MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

IV Estimation of Spatial Dynamic Panels with Interactive Effects: Large Sample Theory and an Application on Bank Attitude Toward Risk

Guowei Cui, Vasilis Sarafidis and Takashi Yamagata

March 2020

Working Paper 11/20

IV Estimation of Spatial Dynamic Panels with Interactive Effects: Large Sample Theory and an Application on Bank Attitude Toward Risk

Guowei Cui
HUST
and
Monash University

Vasilis Sarafidis*
Monash University

Takashi Yamagata†
University of York
and
Osaka University

March 27, 2020

Abstract

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. For this class of models, the only approaches available in the literature are based on quasi-maximum likelihood estimation. The approach put forward in this paper is appealing from both a theoretical and a practical point of view for a number of reasons. Firstly, the proposed IV estimator is linear in the parameters of interest and it is computationally inexpensive. Secondly, the IV estimator is free from asymptotic bias. In contrast, existing QML estimators suffer from incidental parameter bias, depending on the magnitude of unknown parameters. Thirdly, the IV estimator retains the attractive feature of Method of Moments estimation in that it can accommodate endogenous regressors, so long as external exogenous instruments are available. The IV estimator is consistent and asymptotically normal as $N, T \to \infty$, with $N/T^2 \to 0$ and $T/N^2 \to 0$. The proposed methodology is employed to study the determinants of risk attitude of banking institutions. The results of our analysis provide evidence that the more risk-sensitive capital regulation that was introduced by the Basel III framework in 2011 has succeeded in influencing banks' behaviour in a substantial manner.

JEL classification: C33; C36; C38; C55.

Key Words: Panel data, instrumental variables, state dependence, social interactions, common factors, large N and T asymptotics.

*Corresponding author. Department of Econometrics and Business Statistics, Monash University, VIC 3145, Australia. E-mail address: vasilis.sarafidis@monash.edu.

This author gratefully acknowledges financial support from the Australian Research Council, under research grant number DP-170103135.

†Department of Economics and Related Studies, University of York, York YO10 5DD, UK. E-mail address: ty509@york.ac.uk.

This author gratefully acknowledges the financial support by JSPS KAKENHI Grant Number 18K01545.

1 Introduction

Economic behavior 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, regulatory constraints, and economic slack, among other factors. The importance of intertemporal dependence has been widely recognised in the panel data literature since its infancy.[1]

More recently, it has been forcefully pointed out that, in addition to state dependence, economic behaviour is also subject to network effects, and social interactions among individual agents (see e.g. the pioneering work of Case (1991) and Manski (1993)). At the same time, economic agents inhabit common economic environments, and therefore their behaviour is subject to aggregate shocks, which may be due to shifts in technology and productivity, changes in preferences and tastes, to mention only a few. In the former case, economic agents' own behaviour is influenced by the behaviour of other agents, possibly their peers. In the latter case, economic agents' own behaviour is influenced by aggregate, economy-wide shocks.

In panel data analysis, state dependence is commonly characterised using dynamic models, whereas social interactions are formulated based on spatial econometric techniques, as described e.g. in Kelejian and Piras (2017). Finally, aggregate unobserved shocks are typically represented by common factors, also known as "interactive effects" (Sarafidis and Wansbeek (2012)).

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, the only approaches available in the literature are based on quasi-maximum likelihood estimation (QMLE); see Shi and Lee (2017) and Bai and Li (2018). 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.

Firstly, 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 Section 12.3.2 in Lee and Yu (2015) for an overview of potential issues related to MLE of spatial panels with large N. To provide some indication of the likely computational gains of our method, it is worth pointing out in the Monte Carlo section of this paper we found that the total length of time taken to estimate 2,000 replications of the model when $N = T = 200$, was roughly 4.5 minutes for IV and 4.5 hours for QMLE. Therefore, QMLE was 60 times slower than IV in this specific design.[2]

Secondly, the proposed IV approach is free from asymptotic bias. In contrast, existing QML estimators suffer from incidental parameter bias, depending on the sample size as well as on the magnitude of unknown parameters of the data generating process (DGP). Unfortunately, approximate procedures aiming to re-center 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 it is confirmed in the Monte Carlo study presented in this paper.

[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).

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

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. Even in cases where such instruments are not easy to find, our approach at least provides a natural framework for testing for endogeneity of the regressors, based on the overidentifying restrictions test statistic. In contrast, the restriction of exogeneity of the regressors is difficult to verify within the MLE framework and therefore it is typically taken for granted.

There is substantial literature on dynamic panel data models 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 shocks, common factors and strong cross-sectional dependence; see e.g. Chudik and Pesaran (2015), Everaert and De Groote (2016), Moon and Weidner (2017), Juodis et al. (2020) and Norkute et al. (2020). 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 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 estimator: in the first step, the common factors in the exogenous covariates are projected out using principal components analysis, as in Bai (2009). Next, the slope parameters of the model are estimated using standard IV regression, which makes use of instruments constructed from the defactored regressors. In the second step, the factors that enter into the error term of the model are extracted from the first-step IV residuals. Subsequently, the entire model is defactored, and the model parameters are re-estimated using the same instruments as in step one.

The strategy above requires that the covariates used to construct instruments are strictly exogenous with respect to the purely idiosyncratic error term. That is, endogeneity arises primarily due to non-zero correlations between the regressors and the common factor component. Otherwise, the proposed approach requires the use of external instruments, which are exogenous with respect to the idiosyncratic disturbance, although they can be potentially correlated with the common factor component.

The proposed IV estimator is consistent and asymptotically normally distributed as $N, T \to \infty$, with $N/T^2 \to 0$ and $T/N^2 \to 0$. Moreover, the proposed 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 of our procedure. In addition, the proposed estimator is not subject to "Nickell bias" (bias of order $O(T^{-1})$) that arises with least squares and QML-type estimators in dynamic panel data models.

The underlying assumption behind our approach is that the covariates of the model are subject to a linear common factor structure. While this poses certain restrictions on the data generating process from a statistical point of view, there exist several economic theories and plenty of evidence that provide support for such assumption (see e.g. Favero et al. (2005) and Heckman et al. (2006)). Furthermore, this assumption has been frequently employed in both econometrics and statistics literature (see e.g. Pesaran et al. (2013), Bai and Li (2013) and Hansen and Liao (2018), among many others.) Finally, it is worth noting that we do not restrict the common factors that hit the regressors to be identical to the factors that enter into the regression disturbance.

We apply our methodology to study the determinants of risk attitude of banking institutions, with emphasis on the impact of increased capital regulation. To the best of our knowledge, this is the first paper in the banking literature that estimates state dependence and endogenous network effects, while controlling at the same time for unobserved aggregate shocks. The results of our analysis bear important policy implications and provide evidence that the more risk-sensitive capital regulation that was introduced by the Basel III framework in 2011 has succeeded in influencing banks' behaviour in a substantial manner.

The remainder of this paper is organised as follows. Section 2 describes the model and the main idea behind the proposed method. Section 3 lists the set of assumptions employed and derives the large sample properties of the proposed IV estimator. Section 4 examines the finite sample performance of the estimator and confirms that it performs well. Section 5 presents the empirical illustration. A final section concludes. Proofs of the main results of the paper are documented in the Online Appendix of the paper.

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 Main Idea

We consider the following spatial dynamic panel data model with exogenous covariates:

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

$i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, 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 denoted as $\boldsymbol{\beta}$. 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, due to habit formation, costs of adjustment, and state dependence.[3]

[3]It is straightforward enough to extend this model by adding a spatial-time lag, as e.g. in Shi and Lee (2017). We do not explicitly consider this specification here in order to simplify the exposition. The theory developed in the present paper remains valid for this case, with only minor modifications. Simulation results for this specification are reported in Section 4. Furthermore, exogenous network effects, e.g. through an additional term $\sum_{j=1}^{N} w_{ij}\mathbf{x}_{jt}'\boldsymbol{\delta}$, and further lagged values of y_{it} can also be allowed in a straightforward manner without affecting the main derivations of the paper.

The error term of the model is composite, in particular, $\mathbf{h}_t^0$ denotes an $r_y \times 1$ vector of latent factors, $\boldsymbol{\varphi}_i^0$ denotes the corresponding factor loadings, and ε_{it} is a purely idiosyncratic error component.

In order to ensure that the covariates are endogenous with respect 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 a $r_x \times 1$ 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 order $k \times 1$. 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.

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 Eq. (2.1) allows for dynamics through the presence of a lagged dependent variable. In addition, the covariates in Eq. (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, in practice, the common shocks that hit y and X may not be identical.

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. 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 Eq. (2.3) can be written more succinctly as follows:

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

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

Suppose for the moment that both $\mathbf{H}^0$ and $\mathbf{F}^0$ are observed. Let $\mathbf{F}_{-1}^0 \equiv L^1\mathbf{F}^0$, where L^1 denotes the time series lag operator. Let also $\mathbf{M}_{\mathbf{F}^0}$ and $\mathbf{M}_{\mathbf{F}_{-1}^0}$ denote $T \times T$ matrices that project onto the orthogonal complement of $\mathbf{F}^0$ and $\mathbf{F}_{-1}^0$, respectively.

Our approach involves two steps. In the first step, we eliminate the common factors in $\mathbf{X}_i$ and $\mathbf{X}_{i,-1}$ ($\equiv L^1\mathbf{X}_i$) by pre-multiplying these matrices by $\mathbf{M}_{\mathbf{F}^0}$ and $\mathbf{M}_{\mathbf{F}_{-1}^0}$, respectively. Note that by construction, $\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i = \mathbf{M}_{\mathbf{F}^0}\mathbf{V}_i$ and $\mathbf{M}_{\mathbf{F}_{-1}^0}\mathbf{X}_{i,-1} = \mathbf{M}_{\mathbf{F}_{-1}^0}\mathbf{V}_{i,-1}$.

Therefore, assuming that $\mathbf{V}_i$ is independent of $\mathbf{u}_i$, it is straightforward to see that $E\left[\mathbf{X}_i\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i\right] = E\left[\mathbf{V}_i\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i\right] = \mathbf{0}$ and $E\left[\mathbf{X}_{i,-1}\mathbf{M}_{\mathbf{F}_{-1}^0}\mathbf{u}_i\right] = E\left[\mathbf{V}_{i,-1}\mathbf{M}_{\mathbf{F}_{-1}^0}\mathbf{u}_i\right] = \mathbf{0}$.

Thus, under certain regularity conditions to be specified shortly, the following matrix of instruments

$$\mathbf{Z}_i = \left(\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i,\ \ \mathbf{M}_{\mathbf{F}_{-1}^0}\mathbf{X}_{i,-1},\ \ \mathbf{M}_{\mathbf{F}^0}\sum_{j=1}^{N} w_{ij}\mathbf{X}_j\right) \tag{2.5}$$

satisfies the orthogonality condition $E\left[\mathbf{Z}_i'\mathbf{u}_i\right] = \mathbf{0}$. Moreover, provided that $E\left[\mathbf{Z}_i'\mathbf{C}_i\right] \neq \mathbf{0}$, $\boldsymbol{\theta}$ can be estimated consistently by linear IV regression using $\mathbf{Z}_i$.

