

Tilburg University

## **GMM Estimation of Fixed Effects Dynamic Panel Data Models with Spatial Lag and Spatial Errors (Revised version of CentER DP 2011-134)**

Cizek, P.; Jacobs, J.; Ligthart, J.E.; Vrijburg, H.

*Publication date:*  
2015

*Document Version*  
Early version, also known as pre-print

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Cizek, P., Jacobs, J., Ligthart, J. E., & Vrijburg, H. (2015). *GMM Estimation of Fixed Effects Dynamic Panel Data Models with Spatial Lag and Spatial Errors (Revised version of CentER DP 2011-134)*. (CentER Discussion Paper; Vol. 2015-003). CentER, Center for Economic Research.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2014-003

**GMM ESTIMATION OF FIXED EFFECTS DYNAMIC  
PANEL DATA MODELS WITH SPATIAL LAG  
AND SPATIAL ERRORS**

By

Pavel Čížek, Jan P.A.M. Jacobs,  
Jenny Ligthart, Hendrik Vrijburg

20 January, 2015

This is a revised version of CentER Discussion Paper

No. 2011-134

ISSN 0924-7815  
ISSN 2213-9532

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

Pavel Čížek

CentER and Department of Econometrics and Operations Research, Tilburg University

Jan P.A.M. Jacobs<sup>1</sup>

University of Groningen, University of Tasmania, CAMA and CIRANO

Jenny Ligthart

CentER and Department of Economics, Tilburg University

Hendrik Vrijburg

Erasmus School of Economics, Erasmus University Rotterdam

This version: November 2014  
(A revised version of CentER  
DP2011-134 from December 2011)

<sup>1</sup>Corresponding author: Jan P.A.M. Jacobs, Faculty of Economics and Business, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands, Phone: +31-50-363-3681, Fax: +31-50-363-7337. [j.p.a.m.jacobs@rug.nl]

## Abstract

The three-step generalized methods of moments (GMM) approach of Kapoor, Kelejian and Prucha (2007), which corrects for spatially correlated errors in static panel data models, is extended by introducing fixed effects, a spatial lag, and a one-period lag of the dependent variable as additional explanatory variables. Combining this approach with the dynamic panel-data 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. The proposed spatially corrected GMM estimates are based on a spatial lag and a transformation correcting for the spatial error correlation. We prove their consistency and asymptotic normality for a large number of spatial units and a fixed number of time periods. Feasible spatial correction based on estimated spatial error correlation is shown to lead to estimators that are asymptotically equivalent to the infeasible estimators based on a known spatial error correlation. Monte Carlo simulations show that the root mean squared error of spatially corrected GMM estimates is generally smaller than that of corresponding spatial GMM estimates in which spatial error correlation is ignored.

**JEL codes:** C15, C21, C22, C23

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

# 1 Introduction

The fields of dynamic panel data models and spatial econometric models have matured rapidly and have reached (graduate) textbooks during the last decade.<sup>1</sup> Panel data may feature state dependence, i.e., the dependent variable is correlated over time, as well as display spatial dependence, i.e., 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.<sup>2</sup> For this model class, Elhorst (2005, 2008, 2010, 2014), 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, but it has not received much attention in the spatial econometrics literature. Lee and Liu (2010), Lin and Lee (2010), and Liu, Lee, and Bollinger (2010) study spatial GMM estimators for static panels. In a recent paper, Lee and Yu (2014) investigate efficient GMM estimation for spatial dynamic panel data with fixed effects.<sup>3</sup> 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 spatial lag and spatial errors. Contrary to Lee and Yu (2014), we consider only estimation based on the moment conditions linear in parameters, but allowing for spatially correlated errors.

Many economic interactions among agents can be characterized by a spatially lagged dependent variable or observations on the dependent variable in other locations than the 'home'

---

<sup>1</sup>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.

<sup>2</sup>Badinger et al. (2004), Foucault et al. (2008), Jacobs, Ligthart and Vrijburg (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.

<sup>3</sup>Using a Monte Carlo simulation study, Kukenova and Monteiro (2009) and Elhorst (2010) explore GMM in a spatial dynamic panel data framework. Kukenova and Monteiro (2009) analyze a spatial system GMM estimator and include an exogenous 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.

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, Rosen, and Hines, 1993). In the trade literature, foreign direct investment (FDI) inflows into the host country depend on FDI inflows into proximate host countries (cf. Blonigen, Davies, Waddell, and Naughton, 2007). The spatial lag structure allows explicit measurement of 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.<sup>4</sup> 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, Pfaffermayr, and Winner (2005) find evidence that spatial error dependence may exist above and beyond the theoretically motivated spatial lag structure.<sup>5</sup>

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. In dynamic panels with unobserved heterogeneity, Nickell (1981) shows that the standard least squares dummy variable estimator is biased and

---

<sup>4</sup>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.

<sup>5</sup>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.

inconsistent for large  $N$  and fixed small  $T$ .<sup>6</sup> The standard Arellano-Bond estimator is known to be inefficient 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 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 extends these estimators by 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), and standard instruments for dynamic panel data models.

As the structure of the spatial error correlation might not be known, we propose first estimators that are robust to the misspecification of the spatial error correlation and do not even require its knowledge. If the spatial error structure can be assumed or is known, the estimators can take it into account and correct 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.<sup>7</sup> If a spatial GMM estimator corrects for spatial error correlation, we speak of ‘spatially corrected’ GMM estimators. Recently, Kapoor, Kelejian, and Prucha (2007) designed a GMM procedure to deal with spatial error correlation in static random effects 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. This is achieved by modifying 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  $T$ . Possible extensions are discussed as well, including the spatio-temporal model and the use of pre-determined variables.<sup>8</sup>

---

<sup>6</sup>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.

<sup>7</sup>Anselin, Gallo, and Gayet (2008) call this model class a ‘time-space simultaneous model.’

<sup>8</sup>Yu, De Jong, and Lee (2008) and Pesaran and Tosetti (2011) study the properties of ML estimators

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, 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 a strong positive spatial error correlation. Moreover, the RMSE of the spatially corrected GMM estimates is not much affected by the size of the spatial lag parameter. We also find that the spatial Blundell-Bond estimators outperform the spatial Arellano-Bond estimators, which are however consistent under more general assumptions. 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, whereas Section 5 discusses possible extensions of the main model. Section 6 presents Monte Carlo simulation outcomes. Finally, Section 7 concludes. The proofs are in the Appendix.

## 2 The Spatial Dynamic Panel Data Model

Consider a panel with  $i = 1, \dots, N$  spatial units and  $t = 1, \dots, T$  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:

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

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,  $\mathbf{X}_N(t)$

---

in the context of dynamic, possibly nonstationary, panels with fixed effects and spatial error correlation, assuming both  $N$  and  $T$  large.



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. 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,  $\mathbf{y}_N = [\mathbf{y}_N^\top(1), \dots, \mathbf{y}_N^\top(T)]^\top$  or  $\mathbf{X}_N = [\mathbf{X}_N^\top(1), \dots, \mathbf{X}_N^\top(T)]^\top$ , where  $\top$  denotes a transpose. Further, the scalar parameter  $\lambda$  is the coefficient of the lagged dependent variable,  $\delta$  is the spatial autoregressive coefficient, which 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). Nevertheless, the row normalization is standard in spatial applications, and therefore, we use it in the simulations of Section 6.

The reduced form of equation (1) amounts to:

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

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

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

where  $\omega_L$  and  $\omega_U$  denote the smallest (i.e., the most negative) and largest characteristic roots of  $\mathbf{W}_N$ , respectively (cf. Elhorst, 2008). If  $\mathbf{W}_N$  is row normalized,  $\omega_U = 1$ .<sup>9</sup> Equation (3) yields a

---

<sup>9</sup>No general results hold for the smallest characteristic root of the matrix of spatial weights. The lower bound  $\omega_L$  is typically less than  $-1$ ; see Elhorst (2008, p. 422).

tradeoff between the size of  $\lambda$  and  $\delta$ .

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

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

where  $\mathbf{M}_N \mathbf{u}_N(t)$  is the spatial error term,  $\rho$  is a (second) spatially autoregressive coefficient, and  $\boldsymbol{\varepsilon}_N(t)$  denotes the vector of innovations.<sup>10</sup> The interpretation of the ‘nuisance’ parameter  $\rho$  is very different from  $\delta$  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  $\delta$  and  $\rho$  in the absence of exogenous variables and a dynamic lag. The spatial error process in the 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 and thus yields feedback effects that are bounded.

The vector of innovations is defined as:

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

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  $\sigma_v^2$ , 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.

---

<sup>10</sup>As  $\boldsymbol{\varepsilon}_N(t)$  contains the individual effects, see (5),  $\mathbf{u}_N(t)$  defined in (4) contains a weighted average of individual specific effects. This is closely related to the common correlated effects of Pesaran and Tosetti (2010).

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

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

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

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 and  $\boldsymbol{\theta} = [\lambda, \delta, \boldsymbol{\beta}^\top]^\top$  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  $\delta > 0$ , the model reduces to the familiar spatial lag model (also known as the mixed regressive-spatial autoregressive model; see Anselin, 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), who explicitly correct for spatial error correlation, to include both a time lag and a spatial lag of the dependent variable, and additionally, account for the fixed effects. We apply a panel GMM procedure and propose a set of suitable 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 two stages.

#### 3.1 Infeasible GMM estimators

To construct a GMM estimator for model (6)–(7), two transformations are necessary. The classical approach to the GMM estimation of the fixed-effects dynamic panel-data models relies on differencing that eliminates the time-invariant individual effects. The serial error correlation cre-

ated by such a transformation is taken into account by means of the GMM weighting matrix as discussed later.

In model (6)–(7), there is one additional type of error correlation: the spatial correlation of errors. As the moment conditions in dynamic panels have to be defined as cross-sectional averages, the spatial correlation cannot be accounted for by the GMM weighting matrix, which accommodates only the correlation among moment conditions. On the other hand, many estimation procedures typically perform best if errors are uncorrelated and homoscedastic. Since the spatial correlation cannot be taken into account by a GMM estimator based on cross-sectional averages, we base estimation on the transformed model:

$$\mathbf{B}_N \mathbf{y}_N(t) = \mathbf{B}_N \mathbf{Z}_N(t) \boldsymbol{\theta} + \mathbf{B}_N \mathbf{u}_N(t), \quad t = 2, \dots, T, \quad (8)$$

where  $\mathbf{B}_N$  is a non-singular  $N \times N$  transformation matrix. Denoting  $\tilde{\mathbf{y}}_N(t) = \mathbf{B}_N \mathbf{y}_N(t)$ ,  $\tilde{\mathbf{Z}}_N(t) = \mathbf{B}_N \mathbf{Z}_N(t)$ , and  $\tilde{\mathbf{u}}_N(t) = \mathbf{B}_N \mathbf{u}_N(t)$ , the model can be concisely rewritten as  $\tilde{\mathbf{y}}_N(t) = \tilde{\mathbf{Z}}_N(t) \boldsymbol{\theta} + \tilde{\mathbf{u}}_N(t)$ . Given error structure (7), the errors in model (8) become uncorrelated across individuals if  $\mathbf{B}_N = \mathbf{I}_N - \rho \mathbf{M}_N$ . For this particular choice  $\mathbf{B}_N$  or an estimate thereof, the model (and the corresponding estimators) will be called spatially corrected.

We will first construct two infeasible GMM estimators for the parameters of model (8) under the assumption that  $\mathbf{B}_N$  is known. The moment conditions of the proposed estimators will however not rely on the assumption of spatially uncorrelated errors as the matrix  $\mathbf{B}_N = \mathbf{I}_N - \rho \mathbf{M}_N$  is generally not known due to its dependence on the parameter  $\rho$  and will have to be estimated as discussed in Section 3.2. Hence, even the choice  $\mathbf{B}_N = \mathbf{I}_N$  representing no spatial correlation or the lack of knowledge about the spatial error structure is valid.

### 3.1.1 Arellano-Bond Estimator

To estimate  $\boldsymbol{\theta}$  in (8), we employ a GMM estimator defined by a set of linear moment conditions for the error term  $\tilde{\mathbf{u}}_N(t)$ . Later, equations identifying  $\boldsymbol{\theta}$  will be obtained by substituting for the error term from the model equation,  $\tilde{\mathbf{u}}_N(t) = \tilde{\mathbf{y}}_N(t) - \tilde{\mathbf{Z}}_N(t) \boldsymbol{\theta}$ .

First, to eliminate the unit-specific fixed effects  $\boldsymbol{\eta}_N$  contained in  $\tilde{\mathbf{u}}_N(t) = \mathbf{B}_N \mathbf{u}_N(t)$  due to (7), the first differences of (8) are taken analogously to Arellano and Bond (1991):

$$\Delta \tilde{\mathbf{y}}_N(t) = \Delta \tilde{\mathbf{Z}}_N(t) \boldsymbol{\theta} + \Delta \tilde{\mathbf{u}}_N(t), \quad t = 3, \dots, T, \quad (9)$$

where  $\Delta \mathbf{q}_N(t) \equiv \mathbf{q}_N(t) - \mathbf{q}_N(t-1)$  for  $\mathbf{q}_N(t) \in \{\tilde{\mathbf{y}}_N(t), \tilde{\mathbf{Z}}_N(t), \tilde{\mathbf{u}}_N(t), \boldsymbol{\varepsilon}_N(t), \mathbf{v}_N(t)\}$  and  $\Delta \tilde{\mathbf{u}}_N(t) = \mathbf{B}_N (\mathbf{I}_N - \rho \mathbf{M}_N)^{-1} \Delta \boldsymbol{\varepsilon}_N(t) = \mathbf{B}_N (\mathbf{I}_N - \rho \mathbf{M}_N)^{-1} \Delta \mathbf{v}_N(t)$  does not contain  $\boldsymbol{\eta}_N$  anymore. 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, these 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  $\Delta \tilde{\mathbf{u}}_N(t)$  for each  $t = 3, \dots, T$ . First, the moment conditions identifying the coefficients of the strictly exogenous variables are:

$$\mathbb{E}[\Delta \tilde{\mathbf{X}}_N^\top(t) \Delta \tilde{\mathbf{u}}_N(t)] = 0, \quad t = 3, \dots, T. \quad (10)$$

Next, Arellano and Bond (1991) propose to use the levels of the dependent variable,  $\tilde{\mathbf{y}}_N(t - 2), \dots, \tilde{\mathbf{y}}_N(1)$ , as instruments for the time lag of the dependent variable in first differences (i.e.,  $\Delta \tilde{\mathbf{y}}_N(t - 1)$ ). The instruments are correlated with the time lag of the dependent variable in first differences  $\Delta \tilde{\mathbf{y}}_N(t - 1)$ , but are uncorrelated with the ‘future’ error term in first differences,  $\Delta \tilde{\mathbf{u}}_N(t)$ , since the unit-specific effects are eliminated from the differenced variables. This property holds even in the spatial model defined by (8) and (7) because the spatial correlation applies only within a given time period  $t$ , and hence,  $\tilde{\mathbf{y}}_N(t - 2)$  is correlated with  $\tilde{\mathbf{u}}_N(t - 2), \dots, \tilde{\mathbf{u}}_N(1)$ , but cannot be correlated with  $\tilde{\mathbf{u}}_N(t)$  and  $\tilde{\mathbf{u}}_N(t - 1)$ . Consequently, we impose the following moment conditions to identify  $\lambda$ :

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

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

For the spatial lag, we consider two alternative sets 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  $\delta$ :

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

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). The expected value of the spatial lag  $\mathbf{W}_N \tilde{\mathbf{y}}_N(t)$  depends on the spatial lags of  $\mathbf{W}_N \tilde{\mathbf{X}}_N(t) \boldsymbol{\beta}$  (see the reduced form of (1)); the first differences of  $\mathbf{W}_N \tilde{\mathbf{y}}_N(t)$  are thus correlated with the differences of  $\mathbf{W}_N \tilde{\mathbf{X}}_N(t)$ , which are proposed as instruments:  $\mathbf{W}_N \Delta \tilde{\mathbf{X}}_N(t)$ . As the strictly exogenous variables  $\Delta \tilde{\mathbf{X}}_N(t)$  are not correlated with the error term  $\Delta \tilde{\mathbf{u}}_N(t)$ , the instruments satisfy the following moment conditions:<sup>11</sup>

$$E[\{\mathbf{W}_N^l \Delta \tilde{\mathbf{X}}_N(t)\}^\top \Delta \tilde{\mathbf{u}}_N(t)] = 0, \quad t = 3, \dots, T, \quad l = 1, \dots, L. \quad (13)$$

Note that the moment conditions specified for the spatial autoregressive parameter  $\delta$  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  $\lambda$  and  $\delta$  or large  $\beta$  imply stronger correlation of  $\mathbf{W}_N \tilde{\mathbf{y}}_N(t)$  with the instruments given in (12) or (13) for  $s \geq 1$  and  $l \geq 1$ , respectively.

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

---

<sup>11</sup>In these moment conditions, the averages can also be taken across time periods due to the strict exogeneity of  $\Delta \tilde{\mathbf{X}}_N(t)$ .

$\tilde{\mathbf{y}}_N(t) - \tilde{\mathbf{Z}}_N(t)\boldsymbol{\theta}$ , the proposed GMM estimator will minimize

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

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

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_{N,AB} &= \left[ \Delta \tilde{\mathbf{Z}}_N^\top \tilde{\mathbf{H}}_{N,AB} \mathbf{A}_{N,AB} \tilde{\mathbf{H}}_{N,AB}^\top \Delta \tilde{\mathbf{Z}}_N \right]^{-1} \Delta \tilde{\mathbf{Z}}_N^\top \tilde{\mathbf{H}}_{N,AB} \mathbf{A}_{N,AB} \tilde{\mathbf{H}}_{N,AB}^\top \Delta \tilde{\mathbf{y}}_N \\ &= \left[ \tilde{\mathbf{Z}}_{N,AB}^\top \tilde{\mathbf{H}}_{N,AB} \mathbf{A}_{N,AB} \tilde{\mathbf{H}}_{N,AB}^\top \tilde{\mathbf{Z}}_{N,AB} \right]^{-1} \tilde{\mathbf{Z}}_{N,AB}^\top \tilde{\mathbf{H}}_{N,AB} \mathbf{A}_{N,AB} \tilde{\mathbf{H}}_{N,AB}^\top \tilde{\mathbf{y}}_{N,AB}. \end{aligned} \quad (14)$$

The weighting matrix  $\mathbf{A}_{N,AB}$  recommended under the assumption of iid errors  $\tilde{\mathbf{u}}_N$  by Arellano and Bond (1991) is equal to the  $J \times J$  matrix  $\mathbf{A}_{N,AB} = [\tilde{\mathbf{H}}_{N,AB}^\top \mathbf{G}_{N,AB} \tilde{\mathbf{H}}_{N,AB} / N]^{-1}$ , where  $\mathbf{G}_{N,AB} = \mathbf{G} \otimes \mathbf{I}_N$  is an  $N(T-2) \times N(T-2)$  weighting matrix with elements  $(i, j = 1, \dots, T-2)$

$$G_{ij} = \begin{cases} 2 & \text{if } i = j \\ -1 & \text{if } i = j + 1 \\ -1 & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

and  $\otimes$  denotes the Kronecker product. In our model, this weighting matrix is the optimal GMM weighting matrix if the errors  $\mathbf{v}_N(t)$  are independently and identically distributed and if the transformation matrix  $\mathbf{B}_N = \mathbf{I}_N - \rho \mathbf{M}_N$  as discussed later. In the presence of heteroscedasticity or serial autocorrelation in  $\mathbf{v}_n$  or incorrect  $\mathbf{B}_N$ , the standard two-step GMM estimator can be employed, that is, one can re-estimate (14) using the optimal weighting matrix based on the initial estimate  $\hat{\boldsymbol{\theta}}_{N,AB}$ :  $\hat{\mathbf{A}}_{N,AB} = [\widehat{\text{Var}}\{\tilde{\mathbf{H}}_{N,AB}^\top (\tilde{\mathbf{y}}_{N,AB} - \tilde{\mathbf{Z}}_{N,AB} \hat{\boldsymbol{\theta}}_{N,AB})\}]^{-1}$ .

This baseline Arellano-Bond estimator can be further extended by using better instruments for the lagged dependent variable: the traditional ones defined in (11) do not perform well for

values of the autoregressive parameter close to 1. One possible improvement includes the lagged differences of the exogenous variables as instruments:

$$\mathbb{E}[\Delta \tilde{\mathbf{X}}_N^\top(t-1) \Delta \tilde{\mathbf{u}}_N(t)] = 0, \quad t = 3, \dots, T. \quad (16)$$

Another option discussed in the following section consists of the moment conditions suggested by Blundell and Bond (1998), which however require more restrictive assumptions on the fixed effects.

### 3.1.2 Blundell-Bond Estimator

The GMM approach of Blundell and Bond (1998)—often referred to as the system GMM estimator—can be used to extend the Arellano and Bond (1991) conditions by specifying moment conditions also for variables in levels rather than only for their first differences. This typically improves estimation for large values of the autoregressive coefficient, but it is possible only if the individual effects are not correlated with the differences of the response and explanatory variables: the individual effects are allowed to be correlated only with time-independent components of explanatory variables.<sup>12</sup>

The Blundell-Bond estimator for the spatially autoregressive dynamic panel model can be constructed by stacking equation (9) and

$$\tilde{\mathbf{y}}_N(t) = \tilde{\mathbf{Z}}_N(t) \boldsymbol{\theta} + \tilde{\mathbf{u}}_N(t), \quad t = 3, \dots, T. \quad (17)$$

The Blundell and Bond (1998) moment conditions for the level equation (17), which contains individual effects  $\boldsymbol{\eta}_N$  in  $\tilde{\mathbf{u}}_N(t)$ , 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:

$$\mathbb{E}[\Delta \tilde{\mathbf{X}}_N^\top(t) \tilde{\mathbf{u}}_N(t)] = 0, \quad t = 3, \dots, T, \quad (18)$$

---

<sup>12</sup>In this sense, the assumptions of Blundell and Bond (1998) are closer to the correlated random effects than to the fully general fixed effects approach.



which—in contrast to the Arellano-Bond estimator in Section 3.1.1—requires the individual effects to be independent of  $\Delta\tilde{\mathbf{X}}_N(t)$ . The equivalents of the instruments for both the time and spatially lagged dependent variables given in (11), (12), and (13) for model (9) can be analogously specified for model (17) as

$$E[\Delta\tilde{\mathbf{y}}_N^\top(t-s)\tilde{\mathbf{u}}_N(t)] = 0, \quad t = 3, \dots, T, \quad s = 1, \dots, t-2, \quad (19)$$

$$E[\{\mathbf{W}_N^l \Delta\tilde{\mathbf{y}}_N(t-s)\}^\top \tilde{\mathbf{u}}_N(t)] = 0, \quad t = 3, \dots, T, \quad s = 1, \dots, t-2, \quad l = 1, 2, \dots, L, \quad (20)$$

$$E[\{\mathbf{W}_N^l \Delta\tilde{\mathbf{X}}_N(t)\}^\top \tilde{\mathbf{u}}_N(t)] = 0, \quad t = 3, \dots, T, \quad l = 1, 2, \dots, L, \quad (21)$$

respectively.<sup>13</sup> All these moment conditions can be concisely written as  $E[\tilde{\mathbf{H}}_{N,LVL}^\top(t)\tilde{\mathbf{u}}_N(t)] = 0$ , where the columns of  $\tilde{\mathbf{H}}_{N,LVL}(t)$  represent the instruments  $\Delta\tilde{\mathbf{X}}_N(t)$ ,  $\Delta\tilde{\mathbf{y}}_N(t-s)$ ,  $\mathbf{W}_N^l \Delta\tilde{\mathbf{y}}_N(t-s)$ , and  $\mathbf{W}_N \Delta\tilde{\mathbf{X}}_N(t)$  given above (abbreviation LVL refers to the level equation).

