CentER
Discussion Paper

No. 2011-134

GMM Estimation of Fixed Effects Dynamic Panel Data Models with Spatial Lag and Spatial Errors

By

Pavel Cizek, Jan Jacobs, Jenny Ligthart, Hendrik Vrijburg

December 2011

ISSN 0924-7815

Tilburg University

GMM Estimation of Fixed Effects Dynamic Panel Data Models with Spatial Lag and Spatial Errors*

Pavel Čížek† Jan Jacobs‡ Jenny Ligthart§ Hendrik Vrijburg¶

November 2011

Abstract

We extend the three-step generalized methods of moments (GMM) approach of Kapoor et al. (2007), which corrects for spatially correlated errors in static panel data models, by introducing a spatial lag and a one-period lag of the dependent variable as additional explanatory variables. Combining the extended Kapoor et al. (2007) approach with the dynamic panel data model GMM estimators of Arellano and Bond (1991) and Blundell and Bond (1998) and specifying moment conditions for various time lags, spatial lags, and sets of exogenous variables yields new spatial dynamic panel data estimators. We prove their consistency and asymptotic normality for a large number of spatial units N and a fixed small number of time periods T. Monte Carlo simulations demonstrate that the root mean squared error of spatially corrected GMM estimates—which are based on a spatial lag and spatial error correction—is generally smaller than that of corresponding spatial GMM estimates in which spatial error correction is ignored. We show that the spatial Blundell-Bond estimators outperform the spatial Arellano-Bond estimators.

JEL codes: C15, C21, C22, C23

Keywords: Dynamic panel models, spatial lag, spatial error, GMM estimation

*This paper is a substantially revised version of CentER Discussion Paper No. 2009-92. Part of this paper was written when the third author was employed at Georgia State University. The authors would like to thank Rob Alessie, Badi Baltagi, Peter Egger, J. Paul Elhorst, Vincenzo Verardi, Tom Wansbeek, three anonymous referees, seminar participants at Maastricht University, and conference participants at the 3rd World Conference of the Spatial Econometrics Association (Barcelona, July 2009) and the Research Day of the Netherlands Network of Economics (Utrecht, October 2007) for helpful discussions.

†CentER and Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31-13-466-8723, Fax: +31-13-466-3280, E-mail: p.cizek@tilburguniversity.edu

‡Department of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands, Phone: +31-50-363-3681, Fax: +31-50-363-7337, E-mail: j.p.a.m.jacobs@rug.nl

§Corresponding author: CentER and Department of Economics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31-13-466-8755, Fax: +31-13-466-4032, E-mail:j.ligthart@tilburguniversity.edu

¶Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Phone: +31-10-408-1481, Fax: +31-10-408-9166, E-mail: vrijburg@ese.eur.nl

1 Introduction

The separate literatures on dynamic panel data models and spatial econometric models have matured rapidly and have reached (graduate) textbooks during the last decade.[1] Panel data may feature state dependence—i.e., the dependent variable is correlated over time—as well as display spatial dependence, that is, the dependent variable is correlated in space. Applied economists' interest in frameworks that integrate spatial considerations into dynamic panel data models is a fairly recent development, however.[2] For this model class, Elhorst (2005, 2008, 2010), Su and Yang (2008), Yu et al. (2008), Lee and Yu (2010b), and Yu and Lee (2010) have analyzed the properties of maximum likelihood (ML) estimators and combinations of ML and corrected least squares dummy variable estimators. During the last decade, the flexible generalized method of moments (GMM) framework for dynamic panels has gained popularity,[3] but it has not received much attention in the spatial econometrics literature yet. The papers by Lee and Liu (2010), Lin and Lee (2010), and Liu et al. (2010) study spatial GMM estimators for static panels.[4] Our paper integrates the two strands of literature by investigating theoretically and numerically the performance of various spatial GMM estimators for dynamic panel data models with fixed effects.

Many economic interactions among agents are characterized by a spatially lagged dependent variable, which consists of observations on the dependent variable in other locations than the 'home' location. In the public finance literature, for example, local governments take into account the behavior of neighboring governments in setting their tax rates (cf. Wilson, 1999, and Brueckner, 2003) and deciding on the provision of public goods (cf. Case et al., 1993). In the

[1]See Arellano (2003) and Baltagi (2008, Chapter 8) for an analysis of dynamic panel data models and Anselin (1988, 2006) for a treatment of spatial econometrics.

[2]Badinger et al. (2004), Foucault et al. (2008), Jacobs et al. (2010), Brady (2011), and Bartolini and Santolini (2012) provide empirical applications of spatial dynamic panel data models. See Lee and Yu (2010a) and Elhorst (2011) for an overview of dynamic spatial panel models.

[3]The GMM framework can handle multiple endogenous explanatory variables, fixed effects, and unbalanced panels.

[4]Using a Monte Carlo simulation study, Kukenova and Monteiro (2009), and Elhorst (2010) are the only ones exploring GMM in a spatial dynamic panel data framework. Kukenova and Monteiro (2009) analyze a spatial system GMM estimator and include an endogenous covariate in addition to a spatial lag and the time lag of the dependent variable. Elhorst (2010) briefly touches upon difference GMM estimators with a spatial lag in order to compare them to spatial ML estimators. However, both studies do not correct their spatial GMM estimators for potential spatial error correlation.

trade literature, foreign direct investment (FDI) inflows into the host country depend on FDI inflows into proximate host countries (cf. Blonigen et al., 2007). The spatial lag structure allows one to explicitly measure the strength of the spatial interaction. Spatial error dependence is an alternative way of capturing spatial aspects and may arise due to an omitted explanatory variable.[5] Spatially correlated errors can be thought of as analogous to the well-known practice of clustering error terms by groups, which are defined based on some directly observable characteristic of the group. In spatial econometrics, the groups are based on spatial 'similarity,' which is typically captured by some geographic characteristic (e.g., proximity). Spatial panel data applications typically employ either a spatial lag model or a spatial error model. Ignoring spatial error correlation in static panel data models may give rise to a loss of efficiency of the estimates and may thus erroneously suggest that strategic interaction is absent. In contrast, disregarding spatial dependency in the dependent variable comes at a relatively high cost because it gives rise to biased estimates (cf. LeSage and Pace, 2009, p. 158). Rather than using either a spatial lag model or spatial error model, we allow both processes to be simultaneously present. Indeed, in their empirical tax competition model, Egger et al. (2005) find evidence that spatial error dependence may exist above and beyond the theoretically motivated spatial lag structure.[6]

Non-spatial dynamic panel data models are usually estimated using the GMM estimator of Arellano and Bond (1991), which differs from static panel GMM estimators in the set of moment conditions and the matrix of instruments.[7] The standard Arellano-Bond estimator is known to be rather inefficient when instruments are weak (e.g., if time dependency is strong) because it makes use of information contained in first differences of variables only. Alternatively, authors have used Blundell and Bond's (1998) system approach, which consists of both first-differenced

[5]Spatial error correlation may also result from measurement error in variables, a misspecified functional form of the regression equation, the absence of a spatial lag or a misspecified weighting matrix.

[6]Case et al. (1993), Jacobs et al. (2010), Baltagi and Bresson (2011), and Brady (2011) also consider spatial models with both spatial lag and spatial error components. Only the study by Jacobs et al. (2010) uses a spatial dynamic panel data model.

[7]In dynamic panels with unobserved heterogeneity, Nickell (1981) shows that the standard least squares dummy variable estimator is biased and inconsistent for large N and fixed small T. Anderson and Hsiao (1982) suggest simple instrument variable estimators for a first differenced model, which uses the second lag of the dependent variable—either in differences or levels—to instrument the lagged dependent variable.

and level equations and an extended set of internal instruments. In the following, we contribute to the literature by developing spatial variants of the Arellano-Bond and Blundell-Bond estimators. Our new approach involves defining appropriate instruments to control for the endogeneity of the spatial lag and time lag of the dependent variable while correcting for spatial error correlation. For this purpose, we use new spatial instruments—which are based on a combination of several spatial lags and a modification of the approach of Kelejian and Robinson (1993)—combined with standard instruments for dynamic panel data models.

To account for spatial error correlation, we analyze the properties of our estimators first without and later with a correction for spatial error correlation. Throughout the paper, we use the term 'spatial' GMM estimators to refer to GMM estimators for panel data models including a spatial lag with or without correction for spatial error correlation.[8] If a spatial GMM estimator corrects for spatial error correlation, we speak of 'spatially corrected' GMM estimators. Recently, Kapoor et al. (2007) designed a GMM procedure to deal with spatial error correlation in static panels. We extend their three-step spatial procedure to panels with a spatially lagged dependent variable, a one-period time lag of the dependent variable, and unit-specific fixed effects. In addition, we modify their second-stage moment conditions by considering the first differences of errors. We analytically investigate the asymptotic properties of the estimators for large N and fixed small T and briefly discuss the case of large T and small N.[9] Specifically, we show that our spatial GMM estimators are consistent and asymptotically normal in the first case and explain that the number of instruments has to be bounded to obtain consistency in the latter case.

The finite-sample performance of the spatial GMM estimators is investigated by means of Monte Carlo simulations. The simulation experiments indicate that the root mean squared error (RMSE) of spatially corrected GMM estimates—which are based on a spatial lag and spatial error correction—is generally smaller than that of corresponding spatial GMM estimates in which spatial error correlation is ignored, particularly for strong positive error correlation. The RMSE of

[8] Anselin et al. (2008) call this model class a 'time-space simultaneous model.'

[9] Yu et al. (2008) and Pesaran and Tosetti (2011) study the properties of ML estimators in the context of dynamic, possibly nonstationary, panels with fixed effects and spatial error correlation, assuming both N and T large.

the spatial GMM estimates, however, is not much affected by the size of the spatial lag parameter. We also show that the spatial Blundell-Bond estimators outperform the spatial Arellano-Bond estimators. Finally, we find that spatial estimators using spatially weighted endogenous variables as instruments in addition to weighted exogenous variables are more efficient than those based on weighted exogenous variables.

The paper is organized as follows. Section 2 sets out our spatial dynamic panel data model. Section 3 develops the two estimators for spatial dynamic panel data models, that is, the spatially corrected Arellano-Bond and Blundell-Bond estimators. Section 4 proves the consistency and asymptotic normality of the spatial estimators. Section 5 presents Monte Carlo simulation outcomes. Finally, Section 6 concludes. The proofs are in the Appendix.

2 The Spatial Dynamic Panel Data Model

Consider a panel with $i = 1, ..., N$ spatial units and $t = 1, ..., T$ time periods. The focus is on panels with a small number of time periods relative to the number of spatial units. Assume that the data at time t are generated according to the following model:

$$\mathbf{y}_N(t) = \lambda \mathbf{y}_N(t-1) + \delta \mathbf{W}_N \mathbf{y}_N(t) + \mathbf{X}_N(t)\boldsymbol{\beta} + \mathbf{u}_N(t), \qquad t = 2, ..., T, \tag{1}$$

where $\mathbf{y}_N(t)$ is an $N \times 1$ vector of observations on the dependent variable, $\mathbf{y}_N(t-1)$ is a one-period time lag of the dependent variable, $\mathbf{W}_N$ is an $N \times N$ matrix of spatial weights, $\mathbf{X}_N(t)$ is an $N \times K$ matrix of observations on the strictly exogenous explanatory variables (where K denotes the number of covariates), and $\mathbf{u}_N(t)$ is an $N \times 1$ vector of error terms.[10] If we later need to refer to observations from all applicable time periods in a given context, we simply omit the time specification in brackets; here, for example, $\mathbf{y}_N = [\mathbf{y}_N^\top(1), \ldots, \mathbf{y}_N^\top(T)]^\top$ or $\mathbf{X}_N = [\mathbf{X}_N^\top(1), \ldots, \mathbf{X}_N^\top(T)]^\top$, where $\top$ denotes a transpose. Further, the scalar parameter λ is the coefficient of the lagged dependent variable, δ is the spatial autoregressive coefficient, which

[10] Our specification does not include $\mathbf{W}_N \mathbf{y}_N(t-1)$, which yields a so-called spatiotemporal model. See Yu et al. (2008) for such an approach. We leave this extension for future research.

measures the endogenous interaction effect among units, and $\boldsymbol{\beta}$ is a $K \times 1$ vector of (fixed) slope coefficients.

The spatial lag is denoted by $\mathbf{W}_N \mathbf{y}_N(t)$, which captures the contemporaneous correlation between unit i's behavior and a weighted sum of the behavior of units $j \neq i$. The elements of $\mathbf{W}_N$ (denoted by w_{ij}) are exogenously given, non-negative, and zero on the diagonal of the matrix. Note that there is little formal guidance on choosing the 'correct' spatial weights because many definitions of neighbors are possible. The literature usually employs contiguity (i.e., units having common borders) or physical distance between units as weighting factors. We assume the elements of $\mathbf{W}_N$ to be row normalized so that each row sums to one. This is not the only possible normalization, see, for example, Kelejian and Prucha (2010). Row normalization is standard in spatial applications and therefore we use it in the simulations of Section 5.

The reduced form of equation (1) amounts to:

$$\mathbf{y}_N(t) = (\mathbf{I}_N - \delta \mathbf{W}_N)^{-1} \left[\lambda \mathbf{y}_N(t-1) + \mathbf{X}_N(t)\boldsymbol{\beta} + \mathbf{u}_N(t) \right], \tag{2}$$

where $\mathbf{I}_N$ is an identity matrix of dimension $N \times N$. Stationarity of the model does not only require that $|\lambda| < 1$, but also that the characteristic roots of the matrix $\lambda(\mathbf{I}_N - \delta \mathbf{W}_N)^{-1}$ should lie in the unit circle, which is the case if (cf. Elhorst, 2008)

$$|\lambda| + \delta\omega_L < 1 \quad \text{if} \quad \delta < 0 \qquad \text{and} \qquad |\lambda| + \delta\omega_U < 1 \quad \text{if} \quad \delta \geq 0, \tag{3}$$

where ω_L and ω_U denote the smallest (i.e., the most negative) and largest characteristic roots of $\mathbf{W}_N$, respectively. If $\mathbf{W}_N$ is row normalized, we find $\omega_U = 1$.[11] Equation (3) yields a tradeoff between the size of λ and δ.

Spatial error correlation may arise, for example, when omitted variables follow a spatial pattern, yielding a non-diagonal variance-covariance matrix of the error term $\mathbf{u}_N(t)$. In the case of spatial error correlation, the error structure in (1) is a spatially weighted average of the error components of neighbors, where the weights are given by a row-normalized $N \times N$ matrix $\mathbf{M}_N$

[11] No general results hold for the smallest characteristic root of the matrix of spatial weights. The lower bound ω_L is typically less than -1; see Elhorst (2008, p. 422).

of spatial weights (with typical element m_{ij}). More formally, the spatially autoregressive process is given by:

$$\mathbf{u}_N(t) = \rho \mathbf{M}_N \mathbf{u}_N(t) + \boldsymbol{\varepsilon}_N(t), \tag{4}$$

where $\mathbf{M}_N \mathbf{u}_N(t)$ is the spatial error term, ρ is a (second) spatially autoregressive coefficient, and $\boldsymbol{\varepsilon}_N(t)$ denotes a vector of innovations. The interpretation of the 'nuisance' parameter ρ is very different from δ in the spatial lag model in that there is no particular relation to a substantive theoretical underpinning of the spatial interaction. We follow the common practice in the literature by assuming $\mathbf{W}_N \neq \mathbf{M}_N$, which allows us to identify both spatial parameters δ and ρ in the absence of exogenous variables and a dynamic lag.[12] The spatial error process in reduced form is $\mathbf{u}_N(t) = (\mathbf{I}_N - \rho \mathbf{M}_N)^{-1} \boldsymbol{\varepsilon}_N(t)$. If $|\rho| < 1$, the spatial error process is stable thus yielding feedback effects that are bounded.

The vector of innovations is defined as:

$$\boldsymbol{\varepsilon}_N(t) = \boldsymbol{\eta}_N + \mathbf{v}_N(t), \qquad \mathbf{v}_N(t) \sim \text{iid}(0, \sigma_v^2 \mathbf{I}_N), \tag{5}$$

where $\boldsymbol{\eta}_N$ is an $N \times 1$ vector representing unobservable unit-specific fixed effects and $\mathbf{v}_N(t)$ is an $N \times 1$ vector of independently and identically distributed (iid) error terms with variance σ_v^2, which is assumed to be constant across units and time periods. In the following, we consider a specification in which $\boldsymbol{\eta}_N$ is possibly correlated with the regressors.