Remark 2.1 The validity of the procedure above crucially hangs on the assumption that $\mathbf{X}_i$ is strictly exogenous with respect to $\boldsymbol{\varepsilon}_i$. Violations of such restriction are detectable using the overidentifying restrictions test statistic, which is readily available within our framework. 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 Eq. (2.5) replaced by the external instruments. As it is common practice in the literature (e.g. Robertson and Sarafidis (2015) and Kuersteiner and Prucha (2018)), 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 the Monte Carlo study of Section 4.

Remark 2.2 More instruments can be available with respect to further lags of $\mathbf{x}_{it}$, such as $\mathbf{M}_{\mathbf{F}^0_{-\tau}}\mathbf{X}_{i,-\tau}$, for $\tau > 1$, or with respect to lags of the spatial covariates, such as $\mathbf{M}_{\mathbf{F}^0_{-\tau}}\sum_{j=1}^{N} w_{ij}\mathbf{X}_{j,-\tau}$, for $\tau \geq 1$. In addition, instruments constructed from powers of the spatial weights matrix can be available, such as $\mathbf{M}_{\mathbf{F}^0}\sum_{j=1}^{N} w_{ij}^{(\ell)}\mathbf{X}_j$, for $\ell = 2, 3, \dots$, 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.[4] In the present paper we assume that both $\tau \geq 1$ and $\ell \geq 1$ are small and do not depend on T.[5]

The second step of our approach involves extracting $\mathbf{H}^0$ from the residuals obtained by the first-step IV regression. Subsequently, $\mathbf{H}^0$ is eliminated from the entire model using the following transformation:

$$\mathbf{M}_{\mathbf{H}^0}\mathbf{y}_i = \mathbf{M}_{\mathbf{H}^0}\mathbf{C}_i\boldsymbol{\theta} + \mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i. \tag{2.6}$$

Based on similar reasoning as before, we can easily see that $E\left[\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i\right] = \mathbf{0}$. Thus, it is straightforward to apply linear IV regression to the transformed model in Eq. (2.6) using $\mathbf{Z}_i$.[6]

Remark 2.3 Since our approach makes use of the defactored covariates as instruments, identification of the autoregressive and spatial parameters requires that *at least one* element

[4] 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.

[5] The limit behaviour of the estimators when the number of instruments increases with T might be of theoretical interest, however it is beyond the scope of this paper. See Alvarez and Arellano (2003), among others, for a related analysis.

[6] The second-step estimation is equivalent to employing the transformed instrument set, $\mathbf{M}_{\mathbf{H}^0}\mathbf{Z}_i$, to the original model in Eq. (2.4).

of $\boldsymbol{\beta}$ is not equal to zero. Otherwise, it is easily seen from Eq. (2.5) that identification of ρ and ψ is not possible since the lagged and spatial defactored covariates become irrelevant instruments. We believe that this requirement is mild, especially compared to the restriction that all of the elements in $\boldsymbol{\beta}$ are non-zero. Moreover, this restriction is common in 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 non-zero coefficients, since by construction the IV estimation procedure does not require all instruments to be relevant to all endogenous regressors.

In practice, the factor component is typically unobserved. Therefore, we propose estimation of spans of $\mathbf{H}^0$ and $\mathbf{F}^0$, using principal components analysis, as advanced by Bai (2003) and Bai (2009), among many others.[7]

In particular, in what follows we employ a principal component estimator that minimises (say) $N^{-1}\sum_{i=1}^{N}||\mathbf{X}_i - \mathbf{F}\boldsymbol{\Gamma}_i||^2$, subject to the restrictions $\mathbf{F}'\mathbf{F}/T = \mathbf{I}_T$ and $N^{-1}\sum_{i=1}^{N}\text{vec}(\boldsymbol{\Gamma}_i)'\text{vec}(\boldsymbol{\Gamma}_i)$ being diagonal. A similar procedure applies for $\mathbf{F}_{-1}$, based on $L^1\mathbf{X}_i$. Let $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{F}}_{-1}$ be defined as $\sqrt{T}$ times the eigenvectors corresponding to the r_x largest eigenvalues of the $T\times T$ matrices $(NT)^{-1}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{X}_i'$ and $(NT)^{-1}\sum_{i=1}^{N}\mathbf{X}_{i,-1}\mathbf{X}_{i,-1}'$, respectively. In addition, let $\widehat{\boldsymbol{\Gamma}}_i = T^{-1}\widehat{\mathbf{F}}'\mathbf{X}_i$, $\widehat{\mathbf{V}}_i = \mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i$ and $\widehat{\mathbf{V}}_{i,-1} = \mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{i,-1}$. Hence, the matrix of instruments employed in practice is given by $\widehat{\mathbf{Z}}_i = \left(\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i, \mathbf{M}_{\widehat{\mathbf{F}}_{-1}}\mathbf{X}_{i,-1}, \sum_{j=1}^{N} w_{ij}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_j\right)$. The first-step IV estimator of $\boldsymbol{\theta} = (\psi, \rho, \boldsymbol{\beta}')'$, is defined as follows:

$$\widehat{\boldsymbol{\theta}} = \left(\widehat{\mathbf{A}}'\widehat{\mathbf{B}}^{-1}\widehat{\mathbf{A}}\right)^{-1}\widehat{\mathbf{A}}'\widehat{\mathbf{B}}^{-1}\widehat{\mathbf{c}}_y, \tag{2.7}$$

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).

Let $\widehat{\mathbf{H}}$ denote the scaled PC estimator of $\mathbf{H}$, defined as $\sqrt{T}$ times the eigenvectors corresponding to the r_y largest eigenvalues of the $T\times T$ matrices $(NT)^{-1}\sum_{i=1}^{N}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'$.

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

$$\widetilde{\boldsymbol{\theta}} = (\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}})^{-1}\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{c}}_y \tag{2.8}$$

where

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

A particularly useful diagnostic within IV estimation is the so-called overidentifying restrictions (J) test statistic. In our context, this test is expected to pick up potential violations

[7]The present section treats the number of factors, r_y and r_x, as given. In practice, r_x can be straightforwardly estimated from the raw data $\mathbf{x}_{it}$, $t = 0, ..., T$, $i = 1, ..., N$, using either the information criterion approach of Bai and Ng (2002), or the eigenvalue method of Ahn and Horenstein (2013). Similarly, r_y can be estimated from the residual covariance matrix using the aforementioned methods. The results reported in the Monte Carlo section indicate that these methods provide quite an accurate determination of the number of factors in our experimental design.

of exogeneity of the defactored covariates with respect to the idiosyncratic error in the DGP for y, ε_{it}. The J test statistic is given by

$$J = \frac{1}{NT}\left(\sum_{i=1}^{N}\tilde{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right)\widehat{\boldsymbol{\Omega}}^{-1}\left(\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\tilde{\mathbf{u}}_i\right) \tag{2.9}$$

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

3 Asymptotic properties

This section analyses the limiting properties of the proposed two-step IV estimator $\tilde{\boldsymbol{\theta}}$. The following assumptions are employed throughout the paper:

Assumption A (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 B (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 $\|\boldsymbol{\Sigma}_{ij,st}\| \leq \tilde{\sigma}_{st}$ 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.

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

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

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

Assumption E (weighting matrix) *The weights matrix* $\mathbf{W}_N$ *satisfies that*

1. *All diagonal elements of* $\mathbf{W}_N$ *are zeros;*
2. *The matries* $\mathbf{W}_N$ *and* $\mathbf{I}_N - \psi\mathbf{W}_N$ *are invertible;*
3. *The row and column sums of the matrices* $\mathbf{W}_N$ *and* $(\mathbf{I}_N - \psi\mathbf{W}_N)^{-1}$ *are bounded uniformly in absolute value.*
4. $$\sum_{\ell=0}^{\infty}\left\|[\rho(\mathbf{I}_N-\psi\mathbf{W}_N)^{-1}]^\ell\right\|_\infty \leq C; \quad \sum_{\ell=0}^{\infty}\left\|[\rho(\mathbf{I}_N-\psi\mathbf{W}_N)^{-1}]^\ell\right\|_1 \leq C$$

Assumption F (identification) *We assume that*

1. *The matrices* $T^{-1}\mathbf{Z}_i'\mathbf{Z}_i$ *and* $T^{-1}\mathbf{Z}_i'\mathbf{C}_i$ *have full column rank for all* i *and* T*;*
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.

The assumptions above merit some discussion. Assumption A is imposed mainly for simplicity. In particular, in practice ε_{it} can be heterogeneously distributed across both i and t. However, as it commonly the case in a large body of the panel data literature based on Method of Moments estimation, we do not consider such generalizations in order to avoid unnecessary notational complexity.

Assumption B implies that the covariates of the model, $\mathbf{x}_{it}$, are strictly exogenous with respect to ε_{it}, i.e. $E\left(\varepsilon_{it}|\mathbf{x}_{is}'\right) = 0$ for all t and s. This assumption is often employed in the panel data literature with common factor residuals when both N and T are large (see e.g. Pesaran (2006) and Bai (2009)). Assumption B implies that defactored covariates are valid instruments for the endogenous variables of the model. In addition, Assumption B allows for cross-sectional and time series heteroskedasticity, 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 $\mathbf{V}_i$.

Assumptions C and D are standard in the principal components literature; see e.g. Bai (2003), among others. Assumption C permits correlations between $\mathbf{f}_t^0$ and $\mathbf{h}_t^0$, and within each one of them. Assumption D allows for possible non-zero 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 E is standard in the spatial literature, see e.g. Kelejian and Prucha (2010). In particular, Assumption E.1 is just a normalisation of the model and implies that no individual is viewed as its own neighbour. Assumption E.2 implies that there is no dominant unit in the sample, i.e. an individual unit that is asymptotically, for N large, correlated with all remaining individuals. Assumptions E.3-E.4 concern the parameter space of ψ and are discussed in detail 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 E holds true. Moreover, $\mathbf{W}_N$ is not required to be row normalized. Although it is convenient to work with a row-normalised weighting matrix, in some applications, especially those analysing social interactions and network structures, row normalisation might not always be appropriate.

Finally, Assumption F ensures identification based on IV regression, see e.g. Wooldridge (2002, Ch. 5).

The asymptotic properties of the one-step estimator are determined primarily by those of $\widehat{\mathbf{Z}}_i'\mathbf{u}_i/\sqrt{NT}$. Thus, the following proposition is useful:

Proposition 3.1 *Under Assumptions A-F, we have*

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{Z}_i'\mathbf{u}_i + O_p\left(\sqrt{\frac{T}{N}}\right) + O_p\left(\sqrt{\frac{N}{T}}\right) + o_p(1)$$

where $\mathbb{Z}_i = \left(\sum_{j=1}^{N} w_{ij}\mathbf{M}_{\mathbf{F}}\boldsymbol{\mathcal{X}}_j, \mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{F}^0_{-1}}\boldsymbol{\mathcal{X}}_{i,-1}, \mathbf{M}_{\mathbf{F}^0}\boldsymbol{\mathcal{X}}_i\right)$ *with* $\boldsymbol{\mathcal{X}}_i = \mathbf{X}_i - \frac{1}{N}\sum_{\ell=1}^{N}\mathbf{X}_\ell\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_i^0$, $\boldsymbol{\mathcal{X}}_{i,-1} = \mathbf{X}_{i,-1} - \frac{1}{N}\sum_{\ell=1}^{N}\mathbf{X}_{\ell,-1}\boldsymbol{\Gamma}_\ell^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_i^0, \boldsymbol{\Upsilon}^0 = N^{-1}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i^0\boldsymbol{\Gamma}_i^{0\prime}$.