Merging the information from all available time periods again, let  $\tilde{\mathbf{H}}_{N,LVL}$  be a block-diagonal matrix consisting of blocks  $\tilde{\mathbf{H}}_{N,LVL}(t)$  for  $t = 3, \dots, T$ ,  $\tilde{\mathbf{y}}_N = [\tilde{\mathbf{y}}_N^\top(3), \dots, \tilde{\mathbf{y}}_N^\top(T)]^\top$ , and  $\tilde{\mathbf{Z}}_N = [\tilde{\mathbf{Z}}_N^\top(3), \dots, \tilde{\mathbf{Z}}_N^\top(T)]^\top$ . These instruments for the level equation (17) are typically used jointly with the instruments introduced in Section 3.1.1 for the differenced equation (9). To define the Blundell-Bond estimator for the spatially autoregressive dynamic panel model, we thus define merged vectors and matrices for both systems: the vector of responses  $\tilde{\mathbf{y}}_{N,BB} = [\Delta\tilde{\mathbf{y}}_N^\top, \tilde{\mathbf{y}}_N^\top]^\top$ , the matrix of explanatory variables  $\tilde{\mathbf{Z}}_{N,BB} = [\Delta\tilde{\mathbf{Z}}_N^\top, \tilde{\mathbf{Z}}_N^\top]^\top$ , the vector of errors  $\tilde{\mathbf{u}}_{N,BB} = [\Delta\tilde{\mathbf{u}}_N^\top, \tilde{\mathbf{u}}_N^\top]^\top$ , the instruments  $\tilde{\mathbf{H}}_{N,BB} = \text{diag}\{\mathbf{H}_{N,AB}, \mathbf{H}_{N,LVL}\}$ , and the weighting matrices  $\mathbf{G}_{N,BB} = \text{diag}\{\mathbf{G}_{N,AB}, \mathbf{I}_{T-2} \otimes \mathbf{I}_N\}$  and  $\mathbf{A}_{N,BB} = [\tilde{\mathbf{H}}_{N,BB}^\top \mathbf{G}_{N,BB} \tilde{\mathbf{H}}_{N,BB} / N]^{-1}$ .<sup>14</sup> Minimizing

$$\begin{aligned} & \frac{1}{N} (\tilde{\mathbf{H}}_{N,BB}^\top \tilde{\mathbf{u}}_{N,BB})^\top \mathbf{A}_{N,BB} (\tilde{\mathbf{H}}_{N,BB}^\top \tilde{\mathbf{u}}_{N,BB}) \\ & = \frac{1}{N} \left[ \tilde{\mathbf{H}}_{N,BB}^\top (\tilde{\mathbf{y}}_{N,BB} - \tilde{\mathbf{Z}}_{N,BB} \boldsymbol{\theta}) \right]^\top \mathbf{A}_{N,BB} \left[ \tilde{\mathbf{H}}_{N,BB}^\top (\tilde{\mathbf{y}}_{N,BB} - \tilde{\mathbf{Z}}_{N,BB} \boldsymbol{\theta}) \right] \end{aligned}$$

<sup>13</sup>In the moment conditions with strictly exogenous variables, the averages can again be taken across time periods.

<sup>14</sup>Without prior knowledge of  $(\varepsilon_i, \eta_i)$  moments, an asymptotically optimal weighting matrix cannot be constructed in the first step (cf. Blundell and Bond, 1998). See Kiviet (2007) for alternatives.

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

$$\tilde{\boldsymbol{\theta}}_{N,BB} = \left( \tilde{\mathbf{Z}}_{N,BB}^\top \tilde{\mathbf{H}}_{N,BB} \mathbf{A}_{N,BB} \tilde{\mathbf{H}}_{N,BB}^\top \tilde{\mathbf{Z}}_{N,BB} \right)^{-1} \tilde{\mathbf{Z}}_{N,BB}^\top \tilde{\mathbf{H}}_{N,BB} \mathbf{A}_{N,BB} \tilde{\mathbf{H}}_{N,BB}^\top \tilde{\mathbf{y}}_{N,BB}. \quad (22)$$

Similarly to the estimator (14), the two-step GMM estimator can be used, where the optimal weight matrix is estimated by  $\hat{\mathbf{A}}_{N,BB} = [\widehat{\text{Var}}\{\tilde{\mathbf{H}}_{N,BB}^\top (\tilde{\mathbf{y}}_{N,BB} - \tilde{\mathbf{Z}}_{N,BB} \tilde{\boldsymbol{\theta}}_{N,BB})\}]^{-1}$ .

Given that the forms (14) and (22) are identical, we will use for the sake of simplicity only the notation  $\tilde{\boldsymbol{\theta}}_N, \tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N, \tilde{\mathbf{Z}}_N, \tilde{\mathbf{H}}_N$ , etc. from now on, representing the (infeasible) estimates and the vectors and matrices of responses, errors, covariates, instruments and so on used for estimation, both in the case of the spatial Arellano-Bond or Blundell-Bond estimators. The corresponding vectors and matrices used within the feasible estimation procedure in Section 3.2 will be denoted  $\hat{\boldsymbol{\theta}}_N, \hat{\mathbf{y}}_N, \hat{\mathbf{u}}_N, \hat{\mathbf{Z}}_N, \hat{\mathbf{H}}_N$ , and so on.

### 3.2 Feasible GMM estimators

The estimators (14) and (22) proposed in Section 3.1 are applicable for any transformation matrix  $\mathbf{B}_N$  (subject to some regularity conditions) as the moment conditions do not rely on the spatial uncorrelatedness of the error terms. The GMM estimation based on the transformation matrix  $\mathbf{B}_N = \mathbf{I}_N - \rho \mathbf{M}_N$  that completely eliminates the spatial correlation among errors is however infeasible as parameter  $\rho$  is unknown. To construct a feasible estimator, we have to proceed in three steps. First using a known  $\mathbf{B}_N$  such as  $\mathbf{B}_N = \mathbf{I}_N$ , apply (14) or (22) to obtain an initial estimator  $\hat{\boldsymbol{\theta}}_N^0$  of  $\boldsymbol{\theta}$  by using  $\tilde{\boldsymbol{\theta}}_N = \tilde{\boldsymbol{\theta}}_{N,AB}$  or  $\tilde{\boldsymbol{\theta}}_{N,BB}$ . Next, construct residuals  $\hat{\mathbf{u}}_N(t) = \mathbf{y}_N(t) - \mathbf{Z}_N(t) \hat{\boldsymbol{\theta}}_N^0$  and use them to consistently estimate  $\rho$  by  $\hat{\rho}_N$  (an estimator will be described in the following paragraphs). Then the estimate of the spatial transformation matrix  $\hat{\mathbf{B}}_N = \mathbf{I}_N - \hat{\rho}_N \mathbf{M}_N$  can be used to define transformed variables  $\hat{\mathbf{y}}_N(t) = \hat{\mathbf{B}}_N \mathbf{y}_N(t)$ ,  $\hat{\mathbf{Z}}_N(t) = \hat{\mathbf{B}}_N \mathbf{Z}_N(t)$ , and  $\hat{\mathbf{u}}_N(t) = \hat{\mathbf{B}}_N \mathbf{u}_N(t)$  and the transformed model  $\hat{\mathbf{y}}_N(t) = \hat{\mathbf{Z}}_N(t) \boldsymbol{\theta} + \hat{\mathbf{u}}_N(t)$ . Finally, the GMM estimators (14) or (22) are applied again to obtain the final estimate  $\hat{\boldsymbol{\theta}}_N$ —the spatially corrected Arellano-Bond or Blundell-Bond estimator, respectively:

$$\hat{\boldsymbol{\theta}}_N = \left( \hat{\mathbf{Z}}_N^\top \hat{\mathbf{H}}_N \hat{\mathbf{A}}_N \hat{\mathbf{H}}_N^\top \hat{\mathbf{Z}}_N \right)^{-1} \hat{\mathbf{Z}}_N^\top \hat{\mathbf{H}}_N \hat{\mathbf{A}}_N \hat{\mathbf{H}}_N^\top \hat{\mathbf{y}}_N, \quad (23)$$

where  $\hat{\mathbf{H}}_N$  represents the instrument matrix corresponding to the Arellano-Bond or Blundell-Bond estimators as constructed in Section 3.1 and  $\hat{\mathbf{A}}_N = [\hat{\mathbf{H}}_N^\top \mathbf{G}_N \hat{\mathbf{H}}_N / N]^{-1}$  is the feasible GMM weight matrix corresponding to the infeasible  $\mathbf{A}_N$  (thus using  $\hat{\mathbf{H}}_N$  instead of  $\tilde{\mathbf{H}}_N$ ). Again, (23) can be extended by estimating and applying the optimal GMM weight matrix.

To perform this three-step procedure consisting of the initial estimation, estimation of  $\rho$ , and the spatially corrected estimation, the estimates of the parameters describing the error distribution have to be derived. Having the initial estimate  $\hat{\boldsymbol{\theta}}_N^0$  of the regression coefficients, the parameters  $\rho$  and  $\sigma_v^2$  of the error distribution will now be estimated; recall from (7) and (5) that  $\mathbf{u}_N(t) = (\mathbf{I}_N - \rho \mathbf{M}_N)^{-1} [\boldsymbol{\eta}_N + \mathbf{v}_N(t)]$ , where  $\mathbf{v}_N(t) \sim \text{iid}(0, \sigma_v^2 \mathbf{I}_N)$ . To estimate  $\rho$  and  $\sigma_v^2$ , a GMM estimator is again constructed – this time based on errors  $\mathbf{u}_N(t)$ , which are estimated by the regression residuals  $\hat{\mathbf{u}}_N^0(t) = \mathbf{y}_N(t) - \mathbf{Z}_N(t) \hat{\boldsymbol{\theta}}_N^0$ . 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  $\rho$  and  $\sigma_v^2$  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)  $\Delta \boldsymbol{\varepsilon}_N = [\Delta \boldsymbol{\varepsilon}_N^\top(2), \dots, \Delta \boldsymbol{\varepsilon}_N^\top(T)]^\top$  and  $\Delta \mathbf{u}_N = [\Delta \mathbf{u}_N^\top(2), \dots, \Delta \mathbf{u}_N^\top(T)]^\top$  (the information from all the time periods is thus merged). 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

$$\Delta \boldsymbol{\varepsilon}_N \equiv \Delta \mathbf{u}_N - \rho \Delta \bar{\mathbf{u}}_N, \quad \Delta \bar{\boldsymbol{\varepsilon}}_N \equiv \Delta \bar{\mathbf{u}}_N - \rho \Delta \bar{\bar{\mathbf{u}}}_N. \quad (24)$$

The three equations identifying  $\rho$  and  $\sigma_v^2$  are as follows (see Appendix A.1 for a derivation):

$$\mathbb{E} \begin{bmatrix} \frac{1}{N(T-1)} \Delta \boldsymbol{\varepsilon}_N^\top \Delta \boldsymbol{\varepsilon}_N \\ \frac{1}{N(T-1)} \Delta \bar{\boldsymbol{\varepsilon}}_N^\top \Delta \bar{\boldsymbol{\varepsilon}}_N \\ \frac{1}{N(T-1)} \Delta \bar{\bar{\boldsymbol{\varepsilon}}}_N^\top \Delta \boldsymbol{\varepsilon}_N \end{bmatrix} = \begin{bmatrix} 2\sigma_v^2 \\ 2\sigma_v^2 \text{tr}(\mathbf{M}_N^\top \mathbf{M}_N) / N \\ 0 \end{bmatrix}, \quad (25)$$

where  $\text{tr}(\mathbf{M}_N^\top \mathbf{M}_N)$  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 (25) using  $\Delta \mathbf{u}_N$  and  $\Delta \bar{\mathbf{u}}_N$ , see (24), we obtain the following moment conditions:

$$\mathbb{E}[\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N(\rho, \rho^2, \sigma_v^2)^\top] = 0, \quad (26)$$

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

$$\mathbf{\Gamma}_N = \begin{bmatrix} \frac{2}{N(T-1)}\Delta\mathbf{u}_N^\top\Delta\bar{\mathbf{u}}_N & -\frac{1}{N(T-1)}\Delta\bar{\mathbf{u}}_N^\top\Delta\bar{\mathbf{u}}_N & 2 \\ \frac{2}{N(T-1)}\Delta\bar{\mathbf{u}}_N^\top\Delta\bar{\mathbf{u}}_N & -\frac{1}{N(T-1)}\Delta\bar{\mathbf{u}}_N^\top\Delta\bar{\mathbf{u}}_N & \frac{2}{N}\text{tr}(\mathbf{M}_N^\top\mathbf{M}_N) \\ \frac{2}{N(T-1)}[\Delta\bar{\mathbf{u}}_N^\top\Delta\bar{\mathbf{u}}_N + \Delta\bar{\mathbf{u}}_N^\top\Delta\mathbf{u}_N] & -\frac{1}{N(T-1)}\Delta\bar{\mathbf{u}}_N^\top\Delta\bar{\mathbf{u}}_N & 0 \end{bmatrix}. \quad (27)$$

The nonlinear system of equations (26) can be solved by GMM to obtain estimates of  $\rho$  and  $\sigma_v^2$ . Since the  $\Delta\mathbf{u}_N$ 's are not known, they have to be estimated by regression residuals  $\Delta\hat{\mathbf{u}}_N^0(t) = \Delta\mathbf{y}_N(t) - \Delta\mathbf{Z}_N(t)\hat{\boldsymbol{\theta}}_N^0$ , where  $\hat{\boldsymbol{\theta}}_N^0$  is an initial estimator obtained from (14) and (22) using the transformation matrix  $\mathbf{B}_N = \mathbf{I}_N$ , for instance. Denoting the analogs of  $\gamma_N$  and  $\mathbf{\Gamma}_N$  based on the regression residuals  $\Delta\hat{\mathbf{u}}_N^0$  by  $\hat{\gamma}_N$  and  $\hat{\mathbf{\Gamma}}_N$ , respectively, the GMM estimator of  $\rho$  and  $\sigma_v$  based on (26) is defined by

$$(\hat{\rho}_N, \hat{\sigma}_{v,N}) = \arg \min_{\rho, \sigma_v} [\hat{\gamma}_N - \hat{\mathbf{\Gamma}}_N(\rho, \rho^2, \sigma_v^2)^\top]^\top \mathbf{C}_N [\hat{\gamma}_N - \hat{\mathbf{\Gamma}}_N(\rho, \rho^2, \sigma_v^2)^\top], \quad (28)$$

where  $\mathbf{C}_N$  is a GMM weighting matrix. Initially and in Section 6, we use only  $\mathbf{C}_N = \mathbf{I}_3$ . The optimal weight matrix can be derived analogously to Kapoor et al. (2007) and has the same form as in that paper (see Appendix A.1 for verification). Its computation is cumbersome and we recommend, similarly to Kapoor et al. (2007), to use  $\mathbf{C}_N = \mathbf{I}_3$ . Furthermore, simulations in Section 6 indicate that replacing the optimal weighting matrix by the identity matrix hardly influences the precision of estimates.

## 4 Asymptotic Properties of the Estimators

To formulate the asymptotic results for the estimators  $\tilde{\boldsymbol{\theta}}_N$  and  $\hat{\boldsymbol{\theta}}_N^0$  [given in (14) or (22)],  $\hat{\rho}_N$  and  $\hat{\sigma}_{v,N}$  [given in (28)], and  $\hat{\boldsymbol{\theta}}_N$  [given in (23)], let  $\boldsymbol{\theta}^0$ ,  $\rho^0$ , and  $\sigma_v^0$  denote the true parameter values. Recall that  $\tilde{\boldsymbol{\theta}}_N$  and  $\hat{\boldsymbol{\theta}}_N$  can represent here the (spatially-corrected) spatial Arellano-Bond or Blundell-Bond estimator depending on which moment conditions are used. The feasible estimates depend on an estimate  $\hat{\mathbf{B}}_N$  of the  $N \times N$  spatial transformation matrix. To explicitly state that  $\mathbf{B}_N$  depends only on a finite set of parameters, let  $\mathbf{B}_N = \sum_{k=1}^{\varkappa} \phi_k \boldsymbol{\Phi}_{k,N}$ , where  $\phi_k$ 's are (possibly unknown) scalars and  $\boldsymbol{\Phi}_{k,N}$  are known  $N \times N$  matrices for  $k = 1, \dots, \varkappa$  with a

fixed  $\varkappa \in \mathbb{N}$ . Our main interest lies in  $\mathbf{B}_N = \mathbf{I}_N - \rho \mathbf{M}_N$ , which corresponds to  $\varkappa = 2$ ,  $\phi_1 = 1$ ,  $\Phi_{1,N} = \mathbf{I}_N$ ,  $\phi_2 = \rho$ , and  $\Phi_2 = \mathbf{M}_N$  and which contains one unknown parameter  $\rho$  to be estimated.

Further, an extended notation for the spatial matrices aggregated across all time periods is needed: in the case of the Arellano-Bond estimator, let  $\mathbf{I}_{N\otimes} = \mathbf{I}_{T-2} \otimes \mathbf{I}_N$ ,  $\mathbf{M}_{N\otimes} = \mathbf{I}_{T-2} \otimes \mathbf{M}_N$ , and  $\mathbf{W}_{N\otimes} = \mathbf{I}_{T-2} \otimes \mathbf{W}_N$ ; in the case of the Blundell-Bond estimator, let  $\mathbf{I}_{N\otimes} = \mathbf{I}_{2(T-2)} \otimes \mathbf{I}_N$ ,  $\mathbf{M}_{N\otimes} = \mathbf{I}_{2(T-2)} \otimes \mathbf{M}_N$ , and  $\hat{\mathbf{W}}_{N\otimes} = \mathbf{I}_{2(T-2)} \otimes \mathbf{W}_N$ . Similarly,  $\mathbf{B}_{N\otimes} = \mathbf{I}_{T-2} \otimes \mathbf{B}_N$  or  $\mathbf{B}_{N\otimes} = \mathbf{I}_{2(T-2)} \otimes \mathbf{B}_N$  and  $\Phi_{k,N\otimes} = \mathbf{I}_{T-2} \otimes \Phi_{k,N}$  or  $\Phi_{k,N\otimes} = \mathbf{I}_{2(T-2)} \otimes \Phi_{k,N}$  for  $k = 1, \dots, \varkappa$ .

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 spatially corrected GMM estimator are specified. Throughout the section, we assume  $N \rightarrow +\infty$  and  $T = c_0$ , where  $c_0$  is a constant. More specifically, the number of instruments is assumed to be fixed: this is guaranteed by  $T$  being finite for instruments lagged in time and by  $L$  in (13) and (21) being finite for spatially lagged instruments.

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 existence of the fourth moments is made for the convenience of using some auxiliary results of Kelejian and Prucha (2010). Similarly, the assumption of identically distributed errors is made for simplicity and could be relaxed. The only restrictive assumption on the individual effects follows from Blundell and Bond (1998), see Assumption E4 below, which is closer to the (correlated) random effect structure than to the fixed effects assumption and which applies only to the (spatial) Blundell-Bond estimator, but not to the Arellano-Bond estimator. Hence, Assumption E4 is not required for estimation as it is irrelevant in the case of the proposed spatial Arellano-Bond estimator. (The validity of Assumption E4 under various model assumptions is extensively discussed in Blundell, Bond, and

Windmeijer, 2000.) Finally, Assumption E also specifies a weak initial condition on the response variable: in Assumption E1, we state that the responses in the first time period,  $\mathbf{y}_N(1)$ , can be spatially correlated in a general way and that their spatial structure obeys similar rules as the model or errors themselves in that each element of  $\mathbf{y}_N(1)$  is a weighted sum of independent random variables, see equation (2); these random variables are denoted  $\mathbf{v}_N(1)$  since the model (2) and its errors  $\mathbf{v}_N(t)$  are defined only for  $t \geq 2$ .

### Assumption E

1. The initial value  $\mathbf{y}_N(1)$  can be written as  $\mathbf{y}_N(1) = \mathbf{S}_N \mathbf{v}_N(1)$ , where the elements of  $\mathbf{v}_N(1) = [v_1(1), \dots, v_N(1)]^\top$  are independent and identically distributed, have uniformly bounded second moments, and  $\mathbf{S}_N$  is a  $N \times N$  matrix.
2. The error vectors  $\mathbf{v}_N(t) = [v_1(t), \dots, v_N(t)]^\top$  are independent with independently and identically distributed elements for each  $N \in \mathbb{N}$  and  $t = 2, \dots, T$  with zero mean  $E[v_i(t)] = 0$ , a finite variance  $\text{Var}[v_i(t)] = \sigma_v^2$ ,  $i = 1, \dots, N$ , and uniformly bounded fourth moments. Further,  $\mathbf{v}_N(t)$  is assumed to be independent of  $\boldsymbol{\eta}_N$  and of  $\mathbf{X}_N(t)$ ,  $\mathbf{X}_N(s)$ , and  $\mathbf{y}_N(s)$  for any  $s = 1, \dots, t - 1$ ;  $t = 2, \dots, T$ .
3. The fixed effects  $\boldsymbol{\eta}_N = [\eta_1, \dots, \eta_N]^\top$  are mutually independent and have uniformly bounded fourth moments.
4. In the case of the Blundell-Bond estimator,  $\boldsymbol{\eta}_N$  is additionally assumed to be uncorrelated with  $\Delta \mathbf{Z}_N(t)$ :  $E[\Delta \mathbf{Z}_N(t) \eta_i] = 0$  for  $t = 2, \dots, T$ .

