1}^{N}\mathbb{E}\left(\mathbf{V}_{\ell,-1}\mathbf{V}_{\ell,-1}'\right)$.

Combining Lemmas S1.5, S1.6, S1.7, S1.8, and S1.10 in the Online Appendix, we can derive that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\boldsymbol{\mathcal{X}}_j'\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i + \sqrt{\frac{T}{N}}\mathbf{b}_{11} + \sqrt{\frac{N}{T}}\mathbf{b}_{21} + o_p(1), \tag{A.4}$$

where

$$\begin{aligned}
\mathbf{b}_{11} &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{h=1}^{N} w_{ij}\frac{\boldsymbol{\mathcal{V}}_j'\mathbf{V}_h}{T}\boldsymbol{\Gamma}_h^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\left(\frac{\mathbf{F}^{0\prime}\mathbf{F}^0}{T}\right)^{-1}\frac{\mathbf{F}^{0\prime}\mathbf{u}_i}{T}\\
\mathbf{b}_{21} &= -\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\boldsymbol{\Gamma}_j^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\left(\frac{\mathbf{F}^{0\prime}\mathbf{F}^0}{T}\right)^{-1}\mathbf{F}^{0\prime}\boldsymbol{\Sigma}\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i.
\end{aligned} \tag{A.5}$$

From (A.1)–(A.5), we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{Z}_i'\mathbf{u}_i + \sqrt{\frac{T}{N}}\mathbf{b}_1 + \sqrt{\frac{N}{T}}\mathbf{b}_2 + o_p(1),$$

where $\mathbb{Z}_i = \left(\sum_{j=1}^{N} w_{ij}\mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{V}}_j, \mathbf{M}_{\mathbf{F}_{-1}^0}\boldsymbol{\mathcal{V}}_{i,-1}, \mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{V}}_i\right)$, $\mathbf{b}_1 = \left(\mathbf{b}_{11}', \mathbf{b}_{12}', \mathbf{b}_{13}'\right)'$ and $\mathbf{b}_2 = \left(\mathbf{b}_{21}', \mathbf{b}_{22}', \mathbf{b}_{23}'\right)'$. This completes the proof. □