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

Theorem 3.1 *Under Assumptions A-F, as* $N, T \to \infty$, $N/T^2 \to 0$ *and* $T/N^2 \to 0$, *we have*

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

Asymptotic normality follows through by using similar arguments as for $\widetilde{\boldsymbol{\theta}}$ below. To save space, we do not derive this property explicitly here because $\widehat{\boldsymbol{\theta}}$ is mainly used to estimate $\mathbf{H}$.

Since the asymptotic properties of the two-step estimator are determined primarily by those of $\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i/\sqrt{NT}$, in what follows we focus on this particular term. The formal analysis is provided as a proposition below.

Proposition 3.2 *Under Assumptions A-F, 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 centered at zero. Hence, the asymptotic normality result can be readily obtained by applying the central limit theorem for martingale differences in Kelejian and Prucha (2001).

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

Theorem 3.2 *Under Assumptions A-F, as $N,T\to\infty$, $N/T^2\to 0$ and $T/N^2\to 0$, we have*

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

where $\boldsymbol{\Psi}=\sigma_\varepsilon^2\left(\mathbf{A}_0'\mathbf{B}_0^{-1}\mathbf{A}_0\right)^{-1}$, $\mathbf{A}_0=\text{plim}_{N,T\to\infty}\mathbf{A}$, $\mathbf{B}_0=\text{plim}_{N,T\to\infty}\mathbf{B}$, *with*

$$\mathbf{A}=\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{C}_i\,,\mathbf{B}=\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i\,.$$

Moreover, $\widetilde{\boldsymbol{\Psi}}-\boldsymbol{\Psi}\xrightarrow{p}\mathbf{0}$ *as* $N,T\to\infty$, *where*

$$\widetilde{\boldsymbol{\Psi}}=\tilde{\sigma}_\varepsilon^2\left(\widetilde{\mathbf{A}}'\widetilde{\mathbf{B}}^{-1}\widetilde{\mathbf{A}}\right)^{-1}\,.$$

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

The limit distribution of the overidentifying restrictions test statistic is established in the following theorem:

Theorem 3.3 *Under Assumptions A-F, as $N,T\to\infty$, $N/T^2\to 0$ and $T/N^2\to 0$, we have*

$$J\xrightarrow{d}\chi_\nu^2$$

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

4 Monte Carlo Experiments

This section investigates the finite sample behaviour of the proposed approach by means of Monte Carlo experiments. We shall focus on the mean, root mean squared error (RMSE), empirical size and power of the t-test. In addition, we report results on the absolute relative bias (ARB), which is defined as

$$ARB=\frac{|\widehat{\theta}_\ell-\theta_\ell|}{\theta_\ell}\times 100, \tag{4.1}$$

where θ_ℓ denotes the ℓth entry of $\boldsymbol{\theta}=(\psi,\rho,\boldsymbol{\beta}')'$.

4.1 Design

We consider the following spatial dynamic panel data model:

$$y_{it} = \alpha_i + \rho y_{it-1} + \psi \sum_{j=1}^{N} w_{ij} y_{jt} + \sum_{\ell=1}^{k} \beta_\ell x_{\ell it} + u_{it};\ u_{it} = \sum_{s=1}^{r_y} \varphi_{si}^0 f_{s,t}^0 + \varepsilon_{it}, \tag{4.2}$$

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

$$f_{s,t}^0 = \rho_{fs} f_{s,t-1}^0 + (1 - \rho_{fs}^2)^{1/2} \zeta_{s,t}, \tag{4.3}$$

with $\zeta_{s,t} \sim i.i.d.N(0, 1)$ for $s = 1, ..., r_y$. We set $\rho_{fs} = 0.5$ for all s, $k = 2$ and $r_y = 3$. Therefore, there are two regressors and three factors.

The spatial weighting matrix, $\mathbf{W}_N = [w_{ij}]$ is an invertible rook matrix of circular form (see Kappor et al. (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 process for the covariates is given by

$$x_{\ell it} = \mu_{\ell i} + \sum_{s=1}^{r_x} \gamma_{\ell si}^0 f_{s,t}^0 + v_{\ell it};\ i = 1, 2, ..., N; \quad t = -49, -48, ..., T, \tag{4.4}$$

for $\ell = 1, 2$.

We set $r_x = 2$. This implies that the first two factors in u_{it}, f_{1t}^0, f_{2t}^0, also drive the DGP for $x_{\ell it}$, $\ell = 1, 2$. However, f_{3t}^0 does not enter into the DGP of the covariates directly. Observe that, using notation of earlier sections, $\mathbf{f}_t^0 = (f_{1t}^0, f_{2t}^0)'$, and $\mathbf{h}_t^0 = (f_{1t}^0, f_{2t}^0, f_{3t}^0)'$.

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

$$v_{\ell it} = \rho_{v,\ell} v_{\ell it-1} + (1 - \rho_{v,\ell}^2)^{1/2} \varpi_{\ell it};\ \varpi_{\ell it} \sim i.i.d.N(0, \varsigma_v^2), \tag{4.5}$$

for $\ell = 1, 2$. We set $\rho_{v,\ell} = \rho_v = 0.5$ 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

$$\alpha_i \sim i.i.d.N(0, (1 - \rho)^2); \quad \mu_{\ell i} = \rho_{\mu,\ell} \alpha_i + (1 - \rho_{\mu,\ell}^2)^{1/2} \omega_{\ell i}, \tag{4.6}$$

where $\omega_{\ell i} \sim i.i.d.N(0, (1 - \rho)^2)$, for $\ell = 1, 2$. We set $\rho_{\mu,\ell} = 0.5$ for $\ell = 1, 2$.
The factor loadings in u_{it} are generated as $\varphi_{si}^0 \sim i.i.d.N(0, 1)$ for $s = 1, ..., r_y (= 3)$, and the factor loadings in x_{1it} and x_{2it} are drawn as

$$\gamma_{1si}^0 = \rho_{\gamma,1s} \varphi_{3i}^0 + (1 - \rho_{\gamma,1s}^2)^{1/2} \xi_{1si};\ \xi_{1si} \sim i.i.d.N(0, 1); \tag{4.7}$$

$$\gamma^0_{2si} = \rho_{\gamma,2s}\gamma^0_{si} + (1-\rho^2_{\gamma,2s})^{1/2}\xi_{2si};\ \xi_{2si} \sim i.i.d.N(0,1); \tag{4.8}$$

respectively, for $s = 1, ..., r_x = 2$. The process in Eq. (4.7) allows the factor loadings to $f^0_{1,t}$ and $f^0_{2,t}$ 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. $f^0_{3,t}$. On the other hand, Eq. (4.8) ensures that the factor loadings to $f^0_{1,t}$ and $f^0_{2,t}$ in x_{2it} are allowed to be correlated with the factor loadings corresponding to the same factors in u_{it}, $f^0_{1,t}$ and $f^0_{2,t}$. We consider $\rho_{\gamma,11} = \rho_{\gamma,12} \in \{0, 0.5\}$, whilst $\rho_{\gamma,21} = \rho_{\gamma,22} = 0.5$.

It is straightforward to see that the average variance of ε_{it} depends only on ς^2_ε. Let π_u denote the proportion of the average variance of u_{it} that is due to ε_{it}. That is, we define $\pi_u := \varsigma^2_\varepsilon / \left(r_y + \varsigma^2_\varepsilon\right)$. 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^2_ε yields

$$\varsigma^2_\varepsilon = \frac{\pi_u}{(1-\pi_u)} r_y. \tag{4.9}$$

We set ς^2_ε such that $\pi_u \in \{1/4, 3/4\}$.[8]

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

$$SNR := \frac{\text{var}\left[(y_{it} - \varepsilon_{it})\,|\mathcal{L}\right]}{\overline{\text{var}}\left(\varepsilon_{it}\right)} = \frac{\left(\frac{\beta_1^2+\beta_2^2}{1-\rho_\upsilon^2}\right)\varsigma^2_\upsilon + \frac{\varsigma^2_\varepsilon}{1-\rho^2_\upsilon} - \varsigma^2_\varepsilon}{\varsigma^2_\varepsilon}, \tag{4.10}$$

where $\mathcal{L}$ is the information set that contains the factor structure, the individual-specific effects and the spatial lag[9], whereas $\overline{\text{var}}\left(\varepsilon_{it}\right)$ is the overall average of $E\left(\varepsilon^2_{it}\right)$ over i and t. Solving for ς^2_υ yields

$$\varsigma^2_\upsilon = \varsigma^2_\varepsilon \left[SNR - \frac{\rho^2_\upsilon}{1-\rho^2_\upsilon}\right]\left(\frac{\beta_1^2+\beta_2^2}{1-\rho^2_\upsilon}\right)^{-1}. \tag{4.11}$$

We set $SNR = 4$, which lies with the range $\{3, 9\}$ considered by the simulation study of Bun and Kiviet (2006) and Juodis and Sarafidis (2018).

We set $\rho = 0.4$, $\psi = 0.25$, and $\beta_1 = 3$ and $\beta_2 = 1$, following Bai (2009).

In addition to the model provided in Eq. (4.2), we also consider an augmented model that includes a spatial-time lag as an additional covariate, that is,

$$y_{it} = \alpha_i + \rho y_{it-1} + \psi \sum_{j=1}^{N} w_{ij} y_{jt} + \sum_{\ell=1}^{k} \beta_\ell x_{\ell it} + \psi_1 \sum_{j=1}^{N} w_{ij} y_{jt-1} + u_{it}. \tag{4.12}$$

We study the optimal two-step IV estimator, defined in Eq. (2.8), based on the following

[8]These values of π_u are motivated by the results in Sargent and Sims (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.