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—where we assumed row normalized matrices—by allowing various normalizations of spatial weight matrices (see Kelejian and Prucha, 2010). Similar assumptions are also applied to the transformation matrix  $\mathbf{B}_N$  and its decomposition based on  $\boldsymbol{\Phi}_{k,N}$ , which however automatically satisfy them if  $\mathbf{B}_N$  has a form  $\mathbf{B}_N = \mathbf{I}_N - \rho \mathbf{M}_N$  for some  $\rho$  (note that the initial choice  $\mathbf{B}_N = \mathbf{I}_N$  is included for  $\rho = 0$ ).

## Assumption S

1. All diagonal elements of  $\mathbf{W}_N$  and  $\mathbf{M}_N$  are zero; all diagonal elements of  $\mathbf{B}_N$  are equal to one.
2. There exist finite positive constants  $K'_\delta, K''_\delta, K'_\rho,$  and  $K''_\rho$  such that matrices  $\mathbf{I}_N - \delta\mathbf{W}_N,$   $\mathbf{I}_N - \rho\mathbf{M}_N,$  and  $\mathbf{B}_N$  are non-singular for all  $\delta \in (-K'_\delta, K''_\delta)$  and  $\rho \in (-K'_\rho, K''_\rho)$  and any  $N \in \mathbb{N}.$
3. The absolute values of the row and column sums of  $\mathbf{B}_N, \mathbf{S}_N, \mathbf{W}_N, \mathbf{M}_N, (\mathbf{I}_N - \delta^0\mathbf{W}_N)^{-1},$   $(\mathbf{I}_N - \rho^0\mathbf{M}_N)^{-1},$  and  $\Phi_{k,N}, k = 1, \dots, \varkappa,$  are bounded uniformly in  $N \in \mathbb{N}.$

Next, the assumptions concerning the explanatory variables and the imposed instrumental variables are specified. To guarantee identification of the model parameters, the matrix of explanatory variables  $\tilde{\mathbf{Z}}_N$  and of the instrumental matrix  $\tilde{\mathbf{H}}_N$  are generally assumed to have rank at least  $K+2$  so that results apply to various sets of instruments proposed in Section 3.1. To achieve this, for example, in the case of purely dynamic model without explanatory variables, it is sufficient that the parameters linking the instruments and instrumented variables are non-zero, that is,  $\lambda \neq 0$  and  $\delta \neq 0,$  and that the spatial weight matrices  $\mathbf{I}_N, \mathbf{W}_N,$  and  $\mathbf{W}_N^2$  are not linearly dependent. See a detailed discussion of the identification assumptions in this framework is given by Lee and Yu (2013).

Additionally, we require only the existence of various finite moments of the explanatory variables, instruments, and moment conditions as needed for the central limit theorem (see Assumption V4 and V5 below). Since the spatial structure might change with an increasing sample size in a generally unspecified way, these expectations change as well with an increasing sample size; for example  $E[\tilde{\mathbf{H}}_N^\top \tilde{\mathbf{Z}}_N/N]$  changes with  $N$  along with the corresponding spatial matrices. We assume for simplicity of notation that these averages have well-defined limits, for example, that  $\lim_{N \rightarrow \infty} E[\tilde{\mathbf{H}}_N^\top \tilde{\mathbf{Z}}_N/N]$  exists, but this assumption can be relaxed. Finally, 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.  $\mathbf{Z}_N$  has a full rank almost surely.
2.  $\mathbf{H}_N$  has a rank greater or equal to  $K + 2$  almost surely.
3. The expectations  $E(X_{N,ij})^{2+\psi}$ ,  $E(Z_{N,ij})^{2+\psi}$  and  $E(H_{N,ij\varepsilon_{N,k}})^{2+\psi}$  are bounded uniformly in  $i, j, k$  for some  $\psi > 0$ .
4. The limits of matrices  $\lim_{N \rightarrow \infty} E[\mathbf{H}_N^\top \boldsymbol{\Phi}_{k,N^\otimes}^\top \boldsymbol{\Phi}_{k,N^\otimes} \mathbf{Z}_N] = \tilde{\mathbf{Q}}_{k,HZ}$  exist and are non-singular for  $k = 1, \dots, \varkappa$  as well as their linear combination  $\lim_{N \rightarrow \infty} E[\tilde{\mathbf{H}}_N^\top \tilde{\mathbf{Z}}_N] = E[\mathbf{H}_N^\top \mathbf{B}_{N^\otimes}^\top \mathbf{B}_{N^\otimes} \mathbf{Z}_N] = \tilde{\mathbf{Q}}_{HZ}$ , which exists and is non-singular.
5. The limits of variance matrices ( $k = 1, \dots, \varkappa$ )
 
$$\lim_{N \rightarrow \infty} E[\mathbf{H}_N^\top \boldsymbol{\Phi}_{k,N^\otimes}^\top \boldsymbol{\Phi}_{k,N^\otimes} [\mathbf{I}_{N^\otimes} - \rho^0 \mathbf{M}_{N^\otimes}]^{-1} \boldsymbol{\varepsilon}_N \boldsymbol{\varepsilon}_N^\top [\mathbf{I}_{N^\otimes} - \rho^0 \mathbf{M}_{N^\otimes}]^{-1} \boldsymbol{\Phi}_{k,N^\otimes}^\top \boldsymbol{\Phi}_{k,N^\otimes} \mathbf{H}_N / N] = \tilde{\mathbf{Q}}_{k,H\Sigma H}$$
 as well as  $\lim_{N \rightarrow \infty} E[\tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N \tilde{\mathbf{u}}_N^\top \tilde{\mathbf{H}}_N / N] = \lim_{N \rightarrow \infty} E[\mathbf{H}_N^\top \mathbf{B}_{N^\otimes}^\top \mathbf{B}_{N^\otimes} [\mathbf{I}_{N^\otimes} - \rho^0 \mathbf{M}_{N^\otimes}]^{-1} \boldsymbol{\varepsilon}_N \boldsymbol{\varepsilon}_N^\top [\mathbf{I}_{N^\otimes} - \rho^0 \mathbf{M}_{N^\otimes}]^{-1} \mathbf{B}_{N^\otimes}^\top \mathbf{B}_{N^\otimes} \mathbf{H}_N / N] = \tilde{\mathbf{Q}}_{H\Sigma H}$  are finite and positive definite.

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  $\tilde{\mathbf{A}}_N$ ,  $\hat{\mathbf{A}}_N$ ,  $\mathbf{C}_N$ , and  $\boldsymbol{\Gamma}_N$  are non-singular. This assumption is again general to accommodate any choice of the weight matrices; regarding the matrices suggested in Section 3, they are mostly deterministic (e.g., identity matrices) with the exception of  $\mathbf{A}_{N,AB} = [\tilde{\mathbf{H}}_{N,AB}^\top \mathbf{G}_{N,AB} \tilde{\mathbf{H}}_{N,AB} / N]^{-1}$ ; in the case of  $\mathbf{A}_{N,AB}$ , convergence to a well-defined limit follows from Lemma 2 in the Appendix.

### 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  $\mathbf{A}_N$  satisfy  $\text{p-lim}_{N \rightarrow \infty} \mathbf{A}_N = \mathbf{A}$ , where  $\mathbf{A}$  is a finite positive definite matrix.
3. Non-singular symmetric matrices  $\hat{\mathbf{A}}_N$  satisfy  $\text{p-lim}_{N \rightarrow \infty} \hat{\mathbf{A}}_N = \mathbf{A}$ , where  $\mathbf{A}$  is a finite positive definite matrix.
4. The parameter space  $\Phi$  for  $\boldsymbol{\varphi} = (\rho, \sigma_v)^\top$  is a compact subset of  $(-K'_\rho, K''_\rho) \times \mathbb{R}^+$ . Moreover,  $\boldsymbol{\varphi}^0 = (\rho^0, \sigma_v^0)^\top \in \Phi^\circ$ .
5. The smallest eigenvalues of the matrix  $\boldsymbol{\Gamma}_N^\top \boldsymbol{\Gamma}_N$  are uniformly larger than  $\kappa_\Gamma > 0$ .
6. Positive definite matrices  $\mathbf{C}_N$  and  $\hat{\mathbf{C}}_N$  satisfy  $\lim_{N \rightarrow \infty} \mathbf{C}_N = \mathbf{C}$  and  $\text{p-lim}_{N \rightarrow \infty} \hat{\mathbf{C}}_N = \mathbf{C}$ , respectively, where  $\mathbf{C}_N$  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. However, the estimation of  $\rho$  and  $\sigma_v^2$  characterizing the variance and spatial correlation of the errors is nonlinear and it is thus necessary to prove that the parameters are identified and that the finite-sample GMM objective function converges to its population counterpart (cf. Kelejian and Prucha, 2010).

Let us remark at this point that the assumptions specified in Section 4.1 are not sufficient for the general spatial Arellano-Bond or Blundell-Bond estimators. The main reason is that even the unobservable errors exhibit both dependence in space and in time; for example, the differenced error terms  $\Delta \tilde{\mathbf{u}}_N(t)$  in (9) are serially correlated due to differencing and spatially correlated due to cross-sectional correlation (note that this is obviously not the case in the level equation (17)). To derive the asymptotic results presented in this section, one thus needs to impose additional conditions on the spatial dependence such that the error terms have a “short” memory. We can

impose additional conditions on  $\mathbf{W}_N$  that will guarantee that  $\Delta\tilde{\mathbf{u}}_N$  are near-epoch dependent in space, for instance, using the definition of the near-epoch dependence and the law of large numbers of Jenish and Prucha (2012). That means the locations of the cross-sectional units in space have to be defined, the elements of  $\mathbf{W}_N$  and  $\mathbf{M}_N$  have to satisfy additional constraints such as those in Jenish and Prucha (2010, equation (13)), and the proofs could be done analogously to the Jenish and Prucha's (2012) proofs of consistency and asymptotic normality of the cross-sectional GMM estimator.

On the other hand, it is possible to derive the asymptotic properties of the general spatial Arellano-Bond or Blundell-Bond estimators using only the assumptions in Section 4.1 if only a subset of moment conditions is used such that the involved unobservable error terms are not serially correlated. For example, using the moment conditions (11)–(13) of the Arellano-Bond estimator only for odd time periods  $t = 3, 5, \dots, 2\lceil T/2 \rceil - 1$  makes employed  $\Delta\tilde{\mathbf{u}}_N(t)$  independent over time; the same is true if only the moments for the level equation (17) are used. In that case, the near-epoch dependence does not have to be imposed and the asymptotic results can be derived using just the weak Assumption S3 on the spatial weight matrices. Since both sets of assumptions and proofs lead to formally the same results (e.g., the variance matrices can be expressed in the same way in both cases), we provide the proofs under the weaker assumptions of Section 4.1, noting that the other case can be derived analogously using the limit theorems of Jenish and Prucha (2012).

We will show first that the infeasible estimator  $\tilde{\boldsymbol{\theta}}_N$  defined by (14) or (22) for a given sequence of transformation matrices  $\mathbf{B}_N$  is consistent and asymptotically normal. Consequently, the same result applies to the initial estimator  $\hat{\boldsymbol{\theta}}_N^0$  based on  $\mathbf{B}_N = \mathbf{I}_N$ .

**Theorem 1.** *Under Assumptions E, S, V, and G1–G2, the GMM estimator  $\tilde{\boldsymbol{\theta}}_N$  is  $\sqrt{N}$ -consistent and*

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

*in distribution as  $N \rightarrow +\infty$ .*

**Proof.** See Appendix A.3.1.  $\square$

Although the asymptotic distribution of  $\tilde{\boldsymbol{\theta}}_N$  and  $\hat{\boldsymbol{\theta}}_N^0$  is derived in Theorem 1, it is not practically applicable at this stage: the variance matrix  $\tilde{\mathbf{Q}}_{H\Sigma H}$  of  $\tilde{\mathbf{H}}_N \tilde{\mathbf{u}}_N$  generally depends on unknown parameters  $\sigma_v^0$  and  $\rho^0$  (see Assumption V5). This also means that the two-step GMM estimator based on (an estimate of) the optimal weighting matrix  $\mathbf{A} = \tilde{\mathbf{Q}}_{H\Sigma H}^{-1}$  cannot be constructed. To achieve feasible estimation, the variance  $\sigma_v^0$  and spatial autocorrelation parameter  $\rho^0$  has to be estimated first by  $\hat{\sigma}_{v,N}$  and  $\hat{\rho}_N$  defined in equation (28). The consistency of the proposed estimates  $\hat{\sigma}_{v,N}$  and  $\hat{\rho}_N$  is proved in the following theorem.

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

**Proof.** See Appendix A.3.2.  $\square$

Having a consistent estimate  $\hat{\rho}_N$  of  $\rho^0$ , the asymptotic variance in Theorem 1 can be evaluated and the optimal GMM weighting matrix  $\tilde{\mathbf{Q}}_{H\Sigma H}^{-1}$  can be estimated. More importantly,  $\hat{\rho}_N$  can be used to transform the model (8) to obtain (asymptotically) spatially uncorrelated errors. The spatially-corrected infeasible estimator relies on the transformation  $\mathbf{B}_N = \mathbf{I}_N - \rho^0 \mathbf{M}_N$  as described in Section 3. If we now set  $\hat{\mathbf{B}}_N = \mathbf{I}_N - \hat{\rho}_N \mathbf{M}_N$ , Theorem 2 and Assumption S3 guarantee that  $\|\hat{\mathbf{B}}_N - \mathbf{B}_N\|_F = |\hat{\rho}_N - \rho^0| \|\mathbf{M}_N\|_F \rightarrow 0$  in probability as  $N \rightarrow +\infty$ , where  $\|\cdot\|_F$  denotes the Frobenius norm. This asymptotic equivalence of the feasible and infeasible transformations leads to the following theorem: the feasible GMM estimator corresponding to the transformation matrix  $\hat{\mathbf{B}}_N = \sum_{k=1}^{\varkappa} \hat{\phi}_{k,N} \boldsymbol{\Phi}_{k,N}$  is asymptotically equivalent to the infeasible GMM estimator based on  $\mathbf{B}_N = \sum_{k=1}^{\varkappa} \phi_k \boldsymbol{\Phi}_{k,N}$  in general for any transformation such that  $|\hat{\phi}_{k,N} - \phi_k^0| = o_p(1)$  for all  $k = 1, \dots, \varkappa$ , and in particular, for  $\hat{\mathbf{B}}_N = \mathbf{I}_N - \hat{\rho}_N \mathbf{M}_N$  and  $\mathbf{B}_N = \mathbf{I}_N - \rho^0 \mathbf{M}_N$  as  $\hat{\rho}_N - \rho^0 = o_p(1)$  by Theorem 2.

**Theorem 3.** *Under Assumptions E, S, V, and G1–G3, the GMM estimator  $\hat{\boldsymbol{\theta}}_N$  defined in (23) for transformation matrix  $\hat{\mathbf{B}}_N$  is  $\sqrt{N}$ -consistent, asymptotically normal*

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

and asymptotically equivalent to the infeasible estimator,  $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \tilde{\boldsymbol{\theta}}_N) \xrightarrow{p} 0$  as  $N \rightarrow +\infty$ , if  $\mathbf{B}_N = \sum_{k=1}^{\varkappa} \phi_k \boldsymbol{\Phi}_{k,N}$ ,  $\hat{\mathbf{B}}_N = \sum_{k=1}^{\varkappa} \hat{\phi}_{k,N} \boldsymbol{\Phi}_{k,N}$ , and  $|\hat{\phi}_{k,N} - \phi_k^0| = o_p(1)$  as  $N \rightarrow +\infty$  for all  $k = 1, \dots, \varkappa$ .

**Proof.** See Appendix A.3.3.  $\square$

The feasible GMM estimator depends on the choice of two matrices: the transformation  $\mathbf{B}_N$  and the weighting matrix  $\mathbf{A}_N$ . Theorem 3 indicates that the asymptotic distributions of the infeasible and feasible GMM estimators are equivalent as long as  $\hat{\mathbf{B}}_N$  and  $\hat{\mathbf{A}}_N$  are consistent estimates of  $\mathbf{B}_N$  and  $\mathbf{A}_N$ , respectively. First, the transformation  $\mathbf{B}_N = \mathbf{I}_N - \rho^0 \mathbf{M}_N$  has been chosen to remove the spatial error correlation and can be estimated using  $\hat{\rho}_N$  by Theorem 2. It can be shown that this transformation is minimizing the asymptotic variance in special cases; for example in the case of the two-stage least squares estimator corresponding to  $\mathbf{G} = \mathbf{I}$  in Section 3.1, it follows from the Cauchy-Schwartz inequality (Tripathi, 1999) by a similar argument as in Lee (2003).

Next, the optimal choice of the weighting matrix for GMM equals the inverse of the variance of the moment conditions,

$$\begin{aligned} \mathbf{A}_N &= \{E[\tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N \tilde{\mathbf{u}}_N^\top \tilde{\mathbf{H}}_N / N]\}^{-1} \\ &= \{E[\mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} [\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes}]^{-1} \boldsymbol{\varepsilon}_N \boldsymbol{\varepsilon}_N^\top [\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes}]^{-1\top} \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} \mathbf{H}_N / N]\}^{-1} \\ &\rightarrow \tilde{\mathbf{Q}}_{H\Sigma H}^{-1}, \end{aligned}$$

which can be estimated once the estimate  $\hat{\rho}_N$  of  $\rho^0$  is obtained and which results in the asymptotic variance of the spatial GMM estimator being  $[\tilde{\mathbf{Q}}_{HZ}^\top \tilde{\mathbf{Q}}_{H\Sigma H}^{-1} \tilde{\mathbf{Q}}_{HZ}]^{-1}$ . Note that the variance of errors  $E(\boldsymbol{\varepsilon}_N \boldsymbol{\varepsilon}_N^\top)$  depends (i) in the case of the Arellano-Bond estimator only on  $\sigma_v$ , which is consistently estimated by  $\hat{\sigma}_{v,N}$  due to Theorem 2, but (ii) in the case of the Blundell-Bond estimator on variances of all individual effects  $\boldsymbol{\eta}_N$ . More specifically in the case of the spatial Arellano-Bond estimator with  $\mathbf{B}_N = \mathbf{I}_N - \rho^0 \mathbf{M}_N$ , the independence of idiosyncratic shocks  $\mathbf{v}_N(t)$  on covariates and  $\Delta \boldsymbol{\varepsilon}_N(t) = \mathbf{v}_N(t) - \mathbf{v}_N(t-1)$  for  $t = 2, \dots, T$  implies that  $E(\Delta \boldsymbol{\varepsilon}_N \Delta \boldsymbol{\varepsilon}_N^\top) = \mathbf{G}_{N,AB}$

defined in Section 3.1 and

$$\begin{aligned}
\mathbf{A}_N &= \{E[\mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} [\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes}]^{-1} \boldsymbol{\varepsilon}_N \boldsymbol{\varepsilon}_N^\top [\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes}]^{-1\top} \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} \mathbf{H}_N / N]\}^{-1} \\
&= \text{diag}\{E[\mathbf{H}_N^\top(t) \mathbf{B}_N^\top \mathbf{B}_N [\mathbf{I}_N - \rho^0 \mathbf{M}_N]^{-1} \cdot E[\boldsymbol{\varepsilon}_N(t) \boldsymbol{\varepsilon}_N^\top(t) | \tilde{\mathbf{H}}_N(t)] \\
&\quad \cdot [\mathbf{I}_N - \rho^0 \mathbf{M}_N]^{-1\top} \mathbf{B}_N^\top \mathbf{B}_N \tilde{\mathbf{H}}_N(t) / N]\}^{-1} \\
&= \text{diag}\{E[\mathbf{H}_N^\top(t) [\mathbf{I}_N - \rho^0 \mathbf{M}_N]^\top \sigma_v^2 \mathbf{G} [\mathbf{I}_N - \rho^0 \mathbf{M}_N] \mathbf{H}_N(t) / N]\}^{-1} \\
&= \sigma_v^2 E[\tilde{\mathbf{H}}_N^\top \mathbf{G}_{N,AB} \tilde{\mathbf{H}}_N]^{-1}.
\end{aligned}$$

Hence, the proposed first-step weighting matrix of the spatial Arellano-Bond estimator is an estimate of the optimal weighting matrix  $\mathbf{A}_N / \sigma_v^2$  if the errors  $\boldsymbol{\varepsilon}_N$  are homoscedastic.

Consequently, the advantages of the spatial Arellano-Bond estimator are (i) weak identification assumptions allowing for general fixed effects model and (ii) the optimal weighting matrix under homoscedasticity can be constructed a priori. On the other hand, the spatial Blundell-Bond estimator (i) imposes stricter assumptions in that the time changes of covariates cannot be correlated with the individual effects, but (ii) it uses these constrains to construct additional moment conditions that lead to more precise estimation. Irrespective of the chosen estimator, we provide estimators for the matrices entering the asymptotic variance of  $\hat{\boldsymbol{\theta}}_N$  in the final theorem.

**Theorem 4.** *Under the assumptions of Theorems 2 and 3, which guarantee  $\hat{\rho}_N \rightarrow \rho^0$ ,  $\hat{\sigma}_{v,N} \rightarrow \sigma^0$ , and  $\hat{\mathbf{B}}_N \rightarrow \mathbf{B}$  in probability as  $N \rightarrow +\infty$ , it holds that*

$$N^{-1} \hat{\mathbf{H}}_N^\top \hat{\mathbf{Z}}_N = N^{-1} \mathbf{H}_N^\top \hat{\mathbf{B}}_{N\otimes}^\top \hat{\mathbf{B}}_{N\otimes} \mathbf{Z}_N \rightarrow \tilde{\mathbf{Q}}_{HZ}$$

and

$$N^{-1} \hat{\mathbf{H}}_N^\top \hat{\mathbf{u}}_N \hat{\mathbf{u}}_N^\top \hat{\mathbf{H}}_N = N^{-1} \mathbf{H}_N^\top \hat{\mathbf{B}}_{N\otimes}^\top \hat{\mathbf{B}}_{N\otimes} \mathbf{u}_N \mathbf{u}_N^\top \hat{\mathbf{B}}_{N\otimes}^\top \hat{\mathbf{B}}_{N\otimes} \mathbf{H}_N \rightarrow \tilde{\mathbf{Q}}_{H\Sigma H}$$

in probability as  $N \rightarrow +\infty$ .