Equations (1), (4), and (5) can be written concisely as:

$$\mathbf{y}_N(t) = \mathbf{Z}_N(t)\boldsymbol{\theta} + \mathbf{u}_N(t), \tag{6}$$

$$\mathbf{u}_N(t) = (\mathbf{I}_N - \rho \mathbf{M}_N)^{-1} [\boldsymbol{\eta}_N + \mathbf{v}_N(t)], \tag{7}$$

where $\mathbf{Z}_N(t) = [\mathbf{y}_N(t-1), \mathbf{W}_N \mathbf{y}_N(t), \mathbf{X}_N(t)]$ denotes the matrix of regressors, $\boldsymbol{\theta} = [\lambda, \delta, \boldsymbol{\beta}^\top]^\top$ is a vector of $K+2$ parameters. Our general dynamic spatial panel data model embeds various special cases discussed in the literature. If $\lambda = \rho = 0$ and $\delta > 0$, our model reduces to the familiar spatial lag model (also known as the mixed regressive-spatial autoregressive model; see Anselin,

[12] In the simulations of Section 5, we also consider $\mathbf{W}_N = \mathbf{M}_N$.

1988), whereas for $\lambda = \rho = 0$ and $\boldsymbol{\beta} = \mathbf{0}$ we get a pure spatial autoregressive model. If $\lambda = \delta = 0$ and $\rho > 0$, we obtain the spatial error model. If $\lambda > 0$ and $\delta = \rho = 0$, we arrive at Arellano and Bond's dynamic panel data model. Finally, the general spatial dynamic panel data model boils down to a standard static panel data model if $\lambda = \delta = \rho = 0$.

3 Spatial Dynamic Panel Estimators

In this section, the spatial dynamic panel estimators are proposed. We extend the static panel data model of Kapoor et al. (2007)—which explicitly corrects for spatial error correlation—to include both a time lag and a spatial lag of the dependent variable. Because the time lag is endogenous, we apply a panel GMM procedure. We propose sets of instruments for both the time lag and spatial lag of the dependent variable. This procedure yields consistent spatially corrected Arellano-Bond estimators and spatially corrected Blundell-Bond estimators, which will be derived in three stages.

3.1 The First Stage

3.1.1 Arellano-Bond Estimator

To estimate $\boldsymbol{\theta}$, we employ a GMM estimator defined by a set of linear moment conditions for the error term $\mathbf{u}_N(t)$. Later, equations identifying $\boldsymbol{\theta}$ are obtained by substituting for the error term from the model equation, $\mathbf{u}_N(t) = \mathbf{y}_N(t) - \mathbf{Z}_N(t)\boldsymbol{\theta}$.

First, to eliminate the unit-specific fixed effects $\boldsymbol{\eta}_N$ from $\boldsymbol{\varepsilon}_N(t)$, we take first differences of (6) and (7):

$$\Delta \mathbf{y}_N(t) = \Delta \mathbf{Z}_N(t)\boldsymbol{\theta} + \Delta \mathbf{u}_N(t), \tag{8}$$

$$\Delta \mathbf{u}_N(t) = (\mathbf{I}_N - \rho \mathbf{M}_N)^{-1} \Delta \boldsymbol{\varepsilon}_N(t) = (\mathbf{I}_N - \rho \mathbf{M}_N)^{-1} \Delta \mathbf{v}_N(t), \qquad t = 3, ..., T, \tag{9}$$

where $\Delta \mathbf{q}_N(t) \equiv \mathbf{q}_N(t) - \mathbf{q}_N(t-1)$ for $\mathbf{q}_N(t) = \{\mathbf{y}_N(t), \mathbf{Z}_N(t), \mathbf{u}_N(t), \boldsymbol{\varepsilon}_N(t), \mathbf{v}_N(t)\}$. Note that the differenced model is specified only in $T-2$ time periods (and thus $T \geq 3$): one observation is lost due to the first differencing operation and another observation is dropped because of the one-period time lag of the dependent variable.

In the differenced model, both the time lag and the spatial lag of the dependent variable are endogenous. In addition, the two endogenous regressors are correlated with each other. Consistent GMM estimation is possible if there are at least $K+2$ instruments that are correlated with the time lagged, spatially lagged, and exogenous variables and are uncorrelated with the errors $\Delta\mathbf{u}_N(t)$ for each $t = 3, \dots, T$. First, the moment conditions identifying the coefficients of the strictly exogenous variables are

$$\mathrm{E}[\Delta\mathbf{X}_N^\top(t)\Delta\mathbf{u}_N(t)] = 0, \qquad t = 3, ..., T, \tag{10}$$

where E denotes an expectation operator.

Next, Arellano and Bond (1991) propose to use the levels of the dependent variable, $\mathbf{y}_N(t-2), ..., \mathbf{y}_N(1)$, as instruments for the time lag of the dependent variable in first differences (i.e., $\Delta\mathbf{y}_N(t-1)$). The instruments are correlated with the time lag of the dependent variable in first differences $\Delta\mathbf{y}_N(t-1)$, but are uncorrelated with the 'future' error term in first differences, $\Delta\mathbf{u}_N(t)$, since the unit-specific effects are eliminated from the differenced variables. This property holds even in the spatial model defined by (6) and (7) because the spatial correlation applies only within a given time period t and, hence, $\mathbf{y}_N(t-2)$ is correlated with $\mathbf{u}_N(t-2), ..., \mathbf{u}_N(1)$, but cannot be correlated with $\mathbf{u}_N(t)$ and $\mathbf{u}_N(t-1)$. Consequently, we impose the following moment conditions to identify λ:

$$\mathrm{E}[\mathbf{y}_N^\top(t-s)\Delta\mathbf{u}_N(t)] = 0, \qquad t = 3, ..., T, \qquad s = 2, ..., t-1. \tag{11}$$

Equation (11) yields $(T-2)(T-1)/2$ moment conditions for a given N.

For the spatial lag, we consider two alternative set of instruments. The first approach instruments the spatial lag by various time lags of the spatially lagged dependent variable. The validity of such moment conditions follows by the same argument as given in the previous paragraph for equation (11). This approach implies the following moment conditions for δ:

$$\mathrm{E}[\{\mathbf{W}_N^l\mathbf{y}_N(t-s)\}^\top\Delta\mathbf{u}_N(t)] = 0, \quad t = 3, ..., T, \quad s = 2, ..., t-1, \quad l = 1, ..., L, \tag{12}$$

where l indicates various powers of $\mathbf{W}_N$ and the integer L is the maximum spatial lag used for instrumenting. For each power $l \geq 1$, equation (12) yields again $(T-2)(T-1)/2$ moment conditions. The second approach uses instruments based on a modification of Kelejian and Robinson (1993). We expand the expected value of the spatial lag $\mathbf{W}_N \mathbf{y}_N(t)$, which depends on $\mathbf{W}_N \mathbf{X}_N(t)\boldsymbol{\beta}$ [see (1)], and take first differences to propose instruments $\mathbf{W}_N \Delta \mathbf{X}_N(t)$. As the strictly exogenous variables $\Delta \mathbf{X}_N(t)$ are not correlated with the error term $\Delta \mathbf{u}_N(t)$, the instruments satisfy the following moment conditions:

$$\mathrm{E}[\{\mathbf{W}_N \Delta \mathbf{X}_N(t)\}^\top \Delta \mathbf{u}_N(t)] = 0, \qquad t = 3, ..., T. \tag{13}$$

Note that the moment conditions specified for the spatial autoregressive parameter δ for various time lags s, spatial lags l, and sets of exogenous variables will have different precision and power depending on the coefficients in model (1): large λ, δ, and β imply stronger correlation of $\mathbf{W}_N \mathbf{y}_N(t)$ with the instruments given in (12) for $s \geq 1$ and $l \geq 1$.

For each time period, we specified $J \geq K + 2$ moment conditions, which can be concisely written as $\mathrm{E}[\mathbf{H}_{N,AB}^\top(t) \Delta \mathbf{u}_N(t)] = 0$, where the columns of $\mathbf{H}_{N,AB}(t)$ represent the instruments $\Delta \mathbf{X}_N(t), \mathbf{y}_N(t-s), \mathbf{W}_N^l \mathbf{y}_N(t-s)$, and $\mathbf{W}_N \Delta \mathbf{X}_N(t)$ given above. Merging the information from all available time periods, the proposed GMM estimator will minimize

$$\frac{[\mathbf{H}_{N,AB}^\top \Delta \mathbf{u}_N]^\top \mathbf{A}_{N,AB} [\mathbf{H}_{N,AB}^\top \Delta \mathbf{u}_N]}{N} = \frac{[\mathbf{H}_{N,AB}^\top (\Delta \mathbf{y}_N - \Delta \mathbf{Z}_N \boldsymbol{\theta})]^\top \mathbf{A}_{N,AB} [\mathbf{H}_{N,AB}^\top (\Delta \mathbf{y}_N - \Delta \mathbf{Z}_N \boldsymbol{\theta})]}{N}$$

with respect to $\boldsymbol{\theta}$, where $\mathbf{H}_{N,AB}$ is a block-diagonal matrix consisting of blocks $\mathbf{H}_{N,AB}(t), t = 3, \ldots, T$ and $\mathbf{A}_{N,AB}$ is a GMM weighting matrix (recall that here $\Delta \mathbf{y}_N = [\Delta \mathbf{y}_N^\top(3), \ldots, \mathbf{y}_N^\top(T)]^\top$ and $\Delta \mathbf{Z}_N = [\Delta \mathbf{Z}_N^\top(3), \ldots, \Delta \mathbf{Z}_N^\top(T)]^\top$). The resulting first-stage spatial Arellano-Bond estimator then becomes:

$$\hat{\boldsymbol{\theta}}_N = \left[\Delta \mathbf{Z}_N^\top \mathbf{H}_{N,AB} \mathbf{A}_{N,AB} \mathbf{H}_{N,AB}^\top \Delta \mathbf{Z}_N\right]^{-1} \Delta \mathbf{Z}_N^\top \mathbf{H}_{N,AB} \mathbf{A}_{N,AB} \mathbf{H}_{N,AB}^\top \Delta \mathbf{y}_N. \tag{14}$$

The weighting matrix $\mathbf{A}_{N,AB}$ recommended under the assumption of iid errors $\mathbf{u}_N$ by Arellano and Bond (1991) is equal to the $J \times J$ matrix $\mathbf{A}_{N,AB} = [\mathbf{H}_{N,AB}^\top \mathbf{G}_{N,AB} \mathbf{H}_{N,AB}/N]^{-1}$, where

$\mathbf{G}_{N,AB} = \mathbf{G} \otimes \mathbf{I}_N$ is an $N(T-2) \times N(T-2)$ weighting matrix with elements $(i, j = 1, \ldots, T-2)$

$$G_{ij} = \begin{cases} 2 & \text{if } i = j \\ -1 & \text{if } \; i = j+1 \\ -1 & \text{if } \; j = i+1 \\ 0 & \text{otherwise} \end{cases}, \tag{15}$$

and $\otimes$ denotes the Kronecker product. Although not necessarily optimal under the spatial correlation of errors, we do not have a better choice at this stage without knowledge of ρ.

3.1.2 Blundell-Bond Estimator

The standard Arellano-Bond estimator is known to be rather inefficient when instruments are weak because it makes use of information contained in first differences of variables only. To address this shortcoming, the GMM approach of Blundell and Bond (1998)—often referred to as the system GMM estimator—extends the Arellano and Bond (1991) conditions by specifying moment conditions also for variables in levels rather than only for their first differences. The Blundell-Bond estimator for the spatially autoregressive dynamic panel model can be derived by stacking equation (8) and:

$$\mathbf{y}_N(t) = \mathbf{Z}_N(t)\boldsymbol{\theta} + \mathbf{u}_N(t), \qquad t = 3, \ldots, T. \tag{16}$$

The Blundell and Bond (1998) moment conditions for the level equation (16), which contains individual effects $\boldsymbol{\eta}_N$, are constructed using the first-differenced variables as instruments (i.e., using instruments not containing the individual effects). For example, for the strictly exogenous variables

$$\mathrm{E}[\Delta\mathbf{X}_N^\top(t)\mathbf{u}_N(t)] = 0, \qquad t = 3, ..., T, \tag{17}$$

which—in contrast to the estimator in Section 3.1.1—requires the individual effects to be independent of $\Delta\mathbf{X}_N(t)$. The equivalents of the instruments for both the time and spatially lagged dependent variables given in (11), (12), and (13) for model (8) can thus be specified for model

(16) as

$$\begin{aligned}
\mathrm{E}[\Delta\mathbf{y}_N^\top(t-s)\mathbf{u}_N(t)] &= 0, \quad t=3,...,T, \quad s=1,...,t-2,\\
\mathrm{E}[\{\mathbf{W}_N^l\Delta\mathbf{y}_N(t-s)\}^\top\mathbf{u}_N(t)] &= 0, \quad t=3,...,T, \quad s=1,...,t-2, \quad l=1,2,...,L,\\
\mathrm{E}[\{\mathbf{W}_N\Delta\mathbf{X}_N(t)\}^\top\mathbf{u}_N(t)] &= 0, \quad t=3,...,T,
\end{aligned}$$

respectively. These moment conditions can be concisely written as $\mathrm{E}[\mathbf{H}_{N,BB}^\top(t)\mathbf{u}_N(t)]=0$, where the columns of $\mathbf{H}_{N,BB}(t)$ represent the instruments $\Delta\mathbf{X}_N(t), \Delta\mathbf{y}_N(t-s), \mathbf{W}_N^l\Delta\mathbf{y}_N(t-s)$, and $\mathbf{W}_N\Delta\mathbf{X}_N(t)$ given above.

Merging the information from all available time periods, let $\mathbf{H}_{N,BB}$ be a block-diagonal matrix consisting of blocks $\mathbf{H}_{N,BB}(t)$, $\mathbf{y}_N=[\mathbf{y}_N^\top(3),\ldots,\mathbf{y}_N^\top(T)]^\top$, and $\mathbf{Z}_N=[\mathbf{Z}_N^\top(3),\ldots,\mathbf{Z}_N^\top(T)]^\top$ for $t=3,\ldots,T$. These instruments for the level equation (16) are typically used jointly with the instruments introduced in Section 3.1.1 for the differenced equation (8). To define the Blundell-Bond estimator for the spatially autoregressive dynamic panel model, we thus define merged matrices for both systems: the vector of responses $\vec{\mathbf{y}}_N=[\Delta\mathbf{y}_N^\top,\mathbf{y}_N^\top]^\top$, the matrix of explanatory variables $\vec{\mathbf{Z}}_N=[\Delta\mathbf{Z}_N^\top,\mathbf{Z}_N^\top]^\top$, the vector of errors $\vec{\mathbf{u}}_N=[\Delta\mathbf{u}_N^\top,\mathbf{u}_N^\top]^\top$, the instruments $\vec{\mathbf{H}}_N=\mathrm{diag}\{\mathbf{H}_{N,AB},\mathbf{H}_{N,BB}\}$, and the weighting matrices $\vec{\mathbf{G}}_N=\mathrm{diag}\{\mathbf{G}_{N,AB},\mathbf{I}_{T-2}\otimes\mathbf{I}_N\}$ and $\vec{\mathbf{A}}_N=[\vec{\mathbf{H}}_N^\top\vec{\mathbf{G}}_N\vec{\mathbf{H}}_N/N]^{-1}$ (see Kiviet (2007) for alternatives).[13] Minimizing

$$\frac{1}{N}(\vec{\mathbf{H}}_N^\top\Delta\vec{\mathbf{u}}_N)^\top\vec{\mathbf{A}}_N(\vec{\mathbf{H}}_N^\top\Delta\vec{\mathbf{u}}_N)=\frac{1}{N}\left[\vec{\mathbf{H}}_N^\top(\Delta\vec{\mathbf{y}}_N-\Delta\vec{\mathbf{Z}}_N\boldsymbol{\theta})\right]^\top\vec{\mathbf{A}}_N\left[\vec{\mathbf{H}}_N^\top(\Delta\vec{\mathbf{y}}_N-\Delta\vec{\mathbf{Z}}_N\boldsymbol{\theta})\right]$$

with respect to $\boldsymbol{\theta}$ then leads to the first-stage spatial Blundell-Bond estimator:

$$\hat{\boldsymbol{\theta}}_N=\left(\vec{\mathbf{Z}}_N^\top\vec{\mathbf{H}}_N\vec{\mathbf{A}}_N\vec{\mathbf{H}}_N^\top\vec{\mathbf{Z}}_N\right)^{-1}\vec{\mathbf{Z}}_N^\top\vec{\mathbf{H}}_N\vec{\mathbf{A}}_N\vec{\mathbf{H}}_N^\top\vec{\mathbf{y}}_N. \tag{18}$$

Given that the forms (14) and (18) are identical, we will use for the sake of simplicity only the notation $\vec{\mathbf{y}}_N,\vec{\mathbf{u}}_N,\vec{\mathbf{Z}}_N,\vec{\mathbf{H}}_N,\ldots$ from now on, representing the vectors and matrices of responses, errors, covariates, instruments and so on used for estimation, be it in the case of the spatial

[13]Without prior knowledge of (ε_i, η_i) moments, an asymptotically optimal weighting matrix cannot be constructed in the first step (cf. Blundell and Bond, 1998).