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

set of instruments stack in the $(T \times 4k)$ matrix $\hat{\mathbf{Z}}_i$:

$$\hat{\mathbf{Z}}_i = \left(\mathbf{M}_{\widehat{\mathbf{F}}} \underline{\mathbf{X}}_i, \quad \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \underline{\mathbf{X}}_{i,-1}, \quad \mathbf{M}_{\widehat{\mathbf{F}}} \sum_{j=1}^{N} w_{ij} \underline{\mathbf{X}}_j, \quad \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \sum_{j=1}^{N} w_{ij} \underline{\mathbf{X}}_{j,-1} \right). \tag{4.13}$$

$\underline{\mathbf{X}}_i$ is defined as $\underline{\mathbf{X}}_i \equiv \mathbf{X}_i - \overline{\mathbf{X}}$, i.e. the covariates are cross-sectionally demeaned in order to control for individual-specific fixed effects. $\mathbf{M}_{\widehat{\mathbf{F}}}$ is computed based on the estimated factors obtained as $\sqrt{T}$ times the eigenvectors corresponding to the $\hat{r}_x$ largest eigenvalues of the $T \times T$ matrices $(NT)^{-1} \sum_{i=1}^{N} \underline{\mathbf{X}}_i \underline{\mathbf{X}}_i'$, where $\underline{\mathbf{X}}_i = (\underline{\mathbf{x}}_{1i}, \underline{\mathbf{x}}_{2i})$, $\underline{\mathbf{x}}_{\ell i} = \mathbf{x}_{\ell i} - \overline{\mathbf{x}}_{\ell}$, and $\mathbf{x}_{\ell i} = (x_{\ell i1}, \ldots, x_{\ell iT})'$. Similar definitions apply for all remaining projection matrices. Thus, the above set of instruments employs contemporaneous and lagged defactored covariates, as well as contemporaneous and lagged spatial defactored covariates.[10]

In order to allow for cross-section and time series heteroskedasticity, the variance estimator for the two-step IV procedure is given by

$$\widetilde{\boldsymbol{\Psi}} = \left(\widetilde{\mathbf{A}}' \widetilde{\mathbf{B}}^{-1} \widetilde{\mathbf{A}} \right)^{-1} \widetilde{\mathbf{A}}' \widetilde{\mathbf{B}}^{-1} \widehat{\boldsymbol{\Omega}} \widetilde{\mathbf{B}}^{-1} \widetilde{\mathbf{A}} \left(\widetilde{\mathbf{A}}' \widetilde{\mathbf{B}}^{-1} \widetilde{\mathbf{A}} \right)^{-1}, \tag{4.14}$$

with

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{NT} \sum_{i=1}^{N} \hat{\mathbf{Z}}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{M}_{\widehat{\mathbf{F}}} \hat{\mathbf{Z}}_i, \tag{4.15}$$

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

In order to check the performance of the IV estimator with endogenous covariates, as well as the power of the overidentifying restrictions test (defined in Eq. (2.9)), we consider the situation where $\mathbf{x}_{it}$ and ε_{it} are contemporaneously correlated. In particular, for $\ell = 1$ the DGP given by Eq. (4.4) is replaced by the following one:

$$x_{1it} = \mu_{1i} + \sum_{s=1}^{r_x} \gamma_{1si}^0 f_{s,t}^0 + v_{1it} + 0.5\varepsilon_{it}; \; i = 1, 2, ..., N; \quad t = -49, -48, ..., T. \tag{4.16}$$

This implies that x_{1it} is endogenous with respect to ε_{it}, while x_{2it} remains strictly exogenous.[11] We construct a single external instrument, denoted as x_{3it}, which is given by

$$x_{3it} = \mu_{3i} + \sum_{s=1}^{r_x} \gamma_{3si}^0 f_{s,t}^0 + v_{1it} + \vartheta v_{3it}; \; i = 1, 2, ..., N; \quad t = -49, -48, ..., T, \tag{4.17}$$

[10] We have also explored the performance of two additional IV estimators. The first one omits the lagged spatial defactored covariates, i.e. it excludes the last term in Eq. (4.13 from the set of instruments. The second one replaces the lagged spatial defactored covariates with contemporaneous defactored covariates that arise from the square of the spatial weighting matrix, as e.g. in Kelejian and Prucha (2007), p. 143. In the baseline model ($\psi_1 = 0$), all three estimators perform similarly. However, when $\psi_1 \neq 0$, the two IV estimators that do not make use of lagged spatial defactored covariates as instruments, deteriorate substantially in terms of their performance. This is mainly because ψ_1 is weakly identified in this case. In order to facilitate the comparison of the results obtained for the two DGP's in Eq. (4.2) and Eq. (4.12), we shall focus on the IV estimator that makes use of the instruments in Eq. (4.13). The results for the remaining IV estimators are available from the authors upon request.

[11] The power of the J statistic is expected to be higher when both covariates are endogenous.

where μ_{3i}, v_{3it} and γ_{3si}^0 are generated as in Eqs. (4.6), (4.5) and (4.8) respectively. The value of ϑ is set such that the correlation between $(v_{1it} + 0.5\varepsilon_{it})$ in Eq. (4.16) and $(v_{1it} + \vartheta v_{3it})$ in Eq. (4.17) equals 0.5. To obtain a consistent IV estimator, the matrix of instruments is revised as follows:

$$\hat{\mathbf{Z}}_i = \left(\mathbf{M}_{\widehat{\mathbf{F}}} \underline{\widetilde{\mathbf{X}}}_i, \quad \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \underline{\widetilde{\mathbf{X}}}_{i,-1}, \quad \mathbf{M}_{\widehat{\mathbf{F}}} \sum_{j=1}^{N} w_{ij} \underline{\widetilde{\mathbf{X}}}_j, \quad \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} \sum_{j=1}^{N} w_{ij} \underline{\widetilde{\mathbf{X}}}_{j,-1} \right) \tag{4.18}$$

where $\underline{\widetilde{\mathbf{X}}}_i = (\underline{\mathbf{x}}_{3i}, \underline{\mathbf{x}}_{2i})$ is of order $T \times 2$. The projection matrix $\mathbf{M}_{\widehat{\mathbf{F}}}$ is computed based on the estimated factors obtained as $\sqrt{T}$ times the eigenvectors corresponding to the $\hat{r}_x$ largest eigenvalues of the $T \times T$ matrices $(NT)^{-1} \sum_{i=1}^{N} \underline{\widetilde{\mathbf{X}}}_i \underline{\widetilde{\mathbf{X}}}_i'$. A similar definition applies for $\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}$.

As a benchmark, we also consider the bias-corrected QMLE estimator proposed by Shi and Lee (2017).[12] Note that the QMLE estimator is not consistent when the covariates are endogenous with respect to the idiosyncratic error. However, it is of interest to see the extent to which the performance of the estimator is affected under these circumstances.

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/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 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.

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

4.2 Results

Tables 4.1–4.4 report simulation results for the baseline model in Eq. (4.2) for $\pi_u = 3/4$. Results for $\pi_u = 1/4$ can be found in Online Appendix C. "Mean" denotes the average value of the estimated parameters across 2,000 replications. Similarly, "RMSE" represents the average squared deviation of the estimated parameter from its true value across 2,000 samples. "ARB" denotes absolute relative bias, which is defined in Eq. (4.1). Note that the power of the "t-test" is the size-corrected power, for which the 5% critical values used are obtained as the 2.5% and 97.5% quantiles of the empirical distribution of the t-ratio under the null hypothesis.[13]

As we can see from all four tables, 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 shall mainly focus in our discussion on relative RMSE performance, ARB and size properties of the two estimators.

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

[13] The size-adjusted power is employed in the present experimental study because of size distortions, which otherwise make the power comparison between the two estimators difficult.

Table 4.1 presents results for the autoregressive parameter, ρ. QMLE outperforms IV in terms of RMSE, which reflects the higher efficiency of maximum likelihood/least-squares compared to instrumental variables. However, QMLE appears to exhibit substantial ARB and thereby it is severely size-distorted. Note that both ARB and the 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.

Table 4.1: Baseline Model. Results for $\rho = 0.4$, $\pi_u = 3/4$

	IV					QMLE				
	Case I: $N = 100\tau$, $T = 25\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	.400	.017	.082	.065	1.00	.390	.014	2.52	.361	1.00
2	.400	.007	.104	.059	1.00	.396	.006	1.08	.293	1.00
4	.400	.004	.054	.052	1.00	.398	.003	.551	.286	1.00
	Case II: $N = 25\tau$, $T = 100\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	.400	.014	.116	.084	1.00	.400	.008	.169	.138	1.00
2	.400	.007	.033	.066	1.00	.400	.004	.087	.085	1.00
4	.400	.003	.008	.048	1.00	.400	.002	.080	.079	1.00
	Case III: $N = 50\tau$, $T = 50\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	.400	.015	.074	.052	1.00	.397	.009	.870	.146	1.00
2	.400	.007	.026	.056	1.00	.398	.005	.493	.121	1.00
4	.400	.003	.017	.053	1.00	.399	.002	.225	.109	1.00

Table 4.2 focuses on the spatial parameter, ψ. One noticeable difference compared to Table 4.1 is that in this case IV appears to outperform QMLE in terms of RMSE, albeit the difference decreases substantially as the sample size gets larger. As before, IV is subject to some small size distortion when N is small, which however tends to be eliminated quickly as N grows. QMLE is severely size-distorted.

Table 4.2: Baseline Model. Results for $\psi = 0.25$, $\pi_u = 3/4$

	IV					QMLE				
	Case I: $N = 100\tau$, $T = 25\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	.250	.019	.119	.062	1.00	.250	.033	.216	.392	1.00
2	.250	.008	.112	.051	1.00	.250	.015	.159	.258	1.00
4	.250	.004	.080	.052	1.00	.250	.006	.113	.134	1.00
	Case II: $N = 25\tau$, $T = 100\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	.250	.017	.054	.094	1.00	.250	.023	.156	.156	1.00
2	.250	.008	.009	.062	1.00	.250	.010	.226	.078	1.00
4	.250	.004	.016	.054	1.00	.250	.005	.007	.070	1.00
	Case III: $N = 50\tau$, $T = 50\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	.251	.017	.218	.076	1.00	.250	.025	.026	.213	1.00
2	.250	.008	.039	.060	1.00	.250	.011	.092	.139	1.00
4	.250	.004	.012	.054	1.00	.250	.005	.001	.068	1.00

Tables 4.3–4.4 report results for β_1 and β_2, respectively. The results are qualitatively no different from those in Table 4.2, with one exception: when either N or T is small, size-adjusted power appears to be lower compared to the corresponding values in Table 4.2. Moreover, IV often appears to have higher power than QMLE. However, as the sample size increases, power appears to converge towards unity rather quickly.