Proof of Theorem 3.1. We have

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

By Lemma S1.3 in the Online Appendix, $\widehat{\mathbf{A}}-\mathbf{A}_0\xrightarrow{p}\mathbf{0}$ and $\widehat{\mathbf{B}}-\mathbf{B}_0\xrightarrow{p}\mathbf{0}$, and $\widehat{\mathbf{B}}^{-1}-\mathbf{B}_0^{-1}\xrightarrow{p}\mathbf{0}$ by continuous mapping theorem, thus, $\widehat{\mathbf{A}}'\widehat{\mathbf{B}}^{-1}\widehat{\mathbf{A}}-\mathbf{A}_0{}'\mathbf{B}_0^{-1}\mathbf{A}_0\xrightarrow{p}\mathbf{0}$. Under Assumption 3.6, $\mathbf{A}_0{}'\mathbf{B}_0^{-1}\mathbf{A}_0$ is positive definite which implies $\left(\widehat{\mathbf{A}}'\widehat{\mathbf{B}}^{-1}\widehat{\mathbf{A}}\right)^{-1}\widehat{\mathbf{A}}'\widehat{\mathbf{B}}^{-1}=O_p(1)$. By Proposition 3.1 $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{u}_i=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{Z}_i'\left(\mathbf{H}^0\boldsymbol{\varphi}_i^0+\boldsymbol{\varepsilon}_i\right)+\sqrt{\frac{T}{N}}\mathbf{b}_1+\sqrt{\frac{N}{T}}\mathbf{b}_2+o_p(1)$. First, due to the independence between $\boldsymbol{\varepsilon}_i$ and $\mathbb{Z}_i$, a suitable central limit theorem ensures that $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{Z}_i'\boldsymbol{\varepsilon}_i=O_p(1)$. Consider the first component of $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{Z}_i'\mathbf{H}^0\boldsymbol{\varphi}_i^0$. Recalling $\mathbb{Z}_i=\left(\sum_{j=1}^{N}w_{ij}\mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{V}}_j,\mathbf{M}_{\mathbf{F}^0_{-1}}\boldsymbol{\mathcal{V}}_{i,-1},\mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{V}}_i\right)$ where $\boldsymbol{\mathcal{V}}_i=\mathbf{V}_i-\frac{1}{N}\sum_{\ell=1}^{N}\mathbf{V}_\ell\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_i^0$, $\boldsymbol{\mathcal{V}}_{i,-1}=\mathbf{V}_{i,-1}-\frac{1}{N}\sum_{\ell=1}^{N}\mathbf{V}_{\ell,-1}\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_i^0$, and noting $\mathbf{M}_{\mathbf{F}^0}=\mathbf{M}_{\mathbf{F}}=\frac{\mathbf{F}\mathbf{F}'}{T}$, $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{\sum_{j=1}^{N}w_{ij}\boldsymbol{\mathcal{V}}_j'\mathbf{F}}{\sqrt{T}}\frac{\mathbf{F}'\mathbf{H}^0}{T}\boldsymbol{\varphi}_i^0=O_p(1)$ due to the independence between $\boldsymbol{\mathcal{V}}_j$ and $\mathbf{F}$, and cross-sectional independence between $\sum_{j=1}^{N}w_{ij}\boldsymbol{\mathcal{V}}_j$ and $\boldsymbol{\varphi}_i^0$. In a similar manner, we can show that other components in $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{Z}_i'\mathbf{H}^0\boldsymbol{\varphi}_i^0$ and the bias terms $\mathbf{b}_1$ and $\mathbf{b}_2$ are $O_p(1)$. Together with the condition $N/T\to c$ where $0<c<\infty$ as $N,T\to\infty$, we have $\sqrt{NT}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)=O_p(1)$ as required. □

Proof of Proposition 3.2. The term $N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i$, is equal to

$$\begin{pmatrix}N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i\\ N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\mathbf{X}_{i,-1}'\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i\\ N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i\end{pmatrix}. \tag{A.6}$$

Consider the first term in (A.6). By Lemmas S2.4, S2.5, S2.6, and S2.7 in the Online Appendix, and the fact that $\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_j=\mathbf{M}_{\mathbf{F}^0}\mathbf{V}_j$, we can derive that

$$\begin{aligned}&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i\\ &=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}\mathbf{V}_j'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i-\frac{1}{N^{3/2}T^{1/2}}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{j=1}^{N}w_{ij}\boldsymbol{\Gamma}_j^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_\ell^0\mathbf{V}_\ell'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i\\ &\quad+O_p\left(\frac{1}{\delta_{NT}}\right)+O_p\left(\frac{\sqrt{NT}}{\delta_{NT}^3}\right)\\ &=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}\boldsymbol{\mathcal{X}}_j'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i+O_p\left(\frac{1}{\delta_{NT}}\right)+O_p\left(\frac{\sqrt{NT}}{\delta_{NT}^3}\right),\end{aligned}$$

where $\boldsymbol{\mathcal{X}}_j=\mathbf{X}_j-\frac{1}{N}\sum_{\ell=1}^{N}\mathbf{X}_\ell\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_j^0$.

Consider the second term in (A.6). By Lemmas S2.8, S2.9, S2.10, and S2.11 in the Online Appendix, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i,-1}'\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\mathcal{X}}_{i,-1}'\mathbf{M}_{\mathbf{F}^0_{-1}}\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i+O_p\left(\frac{1}{\delta_{NT}}\right)+O_p\left(\frac{\sqrt{NT}}{\delta_{NT}^3}\right).$$