Arellano-Bond or Blundell-Bond estimators.

3.2 The Second Stage

Having derived the first-stage estimate $\hat{\boldsymbol{\theta}}_N$ of regression coefficients, the parameters ρ and σ_v^2 of the error distribution can be estimated. To estimate them, we construct a GMM estimator based on errors $\mathbf{u}_N(t)$, which are in turn replaced by the regression residuals $\hat{\mathbf{u}}_N(t) = \mathbf{y}_N(t) - \mathbf{Z}_N(t)\hat{\boldsymbol{\theta}}_N$. The three proposed moment conditions are a modification of those derived by Kapoor et al. (2007) for random effects static panel models. The main difference is that we base the estimation of ρ and σ_v^2 on the first differences of errors to account for the presence of individual effects.

To define the moment conditions, let us first denote (with a slight abuse of notation) $\Delta\boldsymbol{\varepsilon}_N = [\Delta\boldsymbol{\varepsilon}_N^\top(2), \ldots, \Delta\boldsymbol{\varepsilon}_N^\top(T)]^\top$ and $\Delta\mathbf{u}_N = [\Delta\mathbf{u}_N^\top(2), \ldots, \Delta\mathbf{u}_N^\top(T)]^\top$. Their counterparts spatially transformed by matrix $\mathbf{M}_N$ are $\Delta\bar{\boldsymbol{\varepsilon}}_N = (\mathbf{I}_{T-1} \otimes \mathbf{M}_N)\Delta\boldsymbol{\varepsilon}_N$, $\Delta\bar{\mathbf{u}}_N = (\mathbf{I}_{T-1} \otimes \mathbf{M}_N)\Delta\mathbf{u}_N$, and $\Delta\bar{\bar{\mathbf{u}}}_N = (\mathbf{I}_{T-1} \otimes \mathbf{M}_N)\Delta\bar{\mathbf{u}}_N$, which implies that

$$\Delta\boldsymbol{\varepsilon}_N \equiv \Delta\mathbf{u}_N - \rho\Delta\bar{\mathbf{u}}_N, \qquad \Delta\bar{\boldsymbol{\varepsilon}}_N \equiv \Delta\bar{\mathbf{u}}_N - \rho\Delta\bar{\bar{\mathbf{u}}}_N. \tag{19}$$

The three equations identifying ρ and σ_v^2 are as follows (see Appendix A.1 for a derivation):

$$\mathrm{E}\begin{bmatrix} \frac{1}{N(T-1)}\Delta\boldsymbol{\varepsilon}_N^\top\Delta\boldsymbol{\varepsilon}_N \\ \frac{1}{N(T-1)}\Delta\bar{\boldsymbol{\varepsilon}}_N^\top\Delta\bar{\boldsymbol{\varepsilon}}_N \\ \frac{1}{N(T-1)}\Delta\bar{\boldsymbol{\varepsilon}}_N^\top\Delta\boldsymbol{\varepsilon}_N \end{bmatrix} = \begin{bmatrix} 2\sigma_v^2 \\ 2\sigma_v^2\,\mathrm{tr}(\mathbf{M}_N^\top\mathbf{M}_N)/N \\ 0 \end{bmatrix}, \tag{20}$$

where $\mathrm{tr}(\mathbf{M}_N^\top\mathbf{M}_N)$ denotes the trace of the matrix $\mathbf{M}_N^\top\mathbf{M}_N$. If we now substitute for $\Delta\boldsymbol{\varepsilon}_N$ and $\Delta\bar{\boldsymbol{\varepsilon}}_N$ in (20), using $\Delta\mathbf{u}_N$ and $\Delta\bar{\mathbf{u}}_N$ [see (19)], we obtain the following moment conditions:

$$\mathrm{E}[\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N(\rho, \rho^2, \sigma_v^2)^\top] = 0, \tag{21}$$

where $\boldsymbol{\gamma}_N = [\frac{1}{N(T-1)}\Delta\mathbf{u}_N^\top\Delta\mathbf{u}_N, \frac{1}{N(T-1)}\Delta\bar{\mathbf{u}}_N^\top\Delta\bar{\mathbf{u}}_N, \frac{1}{N(T-1)}\Delta\bar{\mathbf{u}}_N^\top\Delta\mathbf{u}_N]^\top$ and

$$\boldsymbol{\Gamma}_N = \begin{bmatrix} \frac{2}{N(T-1)}\Delta\mathbf{u}_N^\top\Delta\bar{\mathbf{u}}_N & -\frac{1}{N(T-1)}\Delta\bar{\mathbf{u}}_N^\top\Delta\bar{\mathbf{u}}_N & 2 \\ \frac{2}{N(T-1)}\Delta\bar{\mathbf{u}}_N^\top\Delta\bar{\bar{\mathbf{u}}}_N & -\frac{1}{N(T-1)}\Delta\bar{\bar{\mathbf{u}}}_N^\top\Delta\bar{\bar{\mathbf{u}}}_N & \frac{2}{N}\,\mathrm{tr}(\mathbf{M}_N^\top\mathbf{M}_N) \\ \frac{2}{N(T-1)}\left[\Delta\bar{\mathbf{u}}_N^\top\Delta\bar{\mathbf{u}}_N + \Delta\bar{\bar{\mathbf{u}}}_N^\top\Delta\mathbf{u}_N\right] & -\frac{1}{N(T-1)}\Delta\bar{\bar{\mathbf{u}}}_N^\top\Delta\bar{\mathbf{u}}_N & 0 \end{bmatrix}. \tag{22}$$

The nonlinear system of equations (21) can be solved by GMM to obtain estimates of ρ and σ_v^2. Since the $\Delta\mathbf{u}_N$'s are not known, we have to estimate them by regression residuals from (8): $\Delta\hat{\mathbf{u}}_N = \Delta\mathbf{y}_N - \Delta\mathbf{Z}_N\hat{\boldsymbol{\theta}}_N$, where $\hat{\boldsymbol{\theta}}_N$ is an initial estimator obtained in Section 3.1. Denoting the analogs of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ based on the regression residuals $\Delta\hat{\mathbf{u}}_N$ by $\hat{\boldsymbol{\gamma}}_N$ and $\hat{\boldsymbol{\Gamma}}_N$, respectively, the GMM estimator of ρ and σ_v based on (21) is defined by

$$(\hat{\rho}_N, \hat{\sigma}_{v,N}) = \arg\min_{\rho,\sigma_v}[\hat{\boldsymbol{\gamma}}_N - \hat{\boldsymbol{\Gamma}}_N(\rho,\rho^2,\sigma_v^2)^\top]^\top \hat{\mathbf{B}}_N[\hat{\boldsymbol{\gamma}}_N - \hat{\boldsymbol{\Gamma}}_N(\rho,\rho^2,\sigma_v^2)^\top], \tag{23}$$

where $\hat{\mathbf{B}}_N$ is a GMM weighting matrix; in Section 5, we use only $\hat{\mathbf{B}}_N = \mathbf{I}_3$.

3.3 The Third Stage

In the final step, the estimate of ρ can be used to spatially transform the variables in (8) and (16) to yield models with cross-sectionally uncorrelated errors:

$$\begin{aligned} \Delta\tilde{\mathbf{y}}_N(t) &= \Delta\tilde{\mathbf{Z}}_N(t)\boldsymbol{\theta} + \Delta\boldsymbol{\varepsilon}_N(t), & (24)\\ \tilde{\mathbf{y}}_N(t) &= \tilde{\mathbf{Z}}_N(t)\boldsymbol{\theta} + \boldsymbol{\varepsilon}_N(t), & (25) \end{aligned}$$

where $\tilde{\mathbf{p}}_N(t) = (\mathbf{I}_N - \hat{\rho}_N\mathbf{M}_N)\mathbf{p}_N(t)$ for $\mathbf{p}_N = \{\mathbf{y}_N, \mathbf{Z}_N\}$. For this system, we can construct the instruments, moment conditions, and GMM estimator in the same way as in Section 3.1. Note that the moment conditions of Section 3.1 were constructed for any kind of spatial dependence also including the currently proposed errors $(\mathbf{I}_N - \hat{\rho}_N\mathbf{M}_N)\mathbf{u}_N = (\mathbf{I}_N - \hat{\rho}_N\mathbf{M}_N)(\mathbf{I}_N - \rho^0\mathbf{M}_N)^{-1}\boldsymbol{\varepsilon}_N$, where ρ^0 represents the true value of the spatial correlation coefficient. Denoting the matrix of dependent, explanatory, and instrumental variables used in all moment conditions $\vec{\tilde{\mathbf{y}}}_N$, $\vec{\tilde{\mathbf{Z}}}_N$, and $\vec{\tilde{\mathbf{H}}}_N$ as in Section 3.1.2, the final-stage GMM estimator for the spatially transformed model (24) or (24)–(25) equals, analogously to (18),

$$\tilde{\boldsymbol{\theta}}_N = \left[\Delta\vec{\tilde{\mathbf{Z}}}_N^\top \vec{\tilde{\mathbf{H}}}\vec{\tilde{\mathbf{A}}}_N\vec{\tilde{\mathbf{H}}}_N^\top\Delta\vec{\tilde{\mathbf{Z}}}_N\right]^{-1}\Delta\vec{\tilde{\mathbf{Z}}}_N^\top\vec{\tilde{\mathbf{H}}}_N\vec{\tilde{\mathbf{A}}}_N\vec{\tilde{\mathbf{H}}}_N^\top\Delta\vec{\tilde{\mathbf{y}}}_N, \tag{26}$$

where $\vec{\tilde{\mathbf{A}}}_N \equiv \left[\vec{\tilde{\mathbf{H}}}_N^\top\vec{\mathbf{G}}_N\vec{\tilde{\mathbf{H}}}_N/N\right]^{-1}$. We will show in Section 4 that the weighting matrix $\vec{\tilde{\mathbf{A}}}_N$ is the optimal weighting matrix for the Arellano-Bond estimator provided that $\boldsymbol{\varepsilon}_N = (\mathbf{I}_N -$

$\rho^0\mathbf{M}_N)(\mathbf{y}_N - \mathbf{Z}_N\boldsymbol{\theta}^0)$ is homoscedastic, where $\boldsymbol{\theta}^0$ represent the true value of $\boldsymbol{\theta}$ (which requires that $\mathbf{M}_N$ is specified correctly).

Note that we do not attempt to estimate the optimal weighting matrix (see Section 4.2), even though this is certainly possible. Given the size of the weighting matrix (in case of the Blundell-Bond estimator for $T = 5$, up to 50 moment equations are used) and practically relevant sample sizes (e.g., $T = 5$ and $N = 60$, see Section 5), we feel there is little to no benefit in using two-step GMM in these models (especially given the risk of worsening the precision of estimation due to mis-estimation of the weighting matrix); see Appendix A of Blundell and Bond (1998).

4 Asymptotic Properties of the Estimators

To formulate the asymptotic results for the estimators $\boldsymbol{\theta}_N$ [given in (18)], $\hat{\rho}_N$ and $\hat{\sigma}_{v,N}$ [given in (23)] and $\tilde{\boldsymbol{\theta}}_N$ [given in (26)], let $\boldsymbol{\theta}^0$, ρ^0, and σ_v^0 denote their true values. Note that $\boldsymbol{\theta}_N$ ($\tilde{\boldsymbol{\theta}}_N$) can represent here the first (third) stage spatially corrected Arellano-Bond or Blundell-Bond estimator depending on which moment conditions are used. Further, an extended notation for the spatial matrices is needed: in the case of the Arellano-Bond estimator, let $\vec{\mathbf{I}}_N = \mathbf{I}_{T-2} \otimes \mathbf{I}_N$, $\vec{\mathbf{M}}_N = \mathbf{I}_{T-2} \otimes \mathbf{M}_N$, and $\vec{\mathbf{W}}_N = \mathbf{I}_{T-2} \otimes \mathbf{W}_N$; in the case of the Blundell-Bond estimator, let $\vec{\mathbf{I}}_N = \mathbf{I}_{2(T-2)} \otimes \mathbf{I}_N$, $\vec{\mathbf{M}}_N = \mathbf{I}_{2(T-2)} \otimes \mathbf{M}_N$, and $\vec{\mathbf{W}}_N = \mathbf{I}_{2(T-2)} \otimes \mathbf{W}_N$. Additionally, we will extend the '$\vec{\ }$' notation also to the vectors of error terms: $\vec{\mathbf{u}}_N = \vec{\mathbf{y}}_N - \vec{\mathbf{Z}}_N\boldsymbol{\theta}^0$, $\vec{\boldsymbol{\varepsilon}}_N = (\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)(\vec{\mathbf{y}}_N - \vec{\mathbf{Z}}_N\boldsymbol{\theta}^0)$, and $\vec{\boldsymbol{\varepsilon}}_N = (0, \boldsymbol{\eta}_N)^\top + \vec{\mathbf{v}}_N$. In what follows, we will first discuss the imposed assumptions (Section 4.1) and then the derived asymptotic results (Section 4.2).

4.1 Assumptions

First, the assumptions needed for the consistency and asymptotic normality of the three-stage spatially corrected GMM estimator are specified. We use here high-level assumptions so that some strict structure does not have to be imposed on the triangular array of dependent and explanatory variables $\vec{\mathbf{y}}_N$ and $\vec{\mathbf{X}}_N$. These assumptions are conceptually similar to Kapoor et al. (2007), including the assumption of homoscedasticity, and can be relaxed by the method of Kelejian and Prucha (2010), who allow for unknown heteroscedasticity in the innovations.

Throughout the section, we assume $N \to +\infty$ and $T = c_0$, where c_0 is a constant. Although this is a standard setup in the literature, let us note that, if N is fixed and $T \to +\infty$, the proposed estimators will be biased because the number of instruments for some moment conditions is increasing with T (see Alvarez and Arellano, 2003, and Bun and Kiviet, 2006). Bun and Kiviet (2006), however, show that limiting the number of instruments guarantees (asymptotic) unbiasedness of the discussed GMM estimators even if N is fixed and T is large and increasing above any bound. While the theoretical results presented here for a fixed T apply to any number of instruments, we recommend for these theoretical and also practical reasons to limit the number of instruments; in Section 5, the simulation results are obtained using at most three dynamic and three spatial lags.

Now, the first set of assumptions specifies standard assumptions regarding the error terms, which guarantee the validity of the moment conditions specified in Section 3 and the existence of finite second moments for the central limit theorem. The only more restrictive assumption on the individual effects follows from Blundell and Bond (1998), see Assumption E2 below, and the existence of the fourth moments, which is made for the convenience of using some auxiliary results of Kelejian and Prucha (2010).

Assumption E

1. The error vectors $\mathbf{v}_N(t) = [v_{N1}(t), \ldots, v_{NN}(t)]^\top$ are independent and identically distributed for each $N \in \mathbb{N}$ and $t = 1, \ldots, T$ with zero mean $\mathrm{E}[v_{Ni}(t)] = 0$, a finite variance $\mathrm{Var}[v_{Ni}(t)] = \sigma_v^2$, $i = 1, \ldots, N$, and uniformly bounded fourth moments. Further, $\mathbf{v}_N(t)$ is assumed to be independent of $\boldsymbol{\eta}_N$ and $\mathbf{X}_N(t)$ for any $t = 1, \ldots, T$.

2. The fixed effects $\boldsymbol{\eta}_N$ have uniformly bounded fourth moments. In the case of the Blundell-Bond estimator, $\boldsymbol{\eta}_N$ is additionally assumed to be uncorrelated with $\Delta\mathbf{Z}_N(s)$ and explanatory variables have a time-invariant mean, $\mathrm{E}\,\mathbf{Z}_N(t) = \mathrm{E}\,\mathbf{Z}_N(s)$, $s = 1, ..., t-1$ and

$t = 2, \ldots, T$.

3. The variance of $\text{Var}(\vec{\boldsymbol{\varepsilon}}_N|\vec{\mathbf{H}}_N) = \vec{\boldsymbol{\Sigma}}_{\varepsilon,N} = \boldsymbol{\Sigma}_\varepsilon \otimes \vec{\mathbf{I}}_N$, where $\boldsymbol{\Sigma}_\varepsilon$ is a positive-definite matrix.

4. The variance of $\text{Var}(\vec{\mathbf{u}}_N|\vec{\mathbf{H}}_N) = \vec{\boldsymbol{\Sigma}}_{u,N} = (\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1}\vec{\boldsymbol{\Sigma}}_{\varepsilon,N}(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1\top}$.