Table 4.3: Baseline Model. Results for $\beta_1 = 3$, $\pi_u = 3/4$

	IV					QMLE				
	Case I: $N = 100\tau$, $T = 25\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	3.00	.057	.019	.058	.405	3.01	.065	.312	.209	.395
2	3.00	.026	.030	.060	.952	3.01	.026	.204	.087	.982
4	3.00	.013	.022	.051	1.00	3.00	.013	.112	.087	1.00
	Case II: $N = 25\tau$, $T = 100\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	3.00	.052	.021	.082	.465	3.00	.087	.037	.404	.198
2	3.00	.026	.001	.061	.957	3.00	.034	.033	.211	.877
4	3.00	.004	.008	.054	1.00	3.00	.015	.003	.134	1.00
	Case III: $N = 50\tau$, $T = 50\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	3.00	.056	.046	.056	.459	3.00	.065	.144	.236	.370
2	3.00	.027	.004	.067	.931	3.00	.027	.121	.117	.965
4	3.00	.012	.004	.050	1.00	3.00	.012	.045	.085	1.00

Table 4.4: Baseline Model. Results for $\beta_2 = 1$, $\pi_u = 3/4$

	IV					QMLE				
	Case I: $N = 100\tau$, $T = 25\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	1.00	.066	.947	.109	.350	1.06	.093	6.41	.415	.345
2	1.00	.025	.087	.057	.981	1.01	.029	1.35	.142	.982
4	1.00	.012	.039	.055	1.00	1.00	.012	.379	.079	1.00
	Case II: $N = 25\tau$, $T = 100\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	1.00	.049	.043	.086	.465	1.07	.118	7.10	.531	.179
2	1.00	.025	.043	.074	.972	1.02	.040	1.81	.306	.876
4	1.00	.012	.003	.059	1.00	1.00	.014	.361	.134	1.00
	Case III: $N = 50\tau$, $T = 50\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	1.00	.050	.222	.063	.539	1.05	.086	5.18	.369	.311
2	1.00	.024	.023	.052	.985	1.01	.028	1.15	.156	.987
4	1.00	.012	.023	.050	1.00	1.00	.012	.228	.090	1.00

Table 4.5 reports simulation results corresponding to the augmented spatial panel data model in Eq. (4.12). This model includes a spatial-time lag in addition to the remaining covariates. As argued in Section 2, the proposed IV estimator remains consistent and asymptotically normal in this model. Table 4.5 focuses on ψ_1. The results for the remaining coefficients are similar to those obtained for the model without a spatial-time lag and

therefore, to save space, we append these results to Online Appendix C. As we can see, in most cases IV has negligible bias with fairly accurate size properties. On the other hand, QMLE tends to be biased and size-distorted unless both N and T are relatively large. QMLE outperforms IV in terms of RMSE.

Table 4.5: Model with spatial-time lag. Results for $\psi_1 = 0.20$, $\pi_u = 3/4$

	IV					QMLE				
					Case I: $N = 100\tau$, $T = 25\tau$					
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	.200	.030	.128	.059	.920	.204	.017	1.98	.167	1.00
2	.200	.014	.128	.056	1.00	.200	.008	.839	.091	1.00
4	.200	.007	.013	.049	1.00	.200	.004	.612	.095	1.00
					Case II: $N = 25\tau$, $T = 100\tau$					
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	.199	.028	.456	.087	.925	.198	.020	.895	.248	1.00
2	.200	.014	.090	.058	1.00	.200	.009	.028	.173	1.00
4	.200	.007	.150	.060	1.00	.200	.004	.052	.089	1.00
					Case III: $N = 50\tau$, $T = 50\tau$					
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	.199	.028	.434	.070	.922	.201	.016	.475	.137	1.00
2	.200	.014	.069	.067	1.00	.201	.008	.478	.111	1.00
4	.200	.007	.059	.063	1.00	.200	.004	.168	.068	1.00

Table 4.6 reports simulation results for the baseline model in Eq. (4.12) with x_1 endogenous, as in Eq. (4.16). The matrix of instruments is given by Eq. (4.18). To save space, we only report results for the coefficient of the endogenous variable, β_1. As we can see, the IV estimator performs more than satisfactorily. In comparison to Table 4.3, two main differences are noteworthy: firstly, for small values of τ the RMSE of the IV estimator increases by a multiple scalar that roughly ranges between 1.3-1.7; secondly, the power of the t-test decreases substantially for small values of either N or T. These results are not surprising given that the correlation between the defactored regressor x_1 and the instrument drops by a half. In regards to QMLE, ARB ranges between $35\% - 40\%$. Moreover, the size of the t-test equals 1 under all circumstances. These results show that QMLE can be severely distorted in practice when some of the covariates are endogenous. As far as the remaining parameters are concerned (not reported here), the estimate of the autoregressive coefficient appears to be more sensitive to endogeneity than the estimate of the spatial parameter. In particular, the bias of the estimate of ρ fluctuates around 6.7%. Furthermore, the size of the estimator is close to 1 in all cases.[14] This implies that detecting possible violations of exogeneity of the regressors is very important in practice.

[14]These simulation results are available upon request.

Table 4.6: Model with endogenous covariate. Results for $\beta_1 = 3$, $\pi_u = 3/4$

	IV					QMLE				
	Case I: $N = 100\tau$, $T = 25\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	3.01	.077	.269	.075	.312	4.19	1.19	39.5	1.00	.340
2	3.00	.034	.049	.051	.831	4.14	1.14	38.0	1.00	.662
4	3.00	.017	.014	.050	1.00	4.12	1.12	37.4	1.00	.964
	Case II: $N = 25\tau$, $T = 100\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	3.00	.067	.123	.083	.359	4.06	1.06	35.2	1.00	.198
2	3.00	.033	.040	.064	.837	4.07	1.07	35.5	1.00	.356
4	3.00	.017	.028	.055	1.00	4.09	1.09	36.3	1.00	.544
	Case III: $N = 50\tau$, $T = 50\tau$									
τ	Mean	RMSE	ARB	Size	Power	Mean	RMSE	ARB	Size	Power
1	3.00	.069	.184	.065	.359	4.11	1.11	37.0	1.00	.236
2	3.00	.034	.035	.063	.812	4.11	1.11	36.9	1.00	.468
4	3.00	.017	.021	.054	1.00	4.10	1.10	36.7	1.00	.792

Table 4.7 summarises the finite sample behaviour for the overidentifying restrictions test. Panel A and Panel B provide the rejection frequencies corresponding to the IV estimator that makes use of the matrix of instruments given by Eq. (4.13). Columns '**I**', '**II**' and '**III**' correspond to $N = 100\tau$, $T = 25\tau$, $N = 25\tau$, $T = 100\tau$ and $N = 50\tau$, $T = 50\tau$, respectively, for $\tau = 1, 2, 4$.

As we can see, the size of the J-test is close to its nominal level (5%) in most cases, with some minor size distortions when N is small. On the other hand, the J-test appears to have substantial power when the exogeneity of (a subset of) the instruments is violated. For example, the power of the test for $N = T = 50$ and $N = T = 100$ is 30.9% and 88.9% respectively.

Panel C provides results on the empirical size of the J-test when the instruments employed are given by Eq. (5.4). In general, some mild distortions are observed only for N small.

Table 4.7: Size and power performance for the J test statistic

	Panel A			Panel B			Panel C		
	size			power			size		
τ	**I**	**II**	**III**	**I**	**II**	**III**	**I**	**II**	**III**
1	.054	.083	.068	.388	.515	.513	.070	.092	.073
2	.055	.063	.067	.978	.989	.988	.063	.074	.062
4	.049	.053	.051	1.00	1.00	1.00	.049	.058	.048

Finally, a desirable aspect of our approach is that it is linear in the parameters and therefore it is computationally inexpensive. Obviously, the exact computational gains depend on the sample size. As an indication, when $N = T = 50$, the total length of time taken to estimate 2,000 replications of the model was roughly 20 seconds for IV and 740 seconds for QMLE. On the other hand, when $N = T = 200$ the corresponding figures were roughly 4.5

minutes and 4.5 hours for IV and QMLE, respectively.[15] Therefore, QMLE was 37 (60) times slower than IV in these specific designs.

In conclusion, the proposed IV approach performs well in all circumstances.[16] In particular, the finite-sample bias of the estimator is negligible in almost all cases examined, and inferences are fairly 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. Moreover, the RMSE of the IV estimator for the slope coefficients (β_1 and β_2) appears in many cases to be substantially smaller than the corresponding values for QMLE. Thus, considering also the computational simplicity of our methodology, the proposed IV estimator presents an attractive estimation approach in spatial panel data models with interactive effects when both N and T are large.

5 An Analysis of Bank Attitude Towards Risk

In this section we apply our methodology to study the determinants of risk attitude of banking institutions, with emphasis on the impact of increased capital regulation over the past decade or so, via the introduction of Basel III. We employ panel data from a random sample of 350 U.S. banking institutions, each one observed over 56 time periods, namely 2006:Q1-2019:Q4.

There is a large number of empirical studies focusing on the analysis of the main drivers of banks' attitude toward risk. Overall, it is fair to say that the literature has not reached consensus on its main findings. For instance, consider the effect of banking operational efficiency on risk-taking behavior: Kwan and Eisenbeis (1996) find that more efficient banks tend to take on less credit risk, whereas Altunbas et al. (2007) conclude the exact opposite. Similarly, consider the effect of capital adequacy requirements to risk attitude: both Shrieves and Dahl (1992) and Jacques and Nigro (1997) infer that higher capital levels exert a positive influence on the level of risk taken by banks, whereas Stolz et al. (2003) and Roy (2005) report a negative effect on risk appetite.

In this section, we pay focus on the spatial interactions of banking behavior, in order to account for banking linkages and endogenous network effects. Furthermore, we control for unobserved common shocks, such as the recent economic recession that took place during the period 2007-2009, which was accompanied by rapidly falling housing prices. To the best of our knowledge, this is the first paper in the banking literature that estimates state dependence and endogenous network effects, while controlling for the impact of unobserved aggregate shocks.

[15] The simulation experiments have been implemented using Matlab on a single CPU with Core i7-6700 @ 3.40 GHz and 16 GB RAM. The algorithm is currently being written as an ado file in Stata 15 and it will be made available to all Stata users on the web.

[16] Simulation results for the case $\pi_u = 1/4$ are reported in the Online Appendix C of the paper. The results are qualitatively similar in terms of a comparison between IV and QMLE. However, in general, both estimators perform a bit better than in the case where $\pi_u = 3/4$.

5.1 Model Specification

We estimate the following model:

$$NPL_{it} = \rho NPL_{it-1} + \psi \sum_{j=1}^{N} w_{ij} NPL_{jt} + \beta_1 INEFF_{it} + \beta_2 CAR_{it} + \beta_3 SIZE_{it} + \beta_4 BUFFER_{it}$$
$$+ \beta_5 PROFITABILITY_{it} + \beta_6 QUALITY_{it} + \beta_7 LIQUIDITY_{it} + \beta_8 PRESSURE_{it} + u_{it};$$

$$u_{it} = \eta_i + \boldsymbol{\varphi}_i' \mathbf{h}_t^0 + \varepsilon_{it}, \tag{5.1}$$

where $i = 1, \ldots, 350$, and $t = 1, \ldots, 56$. All data are publicly available and they have been downloaded from the Federal Deposit Insurance Corporation (FDIC) website.[17]

NPL_{it} denotes the ratio of non-performing loans to total loans for bank i at time period t. This is a popular measure of bank risk.[18] Higher values of the NPL ratio indicate that banks ex-ante took higher lending risk and therefore they have accumulated ex-post more bad loans;

$INEFF_{it}$ denotes the time-varying operational inefficiency of bank i at period t, which has been constructed based on a cost frontier model using a translog functional form, two outputs and three inputs[19];

CAR_{it} stands for "capital adequacy ratio", which is proxied by the ratio of Tier 1 (core) capital over risk-weighted assets;

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

$BUFFER_{it}$ 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%);

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

$QUALITY_{it}$ represents the quality of banks' assets and is computed as the total amount of loan loss provisions (LLP) expressed as a percentage of assets;

$LIQUIDITY_{it}$ is proxied by the loan-to-deposit (LTD) ratio. The main idea is that

[17] See https://www.fdic.gov/.

[18] An alternative such measure is the ratio of risk-weighted assets to total assets. This involves multiplying the amount of different types of assets by the standardised risk weight associated with each type of assets. However, this measure has been criticised because it can be easily manipulated e.g. by engaging in regulatory capital arbitrage; see Vallascas and Hangendorff (2013).