Similarly, for the third term, we can show that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\mathcal{X}}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i+O_p\left(\frac{1}{\delta_{NT}}\right)+O_p\left(\frac{\sqrt{NT}}{\delta_{NT}^3}\right),$$

which leads to

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i=\begin{pmatrix}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}\boldsymbol{\mathcal{X}}_j'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i\\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\mathcal{X}}_{i,-1}'\mathbf{M}_{\mathbf{F}^0_{-1}}\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i\\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\mathcal{X}}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i\end{pmatrix}+O_p\left(\frac{1}{\delta_{NT}}\right)+O_p\left(\frac{\sqrt{NT}}{\delta_{NT}^3}\right).$$

Noting that $\mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{X}}_i=\mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{V}}_i$ and that $(\mathbf{F}^0,\mathbf{H}^0)$, $\mathbf{V}_i$, $\boldsymbol{\varepsilon}_i$ are all independent of each other, by Lemma S2.12 in the Online Appendix we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i=\begin{pmatrix}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}\mathbf{V}_j'\mathbf{M}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_{i,-1}'\mathbf{M}_{\mathbf{F}^0_{-1}}\boldsymbol{\varepsilon}_i\\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{M}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\end{pmatrix}+O_p\left(\frac{1}{\delta_{NT}}\right)+O_p\left(\frac{\sqrt{NT}}{\delta_{NT}^3}\right),$$

which completes the proof. □

Proof of Theorem 3.2. Substituting $\mathbf{y}_i=\mathbf{C}_i\boldsymbol{\theta}+\mathbf{u}_i$ into $\widetilde{\mathbf{c}}_y$ and multiplying by $\sqrt{NT}$ we have

$$\sqrt{NT}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta})=(\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}})^{-1}\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\cdot N^{-1/2}T^{-1/2}\textstyle\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i.$$

By Lemma S2.2 in the Online Appendix, $\widetilde{\mathbf{A}}-\mathbf{A}_0=o_p(1)\widetilde{\mathbf{B}}-\mathbf{B}_0=o_p(1)$, and together with Proposition 3.2, we obtain

$$\sqrt{NT}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta})=(\mathbf{A}_0'\mathbf{B}_0^{-1}\mathbf{A}_0)^{-1}\mathbf{A}_0'\mathbf{B}_0^{-1}\cdot N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\mathbf{Z}_i'\boldsymbol{\varepsilon}_i+O_p(\delta_{NT}^{-1})+O_p(N^{1/2}T^{1/2}\delta_{NT}^{-3}).$$

The central limit theorem of the martingale difference in Kelejian and Prucha (2001) establishes

$$\sqrt{NT}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta})\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Psi}),$$

since $N,T\to\infty$ with $N/T^2\to0$ and $T/N^2\to0$. Finally, Lemma S2.13 shows that $\tilde{\sigma}_\varepsilon^2-\sigma_\varepsilon^2=O_p(\delta_{NT}^{-1})$, thus, together with Lemma S2.2, $\widetilde{\boldsymbol{\Psi}}-\boldsymbol{\Psi}\xrightarrow{p}\mathbf{0}$, which complete the proof. □

Proof of Theorem 3.3. Noting $\widetilde{\mathbf{u}}_i=\mathbf{u}_i-\mathbf{C}_i\left(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)$ we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widetilde{\mathbf{u}}_i=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i-\widetilde{\mathbf{A}}\sqrt{NT}\left(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\right).$$

Define $\mathbf{L}=\boldsymbol{\Omega}^{-1/2}\mathbf{A}_0$ with $\boldsymbol{\Omega}=\sigma_\varepsilon^2\mathbf{B}_0$. Since $\sqrt{NT}\left(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)=\left(\mathbf{A}_0'\mathbf{B}_0^{-1}\mathbf{A}_0\right)^{-1}\mathbf{A}_0'\mathbf{B}_0^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{Z}_i'\boldsymbol{\varepsilon}_i+o_p(1)$ by Proposition 3.2, we have $\hat{\boldsymbol{\Omega}}^{-1/2}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widetilde{\mathbf{u}}_i=\mathbf{M}_{\mathbf{L}}\boldsymbol{\Omega}^{-1/2}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{Z}_i'\boldsymbol{\varepsilon}_i+o_p(1)$ with $\mathbf{M}_{\mathbf{L}}=\mathbf{I}_{3k}-\mathbf{L}\left(\mathbf{L}'\mathbf{L}\right)^{-1}\mathbf{L}'$ whose rank is $\nu=3k-(k+2)$. Collecting all together, the central limit theorem of Kelejian and Prucha (2001) ensures that $\frac{1}{NT}\sum_{i=1}^{N}\widetilde{\mathbf{u}}'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\hat{\boldsymbol{\Omega}}_{NT}^{-1}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widetilde{\mathbf{u}}\xrightarrow{d}\chi_\nu^2$, as required. □