**Proof.** See Appendix A.3.4.  $\square$

## 5 Extensions

The spatial dynamic panel data model we analysed up to now has a dynamic lag, a spatial lag, and spatial correlated errors, with the exogenous variables  $\mathbf{X}_N(t)$  assumed strictly exogenous. This section briefly discusses two types of extensions: (i) extensions to the model and (ii) extensions to the GMM estimators.

The model introduced in Section 2 can be extended by a spatial-temporal lag,  $\mathbf{W}_N \mathbf{y}_N(t-1)$ , or spatial exogenous variables,  $\mathbf{W}_N \mathbf{X}_N(t)$ :

$$\mathbf{y}_N(t) = \lambda \mathbf{y}_N(t-1) + \delta \mathbf{W}_N \mathbf{y}_N(t) + \gamma \mathbf{W}_N \mathbf{y}_N(t-1) + \mathbf{X}_N(t) \boldsymbol{\beta} + \mathbf{W}_N \mathbf{X}_N(t) \boldsymbol{\theta} + \mathbf{u}_N(t).$$

Consistent GMM estimation is still possible if there are at least  $2K + 3$  valid instruments that are correlated with the time-lagged, spatially lagged, spatio-temporal, exogenous, and spatial exogenous variables and uncorrelated with the errors  $\Delta \tilde{\mathbf{u}}_N(t)$  for each  $t = 3, \dots, T$ . For example, spatial lags of the exogenous variables,  $\mathbf{W}_N \mathbf{X}_N(t-1)$ , can be used to instrument the spatio-temporal variables or the spatially lagged exogenous variables if  $\boldsymbol{\beta} \neq \mathbf{0}$  or values of  $\mathbf{X}_N(t)$  are correlated over time, respectively. Note though that including both the spatio-temporal lag and spatial exogenous variables will likely lead to the problem of weak instruments, especially for smaller absolute values of the autoregressive parameter. Moreover, the identification would require conditions specified in Lee and Yu (2013).

Extensions allowing for higher order dynamic and spatial lags in the model can also be dealt with in a straightforward manner by choosing proper instruments. The same holds for time-varying spatial weights matrices.<sup>15</sup>

The second type of extensions refers to the design of the GMM estimators introduced in Section 3 above. On the one hand, additional or alternative instruments can be employed such as lagged differences of the exogenous variables to instrument dynamic lags of the endogenous variable in the spatial Arellano-Bond estimators as discussed in Section 3.1.1. On the other hand, the assumptions could be relaxed. For example, returning to the original model of Section

---

<sup>15</sup>Lee and Yu (2012) study quasi-ML estimation of spatial models with time-varying weights matrices.

2, we can relax the strict exogeneity assumption on the exogenous variables to  $\mathbf{X}_N(t)$  being pre-determined. In this case the spatio-temporal model of Section 2 stays the same, but the moment conditions (10) in the Arellano-Bond estimator have to be replaced by

$$\mathbb{E}[\tilde{\mathbf{X}}_N^\top(t-s)\Delta\tilde{\mathbf{u}}_N(t)] = 0, \quad t = 3, \dots, T, \quad s = 2, \dots, t-1,$$

and (13) by

$$\mathbb{E}[\{\mathbf{W}_N^l \tilde{\mathbf{X}}_N(t-s)\}^\top \Delta\tilde{\mathbf{u}}_N(t)] = 0, \quad t = 3, \dots, T, \quad s = 2, \dots, t-1, \quad l = 1, \dots, L.$$

In the Blundell-Bond estimator the moment conditions for the pre-determined variables become

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

Provided sufficient moment conditions exist to estimate the  $K + 2$  parameters, our GMM estimators can still be used and are consistent.

## 6 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.

### 6.1 Simulation Design

We report the small sample properties of the estimators using data sets generated based on the spatial dynamic panel data model introduced in 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. The exogenous variable is defined as:

$$\mathbf{X}_N(t) = \boldsymbol{\varsigma} + \boldsymbol{\chi}(t), \quad \boldsymbol{\varsigma} \sim \text{iid } N(0, 1), \quad \boldsymbol{\chi}(t) \sim \text{iid } N(0, 1), \quad (29)$$

where  $\varsigma$  represents the unit-specific component and  $\chi(t)$  denotes a random component; both are drawn from the standard normal distribution.

Thereafter, we use  $\varsigma$  to construct the unit-specific effect. We do this in order to explicitly model the fixed effects that we assume in Section 2. To be specific, we construct  $\boldsymbol{\eta}_N$  as follows

$$\boldsymbol{\eta}_N = \sqrt{\frac{\phi}{2}} [\boldsymbol{\xi} + \varsigma], \text{ with } \boldsymbol{\xi} \sim \text{iid } N(0, 1). \quad (30)$$

where  $\phi$  can be used to increase the variance of the fixed-effect relative to the variance of the idiosyncratic error  $\mathbf{v}_N$ . Combining (30), (4), and (5) with

$$\mathbf{v}_N \sim \text{iid } N(0, \mathbf{I}_{NT}), \quad (31)$$

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  $t = 2, \dots, T$  and  $\mathbf{y}(1) = \boldsymbol{\eta}_N$ . Because it is impossible to analytically establish the conditions for mean-stationarity in the underlying model, and it is unclear how fast a spatial-dynamic model converges numerically, we iterate until the mean of  $\mathbf{y}_N(t)$  converges.

We use different spatial weights 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  $n$  neighbors to each spatial unit  $i$ —while ensuring symmetry—and row normalize the matrices. We generate new matrices  $\mathbf{W}_N$  and  $\mathbf{M}_N$  for each iteration, to make sure that our results do not depend on a specific draw of the weights matrix (but ensuring that all estimators within an iteration follow from the same matrices). In the benchmark scenario, we 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 Bucky ball matrix is shaped like a soccer ball, where the distance from any point to its nearest neighbors is the same for all the points. As the Bucky ball specification assumes  $N = 60$  it cannot be used for varying  $N$ . 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 distance. We randomly generated  $n$  points and compute the Euclidean distance between each pair. Again this weighting matrix is subsequently row-normalized.

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 all estimators are regression equivariant, the coefficient of the exogenous explanatory variable  $\beta$  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  $\rho$  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  $\text{RMSE} = \sqrt{\text{bias}^2 + \left(\frac{q_1 - q_2}{1.35}\right)^2}$ , 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.

## 6.2 Results

Table 1 gives a detailed overview of the estimators considered in the simulation study. We report four different types of spatial GMM estimators all of which instrument the time lag of the dependent variable in addition to addressing spatial aspects. For each estimator, seven instrument-sets have been defined. We consider a spatial Arellano-Bond differenced-based GMM estimator (labeled AB) and a spatial Blundell-Bond system-based GMM estimator (labeled BB). 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 step of the spatial GMM procedure discussed in Section 3.2. 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  $s$  of the spatially lagged dependent variable (indicated by the subscript  $Y$ ), which are labelled spatio-temporal instruments; 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.

Table 2 reports the RMSE of estimating the spatial autoregressive parameter  $\delta$  for the various estimators and different values of  $N$  starting at the benchmark value of  $N = 60$  ( $T = 5$  is fixed). The table shows that the RMSE decreases if  $N$  increases. For those estimators using spatially weighted exogenous variables  $\mathbf{W}_N \Delta \mathbf{X}_N(t)$  as instruments, extending the number of spatial units from 60 to 500 reduces the RMSE by more than 50 % when  $\lambda = 0.3$  and  $\delta = 0.5$ . The decrease in the variance is smaller for all estimators in case  $\lambda = 0.7$  and  $\delta = 0.2$ . Furthermore, estimators using only spatio-temporal instruments ( $Y$ ) have larger RMSEs.

We find that the spatially corrected estimators have generally a smaller RMSE than their non-spatially corrected counterparts. The reduction in the RMSE is more than 5% in the model with  $\lambda = 0.3$  and  $\delta = 0.5$ , but smaller in case  $\lambda = 0.7$  and  $\delta = 0.2$ . Additionally, the (S)BB estimators give rise to a smaller RMSE than the (S)AB estimators in case  $\lambda = 0.7$  and  $\delta = 0.2$ , the improvement in efficiency is more than 50%. For  $\lambda = 0.3$  and  $\delta = 0.5$ , the (S)BB estimators are superior only when the number of spatial units is relatively low ( $N = 60$ ), yielding a reduction in the RMSE of almost 40%. When the number  $N$  of spatial units increases, the superiority of (S)BB estimators is less clear: for lower values of  $\lambda$ , the (S)AB estimators are preferable. This corresponds to the intuition that the advantage of adding the level-equations to the difference-equations is the largest when  $\lambda$  is relatively high or when the number of observations is relatively low.

Note that increasing  $\lambda$  causes an increase in the RMSE for the (S)AB estimator and a decrease for the (S)BB estimator, reflecting that an increase in  $\lambda$  reduces the power of the dynamic instruments in (S)AB instruments relative to the instruments in the levels equation of the (S)BB estimators. With respect to the choice of instrument set, the specifications with both  $X$  and

$Y$  instruments for the spatial lag yield smaller RMSEs than those without exogenous variables in the instrument set for the spatial lag (only  $Y$ ). When comparing estimators using only  $X$  instruments ( $AB_X/BB_X$ ) with those using only spatio-temporal  $Y$  instruments ( $AB_{Y3}/BB_{Y3}$ ) reveals that in 8 out of 12 instances, the former report a lower RMSE. Hence, spatially lagged exogenous variables are generally stronger instruments than spatio-temporal instruments. It is only worthwhile to replace spatially lagged exogenous variables with spatio-temporal instruments for the BB estimator in case  $\lambda = 0.7$  and  $\delta = 0.2$ . The combination of both types of instruments is preferable to using one type of instruments though.

Simple counting based on the lowest RMSE suggests that  $SBB_{XY3}$  is the best estimator in case  $\lambda = 0.7$  and  $\delta = 0.2$ . However, for the case with  $\lambda = 0.3$  and  $\delta = 0.5$ , we see that the best estimator depends on  $N$ :  $SBB_{XY1}$  for  $N = 60$ ,  $SBB_{XY2}$  for  $N = 200$ ; and  $SAB_{XY1}$  for  $N = 500$ . Overall,  $SBB_{XY2}$  performs very well under all configurations. But, as  $AB$  estimators are consistent under more general assumptions, it is interesting to observe that, although none of the estimators comes out on top,  $SAB_{XY1}$  most often features a RMSE that is close to the RMSE of the best estimator amongst the  $SAB$  estimators.

Table 3 presents the RMSE in estimating the parameter  $\delta$  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 might be needed. As before, we limit the lag depth of the dynamic instruments to 3, which corresponds to the maximum number of lags used for  $T = 5$ , guarantees thus comparability across different values of  $T$ , and reduces the RMSE in estimating the spatial lag parameter at higher values of  $T$  (see Jacobs et al., 2009).

Increasing the number of time periods from 5 to 20 in the benchmark case of  $\lambda = 0.3$  and  $\delta = 0.5$  reduces the average RMSE by 20%, where the decline is somewhat stronger for the BB and SBB estimators. In case  $\lambda = 0.7$  and  $\delta = 0.2$ , the fall in RMSE induced by a rise in  $T$  from 5 to 20 for both the (S)AB estimators and (S)BB estimators is on average more than 40%. Overall, when individual effects are not correlated with differences in the response and explanatory variables, the (S)BB estimators are preferred over (S)AB estimators. The preferred

estimator is the  $(S)BB_{XY2}$  for  $\lambda = 0.3$  and  $\delta = 0.5$ , whereas  $(S)BB_{XY3}$  is preferred for  $\lambda = 0.7$  and  $\delta = 0.2$ . Within the SAB estimators,  $SAB_X$  generally reports the lowest RMSE, except in the case of a short panels ( $T = 5$ ) when  $\lambda$  is equal to 0.7; in that case it is advisable to include some spatio-temporal instruments in the instrument matrix.

Furthermore, except for the BB estimators in case  $\lambda = 0.7$  and  $\delta = 0.2$ , spatially corrected estimators are preferred. Note that the RMSE for the BB estimators in case  $\lambda = 0.7$  and  $\delta = 0.2$  is by far the lowest in the sample. Compare for example the RMSE for  $AB_X$  and  $BB_X$  and observe that  $BB_X$  in case of  $\lambda = 0.7$  and  $\delta = 0.2$  shows a 50 % lower RMSE compared to the second lowest RMSE which is found for  $BB_X$  in case  $\lambda = 0.3$ . Maybe the relative efficiency of the BB estimators in this case makes further improvements more difficult.

Finally, the weak instrument problem in these models is nicely illustrated. When only spatio-temporal instruments are used (the  $AB_Y$  and  $BB_Y$  estimators) the simulations suggest it is best to use all of them, hence, all three spatio-temporal lags have explanatory power. However, when exogenous ( $X$ ) instruments are also available, it is best to use only one spatio-temporal lag in the case of the  $AB_{XY}$  estimator. Apparently, although the spatio-temporal instruments have explanatory power, higher lags are relative weak instruments. For the BB estimators it is better to use more spatio-temporal lags, especially when  $\lambda$  and  $\delta$  are high.

Table 4 presents the RMSE of the BB estimators for several values of  $\delta$  in the interval  $[0.3, 0.7]$  and for different values of  $\rho$  (the results for AB estimators are qualitatively similar). We vary  $\rho$  in the interval  $[-0.8, 0.8]$ , where a negative  $\rho$  implies that 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  $\delta$ , we set  $\lambda$  to 0.2. The most prominent feature of the results is that there is a U-shape in the RMSE for the non-spatially corrected BB estimators, which is absent (or reversed) in the case of the spatially corrected BB estimators. That is, we find that generally non-spatially corrected BB estimators have larger RMSEs in estimating  $\lambda$ ,  $\delta$ , and  $\beta$  than their spatially corrected counterparts for  $\rho \neq 0$  (only for  $\lambda$  results are mixed). The reduction in the RMSE of the spatial correction is largest when estimating  $\delta$ . The difference in RMSEs between BB and SBB estimators for large absolute values

of  $|\rho|$  is on average always larger than 16%. The decrease in the RMSE is larger for a large positive  $\rho$  (this yields a gain of more than 60% in estimating  $\delta$ ). In this respect, notice that the final rows of the table show that also  $\rho$  is estimated more accurately when it is larger. We see that the difference between spatially corrected and spatially uncorrected estimators is close to zero for a pure spatial lag model (i.e.,  $\rho = 0$ ).

Table 5 investigates the effect of the specification of the spatial weights matrices on the RMSE from estimating  $\delta$  for various values of  $\rho$ . Reducing the sparsity—fewer zeros through an increase in the number of neighbors as measured by  $n$ —of the random contiguity matrices generally increases the RMSE of all estimators with the exception of the estimators using only  $Y$  instruments. The increase of the RMSE is especially large for the spatially corrected estimators. This can be explained by separating the effect of increasing  $n$  on the spatial and spatially corrected estimators. For the spatial estimators increasing the number of neighbors implies that on the one hand, spatial shocks become more “global”, which makes identification of both  $\delta$  and  $\rho$  more difficult. On the other hand, more neighbors enhances the strength of the spatio-temporal instruments relative to the exogenous instruments which are used to identify  $\delta$ . Intuitively, the increase in  $n$  reduces the variation in  $WX$  instruments as unit-specific shocks in the exogenous variables are more strongly averaged out. For the  $Y$  instruments this effect is less pronounced as they identify  $\delta$  through the time persistency in the model. The spatial lag and its time lags are affected in exactly the same way by the change in the spatial weights matrices  $W$  and  $M$ . As a result, estimators employing only spatio-temporal instruments suffer the least from increasing  $n$ . The RMSE of all spatially corrected estimators is increased due to a reduction of the variance in estimating  $\rho$  (not reported). Note the inverted U-shape: when RMSE is high ( $\rho = |0.8|$  or the shock in the error term dominates) the reduction in the RMSE is relatively small compared to the case where  $\rho = 0.3$ , when spatially dispersed shocks from  $X$  are relatively more important in determining  $Y$ .

For the distance matrix—where weights are computed as  $1/(d_{ij})^p$ , with  $d_{ij}$  the distance between unit  $i$  and  $j$  and  $p \in \{4, 8\}$ —we see that giving more weight to close neighbors (by increasing  $p$ ) improves the efficiency of especially the spatially corrected estimators. This can be explained

from a more accurate estimation of  $\rho$  in addition to any possible efficiency gains in estimating  $\delta$ . The increase in  $p$  causes more variation in the weights which enables better estimation of  $\rho$ . Results for estimating  $\delta$  are mixed, again we see an inverted U-shape when comparing the efficiency gains in estimating  $\delta$  for various values of  $\rho$ .

The Bucky-ball specification—which imposes  $\mathbf{W}_N = \mathbf{M}_N$ , and assumes five or six neighbors—yields a high RMSE (compare to the case  $n = 5$ ). The restriction  $\mathbf{W}_N = \mathbf{M}_N$  makes it harder to separate the processes for  $\rho$  and  $\delta$ . Therefore, efficiency is reduced more for estimators that use fewer or weaker instruments (estimators only using  $Y$  instruments).

The final two tables, Table 6 and 7, show the results of two additional robustness checks for the main set of results. Table 6 shows that using two-step spatial AB/BB estimators or two-step spatially corrected AB/BB estimators instead of one-step estimators does not affect the results at all. Depending on the model, results either improve or deteriorate slightly, where the differences in RMSE are small. Table 7 studies the properties of estimators that uses a subset of the moment conditions which are not serially dependent, following the discussion in Section 4.2. We focus on the *SAB* and *SBB* estimators employing three spatio-temporal lags. Table 4.2 compares the estimators relying on a subset of moment conditions (labeled *ISAB* and *ISBB* respectively, where the *I* stands for an independent subset of moment conditions) with the original estimators using all moment conditions. Using a subset of independent moment conditions generally leads to an increase in the RMSE for all the coefficients compared to the estimators that use all available moment conditions. Only for estimating  $\delta$  in the model with  $\delta = 0.5$ ,  $\lambda = 0.3$  and  $N = 60$ , we find a small reduction in the RMSE. Overall, the efficiency loss is largest for estimating  $\lambda$  and  $\beta$ . The AB estimator shows for both  $\lambda$  and  $\beta$  an efficiency loss of more than 30 % due to using only a subset of the moment conditions, whereas the BB estimator shows an efficiency loss of more than 15 %. The efficiency losses are somewhat smaller for estimating  $\delta$ , on average 15% for AB and 8% for the BB estimators. Overall, increasing  $N$  causes an unambiguous reduction in the RMSE of the *ISAB* and *ISBB* estimators.

## 7 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, and allowing for fixed effects. 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  $T$ . Feasible spatial correction based on estimated spatial error correlation is shown to lead to estimators that are asymptotically equivalent to the infeasible estimators based on a known spatial error correlation. The 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. We show that the spatial Blundell-Bond estimators outperform the spatial Arellano-Bond estimators. However, Blundell-Bond estimators impose stricter assumptions in that time changes of covariates cannot be correlated with the individual effects. Finally, two-step GMM estimators using optimal GMM weighting matrices do not systematically outperform their one-step counterparts.

Table 1: Estimators

	Instruments for $y(-1)$	Instruments for $Wy$	Instruments for $X$
Arellano-Bond estimators			
$(S)AB_X$	Lagged levels (3 lags)	$W\Delta x, W^2\Delta x, W^3\Delta x, W\Delta x(-1), W^2\Delta x(-1), W^3\Delta x(-1)$	$\Delta x, \Delta x(-1)$
$(S)AB_{XY1}$	Lagged levels (3 lags)	$Wy(-1), W\Delta x, W^2\Delta x, W^3\Delta x, W\Delta x(-1), W^2\Delta x(-1), W^3\Delta x(-1)$	$\Delta x, \Delta x(-1)$
$(S)AB_{XY2}$	Lagged levels (3 lags)	$Wy(-1), W^2y(-1), Wy(-2), W^2y(-2), W\Delta x, W^2\Delta x, W^3\Delta x, W\Delta x(-1), W^2\Delta x(-1), W^3\Delta x(-1)$	$\Delta x, \Delta x(-1)$
$(S)AB_{XY3}$	Lagged levels (3 lags)	$Wy(-1), W^2y(-1), W^3y(-1), Wy(-2), W^2y(-2), W^3y(-2), Wy(-3), W^2y(-3), W^3y(-3), W\Delta x, W^2\Delta x, W^3\Delta x, W\Delta x(-1), W^2\Delta x(-1), W^3\Delta x(-1)$	$\Delta x, \Delta x(-1)$
$(S)AB_{Y1}$	Lagged levels (3 lags)	$Wy(-1)$	$\Delta x, \Delta x(-1)$
$(S)AB_{Y2}$	Lagged levels (3 lags)	$Wy(-1), W^2y(-1), Wy(-2), W^2y(-2)$	$\Delta x, \Delta x(-1)$
$(S)AB_{Y3}$	Lagged levels (3 lags)	$Wy(-1), W^2y(-1), W^3y(-1), Wy(-2), W^2y(-2), W^3y(-2), Wy(-3), W^2y(-3), W^3y(-3)$	$\Delta x, \Delta x(-1)$
Blundell-Bond estimators		Instruments for Arellano-Bond estimators plus ...	
$(S)BB_X$	Levels and Dif. (3 lags)	$W\Delta x, W^2\Delta x, W^3\Delta x, W\Delta x(-1), W^2\Delta x(-1), W^3\Delta x(-1)$	$\Delta x, \Delta x(-1)$
$(S)BB_{XY1}$	Levels and Dif. (3 lags)	$W\Delta y(-1), W\Delta x, W^2\Delta x, W^3\Delta x, W\Delta x(-1), W^2\Delta x(-1), W^3\Delta x(-1)$	$\Delta x, \Delta x(-1)$
$(S)BB_{XY2}$	Levels and Dif. (3 lags)	$W\Delta y(-1), W^2\Delta y(-1), W\Delta y(-2), W^2\Delta y(-2), W\Delta x, W^2\Delta x, W^3\Delta x, W\Delta x(-1), W^2\Delta x(-1), W^3\Delta x(-1)$	$\Delta x, \Delta x(-1)$
$(S)BB_{XY3}$	Levels and Dif. (3 lags)	$W\Delta y(-1), W^2\Delta y(-1), W^3\Delta y(-1), W\Delta y(-2), W^2\Delta y(-2), W^3\Delta y(-2), W\Delta y(-3), W^2\Delta y(-3), W^3\Delta y(-3), W\Delta x, W^2\Delta x, W^3\Delta x, W\Delta x(-1), W^2\Delta x(-1), W^3\Delta x(-1)$	$\Delta x, \Delta x(-1)$
$(S)BB_{Y1}$	Levels and Dif. (3 lags)	$W\Delta y(-1)$	$\Delta x, \Delta x(-1)$
$(S)BB_{Y2}$	Levels and Dif. (3 lags)	$W\Delta y(-1), W^2\Delta y(-1), W\Delta y(-2), W^2\Delta y(-2)$	$\Delta x, \Delta x(-1)$
$(S)BB_{Y3}$	Levels and Dif. (3 lags)	$W\Delta y(-1), W^2\Delta y(-1), W^3\Delta y(-1), W\Delta y(-2), W^2\Delta y(-2), W^3\Delta y(-2), W\Delta y(-3), W^2\Delta y(-3), W^3\Delta y(-3)$	$\Delta x, \Delta x(-1)$