[19] In particular, following Altunbas et al. (2007), we specify

$$\begin{aligned} lnTC_{it} = & \sum_{h=1}^{3} \gamma_h lnP_{hit} + \sum_{h=1}^{2} \delta_h lnY_{hit} + 0.5 \sum_{m=1}^{2} \sum_{n=1}^{2} \mu_{mn} lnY_{mit} lnY_{nit} \\ & + \sum_{m=1}^{3} \sum_{n=1}^{3} \pi_{mn} lnP_{mit} lnP_{nit} + \sum_{m=1}^{2} \sum_{n=1}^{3} \xi_{mn} lnY_{mit} lnP_{nit} + \epsilon_i + \tau_t + \upsilon_{it}, \end{aligned} \tag{5.2}$$

where TC represents total cost, Y_1 and Y_2 denote two outputs, net loans and securities, respectively. The former is defined as gross loans minus reserves for loan loss provision. The latter is the sum of securities held to maturity and securities held for sale. P_1, P_2 and P_3 denote three input prices, namely the price of capital, price of labor and price of loanable funds. The model above is estimated using two-way fixed effects regression. The time-varying operational inefficiency component is captured by the sum of the two fixed effects, i.e. $\epsilon_i + \tau_t$.

if this ratio is too high, banks may not have enough liquidity to meet unforeseen funding requirements, and vice versa;

$PRESSURE_{it}$ represents "institutional pressure" and is binary. In specific, 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.

Finally, the error term is composite; η_i captures bank-specific effects, $\mathbf{h}_t^0$ is a $r_y \times 1$ vector of unobserved common shocks with corresponding loadings given by $\boldsymbol{\varphi}_i^0$ and ε_{it} is a purely idiosyncratic error. Note that r_y is unknown.

Some discussion on the interpretation of the parameters that characterise Eq. (5.1) is noteworthy. The autoregressive coefficient, ρ, reflects costs of adjustment that prevent banks from achieving optimal risk levels instantaneously (Shrieves and Dahl (1992)). The coefficient of the spatial lag, ψ, captures endogenous links within a network model of interconnected bank balance sheets.

β_ℓ, for $\ell = 1, \ldots, k$, denote the slope coefficients of the model. β_1 captures the effect of operational inefficiency on problem loans. There are two competing hypotheses that predict opposite scenarios in regards to this effect: the so-called "bad management hypothesis" advocates that lower cost efficiency leads to an increase in the number of problematic loans. In particular, managers' failure to control costs sufficiently, can result in poor monitoring of loans and thereby higher default rates (see e.g. Fiordelisi et al. (2011)). In contrast, the so-called "skimping hypothesis" posits that banks may achieve low costs by under-spending on loan underwriting and monitoring, which brings about a larger volume of problem loans (see e.g. Tan and Floros (2013)). Thus β_1 could be either positive or negative depending on which hypothesis appears to be supported by the data.

β_2 measures the effect of capital adequacy on bank risk. Several theories predict that changes in capital levels and bank risk are positively related to one another. For example, a standard view is that since the value of expected bankruptcy costs is an increasing function of the probability of bankruptcy, banks would tend to increase (decrease) capital levels when they increase (decrease) asset portfolio risk, and conversely.[20]

β_3 measures the effect of size on risk-taking behavior. Under the "too-big-to-fail hypothesis"[21], large banks, knowing that they are systematically very important, may count on public bailout in periods of financial distress. Essentially, this hypothesis reflects the classic moral hazard problem, where one party takes on excessive risk, knowing that it is protected against the risk and that another party will incur the cost.

Capital buffer theory postulates that as banking institutions approach the regulatory minimum capital ratio, they may face incentives to reduce risk in order to avoid the regulatory cost triggered by violations of capital adequacy requirements (see e.g. Furlong and Keely (1989) and Rochet (1992)). Thus, β_4 is expected to be negative. The same argument applies for the coefficient of institutional pressure, β_8.

Finally, the direction of the effects of profitability (ROE), asset quality and liquidity on bank risk behavior, β_5, β_6 and β_7, is ultimately an empirical question. For example, standard theory suggests that higher bank profitability dissuades bank risk-taking because profitable banks stand to lose more shareholder value if downside risks realize (Keeley (1990)). On the

[20] This theory is mainly relevant for banks whose optimum capital ratio is in excess of the regulatory minimum levels. Alternative theories supporting a positive value of β_2 are discussed by Shrieves and Dahl (1992).

[21] This phrase was arguably popularized by George W. Bush's administration during the 2008 financial crisis in order to explain the rationale behind the decision to bail out some large financial companies.

other hand, in the presence of leverage constraints, more profitable banks can borrow more and engage in risky side activities on a larger scale (Martynova et al. (2019)).

The spatial weights matrix has been constructed following the methodology of Fernandez (2011). In particular, let d_{ij} denote the Euclidean distance between a specific financial indicator associated with banks i and j:

$$d_{it} = \sqrt{2\left(1-\rho_{ij}\right)}, \tag{5.3}$$

where ρ_{ij} is Spearman's correlation coefficient. 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 neighbor.

We make use of two financial indicators to construct weights, namely the debt ratio, defined as total liabilities over total assets, and the dividend yield, defined as the dividend over market price per share.

5.2 Estimation

The model in Eq. (5.1) is estimated using the two-step IV estimator put forward in the present paper. $INEFF$ is treated as endogenous with respect to ε_{it} due to reverse causality. Reverse causality 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. To tackle reverse causality we instrument $INEFF$ using the ratio of interest expenses paid on deposits over the value of total deposits. Higher values of this variable indicate lower levels of cost efficiency, all other things being equal.[22]

The remaining covariates are treated as exogenous with respect to ε_{it}. However, these covariates can be potentially correlated with the common factor component, $\boldsymbol{\varphi}_i'\mathbf{h}_t^0$, in which case they are endogenous with respect to the total error term, u_{it}. Therefore, we instrument these covariates using the corresponding defactored regressors. Thus, the matrix of instruments is given by

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

where $\widetilde{\underline{\mathbf{X}}}_i = (\widetilde{\underline{\mathbf{x}}}_{1i}, \underline{\mathbf{x}}_{2i}, \ldots, \underline{\mathbf{x}}_{8i})$ is of order $T \times 8$, with $\underline{\mathbf{x}}_{\ell i} = \mathbf{x}_{\ell i} - \overline{\mathbf{x}}_{\ell}$ and $\mathbf{x}_{\ell i}$ is a vector of order $T \times 1$ that denotes the ℓth covariate corresponding to β_ℓ, for $\ell = 2, \ldots, k$. $\widetilde{\underline{\mathbf{x}}}_{1i}$ denotes the external instrument used to identify the effect of cost inefficiency.

Thus, we make use of 32 moment conditions in total, and we estimate 10 parameters in the baseline model. This means that the number of degrees of freedom equals 22. Such degree of overidentification is important in order to ensure strong identification even if some covariates end up not being statistically significant.

The projection matrix $\mathbf{M}_{\widehat{\mathbf{F}}}$ is computed based on the estimated factors obtained as $\sqrt{T}$ times the eigenvectors corresponding to the $\hat{r}_x$ largest eigenvalues of the $T \times T$ matrices

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

$(NT)^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i'$. $\mathbf{M}_{\widehat{\mathbf{F}}_{-1}}$ and $\mathbf{M}_{\widehat{\mathbf{H}}}=\mathbf{I}_T$ are computed in a similar manner. The number of factors is estimated using the eigenvalue method of Ahn and Horenstein (2013). The variance estimator for the two-step IV procedure is given by Eqs. (4.14)-(4.15).

Following Debarsy et al. (2012), we distinguish between direct, indirect and total effects. In particular, stacking the N observations for each t in Eq. (5.1) yields

$$\begin{aligned} NPL_{(t)} &= \rho NPL_{(t-1)} + \psi \mathbf{W}_N NPL_{(t)} + \sum_{\ell=1}^{k}\beta_\ell \mathbf{x}_{\ell(t)} + \mathbf{u}_{(t)}; \\ \mathbf{u}_{(t)} &= \left(\boldsymbol{\eta} + \boldsymbol{\Phi}^0\mathbf{h}_t^0 + \boldsymbol{\varepsilon}_{(t)}\right), \end{aligned} \tag{5.5}$$

where $NPL_{(t)}$ is $N\times 1$, and similarly for the remaining variables. $\boldsymbol{\Phi}^0 = (\boldsymbol{\varphi}_1^0,\ldots,\boldsymbol{\varphi}_N^0)'$, denotes an $N\times r$ matrix of factor loadings.

Solving the model above yields

$$NPL_{(t)} = (\mathbf{I}_N - \psi\mathbf{W}_N)^{-1}\left(\rho NPL_{(t-1)} + \sum_{\ell=1}^{k}\beta_\ell\mathbf{x}_{\ell(t)}\right) + (\mathbf{I}_N - \psi\mathbf{W}_N)^{-1}\mathbf{u}_{(t)}. \tag{5.6}$$

Thus, the matrix of partial derivatives of the expected value of NPL with respect to the ℓth covariate in period t is given by:

$$\left[\frac{\partial E\left(NPL\right)}{\partial x_{\ell 1}}\ldots\frac{\partial E\left(NPL\right)}{\partial x_{\ell N}}\right]_t = (\mathbf{I}_N - \psi\mathbf{W}_N)^{-1}\left(\beta_\ell\mathbf{I}_N\right). \tag{5.7}$$

Following LeSage and Pace (2009), we define the direct effect as the average of the diagonal elements in the matrix above. The indirect effect is defined as the average of the sum of the column entries other than those on the main diagonal. The total effect is the sum of the two effects. There is no indirect effect when $\psi=0$.

The matrix of partial derivatives of the expected value of NPL_t with respect to the ℓth covariate in the long-run is given by:

$$\left[\frac{\partial E\left(NPL\right)}{\partial x_{\ell 1}}\ldots\frac{\partial E\left(NPL\right)}{\partial x_{\ell N}}\right] = \left[\mathbf{I}_N - (1-\rho)\,\mathbf{I}_N - \psi\mathbf{W}_N\right]^{-1}\left(\beta_\ell\mathbf{I}_N\right). \tag{5.8}$$

As before, the long-run direct effect is defined as the average of the diagonal elements in the matrix above, and the long-run indirect effect is computed as the average of the sum of the column entries other than those on the main diagonal. Note that there is no indirect effect in the long-run when either $\psi=0$ and/or $\beta_\ell=0$.

5.3 Results

Table 5.1 below reports results on the model given in Eq. (5.1) for the entire period of the sample, i.e. 2006:Q1-2019:Q4. Column (1) presents results on the preferred specification, as documented in Section 5.2.[23]

As we can see, the autoregressive and spatial parameters are statistically significant and

[23]Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$. p-values in square brackets.

similar in magnitude, which provides support for the existence of both state dependence that is due to costs of adjustment, and of substantial endogenous network linkages. The magnitude of $\hat{\rho}$ indicates that it takes about 2.5 periods for banks to adjust to their optimal risk levels, all other things being equal.