The spatial structure described by matrices $\mathbf{W}_N$ and $\mathbf{M}_N$ is assumed to follow Assumption S, which is made slightly more general than specified in Section 2—which assumed row normalized matrices—by allowing various normalizations of spatial weight matrices (see Kelejian and Prucha, 2010).

Assumption S

1. All diagonal elements of $\mathbf{W}_N$ and $\mathbf{M}_N$ are zero.

2. There exist finite positive constants $K'_\delta, K''_\delta, K'_\rho$, and K''_ρ such that matrices $\mathbf{I}_N - \delta\mathbf{W}_N$ and $\mathbf{I}_N - \rho\mathbf{M}_N$ are non-singular for all $\delta \in (-K'_\delta, K''_\delta)$ and $\rho \in (-K'_\rho, K''_\rho)$.

3. The absolute values of the row and column sums of $\mathbf{W}_N$, $\mathbf{M}_N$, $(\mathbf{I}_N - \delta^0\mathbf{W}_N)^{-1}$, and $(\mathbf{I}_N - \rho^0\mathbf{M}_N)^{-1}$ are bounded uniformly in $N \in \mathbb{N}$.

The assumptions concerning the explanatory variables and the imposed instrumental variables are high level assumptions, which do not impose a particular structure or distributional assumptions, but require only the existence of particular probability limits and the corresponding expectations needed for the central limit theorem. The latter is used in the proof of the asymptotic normality of $\hat{\boldsymbol{\theta}}_N$. Note that the assumption of the uniformly bounded $(2+\psi)$th moments, see Assumption V3 below, which implies the uniform integrability of the squared moment equations, replaces a more restrictive, though often used condition of bounded nonstochastic regressors (e.g., Kapoor et al., 2007).

Assumption V

1. $\vec{\mathbf{Z}}_N$ has a full rank almost surely.

2. $\vec{\mathbf{H}}_N$ has a rank greater or equal to $K+2$ almost surely.

3. The expectations $\mathrm{E}(\vec{\mathbf{Z}}_{N,ij})^{2+\psi}$ and $\mathrm{E}(\vec{\mathbf{H}}_{N,ij}\vec{\boldsymbol{\varepsilon}}_{N,k})^{2+\psi}$ are bounded uniformly in i,j,k for some $\psi > 0$.

4. The limits of variance matrices $\lim_{N\to\infty} \mathrm{E}(\vec{\mathbf{H}}_N^\top \vec{\boldsymbol{\Sigma}}_{u,N}\vec{\mathbf{H}}_N/N) = \vec{\mathbf{Q}}_{H\Sigma H}$, $\lim_{N\to\infty} \mathrm{E}(\vec{\mathbf{H}}_N^\top \vec{\mathbf{M}}_N^\top \vec{\mathbf{M}}_N \vec{\boldsymbol{\Sigma}}_{u,N}\vec{\mathbf{M}}_N\vec{\mathbf{M}}_N^\top \vec{\mathbf{H}}_N/N) = \vec{\mathbf{Q}}_{HM\Sigma MH}$, $\lim_{N\to\infty} \mathrm{E}[\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)\vec{\boldsymbol{\Sigma}}_{\varepsilon,N}(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^\top \vec{\mathbf{H}}_N/N] = \vec{\mathbf{Q}}_{HEH}$, and $\lim_{N\to\infty} \mathrm{E}(\vec{\mathbf{H}}_N^\top \vec{\mathbf{M}}_N^\top \vec{\boldsymbol{\Sigma}}_{\varepsilon,N}\vec{\mathbf{M}}_N\vec{\mathbf{H}}_N/N) = \vec{\mathbf{Q}}_{HMEMH}$ are finite and positive definite.

5. The limits of matrices $\text{p-lim}_{N\to\infty} N^{-1}\vec{\mathbf{H}}_N^\top \vec{\mathbf{Z}}_N = \vec{\mathbf{Q}}_{HZ}$, $\text{p-lim}_{N\to\infty} N^{-1}\vec{\mathbf{H}}_N^\top \vec{\mathbf{M}}_N\vec{\mathbf{Z}}_N = \vec{\mathbf{Q}}_{HMZ}$, $\text{p-lim}_{N\to\infty} N^{-1}\vec{\mathbf{H}}_N^\top \vec{\mathbf{M}}_N^\top \vec{\mathbf{Z}}_N = \vec{\mathbf{Q}}_{HM^\top Z}$, and $\text{p-lim}_{N\to\infty} N^{-1}\vec{\mathbf{H}}_N^\top \vec{\mathbf{M}}_N^\top \vec{\mathbf{M}}_N\vec{\mathbf{Z}}_N = \vec{\mathbf{Q}}_{HMMZ}$ are finite and have full rank.

6. The matrix $\tilde{\vec{\mathbf{Q}}}_{HZ} = \vec{\mathbf{Q}}_{HZ} - \rho^0(\vec{\mathbf{Q}}_{HMZ} + \vec{\mathbf{Q}}_{HM^\top Z}) + (\rho^0)^2\vec{\mathbf{Q}}_{HMMZ}$ has full rank.

Finally, we have to specify assumptions important for the GMM estimator itself, that is, conditions on the parameter space and the GMM weighting matrices. They mainly guarantee that the spatial correlation matrices $\mathbf{I}_N - \delta\mathbf{W}_N$ and $\mathbf{I}_N - \rho\mathbf{M}_N$ are invertible and GMM matrices $\vec{\mathbf{A}}_N$, $\mathbf{B}_N$, and $\boldsymbol{\Gamma}_N$ are non-singular.

Assumption G

1. The parameter space for $\boldsymbol{\theta} = (\lambda, \delta, \boldsymbol{\beta})^\top$ is $\Theta = (-1,1) \times (-K_\delta', K_\delta'') \times \mathbb{R}^K$.

2. Non-singular symmetric matrices $\vec{\mathbf{A}}_N$ satisfy $\text{p-lim}_{N\to\infty} \vec{\mathbf{A}}_N = \vec{\mathbf{A}}$, where $\vec{\mathbf{A}}$ is a finite positive definite matrix.

3. Non-singular symmetric matrices $\tilde{\vec{\mathbf{A}}}_N$ satisfy $\text{p-lim}_{N\to\infty} \tilde{\vec{\mathbf{A}}}_N = \tilde{\vec{\mathbf{A}}}$, where $\tilde{\vec{\mathbf{A}}}$ is a finite positive definite matrix.

4. The parameter space Φ for $\boldsymbol{\phi} = (\rho, \sigma_v)^\top$ is a compact subset of $(-K_\rho', K_\rho'') \times \mathbb{R}^+$. Moreover, $\boldsymbol{\phi}^0 = (\rho^0, \sigma_v^0)^\top \in \Phi^\circ$.

5. The smallest eigenvalues of the matrix $\mathbf{\Gamma}_N^\top \mathbf{\Gamma}_N$ are uniformly larger than $\kappa_\Gamma > 0$.

6. Positive definite matrices $\hat{\mathbf{B}}_N$ satisfy $\text{p-lim}_{N\to\infty}(\hat{\mathbf{B}}_N - \mathbf{B}_N) = 0$, where $\mathbf{B}_N$ are non-stochastic positive definite matrices with eigenvalues uniformly larger than $\kappa_B > 0$ and uniformly smaller than $K_B > 0$.

4.2 Consistency and Asymptotic Normality

In this section, the asymptotic properties of the proposed estimators are derived. As the regression parameters are estimated by a linear GMM estimator, we only have to account for the spatial error correlation and its estimation to derive the asymptotic distributions in the classical way. On the other hand, the estimation of the spatial error correlation is nonlinear and thus the uniform identification of the parameters has to be verified to guarantee consistency (cf. Kelejian and Prucha, 2010).

We will show first that the initial estimator $\hat{\boldsymbol{\theta}}_N$ defined in (14) or (18) (recall that the latter definition includes the first one as a special case) is consistent.

Theorem 1 *Under Assumptions E, S, V, and G1–G2, the GMM estimator $\hat{\boldsymbol{\theta}}_N$ defined in* (18) *is $\sqrt{N}$-consistent and*

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) \overset{\mathcal{L}}{\to} N(0, [\vec{\mathbf{Q}}_{HZ}^\top \vec{\mathbf{A}} \vec{\mathbf{Q}}_{HZ}]^{-1} \vec{\mathbf{Q}}_{HZ}^\top \vec{\mathbf{A}} \vec{\mathbf{Q}}_{H\Sigma H} \vec{\mathbf{A}}^\top \vec{\mathbf{Q}}_{HZ} [\vec{\mathbf{Q}}_{HZ}^\top \vec{\mathbf{A}} \vec{\mathbf{Q}}_{HZ}]^{-1})$$

as $N \to +\infty$.

Proof. *See Appendix A.2.1.* □

Although the asymptotic distribution of $\hat{\boldsymbol{\theta}}_N$ is derived in Theorem 1, it is not practically applicable at this stage: the variance matrix $\vec{\mathbf{Q}}_{H\Sigma H}$ of $\vec{\mathbf{u}}_N$ depends on an unknown parameter ρ^0 (see Assumption E4). The spatial autocorrelation parameter ρ^0 can be estimated by $\hat{\rho}_N$ defined by (23). Its consistency is proved in the following theorem:

Theorem 2 *Let Assumptions E, S, V3, and G4–G6 hold and $\hat{\boldsymbol{\theta}}_N$ be a $\sqrt{N}$-consistent estimator of $\boldsymbol{\theta}^0$, $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) = O_p(1)$. Then the GMM estimator $\hat{\boldsymbol{\phi}}_N = (\hat{\rho}_N, \hat{\sigma}_{v,N})^\top$ of $\boldsymbol{\phi}^0 = (\rho^0, \sigma_v)^\top$ is consistent, $\hat{\boldsymbol{\phi}}_N \to \boldsymbol{\phi}^0$ in probability as $N \to \infty$.*

Proof. *See Appendix A.2.2.* □

Having a consistent estimate ρ_N of ρ^0, the asymptotic variance in Theorem 1 can be evaluated. More importantly, it can be used to transform the model (24)–(25) to obtain spatially uncorrelated errors as described in Section 3.3. The asymptotic properties of the GMM estimator based on the model (24)–(25) are given in the next theorem.

Theorem 3 *Under Assumptions E, S, V, and G1–G3, the GMM estimator* $\tilde{\boldsymbol{\theta}}_N$ *defined in* (26) *is* $\sqrt{N}$*-consistent and*

$$\sqrt{N}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} N(0, [\vec{\tilde{\mathbf{Q}}}_{HZ}^{\top}\vec{\tilde{\mathbf{A}}}\vec{\tilde{\mathbf{Q}}}_{HZ}]^{-1}\vec{\tilde{\mathbf{Q}}}^{\top}\vec{\tilde{\mathbf{A}}}\vec{\mathbf{Q}}_{HEH}\vec{\tilde{\mathbf{A}}}^{\top}\vec{\tilde{\mathbf{Q}}}_{HZ}[\vec{\tilde{\mathbf{Q}}}_{HZ}^{\top}\vec{\tilde{\mathbf{A}}}\vec{\tilde{\mathbf{Q}}}_{HZ}]^{-1})$$

as $N \to +\infty$*. For* $\vec{\tilde{\mathbf{A}}} = [\vec{\mathbf{Q}}_{HEH}]^{-1}$*, this asymptotic variance matrix reduces to* $[\vec{\tilde{\mathbf{Q}}}_{HZ}^{\top}\vec{\mathbf{Q}}_{HEH}^{-1}\vec{\tilde{\mathbf{Q}}}_{HZ}]^{-1}$.

Proof. *See Appendix A.2.3.* □

It is well-known that the optimal weighting matrix for GMM equals the inverse of the variance of the moment conditions, which in the case of the third-stage estimator still depends on ρ^0 and is asymptotically equal to

$$\vec{\mathbf{Q}}_{HEH} = \lim_{N\to+\infty} \mathrm{E}[\vec{\mathbf{H}}_N^{\top}(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{\top}\vec{\boldsymbol{\Sigma}}_{\varepsilon,N}(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)\vec{\mathbf{H}}_N/N],$$

see Assumption V4 and Theorem 2. Given an estimate $\hat{\rho}_N$, $\vec{\mathbf{Q}}_{HEH}$ can be estimated by

$$\vec{\hat{\mathbf{Q}}}_{HEH,N} = \frac{1}{N}\vec{\mathbf{H}}_N^{\top}(\vec{\mathbf{I}}_N - \hat{\rho}_N\vec{\mathbf{M}}_N)^{\top}\vec{\boldsymbol{\Sigma}}_{\varepsilon,N}(\vec{\mathbf{I}}_N - \hat{\rho}_N\vec{\mathbf{M}}_N)\vec{\mathbf{H}}_N = \frac{1}{N}\vec{\tilde{\mathbf{H}}}_N^{\top}\vec{\boldsymbol{\Sigma}}_{\varepsilon,N}\vec{\tilde{\mathbf{H}}}_N$$

and the weighting matrix proposed in Section 3.3 is thus equal to $\vec{\tilde{\mathbf{A}}}_N = \vec{\hat{\mathbf{Q}}}_{HEH,N}^{-1}$ if $\vec{\mathbf{G}}_N = \vec{\boldsymbol{\Sigma}}_{\varepsilon,N}$, that is, if the errors $\boldsymbol{\varepsilon}_N$ are homoscedastic and the Arellano-Bond estimator is considered.

Given the optimal weighting matrix, it is worth noting that, for $\rho^0 \to 0$, the optimal third-stage GMM estimator converges to the first-stage estimator since $\vec{\tilde{\mathbf{H}}}_N = \vec{\mathbf{H}}_N$ and $\boldsymbol{\varepsilon}_N = \mathbf{u}_N$ for $\hat{\rho}_N = \rho^0 = 0$. A general comparison of the variances of the first-stage and third-stage estimators is, however, difficult as they depend on a general spatial matrix $\vec{\mathbf{M}}_N$ by means of $(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1}$ and $(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)$, respectively. Some indication of the benefits of the spatial error correction are therefore provided by means of simulations in Section 5.

5 Monte Carlo Simulations

To assess the performance of the estimators presented in Section 3, this section reports a Monte Carlo experiment. The design of the Monte Carlo experiment is discussed first before turning to the results.

5.1 Simulation Design

We report the small sample properties of the estimators using data sets generated based on the spatial dynamic panel data model of Section 2. In generating the data, we follow a three-step procedure. First, we generate the vector of covariates, which includes only one exogenous variable. Following Baltagi et al. (2007), the exogenous variable is defined as:

$$\mathbf{X}_N(t) = \boldsymbol{\varsigma} + \boldsymbol{\chi}(t), \qquad \boldsymbol{\varsigma} \sim \text{iid U}[-7.5, 7.5], \qquad \boldsymbol{\chi} \sim \text{iid U}[-5, 5], \tag{27}$$

where $\boldsymbol{\varsigma}$ represents the unit-specific component and $\boldsymbol{\chi}(t)$ denotes a random component; both are drawn from a uniform distribution U defined on a pre-specified interval.

The second step, using (4), (5), and

$$\boldsymbol{\eta}_N \sim \text{iid U}[-1, 1], \qquad \mathbf{v}_N \sim \text{iid N}(0, \mathbf{I}_N), \tag{28}$$

yields the error component $\mathbf{u}_N(t)$. The third step generates data for the dependent variable $\mathbf{y}_N(t)$ and the spatial lag $\mathbf{W}_N\mathbf{y}_N(t)$. The data generation process is given by (6) and (7) for $t = 2, ..., T$ and $\mathbf{y}(1) = \boldsymbol{\eta}_N$. The first $100 - T$ observations of the Monte Carlo runs are discarded to ensure that the results are not unduly affected by the initial values (cf. Hsiao et al., 2002).[14]

Following standard practice in the literature, we use different weight matrices for the spatial lag and spatial error component, that is, $\mathbf{W}_N \neq \mathbf{M}_N$. To accommodate a large N, we generate artificial contiguity matrices. In doing so, we randomly assign a fixed number of neighbors n to each spatial unit i—while maintaining symmetry—and row normalize the matrices. Throughout the simulations, we keep the weights constant for a given N. In the benchmark scenario, we

[14]We have checked the robustness of the results with respect to changes in the initial values.

assume five neighbors of each spatial unit, implying 91.7% zero entries, the so-called sparsity of the weight matrix. As a robustness check, we vary the number of neighbors from 5 to 20 in the random contiguity matrices. In addition, we consider the Bucky ball contiguity specification, which assumes a fixed location of unit i's neighbors. The matrix is shaped like a soccer ball, where the distance from any point to its nearest neighbors is the same for all the points.[15] Depending on whether unit i is a pentagon or hexagon, there are five or six neighbors. Because of its fixed geographic structure, the Bucky ball specification implies $\mathbf{W}_N = \mathbf{M}_N$. Finally, we consider row-normalized weight matrices based on the inverse of squared geographical distance (in kilometers) between the capitals of countries i and j. The bilateral distances are randomly drawn from the distance matrix for 225 countries provided by CEPII.[16]

In the benchmark specification, we use $N = 60$ and $T = 5$. The parameters in (6) and (7) take on the following values in the data generation process. As is standard practice in the literature, the coefficient of the exogenous explanatory variable β is set to unity. We set $\lambda = 0.3$ and $\delta = 0.5$, so that the stationarity conditions (3) are satisfied, and the spatial autocorrelation coefficient ρ equals 0.3. For each experiment, the performance of the estimators is computed based on 1000 replications. Following Kapoor et al. (2007) and others, we measure performance by the $\text{RMSE} = \sqrt{bias^2 + \left(\frac{q_1 - q_2}{1.35}\right)^2}$, where *bias* denotes the difference between the median and the 'true' value of the parameter of interest (i.e., the value imposed in the data-generating process) and $q_1 - q_2$ is the interquantile range (where q_1 is the 0.75 quantile and q_2 is the 0.25 quantile). If the distribution is normal, $(q_1 - q_2)/1.35$ comes close (aside from a rounding error) to the standard deviation of the estimate.