Table 2: RMSE of Spatial GMM Estimators of  $\delta$  for Various Values of  $N$  and  $\lambda$

Estimator	$\lambda = 0.3$			$\lambda = 0.7$		
	$N = 60$	$N = 200$	$N = 500$	$N = 60$	$N = 200$	$N = 500$
$AB_X$	0.173	0.092	0.054	0.222	0.123	0.081
$AB_{XY1}$	0.183	0.091	0.053	0.226	0.124	0.080
$AB_{XY2}$	0.197	0.095	0.054	0.238	0.130	0.082
$AB_{XY3}$	0.216	0.105	0.057	0.239	0.132	0.082
$AB_{Y1}$	0.331	0.290	0.214	0.405	0.376	0.340
$AB_{Y2}$	0.313	0.271	0.219	0.356	0.341	0.275
$AB_{XY3}$	0.303	0.263	0.220	0.334	0.294	0.260
$SAB_X$	0.161	0.085	0.053	0.220	0.116	0.079
$SAB_{XY1}$	0.168	0.083	0.053	0.217	0.118	0.078
$SAB_{XY2}$	0.177	0.088	0.053	0.212	0.122	0.076
$SAB_{XY3}$	0.193	0.093	0.055	0.210	0.123	0.075
$SAB_{Y1}$	0.305	0.263	0.202	0.401	0.348	0.313
$SAB_{Y2}$	0.291	0.245	0.204	0.335	0.312	0.260
$SAB_{Y3}$	0.281	0.240	0.196	0.297	0.265	0.238
$BB_X$	0.145	0.086	0.062	0.085	0.074	0.066
$BB_{XY1}$	0.132	0.086	0.061	0.070	0.057	0.052
$BB_{XY2}$	0.129	0.089	0.063	0.062	0.049	0.046
$BB_{XY3}$	0.136	0.095	0.064	0.059	0.050	0.043
$BB_{Y1}$	0.167	0.145	0.125	0.082	0.070	0.067
$BB_{Y2}$	0.154	0.142	0.121	0.069	0.059	0.057
$BB_{Y3}$	0.154	0.140	0.123	0.064	0.058	0.056
$SBB_X$	0.135	0.082	0.060	0.085	0.067	0.063
$SBB_{XY1}$	0.120	0.079	0.058	0.067	0.055	0.047
$SBB_{XY2}$	0.117	0.081	0.057	0.060	0.050	0.043
$SBB_{XY3}$	0.123	0.084	0.059	0.059	0.048	0.042
$SBB_{Y1}$	0.152	0.136	0.117	0.078	0.067	0.063
$SBB_{Y2}$	0.142	0.131	0.109	0.070	0.058	0.055
$SBB_{Y3}$	0.137	0.129	0.110	0.065	0.057	0.053

Notes: RMSEs based on Monte Carlo simulations with 1000 replications. The other parameters are:  $T = 5$ ,  $n = 5$ ,  $\delta = 0.5$ ,  $\beta = 1$ , and  $\rho = 0.3$ . To meet the stability condition (3),  $\delta$  is set to 0.2 if  $\lambda = 0.7$ .

Table 3: RMSE of Spatial GMM Estimators of  $\delta$  for Various Values of  $T$  and  $\lambda$

Estimator	$\lambda = 0.3$			$\lambda = 0.7$		
	$T = 5$	$T = 10$	$T = 20$	$T = 5$	$T = 10$	$T = 20$
$AB_X$	0.173	0.146	0.138	0.222	0.161	0.130
$AB_{XY1}$	0.183	0.151	0.138	0.226	0.163	0.130
$AB_{XY2}$	0.197	0.161	0.143	0.238	0.170	0.136
$AB_{XY3}$	0.216	0.205	0.199	0.239	0.199	0.167
$AB_{Y1}$	0.331	0.305	0.300	0.405	0.312	0.266
$AB_{Y2}$	0.313	0.308	0.298	0.356	0.313	0.273
$AB_{XY3}$	0.303	0.266	0.246	0.334	0.256	0.204
$SAB_X$	0.161	0.129	0.117	0.220	0.144	0.109
$SAB_{XY1}$	0.168	0.134	0.118	0.217	0.143	0.109
$SAB_{XY2}$	0.177	0.142	0.122	0.212	0.148	0.114
$SAB_{XY3}$	0.193	0.179	0.174	0.210	0.164	0.135
$SAB_{Y1}$	0.305	0.281	0.273	0.401	0.280	0.227
$SAB_{Y2}$	0.291	0.280	0.270	0.335	0.272	0.231
$SAB_{Y3}$	0.281	0.240	0.221	0.297	0.216	0.169
$BB_X$	0.145	0.117	0.102	0.085	0.060	0.040
$BB_{XY1}$	0.132	0.114	0.100	0.070	0.052	0.039
$BB_{XY2}$	0.129	0.112	0.102	0.062	0.046	0.038
$BB_{XY3}$	0.136	0.124	0.119	0.059	0.038	0.028
$BB_{Y1}$	0.167	0.153	0.140	0.082	0.057	0.043
$BB_{Y2}$	0.154	0.143	0.138	0.069	0.049	0.040
$BB_{Y3}$	0.154	0.141	0.138	0.064	0.039	0.029
$SBB_X$	0.135	0.102	0.086	0.085	0.060	0.043
$SBB_{XY1}$	0.120	0.100	0.086	0.067	0.055	0.041
$SBB_{XY2}$	0.117	0.099	0.087	0.060	0.048	0.040
$SBB_{XY3}$	0.123	0.105	0.100	0.059	0.038	0.027
$SBB_{Y1}$	0.152	0.134	0.122	0.078	0.056	0.045
$SBB_{Y2}$	0.142	0.128	0.122	0.070	0.051	0.042
$SBB_{Y3}$	0.137	0.119	0.116	0.065	0.040	0.029

Notes: RMSEs based on Monte Carlo simulations with 1000 replications. The other parameters are:  $N = 60$ ,  $n = 5$ ,  $\delta = 0.5$ ,  $\beta = 1$ , and  $\rho = 0.3$ . To meet the stability condition (3),  $\delta$  is set to 0.2 if  $\lambda = 0.7$ .

Table 4: RMSE of Spatial Blundell-Bond Estimators for Various Values of  $\delta$  and  $\rho$ 

Estimator	Par.	$\delta = 0.3$					$\delta = 0.5$					$\delta = 0.7$				
		$\rho = -0.8$	$\rho = -0.4$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = -0.8$	$\rho = -0.4$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = -0.8$	$\rho = -0.4$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$
$BB_X$	$\lambda$	0.109	0.098	0.092	0.098	0.131	0.096	0.082	0.080	0.081	0.116	0.075	0.074	0.065	0.067	0.135
$BB_{XY1}$	$\lambda$	0.102	0.092	0.088	0.093	0.131	0.095	0.085	0.079	0.080	0.107	0.071	0.065	0.061	0.060	0.119
$BB_{XY3}$	$\lambda$	0.098	0.084	0.083	0.078	0.108	0.080	0.073	0.067	0.062	0.095	0.061	0.051	0.048	0.052	0.114
$SBB_X$	$\lambda$	0.077	0.081	0.083	0.081	0.078	0.070	0.066	0.070	0.072	0.070	0.044	0.053	0.053	0.050	0.051
$SBB_{XY1}$	$\lambda$	0.077	0.081	0.085	0.082	0.082	0.066	0.066	0.067	0.071	0.070	0.044	0.052	0.051	0.050	0.046
$SBB_{XY3}$	$\lambda$	0.077	0.080	0.085	0.082	0.077	0.063	0.068	0.065	0.066	0.068	0.040	0.046	0.048	0.045	0.045
$BB_X$	$\delta$	0.187	0.173	0.152	0.200	0.446	0.164	0.145	0.140	0.171	0.347	0.119	0.092	0.098	0.116	0.228
$BB_{XY1}$	$\delta$	0.187	0.163	0.163	0.200	0.423	0.159	0.138	0.134	0.169	0.328	0.105	0.085	0.088	0.107	0.208
$BB_{XY3}$	$\delta$	0.190	0.171	0.176	0.218	0.434	0.164	0.149	0.145	0.180	0.341	0.107	0.089	0.092	0.117	0.216
$SBB_X$	$\delta$	0.135	0.154	0.160	0.166	0.146	0.109	0.133	0.141	0.147	0.130	0.076	0.081	0.099	0.099	0.090
$SBB_{XY1}$	$\delta$	0.131	0.150	0.166	0.163	0.130	0.104	0.125	0.134	0.141	0.117	0.068	0.076	0.088	0.092	0.080
$SBB_{XY3}$	$\delta$	0.133	0.158	0.179	0.174	0.141	0.103	0.135	0.146	0.147	0.123	0.069	0.078	0.093	0.099	0.085
$BB_X$	$\beta$	0.108	0.094	0.094	0.096	0.148	0.106	0.095	0.091	0.097	0.142	0.113	0.097	0.090	0.099	0.136
$BB_{XY1}$	$\beta$	0.106	0.097	0.095	0.097	0.146	0.107	0.096	0.089	0.097	0.144	0.107	0.093	0.087	0.101	0.135
$BB_{XY3}$	$\beta$	0.112	0.099	0.095	0.098	0.139	0.107	0.097	0.093	0.095	0.135	0.108	0.094	0.087	0.099	0.128
$SBB_X$	$\beta$	0.091	0.087	0.094	0.089	0.089	0.087	0.091	0.090	0.091	0.086	0.085	0.092	0.090	0.089	0.087
$SBB_{XY1}$	$\beta$	0.090	0.089	0.096	0.089	0.088	0.087	0.092	0.089	0.090	0.089	0.080	0.093	0.087	0.089	0.086
$SBB_{XY3}$	$\beta$	0.091	0.087	0.096	0.089	0.092	0.086	0.094	0.092	0.089	0.090	0.082	0.091	0.088	0.088	0.086
$BB_X$	$\rho$	0.141	0.148	0.135	0.107	0.093	0.149	0.155	0.138	0.116	0.095	0.137	0.137	0.128	0.117	0.093
$BB_{XY1}$	$\rho$	0.142	0.146	0.135	0.107	0.091	0.151	0.155	0.138	0.114	0.093	0.137	0.135	0.129	0.118	0.092
$BB_{XY3}$	$\rho$	0.141	0.151	0.137	0.109	0.092	0.145	0.154	0.137	0.114	0.092	0.139	0.137	0.129	0.118	0.089

Notes: RMSEs based on Monte Carlo simulations with 1000 replications. The remaining parameters are:  $N = 60$ ,  $T = 5$ ,  $n = 5$ ,  $\lambda = 0.2$ , and  $\beta = 1$ .

Table 5: RMSE of Spatial GMM Estimators of  $\delta$  for Various Weight Matrices and Values of  $\rho$ 

Estimator	$\rho = -0.8$						$\rho = 0.3$						$\rho = 0.8$					
	Contiguity			Distance <sup>d</sup>	Distance <sup>s</sup>	Bucky	Contiguity			Distance <sup>d</sup>	Distance <sup>s</sup>	Bucky	Contiguity			Distance <sup>d</sup>	Distance <sup>s</sup>	Bucky
	$n = 5$	$n = 10$	$n = 15$				$n = 5$	$n = 10$	$n = 15$				$n = 5$	$n = 10$	$n = 15$			
$AB_X$	0.185	0.201	0.214	0.214	0.180	0.317	0.173	0.233	0.270	0.301	0.231	0.201	0.398	0.438	0.460	0.472	0.454	0.461
$AB_{XY1}$	0.182	0.211	0.237	0.171	0.162	0.358	0.183	0.246	0.292	0.288	0.220	0.207	0.384	0.433	0.456	0.467	0.452	0.464
$AB_{XY2}$	0.190	0.224	0.252	0.187	0.152	0.422	0.197	0.276	0.314	0.273	0.207	0.230	0.388	0.439	0.461	0.459	0.445	0.480
$AB_{XY3}$	0.205	0.248	0.281	0.183	0.152	0.475	0.216	0.297	0.326	0.273	0.205	0.245	0.392	0.441	0.464	0.459	0.446	0.494
$AB_{Y1}$	0.352	0.358	0.364	0.213	0.215	0.756	0.331	0.362	0.385	0.342	0.265	0.436	0.478	0.478	0.485	0.479	0.467	0.553
$AB_{Y2}$	0.307	0.304	0.311	0.182	0.168	0.750	0.313	0.351	0.361	0.285	0.221	0.405	0.454	0.473	0.481	0.464	0.451	0.546
$AB_{Y3}$	0.289	0.299	0.302	0.206	0.159	0.754	0.303	0.337	0.355	0.325	0.210	0.379	0.441	0.466	0.475	0.480	0.449	0.545
$SAB_X$	0.133	0.170	0.188	0.162	0.104	0.170	0.161	0.223	0.264	0.306	0.237	0.203	0.157	0.249	0.332	0.462	0.409	0.426
$SAB_{XY1}$	0.130	0.177	0.204	0.146	0.099	0.171	0.168	0.233	0.280	0.302	0.219	0.211	0.153	0.258	0.353	0.451	0.400	0.434
$SAB_{XY2}$	0.141	0.188	0.213	0.146	0.097	0.204	0.177	0.259	0.305	0.284	0.210	0.231	0.161	0.279	0.372	0.440	0.382	0.455
$SAB_{XY3}$	0.150	0.200	0.236	0.146	0.094	0.246	0.193	0.283	0.321	0.286	0.204	0.256	0.166	0.294	0.382	0.441	0.380	0.473
$SAB_{Y1}$	0.246	0.273	0.272	0.165	0.113	0.765	0.305	0.362	0.389	0.353	0.274	0.484	0.262	0.373	0.437	0.472	0.441	0.546
$SAB_{Y2}$	0.210	0.230	0.247	0.148	0.095	0.758	0.291	0.342	0.359	0.298	0.217	0.453	0.228	0.353	0.423	0.447	0.395	0.541
$SAB_{Y3}$	0.198	0.230	0.243	0.191	0.095	0.724	0.281	0.329	0.351	0.317	0.203	0.429	0.214	0.341	0.412	0.467	0.393	0.541
$BB_X$	0.155	0.163	0.154	0.131	0.129	0.357	0.145	0.171	0.189	0.179	0.144	0.178	0.312	0.345	0.358	0.385	0.387	0.414
$BB_{XY1}$	0.138	0.143	0.144	0.114	0.102	0.354	0.132	0.167	0.182	0.146	0.123	0.165	0.282	0.340	0.357	0.373	0.383	0.404
$BB_{XY2}$	0.140	0.140	0.143	0.119	0.097	0.363	0.129	0.169	0.185	0.139	0.114	0.176	0.289	0.339	0.358	0.363	0.373	0.413
$BB_{XY3}$	0.138	0.143	0.143	0.116	0.097	0.382	0.136	0.177	0.184	0.138	0.113	0.178	0.292	0.339	0.355	0.359	0.369	0.421
$BB_{Y1}$	0.184	0.176	0.171	0.127	0.124	0.495	0.167	0.178	0.187	0.165	0.135	0.236	0.300	0.352	0.358	0.387	0.397	0.436
$BB_{Y2}$	0.176	0.162	0.156	0.113	0.101	0.480	0.154	0.181	0.195	0.153	0.122	0.227	0.308	0.350	0.362	0.378	0.386	0.439
$BB_{Y3}$	0.161	0.157	0.148	0.122	0.099	0.475	0.154	0.187	0.192	0.201	0.121	0.223	0.309	0.348	0.364	0.442	0.381	0.443
$SBB_X$	0.105	0.115	0.123	0.097	0.071	0.140	0.135	0.164	0.188	0.175	0.143	0.167	0.117	0.175	0.229	0.318	0.277	0.338
$SBB_{XY1}$	0.098	0.107	0.116	0.089	0.064	0.140	0.120	0.157	0.174	0.145	0.116	0.153	0.103	0.165	0.221	0.290	0.264	0.324
$SBB_{XY2}$	0.097	0.107	0.114	0.090	0.063	0.140	0.117	0.158	0.178	0.136	0.111	0.159	0.098	0.168	0.224	0.269	0.237	0.334
$SBB_{XY3}$	0.093	0.107	0.113	0.092	0.064	0.154	0.123	0.164	0.179	0.137	0.105	0.153	0.097	0.172	0.224	0.259	0.231	0.344
$SBB_{Y1}$	0.126	0.127	0.123	0.098	0.067	0.266	0.152	0.179	0.188	0.171	0.143	0.239	0.126	0.176	0.233	0.321	0.286	0.382
$SBB_{Y2}$	0.114	0.115	0.120	0.088	0.064	0.238	0.142	0.173	0.189	0.148	0.119	0.224	0.115	0.177	0.235	0.299	0.260	0.384
$SBB_{Y3}$	0.110	0.118	0.120	0.117	0.061	0.238	0.137	0.173	0.184	0.179	0.115	0.211	0.114	0.184	0.235	0.395	0.253	0.392

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.

Table 6: RMSE of one-step versus two-step estimators

Estimator	Parameter	$\lambda = 0.3, \delta = 0.5$						$\lambda = 0.7, \delta = 0.2$					
		$N = 60, T = 5$		$N = 60, T = 20$		$N = 500, T = 5$		$N = 60, T = 5$		$N = 60, T = 20$		$N = 500, T = 5$	
		one-step	two-step	one-step	two-step	one-step	two-step	one-step	two-step	one-step	two-step	one-step	two-step
$SAB_X$	$\delta$	0.161	0.171	0.117	0.112	0.053	0.056	0.220	0.225	0.109	0.104	0.079	0.074
$SAB_{XY_3}$	$\delta$	0.193	0.193	0.174	0.171	0.055	0.057	0.210	0.220	0.135	0.132	0.075	0.077
$SAB_{Y_3}$	$\delta$	0.281	0.286	0.221	0.223	0.196	0.187	0.297	0.324	0.169	0.162	0.238	0.230
$SBB_X$	$\delta$	0.135	0.137	0.086	0.083	0.060	0.054	0.085	0.080	0.043	0.043	0.063	0.048
$SBB_{XY_3}$	$\delta$	0.123	0.126	0.100	0.097	0.059	0.062	0.059	0.058	0.027	0.029	0.042	0.041
$SBB_{Y_3}$	$\delta$	0.137	0.146	0.116	0.113	0.110	0.122	0.065	0.070	0.029	0.030	0.053	0.052
$SAB_X$	$\lambda$	0.100	0.107	0.042	0.042	0.033	0.033	0.173	0.180	0.075	0.074	0.053	0.056
$SAB_{XY_3}$	$\lambda$	0.119	0.128	0.068	0.067	0.035	0.036	0.217	0.219	0.110	0.108	0.060	0.064
$SAB_{Y_3}$	$\lambda$	0.118	0.129	0.074	0.073	0.042	0.043	0.201	0.206	0.112	0.110	0.061	0.063
$SBB_X$	$\lambda$	0.070	0.090	0.037	0.039	0.026	0.028	0.086	0.094	0.068	0.068	0.037	0.040
$SBB_{XY_3}$	$\lambda$	0.068	0.068	0.037	0.038	0.028	0.029	0.083	0.084	0.068	0.067	0.043	0.044
$SBB_{Y_3}$	$\lambda$	0.072	0.077	0.035	0.037	0.030	0.031	0.083	0.091	0.068	0.069	0.041	0.041

Notes: RMSEs based on Monte Carlo simulations with 1000 replications. The remaining parameters are  $\beta = 1$  and  $\rho = 0.3$ .

Table 7: RMSE of Original and Restricted Spatial Corrected Estimators for Various Values of  $N$

Estimator	Parameter	$\lambda = 0.3$ and $\delta = 0.5$			$\lambda = 0.7$ and $\delta = 0.2$		
		$N = 60$	$N = 200$	$N = 500$	$N = 60$	$N = 200$	$N = 500$
$SAB_{XY3}$	$\lambda$	0.119	0.058	0.035	0.217	0.107	0.060
$SBB_{XY3}$	$\lambda$	0.068	0.041	0.028	0.083	0.059	0.043
$ISAB_{XY3}$	$\lambda$	0.165	0.078	0.043	0.302	0.140	0.076
$ISBB_{XY3}$	$\lambda$	0.078	0.049	0.036	0.088	0.067	0.050
$SAB_{XY3}$	$\delta$	0.193	0.093	0.055	0.210	0.123	0.075
$SBB_{XY3}$	$\delta$	0.123	0.084	0.059	0.059	0.048	0.042
$ISAB_{XY3}$	$\delta$	0.191	0.099	0.064	0.265	0.146	0.092
$ISBB_{XY3}$	$\delta$	0.133	0.089	0.064	0.064	0.051	0.046
$SAB_{XY3}$	$\beta$	0.102	0.052	0.032	0.123	0.061	0.039
$SBB_{XY3}$	$\beta$	0.090	0.052	0.033	0.087	0.053	0.036
$ISAB_{XY3}$	$\beta$	0.135	0.067	0.040	0.185	0.088	0.050
$ISBB_{XY3}$	$\beta$	0.100	0.061	0.038	0.094	0.060	0.043
$SAB_{XY3}$	$\rho$	0.123	0.064	0.042	0.124	0.063	0.039
$SBB_{XY3}$	$\rho$	0.117	0.065	0.042	0.120	0.063	0.041
$ISAB_{XY3}$	$\rho$	0.119	0.065	0.037	0.121	0.061	0.039
$ISBB_{XY3}$	$\rho$	0.121	0.067	0.037	0.116	0.064	0.039

*Notes:* RMSEs based on Monte Carlo simulations with 1000 replications. The other parameters are:  $T = 5$ ,  $n = 5$ ,  $\beta = 1$ , and  $\rho = 0.3$ .  $ISAB, ISBB$  denote spatially corrected AB and BB estimators with an independent subset of moment conditions.

# Appendix

## A.1 Derivation of Moment Conditions in Stage Two

To arrive at the moment conditions in (25), 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:

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

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

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

Using the above leads to the following moment conditions:

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

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

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

Dividing (A.3)–(A.5) by  $N(T-1)$  gives the moment conditions in (25).