The coefficient of operational inefficiency is positive and statistically significant, providing support for the "bad management hypothesis" instead of the "skimping hypothesis". This outcome is contrary to the result by Ding and Sickles (2019) but consistent with Williams (2004).

In line with findings from a large body of the literature, the effect of capital adequacy ratio on bank risk is positive and statistically significant at the 5% level.

On the other hand, bank size appears not to be associated with risk attitude from a statistical point of view. This finding per se is in contrast with the "too-big-to-fail hypothesis". However, as we shall shortly see from the discussion of the results documented in Table 5.3, the bank size effect appears to be large and statistically significant when the model is re-estimated during 2006:Q1-2010:Q4 only. This time interval spans the period before and during the financial crisis, and it corresponds to the period before the introduction of the Basel III in 2011.[24] As such, this outcome provides evidence of moral hazard-type behavior of banking institutions before the financial crisis burst. Importantly, such behavior seems to be alleviated during the period where the Basel III applies.

The results for β_4 show that capital buffer has a significant negative effect on risk attitude. That is, banks with low capital buffer are less prone to take on more risk compared to banks with higher levels of capital buffer. This is consistent with the argument that banks with low capital buffers face strong incentives to avoid the regulatory cost triggered by a violation of the capital requirement.

Asset quality appears to have a strong positive effect on risk attitude, which is in line with the findings of Aggarwal and Jacques (2001), who show that banks with higher levels of loan loss provision also have a larger proportion of risky assets in their portfolios. Similarly, liquidity appears to exert a strong positive effect on risk-taking behavior. This is arguably because banks with more liquid assets need less insurance against a possible breach of the minimum capital requirements, all other things being equal. Thus, banks with higher liquidity may be willing to take on more risk.

Profitability does not appear to exert a statistically significant effect on risk attitude. However, this result changes when profitability is proxied using an alternative measure, namely the return on assets (ROA) as opposed to ROE. This outcome is documented in Column (7) of Table 5.2, which will be discussed shortly.

Finally, conditional on capital buffer levels, the effect of institutional pressure is not statistically significant, although the sign of the coefficient is plausible.

The remaining columns of Table 5.1 report results in terms of different specifications and/or different estimation approaches. In particular, Column (2) treats $INEFF$ as strictly exogenous with respect to the idiosyncratic error. Therefore, $INEFF$ is instrumented by the defactored value of the same variable. It is clear from the p-value of the J-test that this particular identification strategy is not valid, an outcome which confirms that $INEFF$ is subject to reverse causality. Hence, an external instrument is required for consistent param-

[24]Basel III is a stricter regulatory framework that incorporates a set of reforms designed to improve the regulation, supervision and risk management within the banking sector. In a nutshell, largely in response to the credit crisis, the Basel III requires banks to maintain proper leverage ratios and meet certain minimum capital requirements.

eter estimation. In comparison to Column (1), major discrepancies in Column (2) include: (i) the estimated autoregressive coefficient appears to be biased upwards, with the difference being statistically significant; (ii) the coefficient of operational inefficiency is not statistically significant, which is counterintuitive. This finding shows the importance of the ability of our approach to potentially allow for general forms of endogeneity.

Column (3) reports results from running an IV regression using the same instruments as in Column (1) but without controlling for a common factor component. That is, essentially in this case $\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} = \mathbf{M}_{\widehat{\mathbf{H}}} = \mathbf{I}_T$. As expected, the model is rejected based on the J-test. Inconsistency of parameter estimation manifests mainly via the estimated autoregressive and spatial lag coefficients, both of which appear to be biased. This bears important implications for long-run direct and indirect estimated effects. Thus, allowing for unobserved common factors appears to be crucial in this application.

Column (4) reports results from a "naive" model based on two-way fixed effects estimation without spatial effects and common factors. In this case, the magnitude of the estimated autoregressive coefficient is more than twice as much as that in Column (1), which implies that bias exceeds 100%. Moreover, the estimated coefficient of operational inefficiency appears to have a negative sign, which is also indicative of large bias due to reverse causality.

Table 5.1: **Baseline results on bank risk-taking model (full sample)**

	(1)	**(2)**	**(3)**	**(4)**
$\widehat{\rho}$ (AR parameter)	0.405***	0.541***	0.654***	0.829***
	(0.060)	(0.067)	(0.045)	(0.004)
$\widehat{\psi}$ (spatial parameter)	0.449***	0.382***	0.301***	–
	(0.104)	(0.089)	(0.044)	
$\widehat{\beta}_1$ (inefficiency)	0.331***	0.486	0.352***	-0.095***
	(0.086)	(0.331)	(0.106)	(0.027)
$\widehat{\beta}_2$ (CAR)	0.011**	0.011**	0.010**	-0.122***
	(0.005)	(0.006)	(0.005)	(0.021)
$\widehat{\beta}_3$ (size)	0.031	0.021	0.027	-0.011
	(0.072)	(0.048)	(0.062)	(0.019)
$\widehat{\beta}_4$ (buffer)	-0.033**	-0.016**	-0.021**	0.124***
	(0.015)	(0.012)	(0.011)	(0.021)
$\widehat{\beta}_5$ (profitability)	-0.002	-0.004**	-0.006	0.010
	(0.002)	(0.002)	(0.002)	(0.011)
$\widehat{\beta}_6$ (quality)	0.224***	0.226***	0.237***	0.213***
	(0.035)	(0.045)	(0.032)	(0.015)
$\widehat{\beta}_7$ (liquidity)	1.438***	0.838***	0.671***	.481***
	(0.213)	(0.188)	(0.157)	(0.061)
$\widehat{\beta}_8$ (inst. pressure)	-0.022	0.044	0.033	0.071***
	(0.041)	(0.041)	(0.041)	(0.024)
$\widehat{r}_y$	1	1	0	–
$\widehat{r}_x$	2	2	0	–
J-test	28.649	44.401	58.677	–
	[0.156]	[0.003]	[0.000]	

Notes: Column (1) presents results on the preferred specification, as documented in Section 5.1. In column (2) $INEFF$ is treated as strictly exogenous with respect to the idiosyncratic error and therefore it is instrumented by the defactored value of $INEFF$. Column (3) reports results obtained by running IV estimation using the same instruments as in (1) but without allowing for a common factor component. That is, in this case $\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{M}_{\widehat{\mathbf{F}}_{-1}} = \mathbf{M}_{\widehat{\mathbf{H}}} = \mathbf{I}_T$. Finally, Column (4) reports results from standard fixed effects estimation without spatial effects and common factors. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. p-values in square brackets.

Table 5.2 below reports additional results on the bank risk-taking model. In particular, Column (5) reports results for the same model as that in Column (1) except that the spatial weighting matrix is computed based on dividend yield as opposed to debt ratio. As we can see, the results are very similar across all coefficients, which indicates that the choice of the spatial weighting matrix is not crucial in this application. Clearly, this is a desirable outcome.

Column (6) adds a spatial-time lag to the specification of the model. The corresponding estimated coefficient, $\hat{\psi}_1$, is not statistically significant at the 10% level. Moreover, the sum of $\hat{\psi}$ and $\hat{\psi}_1$ approximately equals 0.351, which is not statistically different from the spatial coefficient reported in Column (1). This indicates that, conditional on a spatial lag, there is no evidence for a spatial-time lag effect (diffusion).

Finally, Column (7) proxies profitability using the return on assets (ROA), which is defined as income after taxes and extraordinary items (annualized), expressed as a percentage of average total assets. The results between Columns (1) and (7) are almost identical, except this time the effect of profitability appears to be statistically significant at the 5% level.

Table 5.2: Additional results on bank risk-taking model (full sample)

	(1)	**(5)**	**(6)**	**(7)**
$\widehat{\rho}$ (AR parameter)	0.405*** (0.060)	0.403*** (0.059)	0.413*** (0.062)	0.399*** (0.057)
$\widehat{\psi}$ (spatial parameter)	0.449*** (0.104)	0.446*** (0.104)	0.667*** (0.247)	0.450*** (0.102)
$\widehat{\beta}_1$ (inefficiency)	0.331*** (0.086)	0.330*** (0.086)	0.217 (0.138)	0.322*** (0.083)
$\widehat{\beta}_2$ (CAR)	0.011** (0.005)	0.011** (0.004)	0.010** (0.004)	0.011** (0.004)
$\widehat{\beta}_3$ (size)	0.031 (0.072)	0.033 (0.073)	0.056 (0.074)	0.038 (0.072)
$\widehat{\beta}_4$ (buffer)	-0.033** (0.015)	-0.033** (0.015)	-0.031** (0.014)	-0.031** (0.015)
$\widehat{\beta}_5$ (profitability)	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.056** (0.025)
$\widehat{\beta}_6$ (quality)	0.224*** (0.035)	0.224*** (0.035)	0.221*** (0.035)	0.194*** (0.038)
$\widehat{\beta}_7$ (liquidity)	1.438*** (0.213)	1.440*** (0.214)	1.430*** (0.211)	1.455*** (0.214)
$\widehat{\beta}_8$ (inst. pressure)	-0.022 (0.041)	-0.022 (0.041)	-0.021 (0.040)	-0.025 (0.041)
$\widehat{\psi}_1$ (spatial-time lag)	–	–	-0.316 (0.304)	
$\widehat{r}_y$	1	1	1	1
$\widehat{r}_x$	2	2	2	2
J-test	28.649 [0.156]	28.051 [0.174]	30.290 [0.035]	28.937 [0.147]

Notes: Column (1) reports results on the preferred specification, as documented in Section 5.1. Column (5) corresponds to the same model as that in Column (1) except that the spatial *weighting matrix is computed based on dividend yield as opposed to debt ratio. In Column (6) we added a spatial-time lag into the baseline model. Finally, Column (7) proxies the capital adequacy ratio with another variable, defined as the total risk based capital expressed as a percent of risk-weighted assets. Standard errors in parentheses.* *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. *p-values in square brackets.*

Table 5.3 reports 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 financial crisis period (2007-2009). The second subsample corresponds to the Basel III.[25]

Some major differences between the two subperiods are worth noting. First of all, the bank size effect is much larger in magnitude during the period under Basel I-II, and remains statistically significant at the 1% level. This implies that the "too-big-to-fail hypothesis", or moral hazard-type behavior in general, was indeed prevalent before the financial crisis hit, and up to 2010. However, the introduction of Basel III appears to largely alleviate this problem. In particular, the effect of bank size becomes small and is no longer statistically significant. This result is consistent with the findings of Zhu et al. (2020), who show that bank behavior during Basel III provides support to Gibrat's "Law", which postulates that the size of a bank and its growth rate are independent.

Secondly, the effect of operational inefficiency appears to be much larger during the period under the Basel I-II than that under the Basel III. Similarly, quality and liquidity of portfolios exert a much stronger effect on risk-taking behavior during the period under Basel I-II than Basel III. That is, banks with more liquid and higher quality assets are willing to take on more risk during Basel I-II but not so during Basel III. Finally, it appears that more profitable banks are less willing to take on more risk during Basel III, 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 Basel III framework has succeeded in influencing banks' behaviour in a substantial manner. This conclusion is contrary to the findings of Ding and Sickles (2019), who infer that the effectiveness of Basel III may be limited.