5.2 Results

Table 1 reports the RMSE in estimating the spatial autoregressive parameter δ for various estimators and different values of N starting at the benchmark value of $N = 60$ ($T = 5$ is fixed). The rows report four different types of spatial GMM estimators all of which instrument the time lag of

[15]The Bucky ball specification assumes $N = 60$ and therefore cannot be used to vary N.
[16]See http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

the dependent variable in addition to addressing spatial aspects. We consider a spatial Arellano-Bond estimator (labeled AB) and a spatial Blundell-Bond estimator (labeled BB), which do not correct for spatial error correlation and correspond to the first stage of the three-stage spatial GMM procedure. The spatially corrected Arellano-Bond estimator (labeled SAB) and spatially corrected Blundell-Bond estimator (labeled SBB) explicitly correct for spatial error correlation and correspond to the final stage of the three-stage spatial GMM procedure (as discussed in Section 3.3). We use three time lags in instrumenting the one-period time lag of the dependent variable. To instrument the spatial lag, we use various instrument sets: (i) the modified Kelejian and Robinson (1993) instruments (indicated by the subscript X); (ii) time lags s and spatial lags l of the spatially lagged dependent variable (indicated by the subscript Y); and (iii) a combination of the instrument sets X and Y (represented by the subscript XY). The numbers in the subscripts denote the number of time lags s and spatial lags l of $\mathbf{W}_N^l \mathbf{y}_N(t-s)$, where we consider only cases with an equal number of time lags and spatial lags. The instrument set Y captures the case with only endogenous variables as instruments.[17]

We find that the spatially corrected estimators have generally a smaller RMSE than their non-spatially corrected counterparts. Additionally in the benchmark case, the system-based GMM estimators (labeled BB and SBB) give rise to a smaller RMSE than the difference-based GMM estimators (labeled AB and SAB).[18] For models with a large λ, which generates a strong time dependency, the difference in RMSEs between system-based GMM estimators and difference-based GMM estimators becomes larger; the RMSE increases for the SAB estimator and decreases for the SBB estimator. In the benchmark case of $N = 60$ and $\lambda = 0.3$, SBB with instrument set XY and three time and spatial lags shows the smallest RMSE.[19] The bias of this estimator (not shown in the table, see Čížek et al., 2011) amounts to 5.8% of the true parameter value. Note that the specifications without any exogenous variables in the instrument set yields larger RMSEs than those with both X and Y instrument sets or X instrument set, particularly for

[17] The estimator can also be applied if there are no exogenous variables.

[18] Čížek et al. (2011) show that this applies for various values of δ and λ.

[19] This is not a general result. Depending on the parameter values, either one, two, or three spatial lags is optimal. See below.

the difference-based estimators. Finally, using the XY instrument set improves the efficiency of the spatially corrected estimates compared to using only the X instrument set. However, this is not always the case for the AB/BB estimates. The bias of the AB and SAB estimators with X instruments only amounts to 4.2% and 3.95% of δ, respectively (see Čížek et al., 2011), which is much smaller than that found by Elhorst (2010), who reports a value of 20% of δ for his version of the AB estimator.[20]

In line with expectations, the RMSE in estimating δ decreases if N increases and thus yields consistent estimates of the spatial interaction effect. Extending the number of spatial units from 60 to 500 reduces the RMSE by more than 50% in the case of estimators using the spatially weighted exogenous variables $\mathbf{W}_N \Delta \mathbf{X}_N(t)$ as instruments and by 35–40% in the case of estimators using only the spatially weighted lagged dependent variables $\mathbf{W}_N^l \mathbf{y}_N(t-s)$ as instruments. The AB estimator using the Y instrument set exhibits a very large RMSE even for large N, which likely indicates weak instruments. For $N = 200$, the optimal number of spatial lags in the XY instrument set is smaller than three.

Table 2 presents the RMSE in estimating the parameter δ for various estimators and various values of T starting at the benchmark value of $T = 5$ ($N = 60$ is fixed). If the time dimension of the panel rises, techniques to limit the proliferation of instruments are needed. As before, we limit the lag depth of the dynamic instruments to three, which reduces the RMSE in estimating the spatial lag parameter substantially at higher values of T (cf. Jacobs et al., 2009). Increasing the number of time periods from 5 to 20 in the benchmark case of $\lambda = 0.3$ reduces the average RMSE by 34.8%. The decline in RMSE across difference-based estimators and system-based estimators is rather similar. If λ takes on a value of 0.7, the fall in RMSE of the difference-based estimators induced by a rise in T from 5 to 20 is larger than that of the system-based estimators (57% compared with 51%).

Table 3 presents the RMSE of the estimators for several values of δ in the interval $[0.2, 0.7]$ and for different values of ρ. We vary ρ in the interval $[-0.8, 0.8]$, where a negative ρ implies that

[20] Elhorst (2010) does not correct for spatial error correlation, employs slightly different parameters in the Monte Carlo simulations, and uses a different instrument set.

an unobserved positive shock in the equation for spatial unit i decreases the dependent variable in other spatial units $i \neq j$. To make sure the stationarity condition (3) is met for large values of δ, we have set λ to 0.2. We find that non-spatially corrected GMM estimators always have larger RMSEs in estimating λ, δ, and β than their spatially corrected counterparts. The difference in RMSEs between BB and SBB estimators increases for large positive values of ρ, is zero for a pure spatial lag model (i.e., $\rho = 0$), and takes on small positive values for negative ρ values. For intermediate values of ρ, the BB estimators with XY instruments and three spatial and time lags yield smaller RMSEs than the BB estimators with X instruments. However, at high positive values of ρ, in estimating λ and δ, the BB estimators with the set of XY instruments perform less well than those with the set of X instruments. Once we correct for spatial error correlation, the estimators with XY instruments have the lowest RMSE again. Regarding ρ, a larger RMSE is found at negative values of ρ and a smaller RMSE is obtained at positive values of ρ. Finally, the table shows that the RMSE of δ is not affected much by the size of the spatial lag parameter.

Table 4 investigates the effect of the specification of the weight matrices on the RMSE in estimating δ for various values of ρ. Our key result of spatially corrected estimators having a smaller RMSE than non-spatially corrected estimator holds for all investigated specifications of the weight matrix. Reducing the sparsity of the random contiguity matrices increases the RMSE of all estimators with the exception of the first-stage estimators based on the Y instrument set and a negative ρ. In the benchmark case, the Bucky ball specification—which imposes $\mathbf{W}_N = \mathbf{M}_N$ and assumes five or six neighbors—yields a slightly larger RMSE than for the case of $n = 5$. However, the RMSEs of estimators using weights based on physical distance—in which case all cells of the weight matrix are non-zero—are smaller than in the benchmark case of rather sparse random contiguity matrices.[21]

[21]Not only the sparsity of the weight matrix but also the variation in weights affects the RMSEs of the estimators. Note that the coefficient of variation of the weights for both random contiguity matrices is 3.34, whereas it slightly smaller for the distance based matrices (i.e., 3.27 for $\mathbf{W}_N$ and 3.31 for $\mathbf{M}_N$).

6 Conclusion

This paper deals with GMM estimation of spatial dynamic panel data models with fixed effects and spatially correlated errors. We extend the three-step GMM approach of Kapoor et al. (2007), which corrects for spatially correlated errors in static panel data models, by introducing a spatial lag and a one-period lag of the endogenous variable as additional explanatory variables. Combining the extended Kapoor et al. (2007) framework with the dynamic panel data model GMM estimators of Arellano and Bond (1991) and Blundell and Bond (1998) and supplementing the dynamic instruments by various spatial lags and weighted exogenous variables yields new spatial dynamic panel data estimators.

We formally prove the consistency and asymptotic normality of our spatial GMM estimators for the case of large N and fixed small T. For large T and fixed small N, the spatial estimators are consistent if the instrument count per moment condition is bounded from above. Monte Carlo simulations indicate that the RMSE of spatially corrected GMM estimates—which are based on a spatial lag and spatial error correction—is generally smaller than that of the corresponding spatial GMM estimates in which spatial error correlation is ignored, particularly for strong positive spatial error correlation. The RMSE of the spatial GMM estimates, however, is not affected much by the size of the spatial lag parameter in the data generating process. We also show that the spatial Blundell-Bond estimators outperform the spatial Arellano-Bond estimators. Finally, we find that spatial estimators using spatially weighted endogenous variables as instruments in addition to weighted exogenous variables are more efficient than those based on weighted exogenous variables.

In future research, we intend to add a spatially weighted time lag to the model. In addition, we investigate the consequences of replacing a correction for spatial error correlation by spatially weighted covariates in the model.

Table 1: RMSE of Spatial GMM Estimators of δ for Various Values of N and λ

Estimator	$\lambda = 0.3$			$\lambda = 0.7$		
	$N = 60$	$N = 200$	$N = 500$	$N = 60$	$N = 200$	$N = 500$
AB_X	0.0873	0.0450	0.0307	0.1349	0.0728	0.0502
AB_{XY1}	0.0836	0.0447	0.0303	0.1316	0.0710	0.0492
AB_{XY2}	0.0871	0.0447	0.0307	0.1267	0.0677	0.0472
AB_{XY3}	0.0913	0.0490	0.0296	0.1269	0.0686	0.0458
AB_{Y1}	0.2075	0.1738	0.1301	0.3022	0.2586	0.2311
AB_{Y2}	0.1732	0.1551	0.1129	0.2397	0.1977	0.1690
AB_{Y3}	0.1611	0.1350	0.1076	0.2136	0.1701	0.1353
SAB_X	0.0840	0.0462	0.0307	0.1292	0.0734	0.0484
SAB_{XY1}	0.0812	0.0449	0.0300	0.1254	0.0718	0.0478
SAB_{XY2}	0.0808	0.0443	0.0294	0.1177	0.0659	0.0460
SAB_{XY3}	0.0830	0.0456	0.0280	0.1097	0.0670	0.0445
SAB_{Y1}	0.1927	0.1638	0.1205	0.2821	0.2429	0.2210
SAB_{Y2}	0.1609	0.1431	0.1118	0.2051	0.1850	0.1635
SAB_{Y3}	0.1481	0.1280	0.0986	0.1957	0.1620	0.1297
BB_X	0.0789	0.0457	0.0327	0.0706	0.0547	0.0432
BB_{XY1}	0.0604	0.0386	0.0262	0.0516	0.0373	0.0263
BB_{XY2}	0.0598	0.0389	0.0260	0.0443	0.0358	0.0262
BB_{XY3}	0.0597	0.0387	0.0253	0.0417	0.0339	0.0252
BB_{Y1}	0.0894	0.0748	0.0556	0.0567	0.0467	0.0370
BB_{Y2}	0.0812	0.0663	0.0492	0.0482	0.0408	0.0338
BB_{Y3}	0.0829	0.0636	0.0502	0.0446	0.0374	0.0301
SBB_X	0.0759	0.0438	0.0325	0.0714	0.0509	0.0409
SBB_{XY1}	0.0570	0.0359	0.0246	0.0486	0.0373	0.0254
SBB_{XY2}	0.0547	0.0382	0.0240	0.0383	0.0353	0.0242
SBB_{XY3}	0.0533	0.0360	0.0237	0.0376	0.0324	0.0235
SBB_{Y1}	0.0869	0.0748	0.0510	0.0547	0.0462	0.0338
SBB_{Y2}	0.0759	0.0631	0.0472	0.0443	0.0406	0.0312
SBB_{Y3}	0.0714	0.0594	0.0474	0.0408	0.0362	0.0273

Notes: RMSEs based on Monte Carlo simulations with 1000 replications. The parameters in the benchmark scenario are: $N = 60$, $T = 5$, $n = 5$, $\lambda = 0.3$, $\delta = 0.5$, $\beta = 1$, and $\rho = 0.3$. To meet the stability condition (3), δ is set to 0.2 if $\lambda = 0.7$. The labels AB, SAB, BB, and SBB denote the first-stage spatial Arellano-Bond estimator, the spatially corrected Arellano-Bond estimator, the first-stage spatial Blundell-Bond estimator, and the spatially corrected Blundell-Bond estimator, respectively. The subscripts X and Y refer to instrument sets for the spatial lag based on spatially weighted values of $\mathbf{X}_N$ and $\mathbf{y}_N$, respectively. The subscript XY indicates that both instrument sets are employed. The numbers in the subscripts report the number of time lags and spatial lags of the spatially lagged dependent variable used for instrumenting the spatial lag. The one-period time lag of the dependent variable is instrumented by three time lags of the dependent variable.

Table 2: RMSE of Spatial GMM Estimators of δ for Various Values of T and λ

Estimator	$\lambda = 0.3$			$\lambda = 0.7$		
	$T = 5$	$T = 10$	$T = 20$	$T = 5$	$T = 10$	$T = 20$
AB_X	0.0873	0.0628	0.0537	0.1349	0.0891	0.0665
AB_{XY1}	0.0836	0.0605	0.0496	0.1316	0.0782	0.0526
AB_{XY2}	0.0871	0.0696	0.0611	0.1267	0.0799	0.0605
AB_{XY3}	0.0913	0.0827	0.0792	0.1269	0.0893	0.0706
AB_{Y1}	0.2075	0.1419	0.1220	0.3022	0.1513	0.0962
AB_{Y2}	0.1732	0.1333	0.1120	0.2397	0.1363	0.0935
AB_{Y3}	0.1611	0.1289	0.1154	0.2136	0.1294	0.0952
SAB_X	0.0840	0.0582	0.0467	0.1292	0.0827	0.0614
SAB_{XY1}	0.0812	0.0541	0.0434	0.1254	0.0724	0.0484
SAB_{XY2}	0.0808	0.0608	0.0511	0.1177	0.0715	0.0494
SAB_{XY3}	0.0830	0.0710	0.0671	0.1097	0.0770	0.0579
SAB_{Y1}	0.1927	0.1271	0.1048	0.2821	0.1370	0.0833
SAB_{Y2}	0.1609	0.1154	0.0967	0.2051	0.1192	0.0778
SAB_{Y3}	0.1481	0.1108	0.0987	0.1957	0.1110	0.0761
BB_X	0.0789	0.0518	0.0428	0.0706	0.0464	0.0350
BB_{XY1}	0.0604	0.0452	0.0385	0.0516	0.0386	0.0289
BB_{XY2}	0.0598	0.0436	0.0369	0.0443	0.0269	0.0198
BB_{XY3}	0.0597	0.0482	0.0448	0.0417	0.0269	0.0212
BB_{Y1}	0.0894	0.0742	0.0704	0.0567	0.0426	0.0319
BB_{Y2}	0.0812	0.0639	0.0554	0.0482	0.0299	0.0216
BB_{Y3}	0.0829	0.0680	0.0621	0.0446	0.0299	0.0235
SBB_X	0.0759	0.0497	0.0394	0.0714	0.0462	0.0324
SBB_{XY1}	0.0570	0.0426	0.0346	0.0486	0.0365	0.0261
SBB_{XY2}	0.0547	0.0388	0.0314	0.0383	0.0241	0.0173
SBB_{XY3}	0.0533	0.0415	0.0377	0.0376	0.0236	0.0181
SBB_{Y1}	0.0869	0.0710	0.0598	0.0547	0.0406	0.0293
SBB_{Y2}	0.0759	0.0533	0.0470	0.0443	0.0253	0.0187
SBB_{Y3}	0.0714	0.0573	0.0511	0.0408	0.0248	0.0194

Notes: RMSEs based on Monte Carlo simulations with 1000 replications. The parameters in the benchmark scenario are: $N = 60$, $T = 5$, $n = 5$, $\lambda = 0.3$, $\delta = 0.5$, $\beta = 1$, and $\rho = 0.3$. To meet the stability condition (3), δ is set to 0.2 if $\lambda = 0.7$. The labels AB, SAB, BB, and SBB denote the first-stage spatial Arellano-Bond estimator, the spatially corrected Arellano-Bond estimator, the first-stage spatial Blundell-Bond estimator, and the spatially corrected Blundell-Bond estimator, respectively. The subscripts X and Y refer to instrument sets for the spatial lag based on spatially weighted values of $\mathbf{X}_N$ and $\mathbf{y}_N$, respectively. The subscript XY indicates that both instrument sets are employed. The numbers in the subscripts report the number of time lags and spatial lags of the spatially lagged dependent variable used for instrumenting the spatial lag. The one-period time lag of the dependent variable is instrumented by three time lags of the dependent variable.

