J Geogr Syst (2010) 12:241–257
DOI 10.1007/s10109-010-0109-5

ORIGINAL ARTICLE

Spatial models with spatially lagged dependent variables and incomplete data

Harry H. Kelejian · Ingmar R. Prucha

Received: 23 July 2009 / Accepted: 1 February 2010 / Published online: 6 March 2010
© Springer-Verlag 2010

Abstract The purpose of this paper is to suggest estimators for the parameters of spatial models containing a spatially lagged dependent variable, as well as spatially lagged independent variables, and an incomplete data set. The specifications allow for nonstationarity, and the disturbance process of the model is specified non-parametrically. We consider various scenarios concerning the pattern of missing data points. One estimator we suggest is based on a smaller but complete subset of the sample; another is based on a larger but incomplete subset of the sample. We give large sample results for both of these cases.

Keywords Spatial models · Missing data · Instrumental variable estimation

JEL Classification C21 · C31

1 Introduction

Missing data problems often arise in the analysis of spatial models containing spatial lags.[1] In some cases, these data problems arise because of a lack of data on

[1] There is, of course, a large literature on missing data issues. Some important early studies are Anderson (1957), Friedman (1962), Afifi and Elashoff (1966, 1967, 1969), Haitovsky (1968), and Kelejian (1969). A nice, but short, review of some of these early studies is given in Maddala (1977, pp. 201–207); Kmenta (1986, pp. 379–388) also discusses missing data problems, and highlights certain issues. An excellent recent overview of models and procedures is given in Little and Rubin (2002); another interesting text, in a spatial framework, which primarily focuses on the collection of spatial data is Müller (2007). For a review of studies which focus primarily on missing data issues in a spatial framework see Anselin (1988, pp. 172–176) and the references cited there-in. A more extensive review of such studies can be found in Cressie (1993) under the category edge effects in the index of that book.

H. H. Kelejian (✉) · I. R. Prucha
Department of Economics, University of Maryland, College Park, MD 20742, USA
e-mail: kelejian@econ.umd.edu

I. R. Prucha
e-mail: prucha@econ.umd.edu

some variables for units that are defined to be neighbors of other units for which data are available.[2] In some, but not all of these cases, the missing data relate to "edge units"—see e.g., Anselin (1988, pp. 172–173). In other cases, data shortcomings may arise because of either the unavailability of data on certain variables or the reluctance of the researcher to use data on certain variables because of "quality of data" concerns. One example of this would be models that, among other things, require data on the GDPs of the sampled countries, and GDP data are either not available for some countries or are of "dubious" quality and so not used—see e.g., Kelejian et al. (2008).

There are various ways researchers have confronted this problem. One approach is to ignore it if the missing data relate to units that are involved as part of a spatial lag. In this approach, the missing observations are implicitly replaced by zeroes and so, e.g., the spatial lag would be constructed entirely in terms of the available data. Another approach is to complete the sample using available data by estimating in various ways, the observations that are missing. Still another approach is to rely on maximum likelihood estimation.[3] Of course, there are still other approaches, but formal results concerning the distribution of the regression parameter estimators based on incomplete data sets are not available in the framework of a non-stationary spatial model that contains spatial lags in both the dependent and independent variables, as well as a non-parametrically specified disturbance process.[4]

The purpose of this study is to fill this gap. For such non-stationary spatial models, we discuss various scenarios relating to the missing data and the manner in which the sample increases. We suggest regression parameter estimators that are based on a complete subset of the sample, as well as on a larger but incomplete subset of the sample. We give formal large sample results for our suggested estimators in both of these cases. In the context of our model, we give conditions under which the effects of missing observations in the estimation procedure are asymptotically negligible, as well as when they are not! Our asymptotic results account for both of these cases. Finally, we give user friendly suggestions concerning small sample inferences for our suggested estimators.

We specify the model in Sect. 2. That section also contains a discussion of various scenarios as to how the sample may increase. Our suggested estimators are given in Sect. 3, along with theorems that describe their large sample distribution. These theorems relate to the various scenarios concerning the manner in which the sample increases. Conclusions are given in Sect. 4. Technical details are relegated to the "Appendix".