[25] See footnote 24.

Table 5.3: **Results for different subperiods**

	Full	**Basel I-II**	**Basel III**
$\widehat{\rho}$ (AR parameter)	0.405***	0.388***	0.413***
	(0.060)	(0.070)	(0.128)
$\widehat{\psi}$ (spatial parameter)	0.449***	0.255**	0.535**
	(0.104)	(0.109)	(0.270)
$\widehat{\beta}_1$ (inefficiency)	0.331***	0.584**	0.196*
	(0.086)	(0.296)	(0.104)
$\widehat{\beta}_2$ (CAR)	0.011**	0.030***	0.008*
	(0.005)	(0.011)	(0.004)
$\widehat{\beta}_3$ (size)	0.031	0.871***	0.020
	(0.072)	(0.328)	(0.178)
$\widehat{\beta}_4$ (buffer)	-0.033**	-0.028	-0.015
	(0.015)	(0.025)	(0.015)
$\widehat{\beta}_5$ (profitability)	-0.002	-0.002	-0.010**
	(0.002)	(0.004)	(0.004)
$\widehat{\beta}_6$ (quality)	0.224***	0.239***	0.001
	(0.035)	(0.042)	(0.077)
$\widehat{\beta}_7$ (liquidity)	1.438***	2.714***	0.937***
	(0.213)	(0.531)	(0.358)
$\widehat{\beta}_8$ (inst. pressure)	-0.022	0.014	-0.021
	(0.041)	(0.066)	(0.057)
$\widehat{r}_y$	1	1	1
$\widehat{r}_x$	2	1	1
J-test	28.649	30.205	28.937
	[0.156]	[0.114]	[0.147]

Notes: Column "Full" reports results obtained from the full sample, and therefore these are identical to the results of Column (1) in Table 5.1. Column "Basel I-II" reports results for the first subsample that spans 2006:Q1-2010:Q4. This is the period under Basel I-II. Column "Basel III" reports results for the second subsample that spans 2011:Q1-2019:Q4. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. *p-values in square brackets.*

Table 5.4 reports direct, indirect and total effects, which have been computed as described in Eq. (5.7).[26] Total effects are simply the sum of direct and indirect effects. Panel A corresponds to the full sample, i.e. the period spanning 2006:Q1-2019:Q4. In this panel, the direct effects are identical to the estimated coefficients reported in Column (1) of Table 5.1. Direct and indirect effects appear to be of similar magnitude. In particular, roughly speaking, around 55% of the total effects can be attributed to the direct ones, and 45% is due to the indirect effects.

[26]The long-run results are qualitatively identical and so we do not provide them here to save space. They are available upon request.

The results change substantially when the sample is split into two subperiods. In particular, for the first subsample (Panel B), the direct effects appear to be larger, contributing roughly three quarters of the total effect. In contrast, for the second subsample (Panel C), direct effects contribute about 48% of the total effect, which is of similar magnitude with the finding obtained from the full sample.

Table 5.4: Decomposition of effects

Panel A: Full sample

Full sample	**Direct**	**Indirect**	**Total**
inefficiency	0.331	0.265	0.596
CAR	0.011	.008	0.019
size	0.031	0.025	0.056
buffer	-0.033	-0.026	-0.059
profitability	-0.002	-0.002	-0.004
quality	0.224	0.179	0.404
liquidity	1.438	1.157	2.591
inst. pressure	-0.022	-0.018	-0.040

Panel B: Basel I-II

	Direct	**Indirect**	**Total**
inefficiency	0.584	0.197	0.782
CAR	0.030	0.010	0.040
size	0.871	0.197	0.782
buffer	-0.028	-0.009	-0.037
profitability	-0.001	-0.001	-0.002
quality	0.239	0.081	0.320
liquidity	2.714	0.917	3.631
inst. pressure	0.014	0.005	0.019

Panel C: Basel III

	Direct	**Indirect**	**Total**
inefficiency	0.196	0.220	0.417
CAR	0.008	0.008	0.015
size	0.020	0.022	0.042
buffer	-0.015	-0.017	-0.032
profitability	-0.010	-0.011	-0.021
quality	0.001	0.001	0.002
liquidity	0.937	1.052	1.991
inst. pressure	-0.021	-0.023	-0.044

Notes: See the discussion in the main text on the computation of direct and indirect effects. Basel I-II spans the period 2006:Q1-2010:Q4 ($T = 21$). Basel III spans the period 2011:Q1-2019:Q4 ($T = 35$).

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 straightforward to implement. Moreover, it is 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.

Simulation results show that the proposed IV estimator performs well in finite samples, that is, it has negligible bias and produces credible inferences in all cases considered. We have applied our methodology to study the determinants of risk attitude of banking institutions, with emphasis on the impact of increased capital regulation over the past decade or so. The results of our analysis bear important policy implications and provide evidence that the more risk-sensitive capital regulation that was introduced by the Basel III framework in 2011 has succeeded in influencing banks' behaviour in a substantial manner.

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 & Finance*, 25, 1139–1160.

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

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, 49–70.

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

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, 277–297.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

——— (2009): "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

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

——— (2018): "Dynamic spatial panel data models with common shocks," Mimeo.

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 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, 585–612.

Bun, M. J. G. and J. F. Kiviet (2006): "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models," *Journal of Econometrics*, 132, 409–444.

Bun, M. J. G. and V. Sarafidis (2015): "Dynamic Panel Data Models," in *The Oxford Handbook of Panel Data*, ed. by B. H. Baltagi, Oxford: Oxford University Press, chap. 3.

Case, A. (1991): "Spatial Patterns in Household Demand," *Econometrica*, 59, 953–965.

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, 393–420.

Debarsy, N., C. Ertur, and J. P. LeSage (2012): "Interpreting dynamic space time panel data models," *Statistical Methodology*, 9, 158–171.

Ding, D. and R. Sickles (2019): "Capital regulation, efficiency, and risk taking: A spatial panel analysis of US banks," in *Panel Data Econometrics*, ed. by M. Tsionas, London: Elsevier, chap. 13, 405–466.

Everaert, G. and T. De Groote (2016): "Common Correlated Effects Estimation of Dynamic Panels with Cross-Sectional Dependence," *Econometric Reviews*, 35, 428–463.

Favero, C., M. Marcellino, and F. Neglia (2005): "Principal components at work: The empirical analysis of monetary policy with large data sets," *Journal of Applied Econometrics*, 20, 603–620.

Fernandez, V. (2011): "Spatial linkages in international financial markets," *Quantitative Finance*, 11, 237–245.

Fiordelisi, F., D. Marques-Ibanez, and P. Molyneux (2011): "Efficiency and risk in european banking," *Journal of Banking & Finance*, 35, 1315–1326.

Furlong, F. and M. C. Keely (1989): "Capital regulation and bank risk-taking: a note," *Journal of Banking & Finance*, 13, 883–891.

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, 1639–1657.

Hansen, C. and Y. Liao (2018): "The Factor-Lasso and K-Step Bootstrap Approach for Inference in High-Dimensional Economic Applications," *Econometric Theory*, 1–45.

Heckman, J., J. Stixrud, and S. Urzua (2006): "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labour Economics*, 24, 411–4482.

Jacques, K. and P. Nigro (1997): "Risk-based capital, portfolio risk, and bank capital: A simultaneous equations approach," *Journal of Economics & Business*, 49, 533–547.

Juodis, A., H. Karabiyik, and J. Westerlund (2020): "On the Robustness of the Pooled CCE Estimator," *Journal of Econometrics*, (forthcoming).

Juodis, A. and V. Sarafidis (2018): "Fixed T Dynamic Panel Data Estimators with Multifactor Errors," *Econometric Reviews*, 37, 893–929.

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

Keeley, M. C. (1990): "Deposit insurance, risk, and market power in banking," *American Economic Review*, 80, 1183–1200.

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

Kelejian, H. H. and I. R. Prucha (2001): "On the asymptotic distribution of the Moran I test statistic with applications," *Journal of Econometrics*, 140, 219–257.

——— (2007): "HAC estimation in a spatial framework," *Journal of Econometrics*, 140, 131–154.

——— (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, 145–158.

Kuersteiner, G. and I. R. Prucha (2018): "Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity," ArXiv:1802.01755.

Kwan, S. and R. Eisenbeis (1996): "An analysis of inefficiencies in banking: a stochastic cost frontier approach," *Economic Review - Federal Reserve Bank of San Francisco*, 2, 16–26.

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

——— (2015): "Spatial Panel Data Models," in *The Oxford Handbook of Panel Data*, ed. by B. baltagi, Oxford: Oxford University Press, chap. 12, 363–401.

LeSage, J. P. and R. Pace (2009): *Introduction to spatial econometrics*, Chapman and Hall, 1 ed.

Manski, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60, 531 – 542.

Martynova, N., L. Ratnovski, and R. Vlahu (2019): "Bank profitability, leverage constraints, and risk-taking," *Journal of Financial Intermediation*, 1–19.

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

Norkute, M., V. Sarafidis, T. Yamagata, and G. Cui (2020): "Instrumental Variable Estimation of Dynamic Linear Panel Data Models with Defactored Regressors and a Multifactor Error Structure," *Journal of Econometrics*, forthcoming.

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, 967–1012.

Pesaran, M. H., L. V. Smith, and T. Yamagata (2013): "Panel Unit Root Tests in the Presence of a Multifactor Error Structure," *Journal of Econometrics*, 175, 94 – 115.

Robertson, D. and V. Sarafidis (2015): "IV Estimation of Panels with Factor Residuals," *Journal of Econometrics*, 185, 526–541.

Rochet, J. C. (1992): "Capital requirements and the behaviour of commercial banks," *European Economic Review*, 36, 1137–1170.

Roy, P. V. (2005): "The impact of the 1988 basel accord on banks' capital ratios and credit risk-taking: An international study," Mimeo.

Sarafidis, V. and T. J. Wansbeek (2012): "Cross-sectional Dependence in Panel Data Analysis," *Econometric Reviews*, 31, 483–531.

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

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

Shrieves, R. E. and D. Dahl (1992): "The relationship between risk and capital in commercial banks," *Journal of Banking & Finance*, 16, 493–457.

Stolz, S., F. Heid, and D. Porath (2003): "Does capital regulation matter for bank behavior? Evidence for german savings banks," Mimeo.

Tan, Y. and C. Floros (2013): "Risk, capital and efficiency in chinese banking," *Journal of International Financial Markets, Institutions and Money*, 26, 378–393.

Vallascas, F. and J. Hangendorff (2013): "The risk sensitivity of capital requirements: Evidence fro an international sample of large banks," *Review of Finance*, 17, 1947–1988.

Williams, J. (2004): "Determining management behaviour in european banking," *Journal* *of Banking & Finance*, 28, 2427–2460.

Wooldridge, J. (2002): *Econometric Analysis of Cross Section and Panel Data*, MITPress, 1 ed.

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, 118 – 134.

Zhu, H., V. Sarafidis, and M. J. Silvapulle (2020): "A new structural break test for panels with common factors," *Econometrics Journal*, 23, 137–155.