Table 3: RMSE of Spatial Blundell-Bond Estimators for Various Values of δ and ρ

Estimator	Parameter	$\delta = 0.3$ and various ρ					$\delta = 0.5$ and various ρ					$\delta = 0.7$ and various ρ				
		-0.8	-0.4	0	0.4	0.8	-0.8	-0.4	0	0.4	0.8	-0.8	-0.4	0	0.4	0.8
BB_X	λ	0.054	0.047	0.045	0.047	0.086	0.047	0.042	0.042	0.043	0.082	0.041	0.036	0.036	0.041	0.094
BB_{XY3}		0.044	0.040	0.039	0.043	0.083	0.041	0.037	0.036	0.039	0.090	0.037	0.031	0.031	0.039	0.101
SBB_X		0.042	0.044	0.045	0.044	0.042	0.038	0.041	0.041	0.040	0.039	0.031	0.034	0.036	0.034	0.034
SBB_{XY3}		0.033	0.036	0.038	0.037	0.034	0.030	0.034	0.035	0.034	0.033	0.026	0.030	0.031	0.032	0.031
BB_X	δ	0.109	0.096	0.092	0.100	0.197	0.090	0.077	0.076	0.083	0.176	0.060	0.052	0.052	0.060	0.139
BB_{XY3}		0.080	0.067	0.067	0.083	0.256	0.069	0.057	0.056	0.070	0.216	0.050	0.042	0.042	0.055	0.151
SBB_X		0.075	0.085	0.089	0.092	0.090	0.061	0.066	0.075	0.076	0.076	0.040	0.046	0.051	0.054	0.053
SBB_{XY3}		0.053	0.060	0.067	0.067	0.066	0.043	0.051	0.057	0.058	0.059	0.032	0.036	0.042	0.043	0.045
BB_X	β	0.052	0.046	0.045	0.048	0.083	0.052	0.045	0.045	0.050	0.090	0.054	0.048	0.048	0.055	0.105
BB_{XY3}		0.046	0.041	0.039	0.043	0.077	0.048	0.042	0.041	0.045	0.083	0.049	0.043	0.042	0.047	0.090
SBB_X		0.043	0.045	0.045	0.043	0.040	0.044	0.045	0.045	0.043	0.040	0.047	0.049	0.047	0.044	0.043
SBB_{XY3}		0.036	0.039	0.038	0.037	0.036	0.038	0.040	0.041	0.039	0.036	0.039	0.042	0.042	0.039	0.037
SBB_X	ρ	0.134	0.144	0.136	0.110	0.080	0.135	0.147	0.138	0.110	0.085	0.134	0.146	0.140	0.112	0.090
SBB_{XY3}		0.140	0.149	0.140	0.111	0.067	0.140	0.147	0.138	0.111	0.073	0.141	0.147	0.138	0.114	0.086

Notes: RMSEs based on Monte Carlo simulations with 1000 replications. The remaining parameters are: $N = 60$, $T = 5$, $n = 5$, $\lambda = 0.2$, and $\beta = 1$. The labels BB and SBB denote the first-stage spatial Blundell-Bond estimator and the spatially corrected Blundell-Bond estimator, respectively. The subscripts X and Y refer to instrument sets for the spatial lag based on spatially weighted values of $\mathbf{X}_N$ and $\mathbf{y}_N$, respectively. The subscript XY indicates that both instrument sets are employed. The numbers in the subscripts report the number of time lags and spatial lags of the spatially lagged dependent variable used for instrumenting the spatial lag. The one-period time lag of the dependent variable is instrumented by three time lags of the dependent variable.

Table 4: RMSE of Spatial GMM Estimators of δ for Various Weight Matrices and Values of ρ

Estimator	$\rho=-0.8$					$\rho=0.3$					$\rho=0.8$				
	Contiguity				Distance	Contiguity				Distance	Contiguity				Distance
	$n=5$	$n=10$	$n=20$	Bucky		$n=5$	$n=10$	$n=20$	Bucky		$n=5$	$n=10$	$n=20$	Bucky	
AB_X	0.097	0.124	0.153	0.140	0.098	0.087	0.127	0.169	0.081	0.087	0.201	0.268	0.346	0.216	0.196
AB_{XY1}	0.092	0.115	0.123	0.141	0.091	0.084	0.123	0.171	0.081	0.087	0.212	0.292	0.380	0.228	0.197
AB_{XY2}	0.091	0.110	0.123	0.151	0.087	0.087	0.130	0.182	0.087	0.082	0.235	0.318	0.404	0.279	0.220
AB_{XY3}	0.095	0.113	0.127	0.178	0.084	0.091	0.138	0.189	0.094	0.082	0.256	0.329	0.405	0.310	0.226
AB_{Y1}	0.213	0.204	0.193	0.314	0.195	0.207	0.229	0.271	0.263	0.184	0.395	0.413	0.453	0.463	0.355
AB_{Y2}	0.178	0.150	0.145	0.327	0.152	0.173	0.195	0.227	0.229	0.143	0.370	0.403	0.435	0.458	0.324
AB_{Y3}	0.159	0.146	0.134	0.345	0.128	0.161	0.183	0.214	0.208	0.130	0.354	0.388	0.424	0.450	0.309
SAB_X	0.070	0.096	0.114	0.078	0.070	0.084	0.126	0.182	0.085	0.085	0.085	0.131	0.233	0.128	0.079
SAB_{XY1}	0.066	0.091	0.101	0.074	0.067	0.081	0.125	0.176	0.083	0.081	0.079	0.134	0.252	0.134	0.070
SAB_{XY2}	0.067	0.088	0.101	0.078	0.062	0.081	0.126	0.175	0.088	0.077	0.078	0.146	0.278	0.185	0.066
SAB_{XY3}	0.069	0.089	0.109	0.082	0.061	0.083	0.129	0.182	0.087	0.074	0.079	0.156	0.292	0.220	0.069
SAB_{Y1}	0.141	0.152	0.148	0.216	0.124	0.193	0.224	0.292	0.282	0.175	0.182	0.248	0.389	0.442	0.152
SAB_{Y2}	0.119	0.116	0.118	0.195	0.092	0.161	0.185	0.230	0.235	0.135	0.158	0.226	0.356	0.421	0.120
SAB_{Y3}	0.105	0.113	0.113	0.183	0.081	0.148	0.169	0.212	0.205	0.118	0.149	0.214	0.331	0.413	0.115
BB_X	0.088	0.099	0.114	0.125	0.087	0.079	0.102	0.125	0.076	0.071	0.181	0.224	0.272	0.205	0.167
BB_{XY1}	0.066	0.068	0.072	0.103	0.064	0.060	0.075	0.097	0.062	0.053	0.165	0.223	0.283	0.192	0.152
BB_{XY2}	0.063	0.066	0.066	0.101	0.058	0.060	0.075	0.096	0.063	0.053	0.183	0.244	0.296	0.222	0.162
BB_{XY3}	0.065	0.066	0.066	0.106	0.055	0.060	0.078	0.101	0.064	0.054	0.197	0.252	0.297	0.240	0.172
BB_{Y1}	0.101	0.089	0.084	0.159	0.091	0.089	0.100	0.108	0.120	0.080	0.244	0.268	0.307	0.313	0.214
BB_{Y2}	0.088	0.077	0.072	0.142	0.076	0.081	0.093	0.110	0.109	0.071	0.239	0.282	0.314	0.322	0.207
BB_{Y3}	0.084	0.074	0.068	0.140	0.068	0.083	0.094	0.106	0.103	0.068	0.243	0.281	0.309	0.325	0.210
SBB_X	0.060	0.073	0.076	0.068	0.052	0.076	0.107	0.136	0.083	0.073	0.075	0.112	0.182	0.123	0.068
SBB_{XY1}	0.044	0.052	0.060	0.053	0.040	0.057	0.072	0.096	0.062	0.054	0.055	0.082	0.152	0.097	0.051
SBB_{XY2}	0.042	0.050	0.057	0.048	0.038	0.055	0.070	0.092	0.057	0.051	0.054	0.084	0.168	0.116	0.046
SBB_{XY3}	0.041	0.052	0.057	0.048	0.036	0.053	0.074	0.096	0.057	0.048	0.053	0.092	0.174	0.134	0.044
SBB_{Y1}	0.064	0.069	0.066	0.091	0.055	0.087	0.095	0.113	0.122	0.076	0.083	0.109	0.193	0.235	0.073
SBB_{Y2}	0.056	0.060	0.060	0.079	0.046	0.076	0.085	0.102	0.101	0.066	0.072	0.105	0.195	0.243	0.058
SBB_{Y3}	0.052	0.057	0.059	0.069	0.040	0.071	0.087	0.101	0.088	0.061	0.072	0.106	0.188	0.240	0.058

Notes: RMSEs based on Monte Carlo simulations with 1000 replications. The remaining parameters are: $N = 60$, $T = 5$, $\lambda = 0.3$, $\delta = 0.5$, and $\beta = 1$. Bucky refers to the Bucky ball weight matrix and n denotes the number of neighbors in the random contiguity specifications. The labels AB, SAB, BB, and SBB denote the first-stage spatial Arellano-Bond estimator, the spatially corrected Arellano-Bond estimator, the first-stage spatial Blundell-Bond estimator, and the spatially corrected Blundell-Bond estimator, respectively. The subscripts X and Y refer to instrument sets for the spatial lag based on spatially weighted values of $\mathbf{X}_N$ and $\mathbf{y}_N$, respectively. The subscript XY indicates that both instrument sets are employed. The numbers in the subscripts report the time lags and spatial lags of the spatially lagged dependent variable used for instrumenting the spatial lag. The one-period time lag of the dependent variable is instrumented by three time lags of the dependent variable.

Appendix

A.1 Derivation of Moment Conditions in Stage Two

To arrive at the moment conditions in (20), we define the spatially transformed counterpart of $\Delta\boldsymbol{\varepsilon}_N$ by $\Delta\bar{\boldsymbol{\varepsilon}} = (\mathbf{I}_{T-1} \otimes \mathbf{M}_N)\Delta\boldsymbol{\varepsilon}_N$. We make use of the following properties of the error term:

$$\Delta\boldsymbol{\varepsilon}_N = \Delta\mathbf{v}_N, \qquad \mathrm{E}[\mathbf{v}_N\mathbf{v}_N^\top] = \sigma_v^2\mathbf{I}_{N(T-1)}, \tag{A.1}$$

which follows from Assumption E. In addition, we apply $\mathrm{E}[\mathbf{v}_N^\top A\mathbf{v}_N] = \mathrm{tr}(A\,\mathrm{E}\,\mathbf{v}_N\mathbf{v}_N^\top)$, where A is a conformable matrix. Finally, we use:

$$\mathrm{tr}[\mathbf{I}_{T-1} \otimes (\mathbf{M}_N^\top\mathbf{M}_N)] = (T-1)\,\mathrm{tr}(\mathbf{M}_N^\top\mathbf{M}_N), \quad \mathrm{tr}(\mathbf{M}_N^\top) = 0. \tag{A.2}$$

Using the above leads to the following moment conditions:

$$\begin{aligned}
\mathrm{E}[\Delta\boldsymbol{\varepsilon}_N^\top\Delta\boldsymbol{\varepsilon}_N] &= \mathrm{E}[\Delta\mathbf{v}_N^\top\Delta\mathbf{v}_N] \\
&= 2\sigma_v^2\,\mathrm{tr}(\mathbf{I}_{N(T-1)}) = 2\sigma_v^2 N(T-1), & \text{(A.3)}\\
\mathrm{E}[\Delta\bar{\boldsymbol{\varepsilon}}_N^\top\Delta\bar{\boldsymbol{\varepsilon}}_N] &= \mathrm{E}[\Delta\mathbf{v}_N^\top(\mathbf{I}_{T-1} \otimes \mathbf{M}_N^\top\mathbf{M}_N)\Delta\mathbf{v}_N] \\
&= 2\sigma_v^2\,\mathrm{tr}(\mathbf{I}_{T-1}\mathbf{M}_N^\top\mathbf{M}_N) = 2\sigma_v^2(T-1)\,\mathrm{tr}(\mathbf{M}_N^\top\mathbf{M}_N), & \text{(A.4)}\\
\mathrm{E}[\Delta\bar{\boldsymbol{\varepsilon}}_N^\top\Delta\boldsymbol{\varepsilon}_N] &= \mathrm{E}[\Delta\mathbf{v}_N^\top(\mathbf{I}_{T-1} \otimes \mathbf{M}_N^\top)\Delta\mathbf{v}_N] \\
&= 2\sigma_v^2\,\mathrm{tr}(\mathbf{I}_{T-1}\mathbf{M}_N^\top) = 2\sigma_v^2(T-1)\,\mathrm{tr}(\mathbf{M}_N^\top) = 0. & \text{(A.5)}
\end{aligned}$$

Dividing (A.3)–(A.5) by $N(T-1)$ gives the moment conditions in (20).

A.2 Proofs of Asymptotic Properties

A.2.1 Proof of Theorem 1

Definition (18) and models (8) and (16) imply

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_N &= \left[\vec{\mathbf{Z}}_N^\top\vec{\mathbf{H}}_N\vec{\mathbf{A}}_N\vec{\mathbf{H}}_N^\top\vec{\mathbf{Z}}_N\right]^{-1}\vec{\mathbf{Z}}_N^\top\vec{\mathbf{H}}_N\vec{\mathbf{A}}_N\vec{\mathbf{H}}_N^\top\vec{\mathbf{y}}_N \\
&= \boldsymbol{\theta}^0 + \left[\vec{\mathbf{Z}}_N^\top\vec{\mathbf{H}}_N\vec{\mathbf{A}}_N\vec{\mathbf{H}}_N^\top\vec{\mathbf{Z}}_N\right]^{-1}\vec{\mathbf{Z}}_N^\top\vec{\mathbf{H}}_N\vec{\mathbf{A}}_N\vec{\mathbf{H}}_N^\top\vec{\mathbf{u}}_N.
\end{aligned}$$