An optimal GMM estimator for this system of moment conditions relies on an optimal GMM weights-matrix. Following the derivation in Kapoor *et al.* (2007) we arrive at the following matrix for the case of first-differences

$$\mathbf{C}_N = 5 \begin{bmatrix} 2 & \frac{2}{N}\text{tr}(\mathbf{M}_N^\top\mathbf{M}_N) & 0 \\ \frac{2}{N}\text{tr}(\mathbf{M}_N^\top\mathbf{M}_N) & \frac{2}{N}\text{tr}(\mathbf{M}_N^\top\mathbf{M}_N\mathbf{M}_N^\top\mathbf{M}_N) & \frac{1}{N}\text{tr}(\mathbf{M}_N^\top\mathbf{M}_N[\mathbf{M}_N^\top + \mathbf{M}_N]) \\ 0 & \frac{1}{N}\text{tr}(\mathbf{M}_N^\top\mathbf{M}_N[\mathbf{M}_N^\top + \mathbf{M}_N]) & \frac{1}{N}\text{tr}(\mathbf{M}_N^\top\mathbf{M}_N + \mathbf{M}_N^\top\mathbf{M}_N) \end{bmatrix}. \quad (\text{A.6})$$

## A.2 Auxiliary lemmas

In this appendix, the law of large numbers in the presence of spatial correlation is verified for the derivative and variance of the sample moment conditions. Their limits form an inherent part of the asymptotic variance of the corresponding estimators.

Let us first recall the extended notation for the spatial matrices aggregated across all time periods: in the case of the Arellano-Bond estimator, let  $\mathbf{I}_{N\otimes} = \mathbf{I}_{T-2} \otimes \mathbf{I}_N$ ,  $\mathbf{M}_{N\otimes} = \mathbf{I}_{T-2} \otimes \mathbf{M}_N$ , and  $\mathbf{W}_{N\otimes} = \mathbf{I}_{T-2} \otimes \mathbf{W}_N$ ; in the case of the Blundell-Bond estimator, let  $\mathbf{I}_{N\otimes} = \mathbf{I}_{2(T-2)} \otimes \mathbf{I}_N$ ,  $\mathbf{M}_{N\otimes} = \mathbf{I}_{2(T-2)} \otimes \mathbf{M}_N$ , and  $\mathbf{W}_{N\otimes} = \mathbf{I}_{2(T-2)} \otimes \mathbf{W}_N$ . Similarly,  $\mathbf{B}_{N\otimes}$  represents  $\mathbf{I}_{T-2} \otimes \mathbf{B}_N$  or  $\mathbf{I}_{2(T-2)} \otimes \mathbf{B}_N$  and  $\Phi_{k,N\otimes} = \mathbf{I}_{T-2} \otimes \Phi_{k,N}$  or  $\Phi_{k,N\otimes} = \mathbf{I}_{2(T-2)} \otimes \Phi_{k,N}$  are defined for  $k = 1, \dots, \varkappa$ . Finally, recall that the error terms  $\varepsilon_N(t)$  and  $\mathbf{u}_N(t)$  represent – depending on the considered estimator –  $\Delta \mathbf{v}_N(t)$  or  $(\Delta \mathbf{v}_N^\top(t), \boldsymbol{\eta}_N^\top + \mathbf{v}_N^\top(t))^\top$  and their spatially correlated counterparts.

**Lemma 1.** *Under Assumptions E and S and V, it holds that  $N^{-1} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{Z}}_N - \mathbb{E}[N^{-1} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{Z}}_N] \rightarrow 0$  in probability as  $N \rightarrow +\infty$ . Similarly,  $N^{-1} \mathbf{H}_N^\top \Phi_{k,N\otimes}^\top \Phi_{k,N\otimes} \mathbf{Z}_N - \mathbb{E}[N^{-1} \mathbf{H}_N^\top \Phi_{k,N\otimes}^\top \Phi_{k,N\otimes} \mathbf{Z}_N] \rightarrow 0$  in probability as  $N \rightarrow +\infty$  for matrices  $\Phi_{k,N}$ ,  $k = 1, \dots, \varkappa$ .*

*Proof:* Noting that  $\tilde{\mathbf{H}}_N = \mathbf{B}_N \mathbf{H}_N$  and  $\tilde{\mathbf{Z}}_N = \mathbf{B}_N \mathbf{Z}_N$ , the proof is the same for any matrix such as  $\Phi_{k,N}$  satisfying the same Assumptions S3 and V imposed on  $\mathbf{B}_N$  and it is therefore done for  $\mathbf{B}_N$  only.

For the simplicity of notation, let  $\lambda$ ,  $\beta$ ,  $\delta$ , and  $\rho$  represent the true parameter values in this proof. First note that  $N^{-1} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{Z}}_N = N^{-1} \{\mathbf{H}_N^\top(t) \mathbf{B}_N^\top \mathbf{B}_N \mathbf{Z}_N(t)\}_{t=3}^T$  so that the result can be proved for each time period separately. Further,  $\mathbf{Z}_N(t)$  and  $\mathbf{H}_N(t)$  consist of  $\mathbf{X}_N(t)$  and  $\mathbf{y}_N(t)$  and their lags, where  $\mathbf{X}_N(t)$  has independent and identically distributed elements,  $\mathbf{y}_N(t) = (\mathbf{I}_N - \delta \mathbf{W}_N)^{-1} \{\lambda \mathbf{y}_N(t-1) + \mathbf{X}_N(t) \beta + \mathbf{u}_N(t)\} = (\mathbf{I}_N - \delta \mathbf{W}_N)^{-1} \{\lambda^{t-1} \mathbf{y}_N(1) + \lambda^{t-2} \mathbf{X}_N(2) \beta + \dots + \lambda^0 \mathbf{X}_N(t) \beta + \lambda^{t-2} \mathbf{u}_N(2) + \dots + \lambda^0 \mathbf{u}_N(t)\}$  and  $\lambda$  and  $\beta$  represent the true parameter values. Using  $\mathbf{u}_N(t) = (\mathbf{I} - \rho \mathbf{M}_N)^{-1} [\boldsymbol{\eta}_N + \mathbf{v}_N(t)]$  and Assumption V, we can express  $\mathbf{Z}_N(t)$  and  $\mathbf{H}_N(t)$  as linear combinations of  $\mathbf{v}_N(1), \mathbf{v}_N(2), \dots, \mathbf{v}_N(t), \mathbf{X}_{jN}(1), \dots, \mathbf{X}_{jN}(t), \boldsymbol{\eta}_N, j = 1, \dots, K$ , where  $\mathbf{X}_{jN}(t)$  represents the  $j$ th column of  $\mathbf{X}_N(t)$ ; for example,  $\mathbf{y}_N(t) = \lambda^{t-1} (\mathbf{I}_N - \delta \mathbf{W}_N)^{-1} \mathbf{S}_N \mathbf{v}_N(1) + \lambda^{t-2} (\mathbf{I}_N - \delta \mathbf{W}_N)^{-1} (\mathbf{I} - \rho \mathbf{M}_N)^{-1} \mathbf{v}_N(2) + \dots + \lambda^0 (\mathbf{I}_N - \delta \mathbf{W}_N)^{-1} (\mathbf{I} - \rho \mathbf{M}_N)^{-1} \mathbf{v}_N(t) + \beta_1 \lambda^{t-2} (\mathbf{I}_N -$



$\delta \mathbf{W}_N)^{-1} \mathbf{X}_{1N}(1) + \dots + \beta_1 \lambda^0 (\mathbf{I}_N - \delta \mathbf{W}_N)^{-1} \mathbf{X}_{1N}^\top(t) + \dots + \beta_K \lambda^{t-2} (\mathbf{I}_N - \delta \mathbf{W}_N)^{-1} \mathbf{X}_{KN}(1) + \dots + \beta_K \lambda^0 (\mathbf{I}_N - \delta \mathbf{W}_N)^{-1} \mathbf{X}_{KN}(t) + (1 - \lambda^{t-1}) / (1 - \lambda) (\mathbf{I} - \rho \mathbf{M}_N)^{-1} \boldsymbol{\eta}_N \}$ . At the same time, all vectors in these linear combinations consist of independent random variables with uniformly bounded second moments and all matrices in these linear combinations have uniformly bounded row and column sums of absolute values of their elements. Moreover, this matrix property is preserved under matrix multiplication (e.g., Kelejian and Prucha, 1999, footnote 20).

Consider now an arbitrary element of  $N^{-1} \mathbf{H}_N^\top(t) \mathbf{B}_N^\top \mathbf{B}_N \mathbf{Z}_N(t)$ :  $N^{-1} \mathbf{h}_N(t)^\top \mathbf{B}_N^\top \mathbf{B}_N \mathbf{z}_N(t)$ , where  $\mathbf{h}_N(t)$  and  $\mathbf{z}_N(t)$  represent columns of  $\mathbf{H}_N(t)$  and  $\mathbf{Z}_N(t)$ , respectively. Labelling the relevant matrices with bounded row and column sums by  $\mathbf{S}_N^{v,j}$  for  $v \in \{h, z\}$ , let  $\mathbf{h}_N(t) = \mathbf{S}_N^{h,1} \mathbf{v}_N(1) + \mathbf{S}_N^{h,2} \mathbf{v}_N(2) + \dots + \mathbf{S}_N^{h,t} \mathbf{v}_N(t) + \mathbf{S}_N^{h,t+1} \mathbf{X}_{1N}(1) + \dots + \mathbf{S}_N^{h,2t} \mathbf{X}_{1N}(t) + \dots + \mathbf{S}_N^{h,Kt+1} \mathbf{X}_{KN}(1) + \dots + \mathbf{S}_N^{h,(K+1)t} \mathbf{X}_{KN}(t) + \mathbf{S}_N^{h,(K+1)t+1} \boldsymbol{\eta}_N$ , similarly  $\mathbf{z}_N(t) = \mathbf{S}_N^{z,1} \mathbf{v}_N(1) + \mathbf{S}_N^{z,2} \mathbf{v}_N(2) + \dots + \mathbf{S}_N^{z,t} \mathbf{v}_N(t) + \mathbf{S}_N^{z,t+1} \mathbf{X}_{1N}(1) + \dots + \mathbf{S}_N^{z,2t} \mathbf{X}_{1N}(t) + \dots + \mathbf{S}_N^{z,Kt+1} \mathbf{X}_{KN}(1) + \dots + \mathbf{S}_N^{z,(K+1)t} \mathbf{X}_{KN}(t) + \mathbf{S}_N^{z,(K+1)t+1} \boldsymbol{\eta}_N$ , and  $(\boldsymbol{\psi}_N^1, \dots, \boldsymbol{\psi}_N^{(K+1)t+1}) \equiv (\mathbf{v}_N(1), \mathbf{v}_N(2), \dots, \mathbf{v}_N(t), \mathbf{X}_{1N}(1), \dots, \mathbf{X}_{1N}(t), \dots, \mathbf{X}_{KN}(1), \dots, \mathbf{X}_{KN}(t), \boldsymbol{\eta}_N)$ . Consequently,  $N^{-1} \mathbf{h}_N(t)^\top \mathbf{B}_N^\top \mathbf{B}_N \mathbf{z}_N(t) = N^{-1} \left( \sum_{c=1}^{(K+1)t+1} \mathbf{S}_N^{h,c} \boldsymbol{\psi}_N^c \right)^\top \mathbf{B}_N^\top \mathbf{B}_N \left( \sum_{c=1}^{(K+1)t+1} \mathbf{S}_N^{z,c} \boldsymbol{\psi}_N^c \right) = N^{-1} \sum_{c,d=1}^{(K+1)t+1} \boldsymbol{\psi}_N^c \left( \mathbf{S}_N^{h,c} \right)^\top \mathbf{B}_N^\top \mathbf{B}_N \left( \mathbf{S}_N^{z,d} \right) \boldsymbol{\psi}_N^d$ . Given that  $K$  and  $t \leq T$  are fixed and finite, we only have to prove the law of large numbers for  $N^{-1} \boldsymbol{\psi}_N^c \boldsymbol{\Pi}_N \boldsymbol{\psi}_N^d$ , where  $\boldsymbol{\Pi}_N$  is an  $N \times N$  matrix with uniformly bounded row and column sums of absolute values of its elements and  $\boldsymbol{\psi}_N^c$  and  $\boldsymbol{\psi}_N^d$  are vectors of independent random variables with uniformly bounded second moments.

Denoting the elements of  $\boldsymbol{\Pi}_N$ ,  $\boldsymbol{\psi}_N^c$ , and  $\boldsymbol{\psi}_N^d$  by  $\pi_{ij}$ ,  $\psi_i^c$ , and  $\psi_k^d$ , respectively, for  $i, j, k = 1, \dots, N$ , the term  $N^{-1} \boldsymbol{\psi}_N^c \boldsymbol{\Pi}_N \boldsymbol{\psi}_N^d$  can be rewritten as

$$N^{-1} \sum_{i=1}^N \psi_i^c \left[ \sum_{k=1}^N \pi_{ik} \psi_k^d \right] = N^{-1} \sum_{i=1}^n \sum_{k=1}^N \psi_i^c \pi_{ik} \psi_k^d.$$

To show that

$$N^{-1} \sum_{i=1}^N \sum_{k=1}^N \{ \psi_i^c \pi_{ik} \psi_k^d - E(\psi_i^c \pi_{ik} \psi_k^d) \} = o_p(1) \tag{A.7}$$

as  $N \rightarrow +\infty$ , note that the expectations  $\max\{|E(\psi_i^c)|, |E(\psi_i^d)|, |E(\psi_i^c \psi_i^d)|\} \leq C$  and variance  $\text{Var}(\psi_i^c) \leq C$  for any  $i = 1, \dots, N$  and  $N \in \mathbb{N}$ , where  $C$  is a positive constant, due to Assumption E and V. Since  $\boldsymbol{\Pi}_N$  has the row and column sums of absolute values of its elements uniformly

bounded by  $D \geq 0$ , the variance of

$$\begin{aligned}
& \text{Var} \left[ N^{-1} \sum_{i=1}^N \sum_{k=1}^N \{ \psi_i^c \pi_{ik} \psi_k^d - E(\psi_i^c \pi_{ik} \psi_k^d) \} \right] \\
&= N^{-2} \sum_{i,j=1}^N \sum_{k,l=1}^N \text{Cov}(\psi_i^c \pi_{ik} \psi_k^d, \psi_j^c \pi_{jl} \psi_l^d) \\
&= N^{-2} \sum_{i=1}^N \sum_{k,l=1}^N \text{Var}(\psi_i^c) \pi_{ik} E(\psi_k^d) \pi_{il} E(\psi_l^d) + N^{-2} \sum_{i=1}^N \sum_{k,l=1}^N \text{Cov}(\psi_i^c, \psi_i^d) \pi_{ik} E(\psi_k^d) \pi_{li} E(\psi_l^d) \\
&+ N^{-2} \sum_{i=1}^N \sum_{k,l=1}^N \text{Cov}(\psi_i^c, \psi_i^d) \pi_{ki} E(\psi_k^c) \pi_{il} E(\psi_l^d) + N^{-2} \sum_{i=1}^N \sum_{k,l=1}^N \text{Var}(\psi_i^d) \pi_{ki} E(\psi_k^c) \pi_{li} E(\psi_l^c) \\
&\leq N^{-2} C^3 \sum_{i=1}^N \sum_{k,l=1}^N (\pi_{ik} \pi_{il} + \pi_{ik} \pi_{li} + \pi_{ki} \pi_{il} + \pi_{ki} \pi_{li}) \\
&\leq N^{-2} C^3 \sum_{i=1}^N \sum_{k=1}^N \left\{ \pi_{ik} \left[ \sum_{l=1}^N \pi_{il} + \sum_{l=1}^N \pi_{li} \right] + \pi_{ki} \left[ \sum_{l=1}^N \pi_{il} + \sum_{l=1}^N \pi_{li} \right] \right\} \\
&\leq N^{-1} C^3 4D^2
\end{aligned}$$

because the elements of  $\psi_i^v$  and  $\psi_j^w$  are independent for  $i \neq j$  and  $v, w \in \{c, d\}$ . The claim (A.7) now follows from the Chebyshev inequality for  $N \rightarrow +\infty$ .  $\square$

**Lemma 2.** *Let the vector forming the  $i$ th row of  $\mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1}$  be  $\tilde{\mathbf{r}}_{iN}$ ,  $\boldsymbol{\xi}_N = N^{-1/2} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N = N^{-1/2} \mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} \boldsymbol{\varepsilon}_N = N^{-1/2} \sum_{i=1}^{NT^*} \tilde{\mathbf{r}}_{iN} \varepsilon_{iN}$ , and  $\tilde{\mathbf{Q}}_{N, H\Sigma H} = E[N^{-1} \mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} \boldsymbol{\varepsilon}_N \boldsymbol{\varepsilon}_N^\top (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1\top} \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} \mathbf{H}_N]$ , where  $NT^*$  represents the length of  $\boldsymbol{\varepsilon}_N$  and  $\varepsilon_N(2), \dots, \varepsilon_N(T^*)$  are assumed to be mutually independent random vectors. Under Assumptions E and S and V, it holds that  $N^{-1} \sum_{i=1}^{NT^*} \tilde{\mathbf{r}}_{iN} \varepsilon_{iN} \varepsilon_{iN}^\top \tilde{\mathbf{r}}_{iN}^\top - \tilde{\mathbf{Q}}_{N, H\Sigma H} \rightarrow 0$  in probability as  $N \rightarrow +\infty$ .*

*Similarly, let the vector forming the  $i$ th row of  $\mathbf{H}_N^\top \boldsymbol{\Phi}_{k, N\otimes}^\top \boldsymbol{\Phi}_{k, N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1}$  be  $\tilde{\mathbf{r}}_{ik, N}$ ,  $\boldsymbol{\zeta}_{k, N} = N^{-1/2} \mathbf{H}_N^\top \boldsymbol{\Phi}_{k, N\otimes}^\top \boldsymbol{\Phi}_{k, N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} \boldsymbol{\varepsilon}_N = N^{-1/2} \sum_{i=1}^{NT^*} \tilde{\mathbf{r}}_{ik, N} \varepsilon_{iN}$ , and  $\tilde{\mathbf{Q}}_{k, N, H\Sigma H} = E[N^{-1} \mathbf{H}_N^\top \boldsymbol{\Phi}_{k, N\otimes}^\top \boldsymbol{\Phi}_{k, N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} \boldsymbol{\varepsilon}_N \boldsymbol{\varepsilon}_N^\top (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1\top} \boldsymbol{\Phi}_{k, N\otimes}^\top \boldsymbol{\Phi}_{k, N\otimes} \mathbf{H}_N]$ . Under Assumptions E and S and V, it holds that  $N^{-1} \sum_{i=1}^{NT^*} \tilde{\mathbf{r}}_{ik, N} \varepsilon_{iN} \varepsilon_{iN}^\top \tilde{\mathbf{r}}_{ik, N}^\top - \tilde{\mathbf{Q}}_{k, N, H\Sigma H} \rightarrow 0$  in probability as  $N \rightarrow +\infty$ .*

*Proof:* The two results have an identical form and hold for any matrix  $\mathbf{B}_N$  that satisfies Assumptions S3 and V, and in particular, for matrices  $\Phi_{k,N}$ ,  $k = 1, \dots, \varkappa$ . The proof is done for matrix  $\mathbf{B}_N$ .

Given that  $\boldsymbol{\xi}_N = N^{-1/2} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N$  is a finite-dimensional vector, we will prove the result elementwise after introducing an additional notation and decomposition. Consider now an arbitrary element of  $N^{-1/2} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N$ :  $N^{-1/2} \tilde{\mathbf{h}}_{jN}^\top \tilde{\mathbf{u}}_N$ , where  $\tilde{\mathbf{h}}_{jN}(t)$  represents the  $j$ th column of  $\tilde{\mathbf{H}}_N$ . Using the block diagonal structure of  $\tilde{\mathbf{H}}_N$ , it follows that

$$N^{-1/2} \tilde{\mathbf{h}}_{jN}^\top \tilde{\mathbf{u}}_N = N^{-1/2} \sum_{t=3}^{T^*} \mathbf{h}_{jN}^\top(t) \mathbf{B}_N^\top \mathbf{B}_N \mathbf{u}_N(t) = N^{-1/2} \sum_{t=3}^{T^*} \mathbf{h}_{jN}^\top(t) \mathbf{B}_N^\top \mathbf{B}_N (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1} \boldsymbol{\varepsilon}_N(t)$$

and

$$N^{-1} \tilde{\mathbf{h}}_{jN}^\top \tilde{\mathbf{u}}_N \tilde{\mathbf{u}}_N^\top \tilde{\mathbf{h}}_{kN} = \sum_{s,t=3}^{T^*} N^{-1} \mathbf{h}_{jN}^\top(t) \mathbf{B}_N^\top \mathbf{B}_N (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1} \boldsymbol{\varepsilon}_N(t) \boldsymbol{\varepsilon}_N^\top(s) \times (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1\top} \mathbf{B}_N^\top \mathbf{B}_N \mathbf{h}_{kN}(s). \quad (\text{A.8})$$

(Note that  $\mathbf{h}_{jN}(t) \mathbf{B}_N^\top \mathbf{B}_N (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1}$  equals the  $j$ th row of  $\{(\tilde{\mathbf{r}}_{iN})_{i=(t-1)N+1}^{tN}\}$  for some  $t$ .) By the same notation and argument, the  $jk$ th element of variance matrix  $\tilde{\mathbf{Q}}_{N,H\Sigma H} = \mathbb{E}[N^{-1} \mathbf{h}_{jN}^\top \mathbf{B}_N^\top \mathbf{B}_N (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1} \boldsymbol{\varepsilon}_N \boldsymbol{\varepsilon}_N^\top (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1\top} \mathbf{B}_N^\top \mathbf{B}_N \mathbf{h}_{kN}]$  can also be rewritten as

$$\sum_{s,t=3}^{T^*} N^{-1} \mathbb{E}[\mathbf{h}_{jN}^\top(t) \mathbf{B}_N^\top \mathbf{B}_N (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1} \boldsymbol{\varepsilon}_N(t) \boldsymbol{\varepsilon}_N^\top(s) (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1\top} \mathbf{B}_N^\top \mathbf{B}_N \mathbf{h}_{kN}(s)]. \quad (\text{A.9})$$

We can therefore prove that elements of the sum (A.8) converge to the elements of the sum (A.9).