---

[2] As one example, in hedonic models of housing prices involving a spatially lagged dependent variable, the prices of unsold houses would not be known. In some of these cases, the problem would be to predict the prices of the unsold homes given their attributes and the spatial interdependence described by the weighting matrix—see e.g., LeSage and Pace (2004). See also Kelejian and Prucha (2007) for issues relating to such prediction.

[3] See e.g., the review in Anselin (1988, pp. 172–176), the discussions in Cressie (1993) under the category "edge effects" in the index of that book, and the various procedures described by Haining (2003).

[4] We note that the specification of the regression model is typically based on economic theoretical reasoning; the disturbance process is the "residual" in the model, and typically theoretical reasoning does not apply to it. Hence, a non-parametrically specified disturbance process should, perhaps, be considered more often than it is!

2 Model

2.1 Specification

In this section, we specify a model containing spatial lags in the dependent variable, as well as in some of the independent variables. We describe this model in such a way that for **certain** units, the dependent variable and all of the regressors determining that dependent variable, including the spatially lagged variables, are observed. For other units, observations on either the dependent variable and/or all of the regressors determining those dependent variables are not available. More specifically, we divide the units into three groups. For units in groups 1 and 2, the data on the dependent and exogenous variables pertaining to those units are observed. The distinction between units in group 1 and 2 is that for the former all observations needed to formulate spatial lags are also observed, while for the latter, this is not the case. The observations pertaining to units in group 3 are unobserved. As in illustration of such a grouping of the data, consider Fig. 1.

If we assume that spatial interactions can be described by a rook design, then in this illustration, units in group 1 and group 3 are not immediate neighbors, and thus the corresponding block in the spatial weights matrix will be zero. Consequently, we can compute spatial lags of the dependent and exogenous variables for all units in group 1, despite that all variables pertaining to group 3 are unobserved. This is not the case for units in group 2, since those units are immediate neighbors in both groups 2 and 3, and thus spatial lags would depend on observations from units in group 3.

Consistent with the above illustrative example, we assume in the following discussion that the data are ordered such that the (1, 3)-block in the spatial weights

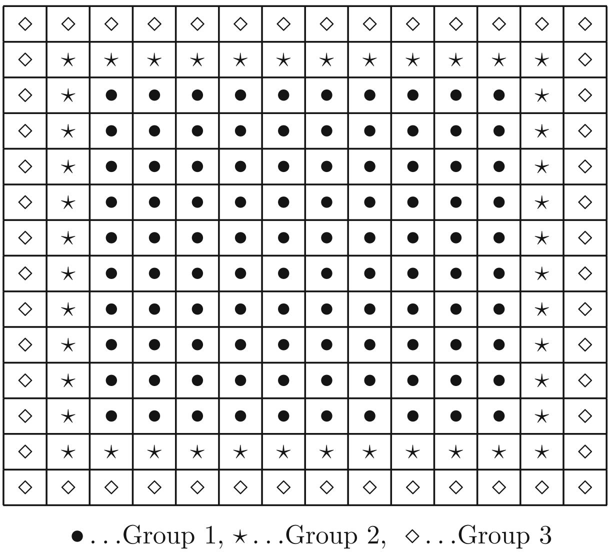

Fig. 1 An exemplary grouping of data on a rectangular grid with neighbors defined by a rook design

matrix corresponding to groups 1 and 3 is zero. Since it has no effect on the derivations relating to our large sample results, we allow for the (3, 1)-block in the spatial weights matrix to be non-zero for purposes of generality, although we expect it to be zero in most applications. In particular, we consider the following model:

$$\begin{pmatrix} y_{n,1} \\ y_{n,2} \\ y_{n,3} \end{pmatrix} = \begin{pmatrix} X_{n,1} \\ X_{n,2} \\ X_{n,3} \end{pmatrix} \beta_1 + \begin{pmatrix} W_{n,11} & W_{n,12} & 0 \\ W_{n,21} & W_{n,22} & W_{n,23} \\ W_{n,31} & W_{n,32} & W_{n,33} \end{pmatrix} \begin{pmatrix} J_{n,1} \\ J_{n,2} \\ J_{n,3} \end{pmatrix} \beta_2 + \lambda \begin{pmatrix} W_{n,11} & W_{n,12} & 0 \\ W_{n,21} & W_{n,22} & W_{n,23} \\ W_{n,31} & W_{n,32} & W_{n,33} \end{pmatrix} \begin{pmatrix} y_{n,1} \\ y_{n,2} \\ y_{n,3} \end{pmatrix} + \begin{pmatrix} u_{n,1} \\ u_{n,2} \\ u_{n,3} \end{pmatrix}, \tag{1}$$

and

$$\begin{pmatrix} u_{n,1} \\ u_{n,2} \\ u_{n,3} \end{pmatrix} = \begin{pmatrix} R_{n,11} & R_{n,12} & R_{n,13} \\ R_{n,21} & R_{n,22} & R_{n,23} \\ R_{n,31} & R_{n,32} & R_{n,33} \end{pmatrix} \begin{pmatrix} \varepsilon_{n,1} \\ \varepsilon_{n,2} \\ \varepsilon_{n,3} \end{pmatrix},$$

where $y_{n,i}$ and $X_{n,i}$ are, respectively, the $n_i \times 1$ vectors of endogenous variables and $n_i \times s$ matrices of exogenous variables corresponding to group i for $i = 1, 2, 3$. The researcher observes $y_{n,1}$, $y_{n,2}$, $X_{n,1}$ and $X_{n,2}$, but not $y_{n,3}$, and $X_{n,3}$. The matrices $W_{n,ij}$ $i, j = 1, 2, 3$ are observed nonstochastic weighting matrices, where $W_{n,13} = 0$. The matrices $J_{n,i}$ are of dimension $n_i \times r$ with $r \leq s$ and represent submatrices of $X_{n,i}$. Hence, the above specification allows for spatial lags in some, but not necessary all, exogenous regressors. Of course, given that the matrices $J_{n,i}$ are submatrices of $X_{n,i}$, it follows that $J_{n,1}$ and $J_{n,2}$ are observed, but not $J_{n,3}$. We conditionalize on the exogenous variables and so take the matrices $X_{n,i}$ and $J_{n,i}$, $i = 1, 2, 3$ as nonstochastic.

For future reference, let $n = n_1 + n_2 + n_3$, and let W_n and R_n be the $n \times n$ matrices

$$W_n = \begin{pmatrix} W_{n,11} & W_{n,12} & 0 \\ W_{n,21} & W_{n,22} & W_{n,23} \\ W_{n,31} & W_{n,32} & W_{n,33} \end{pmatrix}, \quad R_n = \begin{pmatrix} R_{n,11} & R_{n,12} & R_{n,13} \\ R_{n,21} & R_{n,22} & R_{n,23} \\ R_{n,31} & R_{n,32} & R_{n,33} \end{pmatrix} \tag{2}$$

Given the notation in (2), the model in (1) can be written more compactly as

$$\begin{aligned} y_n &= X_n\beta_1 + W_nJ_n\beta_2 + \lambda W_n y_n + u_n \\ &= Z_n\gamma + u_n \\ u_n &= R_n\varepsilon_n \end{aligned} \tag{3}$$

where

$$\begin{aligned} &y'_n = \left(y'_{n,1}, y'_{n,2}, y'_{n,3}\right), \quad X'_n = \left(X'_{n,1}, X'_{n,2}, X'_{n,3}\right), \quad J'_n = \left(J'_{n,1}, J'_{n,2}, J'_{n,3}\right), \\ &Z_n = (X_n, W_nJ_n, W_ny_n); \quad \gamma' = \left(\beta'_1, \beta'_2, \lambda\right), \\ &u'_n = (u'_{n,1}, u'_{n,2}, u'_{n,3}), \quad \varepsilon'_n = \left(\varepsilon'_{n,1}, \varepsilon'_{n,2}, \varepsilon'_{n,3}\right). \end{aligned}$$

We will suggest two instrumental variable estimators for situations where some of the data are missing. These estimators presume that the researcher is able to arrange the data as specified in (1). One of them will be based on the complete portion of the sample, i.e., the portion of the sample for which the dependent variable, the exogenous variables and the spatial lags are observed. The dependent variable corresponding to this estimator is the $n_1 \times 1$ vector $y_{n,1}$. The other estimator will be based on a larger but incomplete portion of the sample, where the dependent and exogenous variables are observed, but where some of the data needed to compute the spatial lags on the r.h.s. are unobserved. The dependent variable corresponding to this estimator is the $n_1 + n_2 \times 1$ vector $[y'_{n,1}, y'_{n,2}]$.