Since Assumptions G and V imply $\vec{\mathbf{A}}_N = \vec{\mathbf{A}} + o_p(1)$ and $N^{-1}\vec{\mathbf{H}}_N^\top\vec{\mathbf{Z}}_N = \vec{\mathbf{Q}}_{HZ} + o_p(1)$ and $\vec{\mathbf{Z}}_N^\top\vec{\mathbf{H}}_N\vec{\mathbf{A}}_N\vec{\mathbf{H}}_N^\top\vec{\mathbf{Z}}_N^\top$ is non-singular, it follows from definition (4) that

$$\begin{aligned}\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) &= \left[\frac{1}{N}\vec{\mathbf{Z}}_N^\top\vec{\mathbf{H}}_N\cdot\vec{\mathbf{A}}_N\cdot\frac{1}{N}\vec{\mathbf{H}}_N^\top\vec{\mathbf{Z}}_N\right]^{-1}\frac{1}{N}\vec{\mathbf{Z}}_N^\top\vec{\mathbf{H}}_N\cdot\vec{\mathbf{A}}_N\cdot\frac{1}{\sqrt{N}}\vec{\mathbf{H}}_N^\top\vec{\mathbf{u}}_N \\ &= \left[\vec{\mathbf{Q}}_{HZ}^\top\vec{\mathbf{A}}\vec{\mathbf{Q}}_{HZ}\right]^{-1}\vec{\mathbf{Q}}_{HZ}^\top\vec{\mathbf{A}}\cdot N^{-1/2}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1}\vec{\boldsymbol{\varepsilon}}_N + o_p(1)\end{aligned}$$

as $N \to +\infty$. The triangular array $\boldsymbol{\xi}_N = N^{-1/2}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1}\vec{\boldsymbol{\varepsilon}}_N$ has zero mean, $\mathrm{E}[N^{-1/2}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1}\vec{\boldsymbol{\varepsilon}}_N] = \mathrm{E}[\vec{\mathbf{H}}_N\vec{\mathbf{u}}_N] = 0$, and a bounded variance matrix since for $N \to +\infty$

$$\mathrm{Var}[N^{-1/2}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1}\vec{\boldsymbol{\varepsilon}}_N] = \mathrm{E}[N^{-1}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1}\vec{\boldsymbol{\Sigma}}_{\varepsilon,N}(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1\top}\vec{\mathbf{H}}_N^\top]$$

$$= \mathrm{E}(N^{-1}\vec{\mathbf{H}}_N^\top\vec{\boldsymbol{\Sigma}}_{u,N}\vec{\mathbf{H}}_N) = \vec{\mathbf{Q}}_{N,H\Sigma H} \to \vec{\mathbf{Q}}_{H\Sigma H}$$

by Assumptions E and V. Since $\vec{\mathbf{Q}}_{N,H\Sigma H}^{-1/2}\boldsymbol{\xi}_N$ forms a triangular array of martingale differences, the finite second moments and uniform integrability of $\vec{\mathbf{Q}}_{N,H\Sigma H}^{-1/2}\boldsymbol{\xi}_N$ (implied by Assumption V3 and the uniform boundedness of $(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1}$ by Assumption S3) allows us to apply the central limit theorem for martingale differences (e.g., Davidson, 1994, Theorems 24.3 and 24.4), which results in the asymptotic normality of $\boldsymbol{\xi}_N$ with the finite asymptotic variance matrix $\vec{\mathbf{Q}}_{H\Sigma H}$. Consequently, $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) = O_p(1)$ and $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} N(0, [\vec{\mathbf{Q}}_{HZ}^\top\vec{\mathbf{A}}\vec{\mathbf{Q}}_{HZ}]^{-1}\vec{\mathbf{Q}}_{HZ}^\top\vec{\mathbf{A}}\vec{\mathbf{Q}}_{H\Sigma H}\vec{\mathbf{A}}^\top\vec{\mathbf{Q}}_{HZ}[\vec{\mathbf{Q}}_{HZ}^\top\vec{\mathbf{A}}\vec{\mathbf{Q}}_{HZ}]^{-1})$ as $N \to +\infty$, where $\mathcal{L}$ denotes convergence in distribution. □

A.2.2 Proof of Theorem 2

The proof is similar to the one of Kelejian and Prucha (2010, Theorem 1). First, the GMM estimator (23) is based on the vector $\boldsymbol{\gamma}_N$ and matrix $\boldsymbol{\Gamma}_N$ defined in (21)–(22). They both have each random element of the form $\Delta\mathbf{u}_N^\top\mathbf{D}_N\Delta\mathbf{u}_N/[N(T-1)]$, where $\mathbf{D}_N = \vec{\mathbf{M}}_N^{k\top}\vec{\mathbf{M}}_N^l$ for $k,l \in \{0,1,2\}$. To derive the limits of $\boldsymbol{\Gamma}_N$ and $\boldsymbol{\gamma}_N$ and also of $\hat{\boldsymbol{\Gamma}}_N$ and $\hat{\boldsymbol{\gamma}}_N$, we will now verify Assumption 4 of Kelejian and Prucha (2010, Lemma C.1) to apply it to $\boldsymbol{\Gamma}_N$ and $\boldsymbol{\gamma}_N$ (Assumptions 1–3 of Kelejian and Prucha, 2010, are implied by Assumptions E, S, and V). This Assumption 4 concerns the estimates $\Delta\hat{\mathbf{u}}_N$ of the error term $\Delta\mathbf{u}_N$, which is equal here to $\Delta\hat{\mathbf{u}}_N = \Delta\mathbf{y}_N - \Delta\mathbf{Z}_N\hat{\boldsymbol{\theta}}_N$. Hence,

$$\Delta\hat{\mathbf{u}}_N - \Delta\mathbf{u}_N = -\Delta\mathbf{Z}_N(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0)$$

and Assumption 4 of Kelejian and Prucha (2010, Lemma C.1) requires that $\Delta\mathbf{Z}_N$ has the uniformly bounded $(2+\psi)$th moments and that $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0)$ is bounded in probability. The first claim follows from Assumption V3 and the Minkowski inequality and the second claim is a consequence of the $\sqrt{N}$-consistence of the initial estimator $\hat{\boldsymbol{\theta}}_N$.

Next, for any $t = 2, \ldots, T$, $\Delta\mathbf{u}_N(t) = (\mathbf{I}_N - \rho^0\mathbf{M}_N)\Delta\boldsymbol{\varepsilon}_N(t)$, where $\Delta\boldsymbol{\varepsilon}_N(t)$ is a vector of independent random variables, and consequently, Lemma C.1(a) of Kelejian and Prucha (2010) can be applied to obtain the following results: $\mathrm{E}[\Delta\mathbf{u}_N^\top(t)\mathbf{D}_N\Delta\mathbf{u}_N(t)]/N$ is uniformly bounded, $\Delta\mathbf{u}_N^\top(t)\mathbf{D}_N\Delta\mathbf{u}_N(t)/N - \mathrm{E}[\Delta\mathbf{u}_N^\top(t)\mathbf{D}_N\Delta\mathbf{u}_N(t)]/N = o_p(1)$, and

$$\frac{1}{N}\Delta\hat{\mathbf{u}}_N^\top(t)\mathbf{D}_N\Delta\hat{\mathbf{u}}_N(t) - \frac{1}{N}\mathrm{E}[\Delta\mathbf{u}_N^\top(t)\mathbf{D}_N\Delta\mathbf{u}_N(t)] = o_p(1)$$

as $N \to +\infty$ for any matrix $\mathbf{D}_N$ with uniformly bounded rows and column sums such as $\mathbf{D}_N = \vec{\mathbf{M}}_N^{k\top}\vec{\mathbf{M}}_N^l$ for $k, l \in \{0, 1, 2\}$. Since $\Delta\mathbf{u}_N = [\Delta\mathbf{u}_N^\top(2), \ldots, \Delta\mathbf{u}_N^\top(T)]^\top$, we proved that $\mathrm{E}\{\Delta\mathbf{u}_N^\top\mathbf{D}_N\Delta\mathbf{u}_N/[N(T-1)]\} = O(1)$ and $\Delta\hat{\mathbf{u}}_N^\top\mathbf{D}_N\Delta\hat{\mathbf{u}}_N/[N(T-1)] - \mathrm{E}\{\Delta\mathbf{u}_N^\top\mathbf{D}_N\Delta\mathbf{u}_N/[N(T-1)]\} = o_p(1)$, and consequently, that $\mathrm{E}\,\boldsymbol{\Gamma}_N$ and $\mathrm{E}\,\boldsymbol{\gamma}_N$ are uniformly bounded and $\boldsymbol{\Gamma}_N - \mathrm{E}\,\boldsymbol{\Gamma}_N = o_p(1)$, $\boldsymbol{\gamma}_N - \mathrm{E}\,\boldsymbol{\gamma}_N = o_p(1)$, $\hat{\boldsymbol{\Gamma}}_N - \mathrm{E}\,\boldsymbol{\Gamma}_N = o_p(1)$, and $\hat{\boldsymbol{\gamma}}_N - \mathrm{E}\,\boldsymbol{\gamma}_N = o_p(1)$. Moreover, due to Assumption G5, $\boldsymbol{\Gamma}_N^\top\boldsymbol{\Gamma}_N$ is non-singular; similarly, Assumption G6 implies that also $\boldsymbol{\Gamma}_N^\top\mathbf{B}_N\boldsymbol{\Gamma}_N$ is non-singular and thus positive definite.

To prove the consistency of the GMM estimator (23), we can use a general result of Pötscher and Prucha (1997, Lemma 3.1), which states that the GMM estimator is consistent if (i) it exists, (ii) the minimum of $J_N(\boldsymbol{\phi}) = \{\mathrm{E}\,\boldsymbol{\gamma}_N - \mathrm{E}\,\boldsymbol{\Gamma}_N\nu(\boldsymbol{\phi})\}^\top\mathbf{B}_N\{\mathrm{E}\,\boldsymbol{\gamma}_N - \mathrm{E}\,\boldsymbol{\Gamma}_N\nu(\boldsymbol{\phi})\}$ at $\boldsymbol{\phi}^0$ is identifiably unique, and (iii) the sample objective function $\hat{J}_N(\boldsymbol{\phi}) = \{\hat{\boldsymbol{\gamma}}_N - \hat{\boldsymbol{\Gamma}}_N\nu(\boldsymbol{\phi})\}^\top\hat{\mathbf{B}}_N\{\hat{\boldsymbol{\gamma}}_N - \hat{\boldsymbol{\Gamma}}_N\nu(\boldsymbol{\phi})\}$ converges uniformly to $J_N(\boldsymbol{\phi})$, where $\nu(\boldsymbol{\phi}) = (\rho, \rho^2, \sigma_v^2)^\top$, $\boldsymbol{\phi} = (\rho, \sigma_v)^\top$, and $\boldsymbol{\phi}^0 = (\rho^0, \sigma_v^0)^\top$. First, the existence of the GMM estimate follows from the continuity of $\hat{J}_N(\boldsymbol{\phi})$: it is continuous in $\boldsymbol{\phi}$ on a compact space Φ and it thus attains its minimum.

Regarding the identification, the objective function $J_N(\boldsymbol{\phi})$ attains its minimum only at $\boldsymbol{\phi}^0 = (\rho^0, \sigma_v^0)^\top$ because $\mathrm{E}\,\boldsymbol{\Gamma}_N \nu(\boldsymbol{\phi}^0) = \mathrm{E}\,\boldsymbol{\gamma}_N$, and by Assumption G,

$$\begin{aligned} J_N(\boldsymbol{\phi}) - J_N(\boldsymbol{\phi}^0) = J_N(\boldsymbol{\phi}) &= \{\nu(\boldsymbol{\phi}) - \nu(\boldsymbol{\phi}^0)\}^\top \mathrm{E}\,\boldsymbol{\Gamma}_N^\top \mathbf{B}_N \,\mathrm{E}\,\boldsymbol{\Gamma}_N \{\nu(\boldsymbol{\phi}) - \nu(\boldsymbol{\phi}^0)\} \\ &\geq \kappa_\Gamma \kappa_B \{\nu(\boldsymbol{\phi}) - \nu(\boldsymbol{\phi}^0)\}^\top \{\nu(\boldsymbol{\phi}) - \nu(\boldsymbol{\phi}^0)\} \\ &\geq \kappa_\Gamma \kappa_B \{(\rho - \rho^0)^2 + [\sigma_v^2 - (\sigma_v^0)^2]\}. \end{aligned}$$

Consequently, for any $\varepsilon > 0$ it holds $\inf_{\{(\rho,\sigma_v)\in\Phi: \|(\rho,\sigma_v)-(\rho^0,\sigma_v^0)\|>\varepsilon\}} J_N(\boldsymbol{\phi}) - J_N(\boldsymbol{\phi}^0) > \kappa_\Gamma \kappa_B \varepsilon^2 > 0$ and $\boldsymbol{\phi}^0 = (\rho^0, \sigma_v^0)^\top$ is identifiably unique.

Finally, $\hat{J}_N(\boldsymbol{\phi})$ can be shown to uniformly converge to $J_N(\boldsymbol{\phi})$ on Φ. Since

$$\begin{aligned} \hat{J}_N(\boldsymbol{\phi}) - J_N(\boldsymbol{\phi}) &= (\boldsymbol{\gamma}_N^\top \hat{\mathbf{B}}_N \boldsymbol{\gamma}_N - \mathrm{E}\,\boldsymbol{\gamma}_N^\top \mathbf{B}_N \,\mathrm{E}\,\boldsymbol{\gamma}_N) - 2(\hat{\boldsymbol{\gamma}}_N^\top \hat{\mathbf{B}}_N \hat{\boldsymbol{\Gamma}}_N - \mathrm{E}\,\boldsymbol{\gamma}_N^\top \mathbf{B}_N \,\mathrm{E}\,\boldsymbol{\Gamma}_N)\nu(\phi) \\ &+ \nu(\boldsymbol{\phi})^\top (\hat{\boldsymbol{\Gamma}}_N^\top \hat{\mathbf{B}}_N \hat{\boldsymbol{\Gamma}}_N - \mathrm{E}\,\boldsymbol{\Gamma}_N^\top \mathbf{B}_N \,\mathrm{E}\,\boldsymbol{\Gamma}_N)\nu(\phi), \end{aligned}$$

and $\boldsymbol{\phi} \in \Phi$, where Φ is compact, $\|\boldsymbol{\phi}\| < K_\phi < +\infty$, we only have to show that the three differences of the type $\hat{\boldsymbol{\Gamma}}_N^\top \hat{\mathbf{B}}_N \hat{\boldsymbol{\Gamma}}_N - \mathrm{E}\,\boldsymbol{\Gamma}_N^\top \mathbf{B}_N \,\mathrm{E}\,\boldsymbol{\Gamma}_N = o_p(1)$ as $N \to +\infty$. This however directly follows from $\boldsymbol{\Gamma}_N - \mathrm{E}\,\boldsymbol{\Gamma}_N = o_p(1)$, $\boldsymbol{\gamma}_N - \mathrm{E}\,\boldsymbol{\gamma}_N = o_p(1)$, $\hat{\boldsymbol{\Gamma}}_N - \mathrm{E}\,\boldsymbol{\Gamma}_N = o_p(1)$, and $\hat{\boldsymbol{\gamma}}_N - \mathrm{E}\,\boldsymbol{\gamma}_N = o_p(1)$ as all these random variables are bounded in probability (see Assumption G), the expectations $\mathrm{E}\,\boldsymbol{\Gamma}_N$ and $\mathrm{E}\,\boldsymbol{\gamma}_N$ were shown to be uniformly bounded, and

$$\begin{aligned} \hat{\boldsymbol{\Gamma}}_N^\top \hat{\mathbf{B}}_N \hat{\boldsymbol{\Gamma}}_N - \mathrm{E}\,\boldsymbol{\Gamma}_N^\top \mathbf{B}_N \,\mathrm{E}\,\boldsymbol{\Gamma}_N &= (\hat{\boldsymbol{\Gamma}}_N^\top - \mathrm{E}\,\boldsymbol{\Gamma}_N^\top)\hat{\mathbf{B}}_N \hat{\boldsymbol{\Gamma}}_N + \mathrm{E}\,\boldsymbol{\Gamma}_N^\top \hat{\mathbf{B}}_N (\hat{\boldsymbol{\Gamma}}_N - \mathrm{E}\,\boldsymbol{\Gamma}_N) \\ &+ \mathrm{E}\,\boldsymbol{\Gamma}_N^\top (\hat{\mathbf{B}}_N - \mathbf{B}_N)\,\mathrm{E}\,\boldsymbol{\Gamma}_N). \end{aligned}$$

Hence, Lemma 3.1 of Pötscher and Prucha (1997) implies consistency of the estimate (23). □

A.2.3 Proof of Theorem 3

Definition (18) and models (24) and (25) imply

$$\tilde{\boldsymbol{\theta}}_N = \left[\vec{\vec{\mathbf{Z}}}_N^\top \vec{\vec{\mathbf{H}}}_N \vec{\vec{\mathbf{A}}}_N \vec{\vec{\mathbf{H}}}_N^\top \vec{\vec{\mathbf{Z}}}_N\right]^{-1} \vec{\vec{\mathbf{Z}}}_N^\top \vec{\vec{\mathbf{H}}}_N \vec{\vec{\mathbf{A}}}_N \vec{\vec{\mathbf{H}}}_N^\top \vec{\vec{\mathbf{y}}}_N = \boldsymbol{\theta}^0 + \left[\vec{\vec{\mathbf{Z}}}_N^\top \vec{\vec{\mathbf{H}}}_N \vec{\vec{\mathbf{A}}}_N \vec{\vec{\mathbf{H}}}_N^\top \vec{\vec{\mathbf{Z}}}_N\right]^{-1} \vec{\vec{\mathbf{Z}}}_N^\top \vec{\vec{\mathbf{H}}}_N \vec{\vec{\mathbf{A}}}_N \vec{\vec{\mathbf{H}}}_N^\top \vec{\vec{\varepsilon}}_N,$$

where $\vec{\tilde{\varepsilon}}_N = \vec{\tilde{\mathbf{y}}}_N - \vec{\tilde{\mathbf{Z}}}_N\boldsymbol{\theta}^0 = (\vec{\mathbf{I}}_N - \hat{\rho}_N\vec{\mathbf{M}}_N)(\vec{\mathbf{y}}_N - \vec{\mathbf{Z}}_N\boldsymbol{\theta}^0)$. First, note that the consistency of $\hat{\rho}_N \to \rho^0$ and Assumption V imply

$$
\begin{aligned}
N^{-1}\vec{\tilde{\mathbf{H}}}_N^\top\vec{\tilde{\mathbf{Z}}}_N &= N^{-1}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \hat{\rho}_N\vec{\mathbf{M}}_N)^\top(\vec{\mathbf{I}}_N - \hat{\rho}_N\vec{\mathbf{M}}_N)\vec{\mathbf{Z}}_N \\
&= N^{-1}\vec{\mathbf{H}}_N^\top\vec{\mathbf{Z}}_N - \hat{\rho}_N N^{-1}\vec{\mathbf{H}}_N^\top\vec{\mathbf{M}}_N\vec{\mathbf{Z}}_N \\
&- \hat{\rho}_N N^{-1}\vec{\mathbf{H}}_N^\top\vec{\mathbf{M}}_N^\top\vec{\mathbf{Z}}_N + \hat{\rho}_N^2 N^{-1}\vec{\mathbf{H}}_N^\top\vec{\mathbf{M}}_N^\top\vec{\mathbf{M}}_N\vec{\mathbf{Z}}_N \\
&= \vec{\mathbf{Q}}_{HZ} - \rho^0(\vec{\mathbf{Q}}_{HMZ} + \vec{\mathbf{Q}}_{HM^\top Z}) + (\rho^0)^2\vec{\mathbf{Q}}_{HMMZ} + o_p(1).
\end{aligned}
$$