Consider now arbitrary, but fixed  $s, t = 3, \dots, T^*$ , and for the simplicity of notation, let  $\lambda, \beta, \delta$ , and  $\rho$  now represent the true parameter values in this proof. We have shown in Lemma 1 that one can express  $\mathbf{h}_{jN}(t) = \mathbf{S}_N^{j,1} \mathbf{v}_N(1) + \mathbf{S}_N^{j,2} \mathbf{v}_N(2) + \dots + \mathbf{S}_N^{j,t} \mathbf{v}_N(t) + \mathbf{S}_N^{j,t+1} \mathbf{X}_{1N}(1) + \dots + \mathbf{S}_N^{j,2t} \mathbf{X}_{1N}(t) + \dots + \mathbf{S}_N^{j,Kt+1} \mathbf{X}_{KN}(1) + \dots + \mathbf{S}_N^{j,(K+1)t} \mathbf{X}_{KN}(t) + \mathbf{S}_N^{j,(K+1)t+1} \boldsymbol{\eta}_N$  using  $N \times N$  matrices  $\mathbf{S}_N^{j,s}$  with bounded row and column sums of the absolute values of their elements. Let us again denote  $(\boldsymbol{\psi}_N^1, \dots, \boldsymbol{\psi}_N^{(K+1)t+1}) \equiv (\mathbf{v}_N(1), \mathbf{v}_N(2), \dots, \mathbf{v}_N(t), \mathbf{X}_{1N}(1), \dots, \mathbf{X}_{1N}(t), \dots, \mathbf{X}_{KN}(1), \dots, \mathbf{X}_{KN}(t), \boldsymbol{\eta}_N)$ , where the elements of each vector  $\boldsymbol{\psi}_N^j$  are mutually independent, though not necessarily identically distributed. The same property holds also for  $\boldsymbol{\varepsilon}_N(t)$ , which equals  $\Delta \mathbf{v}_N(t)$  or  $(\Delta \mathbf{v}_n^\top(t), \boldsymbol{\eta}_N^\top +$

$\mathbf{v}_N^\top(t)^\top$  depending on the estimator and which thus has independent random elements. The scalar  $N^{-1/2}\mathbf{h}_{jN}^\top(t)\mathbf{B}_N^\top\mathbf{B}_N(\mathbf{I}_N - \rho^0\mathbf{M}_N)^{-1}\boldsymbol{\varepsilon}_N(t)$  can thus be rewritten as

$$N^{-1/2} \sum_{c=1}^{(K+1)t+1} \boldsymbol{\psi}_N^c \left( \mathbf{S}_N^{j,c} \right)^\top \mathbf{B}_N^\top \mathbf{B}_N (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1} \boldsymbol{\varepsilon}_N(t).$$

Recall that  $N^{-1/2}\mathbf{h}_{jN}^\top(t)\mathbf{B}_N^\top\mathbf{B}_N(\mathbf{I}_N - \rho^0\mathbf{M}_N)^{-1}\boldsymbol{\varepsilon}_N(t)$  is equal to the  $j$ th row of  $\{(\tilde{\mathbf{r}}_{iN})_{i=(t-1)N+1}^{tN}\}$ , that is, to  $\{(r_{ijN})_{i=(t-1)N+1}^{tN}\}$  for some  $t$ . Thus, the  $j$ th element of  $N^{-1/2} \sum_{i=1}^{NT^*} \tilde{\mathbf{r}}_{iN} \boldsymbol{\varepsilon}_{iN}$  can be expressed as

$$\begin{aligned} N^{-1/2} \sum_{i=1}^{NT^*} \tilde{r}_{ijN} \boldsymbol{\varepsilon}_{iN} &= N^{-1/2} \sum_{t=1}^{T^*} \sum_{i=1}^N \tilde{r}_{[(t-1)N+i]jN} \boldsymbol{\varepsilon}_{iN} \\ &= N^{-1/2} \sum_{t=1}^{T^*} \sum_{c=1}^{(K+1)t+1} \boldsymbol{\psi}_N^c \left( \mathbf{S}_N^{j,c} \right)^\top \mathbf{B}_N^\top \mathbf{B}_N (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1} \boldsymbol{\varepsilon}_N(t) \end{aligned}$$

and

$$\begin{aligned} N^{-1/2} \sum_{i=1}^{NT^*} \tilde{r}_{ijN} \boldsymbol{\varepsilon}_{iN} \boldsymbol{\varepsilon}_{iN}^\top \tilde{r}_{ikN} & \tag{A.10} \\ &= N^{-1/2} \sum_{t=1}^{T^*} \sum_{c,d=1}^{(K+1)t+1} \boldsymbol{\psi}_N^{c^\top} \left( \mathbf{S}_N^{j,c} \right)^\top \mathbf{B}_N^\top \mathbf{B}_N (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1} \boldsymbol{\varepsilon}_N(t) \boldsymbol{\varepsilon}_N^\top(t) \times \\ & \quad \times (\mathbf{I}_N - \rho^0 \mathbf{M}_N)^{-1\top} \mathbf{B}_N^\top \mathbf{B}_N \left( \mathbf{S}_N^{j,d} \right) \boldsymbol{\psi}_N^d. \end{aligned}$$

Given that  $K$  and  $T^* \leq 2T$  are fixed and finite, we only have to prove the law of large numbers for each  $t$  and  $c = 1, \dots, (K+1)t+1$  separately, that is, only for the sequences of random variables  $\boldsymbol{\psi}_N^c \boldsymbol{\Pi}_N^c \boldsymbol{\varepsilon}_N(t)$ , where  $\boldsymbol{\Pi}_N$  is again an  $N \times N$  matrix with uniformly bounded row and column sums of the absolute values of its elements and  $\boldsymbol{\psi}_N^c$  and  $\boldsymbol{\varepsilon}_N(t)$  are vectors of independent random variables with uniformly bounded second moments.

Denoting the elements of  $\boldsymbol{\Pi}_N^c$ ,  $\boldsymbol{\psi}_N^c$ , and  $\boldsymbol{\varepsilon}_N(t)$  by  $\pi_{ij}^c$ ,  $\psi_k^c$ , and  $\varepsilon_k(t)$ , respectively, for  $i, j, k = 1, \dots, N$ , the term  $N^{-1/2}\boldsymbol{\psi}_N^c \boldsymbol{\Pi}_N^c \boldsymbol{\varepsilon}_N(t)$  can be rewritten as

$$N^{-1/2} \sum_{i=1}^N \psi_i^c \left[ \sum_{k=1}^N \pi_{ik}^c \varepsilon_k(t) \right] = N^{-1/2} \sum_{i=1}^N \sum_{k=1}^N \psi_i^c \pi_{ik}^c \varepsilon_k(t)$$

and the term corresponding to (A.10) as

$$N^{-1} \sum_{i=1}^N \sum_{k,l=1}^N \psi_i^c \pi_{ik}^c \varepsilon_k(t) \cdot \varepsilon_l(t) \pi_{il}^d \psi_i^d.$$

To show that

$$N^{-1} \sum_{i=1}^N \sum_{k,l=1}^N \{\psi_i^c \pi_{ik}^c \varepsilon_k(t) \cdot \varepsilon_l(t) \pi_{il}^d \psi_i^d - E(\psi_i^c \pi_{ik}^c \varepsilon_k(t) \cdot \varepsilon_l(t) \pi_{il}^d \psi_i^d)\} = o_p(1), \quad (\text{A.11})$$

note that the expectations  $\max\{|E(\psi_i^c)|, |E(\psi_i^d)|, |E(\psi_i^c \psi_i^d)|, |E(\varepsilon_i(t))|\} \leq C$  and variances  $\max\{\text{Var}(\varepsilon_i(t)), \text{Var}(\psi_i^c), \text{Var}(\psi_i^d)\} \leq C$  are bounded by a positive constant  $C$  for any  $i = 1, \dots, N$ ,  $t = 3, \dots, T$ , and  $N \in \mathbb{N}$  due to Assumption E and V. Since  $\mathbf{\Pi}_N^c$  and  $\mathbf{\Pi}_N^d$  have the row and column sums of absolute values of its elements uniformly bounded by  $D \geq 0$ , we can express and bound the variance of (A.11)

$$\begin{aligned} & \text{Var} \left[ N^{-1} \sum_{i=1}^N \sum_{k,l=1}^N \{\psi_i^c \pi_{ik}^c \varepsilon_k(t) \cdot \varepsilon_l(t) \pi_{il}^d \psi_i^d - E(\psi_i^c \pi_{ik}^c \varepsilon_k(t) \cdot \varepsilon_l(t) \pi_{il}^d \psi_i^d)\} \right] \\ &= N^{-2} \sum_{i,i'=1}^N \sum_{k,l,k',l'=1}^N \text{Cov}(\psi_i^c \pi_{ik}^c \varepsilon_k(t) \cdot \varepsilon_l(t) \pi_{il}^d \psi_i^d, \psi_{i'}^c \pi_{i'k'}^c \varepsilon_{k'}(t) \cdot \varepsilon_{l'}(t) \pi_{i'l'}^d \psi_{i'}^d). \end{aligned}$$

Using the fact that the covariances within the sums can be nonzero only if  $k = k'$  and  $l = l'$  or  $k = l$  and  $k' = l'$  (errors  $\{\varepsilon_k(t)\}_{k=1}^N$  being independent with zero mean) and  $i = i'$  (all vectors  $\psi_i^j$

have mutually independent components), we can further simplify this as

$$\begin{aligned}
&= N^{-2} \sum_{i,i'=1}^N \sum_{k,l=1}^N \text{Cov}(\psi_i^c \pi_{ik}^c \varepsilon_k(t) \cdot \varepsilon_l(t) \pi_{il}^d \psi_i^d, \psi_{i'}^c \pi_{i'k}^c \varepsilon_k(t) \cdot \varepsilon_l(s) \pi_{i'l}^d \psi_{i'}^d) \\
&+ N^{-2} \sum_{i,i'=1}^N \sum_{k,k'=1}^N \text{Cov}(\psi_i^c \pi_{ik}^c \varepsilon_k(t) \cdot \varepsilon_k(t) \pi_{ik}^d \psi_i^d, \psi_{i'}^c \pi_{i'k'}^c \varepsilon_{k'}(t) \cdot \varepsilon_{k'}(s) \pi_{i'k'}^d \psi_{i'}^d) \\
&\leq C \cdot N^{-2} \sum_{i=1}^N \sum_{k,l=1}^N \pi_{ik}^c \pi_{il}^d \pi_{ik}^c \pi_{il}^d + C \cdot N^{-2} \sum_{i=1}^N \sum_{k,k'=1}^N \pi_{ik}^c \pi_{ik}^d \pi_{i'k'}^c \pi_{i'k'}^d \\
&= C \cdot N^{-2} \sum_{i=1}^N \sum_{k=1}^N (\pi_{ik}^c)^2 \sum_{l=1}^N (\pi_{il}^d)^2 + C \cdot N^{-2} \sum_{i=1}^N \sum_{k=1}^N \pi_{ik}^c \pi_{ik}^d \sum_{k'=1}^N \pi_{i'k'}^c \pi_{i'k'}^d \\
&\leq CD^2 \cdot N^{-2} \sum_{i=1}^N \sum_{k=1}^N (\pi_{ik}^c)^2 + CD^2 \cdot N^{-2} \sum_{i=1}^N \sum_{k=1}^N \pi_{ik}^c \pi_{ik}^d \\
&\leq CD^2 \cdot N^{-2} \cdot ND^2 + CD^2 \cdot N^{-2} \cdot ND^2 = N^{-1} CD^4,
\end{aligned}$$

where the last inequalities used the Cauchy-Schwartz inequality and Assumption S. The claim (A.7) now follows from the Chebyshev inequality for  $N \rightarrow +\infty$ .  $\square$

### A.3 Proofs of Asymptotic Properties

#### A.3.1 Proof of Theorem 1

Definitions (14) or (22) and models (9) or (17) imply

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

the expression depends implicitly on the transformation matrix  $\mathbf{B}_N$ , which is however fixed for any given  $N \in \mathbb{N}$ . Since Assumptions G and Lemma 1 imply  $\mathbf{A}_N = \mathbf{A} + o_p(1)$  and  $N^{-1} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{Z}}_N = N^{-1} \mathbf{H}_N^\top \mathbf{B}_N^\top \mathbf{B}_N \mathbf{Z}_N = \tilde{\mathbf{Q}}_{HZ} + o_p(1)$ , matrix  $\tilde{\mathbf{Z}}_N^\top \tilde{\mathbf{H}}_N \mathbf{A}_N \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{Z}}_N$  is thus non-singular for a sufficiently

large  $N$ . It follows from definition (4) that

$$\begin{aligned}\sqrt{N}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) &= \left[ \frac{1}{N} \tilde{\mathbf{Z}}_N^\top \tilde{\mathbf{H}}_N \cdot \mathbf{A}_N \cdot \frac{1}{N} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{Z}}_N \right]^{-1} \frac{1}{N} \tilde{\mathbf{Z}}_N^\top \tilde{\mathbf{H}}_N \cdot \mathbf{A}_N \cdot \frac{1}{\sqrt{N}} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N \\ &= \left[ \tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \tilde{\mathbf{Q}}_{HZ} \right]^{-1} \tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \cdot \left\{ N^{-1/2} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N \right\} + o_p(1)\end{aligned}\quad (\text{A.12})$$

as  $N \rightarrow +\infty$ . Let  $\boldsymbol{\xi}_N = N^{-1/2} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N = N^{-1/2} \mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} \boldsymbol{\varepsilon}_N$ , where  $\boldsymbol{\varepsilon}_N = \Delta \mathbf{v}_N$  or  $\boldsymbol{\varepsilon}_N = (\Delta \mathbf{v}_N^\top, \boldsymbol{\eta}_N^\top + \mathbf{v}_N^\top)^\top$  depending on the estimator; the length of  $\boldsymbol{\varepsilon}_N$  is denoted  $T^*$ . Denoting the vector forming the  $i$ th column of  $\mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1}$  by  $\tilde{\mathbf{r}}_{iN}$ ,  $\boldsymbol{\xi}_N = N^{-1/2} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N = N^{-1/2} \sum_{i=1}^{NT^*} \tilde{\mathbf{r}}_{iN} \varepsilon_{iN}$  and the triangular array  $\{\tilde{\mathbf{r}}_{iN} \varepsilon_{iN}\}_{i=1}^{NT^*}$  has zero mean since the instruments are constructed so that  $E[\tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N] = 0$  and thus  $E[\tilde{\mathbf{r}}_{iN} \varepsilon_{iN}] = 0$ . Suppose now that  $\{\varepsilon_{iN}\}_{i=1}^{NT^*}$  are ordered in the following way:  $(\Delta \mathbf{v}_N^\top(3), \dots, \Delta \mathbf{v}_N^\top(2\lceil T/2 \rceil - 1))^\top$  in the case of the Arellano-Bond estimator (only odd times are used),  $(\mathbf{v}_N^\top(3), \dots, \mathbf{v}_N^\top(T))^\top$  in the case of the estimator based on the level equation only (all times are used), or  $(\Delta \mathbf{v}_N^\top(3), \boldsymbol{\eta}_N^\top + \mathbf{v}_N^\top(4), \Delta \mathbf{v}_N^\top(6), \dots, \boldsymbol{\eta}_N^\top + \mathbf{v}_N^\top(T))^\top$  in the case of the Blundell-Bond estimator, for instance. Noting that  $\mathbf{v}_1(3), \dots, \mathbf{v}_N(T)$  are independent by Assumption E,  $\boldsymbol{\varepsilon}_N(1), \dots, \boldsymbol{\varepsilon}_N(T^*)$  are independent as well and the triangular array  $\{\tilde{\mathbf{r}}_{iN} \varepsilon_{iN}\}_{i=1}^{NT^*}$  forms a sequence of martingale differences: first, by the construction of the instruments,  $E[\boldsymbol{\varepsilon}_N(t) | \tilde{\mathbf{H}}_N^\top(t), \dots, \tilde{\mathbf{H}}_N^\top(3)] = 0$ , which implies  $\tilde{\mathbf{H}}_N^\top(t) \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} E[\boldsymbol{\varepsilon}_N(t) | \tilde{\mathbf{H}}_N^\top(t), \dots, \tilde{\mathbf{H}}_N^\top(3)] = 0$  (and similarly for  $\boldsymbol{\eta}_N$ ); and second, random vectors  $\boldsymbol{\eta}_N$ ,  $\mathbf{v}_N(t)$ , and  $\Delta \mathbf{v}_N(t)$  have mutually independent elements, which implies independence of elements of  $\boldsymbol{\varepsilon}_N(t)$ . Consequently,  $\text{Var}(\boldsymbol{\xi}_N) = N^{-1/2} \sum_{i=1}^N \text{Var}(\tilde{\mathbf{r}}_{iN} \varepsilon_{iN})$  and  $\boldsymbol{\xi}_N$  thus has a bounded variance matrix since for  $N \rightarrow +\infty$

$$\begin{aligned}\text{Var}[N^{-1/2} \mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} \boldsymbol{\varepsilon}_N] \\ &= E[N^{-1} \mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} \boldsymbol{\varepsilon}_N \boldsymbol{\varepsilon}_N^\top (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1\top} \mathbf{B}_{N\otimes}^\top \mathbf{B}_{N\otimes} \mathbf{H}_N] \\ &= \tilde{\mathbf{Q}}_{N,H\Sigma H} \rightarrow \tilde{\mathbf{Q}}_{H\Sigma H}\end{aligned}\quad (\text{A.13})$$

by Assumption V. Finally,  $N^{-1} \sum_{i=1}^{N(T-2)} \tilde{\mathbf{r}}_{iN} \varepsilon_{iN} \varepsilon_{iN} \tilde{\mathbf{r}}_{iN}^\top - \tilde{\mathbf{Q}}_{N,H\Sigma H} \xrightarrow{P} 0$  for  $N \rightarrow +\infty$  by Lemma 2. Consequently, the finite second moments and uniform integrability of  $\hat{\mathbf{Q}}_{N,H\Sigma H}^{-1/2} \boldsymbol{\xi}_N$  (implied

by Assumptions V3 and Assumption S3 by the same argument as in Lemma 2) allows us to apply the central limit theorem for martingale differences (e.g., Davidson, 2000, Theorem 6.2.3, or Davidson, 1994, Theorems 24.3 and 24.4), which implies that  $\boldsymbol{\xi}_N$  converges in distribution to the Gaussian distribution with zero mean and the finite asymptotic variance matrix  $\tilde{\mathbf{Q}}_{H\Sigma H}$  as  $n \rightarrow +\infty$ . Consequently,  $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) = O_p(1)$  and

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

as  $N \rightarrow +\infty$ , where  $\mathcal{L}$  denotes convergence in distribution.  $\square$

### A.3.2 Proof of Theorem 2

The proof is similar to the one of Kelejian and Prucha (2010, Theorem 1). First, the GMM estimator (28) is based on the vector  $\boldsymbol{\gamma}_N$  and matrix  $\boldsymbol{\Gamma}_N$  defined in (26)–(27). 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  $\mathbf{D}_N = \hat{\mathbf{M}}_N^{k\top} \hat{\mathbf{M}}_N^l$  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,

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

Assumption 4 of Kelejian and Prucha (2010, Lemma C.1) requires that  $\Delta \mathbf{Z}_N$  has 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  $t = 2, \dots, T$ ,  $\Delta \mathbf{u}_N(t) = (\mathbf{I}_N - \rho^0 \mathbf{M}_N) \Delta \boldsymbol{\varepsilon}_N(t)$ , 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:  $E[\Delta \mathbf{u}_N^\top(t) \mathbf{D}_N \Delta \mathbf{u}_N(t)]/N$  is uniformly bounded,



$\Delta \mathbf{u}_N^\top(t) \mathbf{D}_N \Delta \mathbf{u}_N(t) / N - \mathbb{E}[\Delta \mathbf{u}_N^\top(t) \mathbf{D}_N \Delta \mathbf{u}_N(t)] / N = o_p(1)$ , and

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

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

To prove the consistency of the GMM estimator (28), 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  $J_N(\boldsymbol{\varphi}) = \{\mathbb{E} \gamma_N - \mathbb{E} \mathbf{\Gamma}_N \nu(\boldsymbol{\varphi})\}^\top \mathbf{C}_N \{\mathbb{E} \gamma_N - \mathbb{E} \mathbf{\Gamma}_N \nu(\boldsymbol{\varphi})\}$  at  $\boldsymbol{\varphi}^0$  is identifiably unique, and (iii) the sample objective function  $\hat{J}_N(\boldsymbol{\varphi}) = \{\hat{\gamma}_N - \hat{\mathbf{\Gamma}}_N \nu(\boldsymbol{\varphi})\}^\top \hat{\mathbf{C}}_N \{\hat{\gamma}_N - \hat{\mathbf{\Gamma}}_N \nu(\boldsymbol{\varphi})\}$  converges uniformly to  $J_N(\boldsymbol{\varphi})$ , where  $\nu(\boldsymbol{\varphi}) = (\rho, \rho^2, \sigma_v^2)^\top$ ,  $\boldsymbol{\varphi} = (\rho, \sigma_v)^\top$ , and  $\boldsymbol{\varphi}^0 = (\rho^0, \sigma_v^0)^\top$ . First, the existence of the GMM estimate follows from the continuity of  $\hat{J}_N(\boldsymbol{\varphi})$ : it is continuous in  $\boldsymbol{\varphi}$  on a compact space  $\Phi$  and it thus attains its minimum.