Proof of Corollary 3.1. In a similar line of the proof for Proposition 3.1, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{u}_i=\frac{1}{\sqrt{NT}}\mathbb{Z}_i'\mathbf{u}_i+\sqrt{\frac{T}{N}}\mathbf{b}_1+\sqrt{\frac{N}{T}}\mathbf{b}_2+o_p(1),$$

where $\mathbb{Z}_i=\left(\sum_{j=1}^{N}w_{ij}\mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{V}}_j,\mathbf{M}_{\mathbf{F}^0_{-1}}\boldsymbol{\mathcal{V}}_{i,-1},\mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{V}}_i,\sum_{j=1}^{N}w_{ij}\mathbf{M}_{\mathbf{F}^0_{-1}}\boldsymbol{\mathcal{V}}_{j,-1}\right)$, $\mathbf{b}_1=\left(\mathbf{b}_{11}',\mathbf{b}_{12}',\mathbf{b}_{13}',\mathbf{b}_{14}'\right)'$, $\mathbf{b}_2=\left(\mathbf{b}_{21}',\mathbf{b}_{22}',\mathbf{b}_{23}',\mathbf{b}_{24}'\right)'$, with $\mathbf{b}_{14}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{h=1}^{N}w_{ij}\frac{\boldsymbol{\mathcal{V}}_{j,-1}'\mathbf{V}_{i,-1}}{T}\Gamma_h^{0\prime}(\Upsilon^0)^{-1}\left(\frac{\mathbf{F}_{-1}^{0\prime}\mathbf{F}_{-1}^0}{T}\right)^{-1}\frac{\mathbf{F}_{-1}^{0\prime}\mathbf{u}_i}{T}$ and $\mathbf{b}_{24}=-\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{h=1}^{N}w_{ij}\Gamma_j^{0\prime}(\Upsilon^0)^{-1}\left(\frac{\mathbf{F}_{-1}^{0\prime}\mathbf{F}_{-1}^0}{T}\right)^{-1}\mathbf{F}_{-1}^{0\prime}\Sigma_{-1}\mathbf{M}_{\mathbf{F}^0_{-1}}\mathbf{u}_i$. Noting that $\mathbf{b}_1$ and $\mathbf{b}_2$ are bounded in probability, $\sqrt{NT}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)=O_p(1)$ immediately follows, where $\boldsymbol{\theta}=(\psi,\rho,\varrho,\boldsymbol{\beta}')'$ and $\widehat{\boldsymbol{\theta}}$ is the first-step estimator. Following the same line of the proof of Theorem 3.2, it is straightforward to show that $\sqrt{NT}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta})=(\mathbf{A}_0'\mathbf{B}_0^{-1}\mathbf{A}_0)^{-1}\mathbf{A}_0'\mathbf{B}_0^{-1}\cdot N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\mathbf{Z}_i'\boldsymbol{\varepsilon}_i+o_p(1)$, and that $\sqrt{NT}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta})\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Psi})$ and $\widetilde{\boldsymbol{\Psi}}-\boldsymbol{\Psi}=o_p(1)$. Finally, noting that $\mathbf{Z}_i$ is $T\times 4k$ and $\boldsymbol{\theta}$ is $k+3$, it is straightforward to see $J\xrightarrow{d}\chi^2_\nu$ with $\nu=4k-(k+3)$, which completes the proof. □