Matrix $N^{-1}\vec{\tilde{\mathbf{H}}}_N^\top\vec{\tilde{\mathbf{Z}}}_N$, which is non-singular by Assumptions S2, V, and G4, thus converges to a non-singular matrix $\vec{\tilde{\mathbf{Q}}}_{HZ}$ in probability. Assumptions G and V further imply that $\vec{\tilde{\mathbf{A}}}_N = \vec{\tilde{\mathbf{A}}} + o_p(1)$ and that $\vec{\tilde{\mathbf{Z}}}_N^\top\vec{\tilde{\mathbf{H}}}_N\tilde{\mathbf{A}}_N\vec{\tilde{\mathbf{H}}}_N^\top\vec{\tilde{\mathbf{Z}}}_N^\top$ is non-singular. Using definition (4), $\vec{\mathbf{u}}_N - \rho^0\vec{\mathbf{M}}_N\vec{\mathbf{u}}_N = \vec{\varepsilon}_N$, results in $\vec{\tilde{\varepsilon}}_N = (\vec{\mathbf{I}}_N - \hat{\rho}_N\vec{\mathbf{M}}_N)\vec{\mathbf{u}}_N = \vec{\mathbf{u}}_N - \hat{\rho}_N\vec{\mathbf{M}}_N\vec{\mathbf{u}}_N = \vec{\varepsilon}_N + (\rho^0 - \hat{\rho}_N)\vec{\mathbf{M}}_N\vec{\mathbf{u}}_N$. We can thus write

$$
\begin{aligned}
\sqrt{N}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) &= \left[\frac{1}{N}\vec{\tilde{\mathbf{Z}}}_N^\top\vec{\tilde{\mathbf{H}}}_N \cdot \vec{\tilde{\mathbf{A}}}_N \cdot \frac{1}{N}\vec{\tilde{\mathbf{H}}}_N^\top\vec{\tilde{\mathbf{Z}}}_N\right]^{-1}\frac{1}{N}\vec{\tilde{\mathbf{Z}}}_N\vec{\tilde{\mathbf{H}}}_N^\top\vec{\tilde{\mathbf{A}}}_N\frac{1}{\sqrt{N}}\vec{\tilde{\mathbf{H}}}_N^\top\vec{\varepsilon}_N \\
&- \left[\frac{1}{N}\vec{\tilde{\mathbf{Z}}}_N^\top\vec{\tilde{\mathbf{H}}}_N \cdot \vec{\tilde{\mathbf{A}}}_N \cdot \frac{1}{N}\vec{\tilde{\mathbf{H}}}_N^\top\vec{\tilde{\mathbf{Z}}}_N\right]^{-1}\frac{1}{N}\vec{\tilde{\mathbf{Z}}}_N^\top\vec{\tilde{\mathbf{H}}}_N\vec{\tilde{\mathbf{A}}}_N\frac{1}{\sqrt{N}}\vec{\tilde{\mathbf{H}}}_N^\top(\hat{\rho}_N - \rho^0)\vec{\mathbf{M}}_N\vec{\mathbf{u}}_N \\
&= \left[\vec{\tilde{\mathbf{Q}}}_{HZ}^\top\vec{\tilde{\mathbf{A}}}\vec{\tilde{\mathbf{Q}}}_{HZ}\right]^{-1}\vec{\tilde{\mathbf{Q}}}_{HZ}^\top\vec{\tilde{\mathbf{A}}} \cdot \frac{1}{\sqrt{N}}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^\top\vec{\varepsilon}_N + o_p(1) &\text{(A.6)} \\
&- \left[\vec{\tilde{\mathbf{Q}}}_{HZ}^\top\vec{\tilde{\mathbf{A}}}\vec{\tilde{\mathbf{Q}}}_{HZ}\right]^{-1}\vec{\tilde{\mathbf{Q}}}_{HZ}^\top\vec{\tilde{\mathbf{A}}} \cdot \frac{1}{\sqrt{N}}\vec{\mathbf{H}}_N^\top(\hat{\rho}_N - \rho^0)\vec{\mathbf{M}}_N^\top\vec{\varepsilon}_N + o_p(1) &\text{(A.7)} \\
&- \left[\vec{\tilde{\mathbf{Q}}}_{HZ}^\top\vec{\tilde{\mathbf{A}}}\vec{\tilde{\mathbf{Q}}}_{HZ}\right]^{-1}\vec{\tilde{\mathbf{Q}}}_{HZ}^\top\vec{\tilde{\mathbf{A}}} \\
&\cdot \frac{1}{\sqrt{N}}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \hat{\rho}_N\vec{\mathbf{M}}_N)^\top(\hat{\rho}_N - \rho^0)\vec{\mathbf{M}}_N\vec{\mathbf{u}}_N + o_p(1) &\text{(A.8)}
\end{aligned}
$$

1. Let us again consider the triangular array $\boldsymbol{\xi}_N = N^{-1/2}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^\top\vec{\varepsilon}_N$ in (A.6), which has zero mean $\mathrm{E}\{\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^\top\vec{\varepsilon}_N\} = 0$ and bounded variance since

$$
\begin{aligned}
\mathrm{Var}[N^{-1/2}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^\top\vec{\varepsilon}_N] &= \mathrm{E}[N^{-1}\vec{\mathbf{H}}_N^\top(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^\top\vec{\boldsymbol{\Sigma}}_{\varepsilon,N}(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)\vec{\mathbf{H}}_N] \\
&= \vec{\mathbf{Q}}_{N,HEH} \to \vec{\mathbf{Q}}_{HEH}
\end{aligned}
$$

as $N \to +\infty$ by Assumptions E and V. Since $\vec{\mathbf{Q}}_{N,HEH}^{-1/2}\boldsymbol{\xi}_N$ forms a triangular array of martingale differences, the finite second moments and uniform integrability (implied by Assumption V3 and the uniform boundedness of $(\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^\top$ by Assumptions S3 and G4) allows us to apply the central limit theorem for martingale differences (e.g., Davidson, 1994, Theorems 24.3 and 24.4), which results in the asymptotic normality of $\boldsymbol{\xi}_N$ with the finite asymptotic variance matrix $\vec{\mathbf{Q}}_{HEH}$.

2. Now, we only have to show that the remaining terms in (A.6)–(A.8) are negligible in probability (knowing that $\hat{\rho}_N - \rho^0 = o_p(1)$ as $N \to +\infty$). For $N \to +\infty$, the first term

$$(\hat{\rho}_N - \rho^0) \cdot \left[\vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \vec{\tilde{\mathbf{Q}}}_{HZ}\right]^{-1} \vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \cdot N^{-1/2} \vec{\mathbf{H}}_N^\top \vec{\mathbf{M}}_N^\top \vec{\boldsymbol{\varepsilon}}_N = o_p(1)$$

because $\hat{\rho}_N - \rho^0 = o_p(1)$ and the second part of the product is asymptotically normal (i.e., bounded in probability) by the same argument as in point 1 (see Assumption V).

The same argument can be used also for the last term (A.8) after rewriting it as

$$\begin{aligned}
&(\hat{\rho}_N - \rho^0) \cdot \left[\vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \vec{\tilde{\mathbf{Q}}}_{HZ}\right]^{-1} \vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \cdot N^{-1/2} \vec{\mathbf{H}}_N^\top (\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^\top \vec{\mathbf{M}}_N (\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1} \vec{\boldsymbol{\varepsilon}}_N \\
-&(\hat{\rho}_N - \rho^0)^2 \cdot \left[\vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \vec{\tilde{\mathbf{Q}}}_{HZ}\right]^{-1} \vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \cdot N^{-1/2} \vec{\mathbf{H}}_N^\top \vec{\mathbf{M}}_N^\top \vec{\mathbf{M}}_N (\vec{\mathbf{I}}_N - \rho^0\vec{\mathbf{M}}_N)^{-1} \vec{\boldsymbol{\varepsilon}}_N;
\end{aligned}$$

that is, each element of the sum is a product of an asymptotically normal random variable and a random variable negligible in probability as $N \to +\infty$ and behaves thus as $o_p(1)$.

3. Because we proved $\sqrt{N}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) = [\vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \vec{\tilde{\mathbf{Q}}}_{HZ}]^{-1} \vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \cdot \boldsymbol{\xi}_N + o_p(1)$, $\sqrt{N}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0)$ is asymptotically normally distributed with a zero mean and finite asymptotic variance matrix

$$\mathbf{V}_{SGMM} = \left[\vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \vec{\tilde{\mathbf{Q}}}_{HZ}\right]^{-1} \vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \vec{\mathbf{Q}}_{HEH} \vec{\tilde{\mathbf{A}}} \vec{\tilde{\mathbf{Q}}}_{HZ} \left[\vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\tilde{\mathbf{A}}} \vec{\tilde{\mathbf{Q}}}_{HZ}\right]^{-1}.$$

For the weighting matrix $\vec{\tilde{\mathbf{A}}} = [\vec{\mathbf{Q}}_{HEH}]^{-1}$, this clearly reduces to $[\vec{\tilde{\mathbf{Q}}}_{HZ}^\top \vec{\mathbf{Q}}_{HEH}^{-1} \vec{\tilde{\mathbf{Q}}}_{HZ}]^{-1}$. □

References

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

Anderson, T., and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Anselin, L. (1988): *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht.

——— (2006): "Spatial Econometrics," in *Palgrave Handbook of Econometrics, Vol. 1*, ed. by T. C. Mills, and K. Patterson, pp. 901–969. Palgrave Macmillan, Basingstoke.

Anselin, L., J. L. Gallo, and H. Jayet (2008): "Spatial Panel Econometrics," in *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice*, ed. by L. Matyas, and P. Sevestre, pp. 627–662. Kluwer, Dordrecht.

Arellano, M. (2003): *Panel Data Econometrics*. Oxford University Press, Oxford, 1st edn.

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.

Badinger, H., W. G. Mueller, and G. Tondl (2004): "Regional Convergence in the European Union, 1985–1999: A Spatial Dynamic Panel Analysis," *Regional Studies*, 38, 241–253.

Baltagi, B. (2008): *Econometric Analysis of Panel Data*. John Wiley and Sons, West Sussex, 4th edn.

Baltagi, B. H., and G. Bresson (2011): "Maximum Likelihood Estimation and Lagrange Multiplier Tests for Panel Seemingly Unrelated Regressions with Spatial Lag and Spatial Errors: An Application to Hedonic Housing Prices in Paris," *Journal of Urban Economics*, 69, 24–42.

Baltagi, B. H., S. Song, B. Jung, and W. Koh (2007): "Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data," *Journal of Econometrics*, 40, 5–51.

Bartolini, D., and R. Santolini (2012): "Political Yardstick Competition among Italian Municipalities on Spending Decisions," *Annals of Regional Science*, forthcoming.

Blonigen, B. A., R. B. Davies, G. R. Waddell, and H. T. Naughton (2007): "FDI in Space: Spatial Autoregressive Relationships in Foreign Direct Investment," *European Economic Review*, 51, 1303–1325.

Blundell, R., and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

Brady, R. R. (2011): "Measuring the Diffusion of Housing Prices Across Space and Time," *Journal of Applied Econometrics*, 26, 213–231.

Brueckner, J. (2003): "Strategic Interaction Among Governments: An Overview of Theoretical Studies," *International Regional Science Review*, 26, 175–188.

Bun, M. J., 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.

Case, A. C., H. S. Rosen, and J. R. Hines (1993): "Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States," *Journal of Public Economics*, 52, 285–307.

Čížek, P., J. P. A. M. Jacobs, J. E. Ligthart, and H. Vrijburg (2011): "Web Appendix: GMM Estimation of Fixed Effects Dynamic Panel Data Models with Spatial Lag and Spatial Errors," mimeo, Tilburg University, Tilburg.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press, Oxford.

Egger, P., M. Pfaffermayr, and H. Winner (2005): "An Unbalanced Spatial Panel Data Approach to US State Tax Competition," *Economics Letters*, 88, 329–335.

Elhorst, J. P. (2005): "Unconditional Maximum Likelihood Estimation of Linear and Log-Linear Dynamic Models for Spatial Panels," *Geographic Analysis*, 37, 85–106.

——— (2008): "Serial and Spatial Error Correlation," *Economics Letters*, 100, 422–424.

——— (2010): "Estimation of Dynamic Panels with Endogenous Interaction Effects When T is Small," *Regional Science and Urban Economics*, 40, 272–282.

——— (2011): "Dynamic Spatial Panels: Models, Methods, and Inferences," *Journal of Geographical Systems*, 42, 338–355.

Foucault, M., T. Madies, and S. Paty (2008): "Public Spending Interactions and Local Politics: Empirical Evidence from French Municipalities," *Public Choice*, 137, 57–80.

Hsiao, C., M. H. Pesaran, and A. Tahmiscioglu (2002): "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics*, 109, 107–150.

Jacobs, J. P. A. M., J. E. Ligthart, and H. Vrijburg (2009): "Dynamic Panel Data Models Featuring Endogenous Interaction and Spatially Correlated Errors," CentER Discussion Paper, No. 2009-92, Tilburg University, Tilburg.

——— (2010): "Consumption Tax Competition Among Governments: Evidence from the United States," *International Tax and Public Finance*, 17, 271–294.

Kapoor, M., H. H. Kelejian, and I. R. Prucha (2007): "Panel Data Models with Spatially Correlated Error Components," *Journal of Econometrics*, 140, 97–130.

Kelejian, H. H., and I. R. Prucha (2010): "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances," *Journal of Econometrics*, 157, 53–67.

KELEJIAN, H. H., AND D. P. ROBINSON (1993): "A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a County Expenditure Model," *Papers in Regional Science*, 72, 297–312.

KIVIET, J. (2007): "Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models," in *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, ed. by G. D. A. Phillips, and E. Tzavalis, pp. 282–318. Cambridge University Press, Cambridge.

KUKENOVA, M., AND J.-A. MONTEIRO (2009): "Spatial Dynamic Panel Model and System GMM: A Monte Carlo Investigation," mimeo, University of Lausanne.

LEE, L.-F., AND X. LIU (2010): "Efficient GMM Estimation of High Order Spatial Autoregressive Models with Autoregressive Disturbances," *Econometric Theory*, 26, 187–230.

LEE, L.-F., AND J. YU (2010a): "Some Recent Developments in Spatial Panel Data Models," *Regional Science and Urban Economics*, 154, 165–185.

——— (2010c): "Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects," *Journal of Econometrics*, 154, 165–185.

LESAGE, J., AND R. K. PACE (2009): *Introduction to Spatial Econometrics*. Taylor and Francis, London.

LIN, X., AND L.-F. LEE (2010): "GMM Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity," *Journal of Econometrics*, 157, 34–52.

LIU, X., L.-F. LEE, AND C. R. BOLLINGER (2010): "An Efficient GMM Estimator of Spatial Autoregressive Models," *Journal of Econometrics*, 159, 303–319.

NICKELL, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

Pesaran, M. H., and E. Tosetti (2010): "Large Panels with Common Factors and Spatial Correlation," *Journal of Econometrics*, 161, 182–202.

Pötscher, B., and I. R. Prucha (1997): *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer Verlag, New York.

Su, L., and Z. Yang (2008): "QML Estimation of Dynamic Panel Data Models with Spatial Errors," mimeo, Peking University, Beijing.

Wilson, J. D. (1999): "Theories of Tax Competition," *National Tax Journal*, 52, 269–304.

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.

Yu, J., and L.-F. Lee (2010): "Estimation of Unit Root Spatial Dynamic Panel Data Models," *Econometric Theory*, 26, 1332–1362.