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

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

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

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

$$\begin{aligned}\hat{J}_N(\varphi) - J_N(\varphi) &= (\gamma_N^\top \hat{\mathbf{C}}_N \gamma_N - \mathbf{E} \gamma_N^\top \mathbf{C}_N \mathbf{E} \gamma_N) - 2(\hat{\gamma}_N^\top \hat{\mathbf{C}}_N \hat{\Gamma}_N - \mathbf{E} \gamma_N^\top \mathbf{C}_N \mathbf{E} \Gamma_N) \nu(\varphi) \\ &+ \nu(\varphi)^\top (\hat{\Gamma}_N^\top \hat{\mathbf{C}}_N \hat{\Gamma}_N - \mathbf{E} \Gamma_N^\top \mathbf{C}_N \mathbf{E} \Gamma_N) \nu(\varphi),\end{aligned}$$

and  $\varphi \in \Phi$ , where  $\Phi$  is compact,  $\|\varphi\| < K_\varphi < +\infty$ , we only have to show that the three differences of the type  $\hat{\Gamma}_N^\top \hat{\mathbf{C}}_N \hat{\Gamma}_N - \mathbf{E} \Gamma_N^\top \mathbf{C}_N \mathbf{E} \Gamma_N = o_p(1)$  as  $N \rightarrow +\infty$ . This however directly follows from our previous results: we showed that  $\Gamma_N - \mathbf{E} \Gamma_N = o_p(1)$ ,  $\gamma_N - \mathbf{E} \gamma_N = o_p(1)$ ,  $\hat{\Gamma}_N - \mathbf{E} \Gamma_N = o_p(1)$ , and  $\hat{\gamma}_N - \mathbf{E} \gamma_N = o_p(1)$ , all these random variables are bounded in probability (see Assumption G), the expectations  $\mathbf{E} \Gamma_N$  and  $\mathbf{E} \gamma_N$  were shown to be uniformly bounded,  $\hat{\mathbf{C}}_N - \mathbf{C}_N = o_p(1)$  by Assumptions G6, and therefore, the claim follows from the equality

$$\begin{aligned}\hat{\Gamma}_N^\top \hat{\mathbf{C}}_N \hat{\Gamma}_N - \mathbf{E} \Gamma_N^\top \mathbf{C}_N \mathbf{E} \Gamma_N &= (\hat{\Gamma}_N^\top - \mathbf{E} \Gamma_N^\top) \hat{\mathbf{C}}_N \hat{\Gamma}_N + \mathbf{E} \Gamma_N^\top \hat{\mathbf{C}}_N (\hat{\Gamma}_N - \mathbf{E} \Gamma_N) \\ &+ \mathbf{E} \Gamma_N^\top (\hat{\mathbf{C}}_N - \mathbf{C}_N) \mathbf{E} \Gamma_N.\end{aligned}$$

Hence, Lemma 3.1 of Pötscher and Prucha (1997) implies consistency of the estimate (28).  $\square$

### A.3.3 Proof of Theorem 3

Definition (23),  $\hat{\mathbf{y}}_N(t) = \hat{\mathbf{B}}_N \mathbf{y}_N(t)$ ,  $\hat{\mathbf{z}}_N(t) = \hat{\mathbf{B}}_N \mathbf{z}_N(t)$ , and models (9) or (17) imply

$$\hat{\boldsymbol{\theta}}_N = \left[ \hat{\mathbf{z}}_N^\top \hat{\mathbf{H}}_N \hat{\mathbf{A}}_N \hat{\mathbf{H}}_N^\top \hat{\mathbf{z}}_N \right]^{-1} \hat{\mathbf{z}}_N^\top \hat{\mathbf{H}}_N \hat{\mathbf{A}}_N \hat{\mathbf{H}}_N^\top \hat{\mathbf{y}}_N = \boldsymbol{\theta}^0 + \left[ \hat{\mathbf{z}}_N^\top \hat{\mathbf{H}}_N \hat{\mathbf{A}}_N \hat{\mathbf{H}}_N^\top \hat{\mathbf{z}}_N \right]^{-1} \hat{\mathbf{z}}_N^\top \hat{\mathbf{H}}_N \hat{\mathbf{A}}_N \hat{\mathbf{H}}_N^\top \hat{\mathbf{u}}_N,$$

where  $\hat{\mathbf{u}}_N = \hat{\mathbf{y}}_N - \hat{\mathbf{Z}}_N \boldsymbol{\theta} = \hat{\mathbf{B}}_{N \otimes} \mathbf{u}_N$ . First, note that the consistency of  $\hat{\rho}_N \rightarrow \rho^0$  and Assumption V imply

$$\begin{aligned}
N^{-1} \hat{\mathbf{H}}_N^\top \hat{\mathbf{Z}}_N &= N^{-1} \mathbf{H}_N^\top \hat{\mathbf{B}}_{N \otimes}^\top \hat{\mathbf{B}}_{N \otimes} \mathbf{Z}_N \\
&= N^{-1} \mathbf{H}_N^\top \left( \sum_{k=1}^{\infty} \hat{\phi}_{k,N} \boldsymbol{\Phi}_{k,N \otimes} \right)^\top \left( \sum_{k=1}^{\infty} \hat{\phi}_{k,N} \boldsymbol{\Phi}_{k,N \otimes} \right) \mathbf{Z}_N \\
&= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \hat{\phi}_{k,N} \hat{\phi}_{l,N} \cdot N^{-1} \mathbf{H}_N^\top \boldsymbol{\Phi}_{k,N \otimes}^\top \boldsymbol{\Phi}_{l,N \otimes} \mathbf{Z}_N \\
&= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \phi_k^0 \phi_l^0 \cdot \tilde{\mathbf{Q}}_{k,HZ} + o_p(1) = \tilde{\mathbf{Q}}_{HZ} + o_p(1), \tag{A.14}
\end{aligned}$$

where the last equality follows from Lemma 1. Matrix  $N^{-1} \hat{\mathbf{H}}_N^\top \hat{\mathbf{Z}}_N$ , which is non-singular by Assumptions S2, V, and G4, thus converges to a non-singular matrix  $\tilde{\mathbf{Q}}_{HZ}$  in probability. Assumptions G and V further imply that  $\hat{\mathbf{A}}_N = \mathbf{A} + o_p(1)$  and that  $\hat{\mathbf{Z}}_N^\top \hat{\mathbf{H}}_N \hat{\mathbf{A}}_N \hat{\mathbf{H}}_N^\top \hat{\mathbf{Z}}_N$  is non-singular for any sufficiently large  $N$ .

Next, using definition (4) results in  $\hat{\mathbf{u}}_N = \hat{\mathbf{B}}_{N \otimes} \mathbf{u}_N = \mathbf{B}_{N \otimes} \mathbf{u}_N + (\hat{\mathbf{B}}_{N \otimes} - \mathbf{B}_{N \otimes}) \mathbf{u}_N = \mathbf{B}_{N \otimes} \mathbf{u}_N + \sum_{k=1}^{\infty} (\hat{\phi}_{k,N} - \phi_k^0) \boldsymbol{\Phi}_{k,N \otimes} \mathbf{u}_N$ . We can thus write

$$\begin{aligned}
\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) &= \left[ \frac{1}{N} \hat{\mathbf{Z}}_N^\top \hat{\mathbf{H}}_N \cdot \hat{\mathbf{A}}_N \cdot \frac{1}{N} \hat{\mathbf{H}}_N^\top \hat{\mathbf{Z}}_N \right]^{-1} \frac{1}{N} \hat{\mathbf{Z}}_N \hat{\mathbf{H}}_N^\top \hat{\mathbf{A}}_N \frac{1}{\sqrt{N}} \hat{\mathbf{H}}_N^\top \hat{\mathbf{u}} \\
&= \left[ \tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \tilde{\mathbf{Q}}_{HZ} \right]^{-1} \tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \cdot \frac{1}{\sqrt{N}} \hat{\mathbf{H}}_N^\top \hat{\mathbf{u}}_N + o_p(1), \tag{A.15}
\end{aligned}$$

where the last element of the product can be decomposed as

$$\begin{aligned}\frac{1}{\sqrt{N}}\hat{\mathbf{H}}_N^\top\hat{\mathbf{u}}_N &= \frac{1}{\sqrt{N}}\mathbf{H}_N^\top\hat{\mathbf{B}}_{N\otimes}^\top\hat{\mathbf{B}}_{N\otimes}\mathbf{u}_N \\ &= \frac{1}{\sqrt{N}}\mathbf{H}_N^\top\{\mathbf{B}_{N\otimes} + (\hat{\mathbf{B}}_{N\otimes} - \mathbf{B}_{N\otimes})\}^\top\{\mathbf{B}_{N\otimes} + (\hat{\mathbf{B}}_{N\otimes} - \mathbf{B}_{N\otimes})\}\mathbf{u}_N \\ &= \frac{1}{\sqrt{N}}\mathbf{H}_N^\top\mathbf{B}_{N\otimes}^\top\mathbf{B}_{N\otimes}\mathbf{u}_N\end{aligned}\tag{A.16}$$

$$+ \frac{1}{\sqrt{N}}\mathbf{H}_N^\top\mathbf{B}_{N\otimes}^\top(\hat{\mathbf{B}}_{N\otimes} - \mathbf{B}_{N\otimes})\mathbf{u}_N\tag{A.17}$$

$$+ \frac{1}{\sqrt{N}}\mathbf{H}_N^\top(\hat{\mathbf{B}}_{N\otimes} - \mathbf{B}_{N\otimes})^\top\mathbf{B}_{N\otimes}\mathbf{u}_N\tag{A.18}$$

$$+ \frac{1}{\sqrt{N}}\mathbf{H}_N^\top(\hat{\mathbf{B}}_{N\otimes} - \mathbf{B}_{N\otimes})^\top(\hat{\mathbf{B}}_{N\otimes} - \mathbf{B}_{N\otimes})\mathbf{u}_N.\tag{A.19}$$

1. First, let us consider the term (A.16), that is, the triangular array  $\boldsymbol{\xi}_N = N^{-1/2}\mathbf{H}_N^\top\mathbf{B}_{N\otimes}^\top\mathbf{B}_{N\otimes}\mathbf{u}_N = N^{-1/2}\tilde{\mathbf{H}}_N^\top\tilde{\mathbf{u}}_N$ . In the proof of Theorem 1, it was shown that  $\boldsymbol{\xi}_N \rightarrow N(\mathbf{0}, \tilde{\mathbf{Q}}_{H\Sigma H})$  in distribution as  $n \rightarrow +\infty$ .
2. Now, we only have to show that the remaining terms in (A.17)–(A.19) are negligible in probability given that  $\mathbf{B}_N = \sum_{k=1}^{\varkappa}\phi_k^0\boldsymbol{\Phi}_k$ ,  $\hat{\mathbf{B}}_N = \sum_{k=1}^{\varkappa}\hat{\phi}_{k,N}\boldsymbol{\Phi}_{k,N}$ , and  $\hat{\phi}_{k,N} \rightarrow \phi_k^0$  in probability as  $n \rightarrow +\infty$  for all  $k = 1, \dots, \varkappa$ . The proof is analogous for all the terms, so we prove it just for (A.17). Let us thus consider the triangular array(s)

$$\begin{aligned}\zeta_N &= \frac{1}{\sqrt{N}}\mathbf{H}_N^\top\mathbf{B}_{N\otimes}^\top(\hat{\mathbf{B}}_{N\otimes} - \mathbf{B}_{N\otimes})\mathbf{u}_N \\ &= \frac{1}{\sqrt{N}}\mathbf{H}_N^\top\mathbf{B}_{N\otimes}^\top\sum_{k=1}^{\varkappa}(\hat{\phi}_{k,N} - \phi_k^0)\boldsymbol{\Phi}_{k,N\otimes}\mathbf{u}_N \\ &= \sum_{k=1}^{\varkappa}(\hat{\phi}_{k,N} - \phi_k^0) \cdot \frac{1}{\sqrt{N}}\mathbf{H}_N^\top\mathbf{B}_{N\otimes}^\top\boldsymbol{\Phi}_{k,N\otimes}\mathbf{u}_N \\ &= \sum_{k=1}^{\varkappa}(\hat{\phi}_{k,N} - \phi_k^0) \cdot \zeta_{k,N},\end{aligned}$$

where  $\zeta_{k,N} = N^{-1/2}\mathbf{H}_N^\top\mathbf{B}_{N\otimes}^\top\boldsymbol{\Phi}_{k,N\otimes}\mathbf{u}_N$ . Since  $\varkappa$  is finite and  $\hat{\phi}_{k,N} - \phi_k^0 = o_p(1)$  as  $n \rightarrow +\infty$  for all  $k = 1, \dots, \varkappa$ ,  $\zeta_N$  is negligible in probability if it is verified that  $\zeta_{k,N}$  converges in distribution to the normal distribution with a finite variance matrix. Similar to the proof

of Theorem 1,

$$\zeta_{k,N} = N^{-1/2} \mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \Phi_{k,N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} \varepsilon_N,$$

where  $\varepsilon_N = \Delta \mathbf{v}_N$  or  $\varepsilon_N = (\Delta \mathbf{v}_N^\top, \boldsymbol{\eta}_N^\top + \mathbf{v}_N^\top)^\top$  depending on the estimator; the length of  $\varepsilon_N$  is denoted  $T^*$ . Denoting the vector forming the  $i$ th column of  $\mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \Phi_{k,N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1}$  by  $\tilde{\mathbf{r}}_{ik,N}$ ,  $\zeta_{k,N} = N^{-1/2} \sum_{i=1}^{NT^*} \tilde{\mathbf{r}}_{ik,N} \varepsilon_{iN}$  and the triangular array  $\{\tilde{\mathbf{r}}_{ik,N} \varepsilon_{iN}\}_{i=1}^{NT^*}$  has zero mean since the instruments are constructed so that  $E[\tilde{\mathbf{H}}_N^\top \Phi \tilde{\mathbf{u}}_N] = 0$  for any square matrix  $\Phi$  including  $\Phi_{k,N}$ ,  $k = 1, \dots, \varkappa$ , and thus  $E[\tilde{\mathbf{r}}_{ik,N} \varepsilon_{iN}] = 0$ . Suppose now that  $\{\varepsilon_{iN}\}_{i=1}^{NT^*}$  are ordered in the following way:  $(\Delta \mathbf{v}_N^\top(3), \dots, \Delta \mathbf{v}_N^\top(2\lfloor T/2 \rfloor - 1))^\top$  in the case of the Arellano-Bond estimator (only odd times are used),  $(\mathbf{v}_N^\top(3), \dots, \mathbf{v}_N^\top(T))^\top$  in the case of the estimator based on the level equation only (all times are used),  $(\Delta \mathbf{v}_N^\top(3), \boldsymbol{\eta}_N^\top + \mathbf{v}_N^\top(4), \Delta \mathbf{v}_N^\top(6), \dots, \boldsymbol{\eta}_N^\top + \mathbf{v}_N^\top(T))^\top$  in the case of the Blundell-Bond estimator, for instance. Noting that  $\mathbf{v}_1(3), \dots, \mathbf{v}_N(T)$  are independent by Assumption E,  $\varepsilon_N(1), \dots, \varepsilon_N(T^*)$  are independent as well and the triangular array  $\{\tilde{\mathbf{r}}_{ik,N} \varepsilon_{iN}\}_{i=1}^{NT^*}$  forms a sequence of martingale differences: first, by construction of the instruments,  $E[\varepsilon_N(t) | \tilde{\mathbf{H}}_N^\top(t), \dots, \tilde{\mathbf{H}}_N^\top(3)] = 0$ , which implies  $\tilde{\mathbf{H}}_N^\top(t) \mathbf{B}_{N\otimes}^\top \Phi_{k,N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} E[\varepsilon_N(t) | \tilde{\mathbf{H}}_N^\top(t), \dots, \tilde{\mathbf{H}}_N^\top(3)] = 0$  (and similarly for  $\boldsymbol{\eta}_N^\top$ ); and second, random vectors  $\boldsymbol{\eta}_N$ ,  $\mathbf{v}_N(t)$ , and  $\Delta \mathbf{v}_N(t)$  have mutually independent elements, which implies independence of elements of  $\varepsilon_N(t)$ . Consequently,  $\text{Var}(\zeta_{k,N}) = N^{-1/2} \sum_{i=1}^N \text{Var}(\tilde{\mathbf{r}}_{ik,N} \varepsilon_{iN})$  and  $\zeta_{k,N}$  thus has a bounded variance matrix since for  $N \rightarrow +\infty$

$$\begin{aligned} & \text{Var}[N^{-1/2} \mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \Phi_{k,N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} \varepsilon_N] \\ &= E[N^{-1} \mathbf{H}_N^\top \mathbf{B}_{N\otimes}^\top \Phi_{k,N\otimes} (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1} \varepsilon_N \varepsilon_N^\top (\mathbf{I}_{N\otimes} - \rho^0 \mathbf{M}_{N\otimes})^{-1\top} \Phi_{k,N\otimes}^\top \mathbf{B}_{N\otimes} \mathbf{H}_N] \\ &= \tilde{\mathbf{Q}}_{k,N,H\Sigma H} \rightarrow \tilde{\mathbf{Q}}_{k,H\Sigma H}. \end{aligned}$$

Finally,  $N^{-1} \sum_{i=1}^{N(T-2)} \tilde{\mathbf{r}}_{ik,N} \varepsilon_{iN} \varepsilon_{iN} \tilde{\mathbf{r}}_{ik,N}^\top - \tilde{\mathbf{Q}}_{k,N,H\Sigma H} \xrightarrow{P} 0$  in probability for  $N \rightarrow +\infty$  by Lemma 2. Consequently, the finite second moments and uniform integrability of  $\hat{\mathbf{Q}}_{k,N,H\Sigma H}^{-1/2} \zeta_{k,N}$  (implied by Assumptions V3 and Assumption S3 by the same argument as in Lemma 2) allows us to apply the central limit theorem for martingale differences (e.g., Davidson,

2000, Theorem 6.2.3), which implies that  $\zeta_{k,N}$  converges in distribution to the Gaussian distribution with zero mean and the finite asymptotic variance matrix  $\tilde{\mathbf{Q}}_{k,H\Sigma H}$  as  $n \rightarrow +\infty$  for any  $k = 1, \dots, \varkappa$ . Consequently, it follows  $\zeta_N = o_p(1)$  as  $n \rightarrow +\infty$ .

3. Because we proved  $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0) = [\tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \tilde{\mathbf{Q}}_{HZ}]^{-1} \tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \cdot N^{-1/2} \hat{\mathbf{H}}_N^\top \hat{\mathbf{u}}_N + o_p(1)$ , where  $N^{-1/2} \hat{\mathbf{H}}_N^\top \hat{\mathbf{u}}_N$  is asymptotically normally distributed as shown in points 1 and 2,  $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}^0)$  is asymptotically normally distributed with a zero mean and finite asymptotic variance matrix

$$\mathbf{V}_{SGMM} = [\tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \tilde{\mathbf{Q}}_{HZ}]^{-1} \tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \tilde{\mathbf{Q}}_{H\Sigma H} \mathbf{A} \tilde{\mathbf{Q}}_{HZ} [\tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \tilde{\mathbf{Q}}_{HZ}]^{-1}.$$

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

4. The claim  $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \tilde{\boldsymbol{\theta}}_N) \xrightarrow{p} 0$  now follows from the point 2 of the proof once we compare equations (A.12) and (A.15): their difference leads to

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \tilde{\boldsymbol{\theta}}_N) = [\tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \tilde{\mathbf{Q}}_{HZ}]^{-1} \tilde{\mathbf{Q}}_{HZ}^\top \mathbf{A} \cdot \left[ \frac{1}{\sqrt{N}} \hat{\mathbf{H}}_N^\top \hat{\mathbf{u}}_N - \frac{1}{\sqrt{N}} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N \right] + o_p(1).$$

Recalling the expansion (A.16)–(A.19), the term  $N^{-1/2} \hat{\mathbf{H}}_N^\top \hat{\mathbf{u}}_N - N^{-1/2} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N$  is however equal to the sum of (A.17)–(A.19) and all these terms were shown to be negligible in probability as  $N \rightarrow +\infty$  in point 2.  $\square$

### A.3.4 Proof of Theorem 4

The fact that  $N^{-1} \hat{\mathbf{H}}_N^\top \hat{\mathbf{Z}}_N$  converges to  $\tilde{\mathbf{Q}}_{HZ}$  in probability has been shown in the proof of Theorem 3, see equation (A.14).

Regarding the second claim, we have shown in equations (A.16)–(A.19) that  $N^{-1/2} \hat{\mathbf{H}}_N^\top \hat{\mathbf{u}}_N = N^{-1/2} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N + o_p(1)$ . Hence,  $N^{-1} \hat{\mathbf{H}}_N^\top \hat{\mathbf{u}}_N \hat{\mathbf{u}}_N^\top \hat{\mathbf{H}}_N = N^{-1} \tilde{\mathbf{H}}_N^\top \tilde{\mathbf{u}}_N \tilde{\mathbf{u}}_N^\top \tilde{\mathbf{H}}_N + o_p(1) = \tilde{\mathbf{Q}}_{N,H\Sigma H} + o_p(1)$  as shown Theorem 1 by Lemma 2 and the claim follows from equation (A.13).  $\square$

## Acknowledgements

This paper is a substantially revised version of CentER Discussion Paper No. 2011-134. A few weeks after the first meeting on the revision of this paper our colleague, co-author and dear friend Jenny Ligthart passed away. This paper could never have been written without her, and the final result would definitely have been better with her input.

The authors would like to thank Rob Alessie, Badi Baltagi, Peter Egger, J. Paul Elhorst, Vincenzo Verardi, Tom Wansbeek, seminar participants at the Universities of Melbourne, Maastricht and Groningen, and conference participants at the 3rd World Conference of the Spatial Econometrics Association (Barcelona, July 2009) and 18th International Panel Data Conference (Paris, July 2012) for helpful comments and discussions. We are grateful to Cheng Hsiao for encouraging us to continue with the topic and an associate editor and three anonymous referees for comments and suggestions.

## References

- 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.
- 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.
- BLUNDELL, R., S. BOND, AND F. WINDMEIJER (2000): “Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator,” in *Advances in Econometrics 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, ed. by B. Baltagi. Elsevier, Amsterdam.
- 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.
- 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.
- DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press, Oxford.
- (2000): *Econometric Theory*. Blackwell Publishing, 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.
- (2014): *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. Springer: Berlin, New York, Dordrecht, London.
- FOUCAULT, M., T. MADIES, AND S. PATY (2008): “Public Spending Interactions and Local Politics: Empirical Evidence from French Municipalities,” *Public Choice*, 137, 57–80.
- 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.
- JENISH, N., AND I. R. PRUCHA (2010): “On spatial processes and asymptotic inference under near-epoch dependence,” Discussion paper.
- (2012): “On spatial processes and asymptotic inference under near-epoch dependence,” *Journal of Econometrics*, 170, 178–190.
- 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 (1999): “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model,” *International Economic Review*, 40, 509–533.
- (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. (2003): “Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances,” *Econometric Reviews*, 72, 297–312.
- 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.
- (2010b): “A Spatial Dynamic Panel Data Model with Both Time and Individual Fixed Effects,” *Econometric Theory*, 26, 564–597.
- (2012): “QML Estimation of Spatial Dynamic Panel Data Models with Time Varying Spatial Weights Matrices,” *Spatial Economic Analysis*, 7, 31–74.
- (2013): “Identification of the spatial Durbin panel models,” Discussion paper.
- (2014): “Efficient GMM estimation of spatial dynamic panel data models with fixed effects,” *Journal of Econometrics*, 180, 174–197.
- 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.
- TRIPATHI, G. (1999): “A matrix extension of the Cauchy-Schwarz inequality,” *Economics Letters*, 66, 1–3.
- 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.