For notational convenience, we define

$$y_{n,1+2} = \left[y'_{n,1}, y'_{n,2}\right]' \tag{4}$$

and correspondingly

$$\begin{aligned} X_{n,1+2} &= \left[X'_{n,1}, X'_{n,2}\right]' \\ J_{n,1+2} &= \left[J'_{n,1}, J'_{n,2}\right]' \\ u_{n,1+2} &= \left[u'_{n,1}, u'_{n,2}\right]'. \end{aligned} \tag{5}$$

We also define $W_{n,1}$ as that portion of the weighting matrix, which determines the spatial lags in the sub-model for $y_{n,1}$ as implied by (1), namely

$$W_{n,1} = \left[W_{n,11} \quad W_{n,12}\right]. \tag{6}$$

In a similar way, $W_{n,1+2}$ is defined as that portion of the weighting matrix, which determines the spatial lags in the sub-model for $y_{n,1+2}$ as implied by (1), namely

$$W_{n,1+2} = \begin{bmatrix} W_{n,11} & W_{n,12} & 0 \\ W_{n,21} & W_{n,22} & W_{n,23} \end{bmatrix}. \tag{7}$$

This notation is extended to submatrices of R_n as defined in (2)—e.g.,

$$\begin{aligned} R_{n,1} &= \begin{pmatrix} R_{n,11} & R_{n,12} & R_{n,13} \end{pmatrix}, \\ R_{n,1+2} &= \begin{pmatrix} R_{n,11} & R_{n,12} & R_{n,13} \\ R_{n,21} & R_{n,22} & R_{n,23} \end{pmatrix}. \end{aligned} \tag{8}$$

In terms of this notation, and borrowing other notation from Kelejian and Prucha (1998) by letting

$$\bar{J}_{n,1} = W_{n,1} J_{n,1+2} \quad \text{and} \quad \bar{y}_{n,1} = W_{n,1} y_{n,1+2},$$

the model in (1) implies that $y_{n,1}$ is determined as

$$\begin{aligned} y_{n,1} &= X_{n,1}\beta_1 + \bar{J}_{n,1}\beta_2 + \lambda \bar{y}_{n,1} + u_{n,1} \\ &\equiv Z_{n,1}\gamma + u_{n,1} \end{aligned} \tag{9}$$

where

$$Z_{n,1} = \left(X_{n,1}, \bar{J}_{n,1}, \bar{y}_{n,1}\right); \quad \gamma' = \left(\beta_1', \beta_2', \lambda\right).$$

The model in (1) implies furthermore that $y_{n,1+2}$ is determined as

$$y_{n,1+2} = X_{n,1+2}\beta_1 + W_{n,1+2}J_n\beta_2 + \lambda W_{n,1+2}y_n + u_{n,1+2}. \tag{10}$$

Assuming $W_{n,23} \neq 0$, the regressors in the (10) are not all observed. Part of our large sample results reported below relate to conditions under which the unobserved parts of (10) are asymptotically negligible. That analysis is facilitated by separating the observed portions of the regressors in (10) from the unobserved portions. Hence, we define $W^o_{n,1+2}$ and $W^u_{n,1+2}$ as, respectively, those parts of $W_{n,1+2}$ that relate to the observable and unobservable regressors, namely

$$W^o_{n,1+2} = \begin{bmatrix} W_{n,11} & W_{n,12} \\ W_{n,21} & W_{n,22} \end{bmatrix}, \qquad W^u_{n,1+2} = \begin{bmatrix} 0 \\ W_{n,23} \end{bmatrix}$$

Given this notation, and again borrowing notation from Kelejian and Prucha (1998) by letting

$$\bar{J}_{n,1+2} = W^o_{n,1+2}J_{n,1+2} \quad \text{and} \quad \bar{y}_{n,1+2} = W^o_{n,1+2}y_{n,1+2}$$

the model in (10) can be rewritten as

$$\begin{aligned} y_{n,1+2} &= X_{n,1+2}\beta_1 + \bar{J}_{n,1+2}\beta_2 + \lambda\bar{y}_{n,1+2} + \xi_n \\ &= Z_{n,1+2}\gamma + \xi_n \end{aligned} \tag{11}$$

where

$$\begin{aligned} Z_{n,1+2} &= \left(X_{n,1+2}, \bar{J}_{n,1+2}, \bar{y}_{n,1+2}\right) \\ \xi_n &= \left[W^u_{n,1+2}J_{n,3}\beta_2 + \lambda W^u_{n,1+2}y_{n,3} + u_{n,1+2}\right] \end{aligned}$$

and where ξ_n in (11) can be viewed as a "contaminated" error vector.

Both (9) and (11) contain spatial lags in the dependent variable, and so our suggested estimators will involve instrument matrices. Consistent with our earlier notation, the instrument matrices for the estimation of (9) and (11) are, respectively, the $n_1 \times q_1$ and $n_1 + n_2 \times q_2$ matrices $H_{n,1}$ and $H_{n,1+2}$. Ideally, we would like to use the conditional means of $\bar{y}_{n,1}$ and $\bar{y}_{n,1+2}$ as instruments. These conditional means depend on $X_{3,n}$ and are hence unobserved. Guided in part by suggestions in Kelejian and Prucha (1998), our recommendation here is to take as instruments the linearly independent columns of

$$\begin{aligned} H_{n,1} &= \left[X_{n,1}, W_{n,11}X_{n,1}, \ldots W^s_{n,11}X_{n,1}, \bar{J}_{n,1}, W_{n,11}\bar{J}_{n,1}, \ldots W^s_{n,11}\bar{J}_{n,1}\right] \\ H_{n,1+2} &= \left[X_{n,1+2}, W^o_{n,1+2}X_{n,1+2}, \ldots (W^o_{n,1+2})^s X_{n,1+2}\right] \end{aligned} \tag{12}$$

where typically $s \leq 2$. Note that since (9) contains the $n_1 \times 1$ endogenous regressor $\bar{y}_{n,1} = [W_{n,11}y_{n,1} + W_{n,12}y_{n,2}]$, one might also add $W_{n,12}X_{n,2}$ to the columns of the instrument matrix $H_{n,1}$ to help account for the mean of $W_{n,12}\, y_{n,2}$. Because $W_{n,12}$ is an $n_1 \times n_2$ matrix and $X_{n,1}$ is an $n_1 \times s$ matrix, $W_{n,12}$ is not conformable for postmultiplication by $X_{n,1}$. Typically, n_2 will be less than n_1; we do not suggest constructing further instruments by postmultiplying "augmented" versions of $W_{n,12}$ by $X_{n,1}$.

2.2 The sample configurations

Our large sample results below relate to the relative sample sizes n_i, $i = 1, 2, 3$. We consider three cases. Specifically, we assume the configuration of the space containing the n units expands in such a way that

Case 1: $n_1 \to \infty$
Case 2: $n_1 \to \infty$, and $\frac{n_3}{(n_1+n_2)^{1/2}} \to 0$
Case 3: $n_1 \to \infty$, and $\frac{n_3}{(n_1+n_2)^{1/2}} \to c > 0$.

Case 1 relates to the use of $y_{n,1}$ as the dependent variable. In this case, there is no assumption concerning the magnitudes of n_2 or n_3 and, hence, they both could increase of (any order) beyond limit! As a preview, the reason these magnitudes do not matter relates to our assumptions below concerning the weighting matrix, W_n. Cases 2 and 3 relate to the use of $y_{n,1+2}$ as the dependent variable. Note that case 2 does not rule out $n_2 \to \infty$, $n_3 \to \infty$, or even $n_i / n_1 \to \infty$, $i = 2, 3$. Case 3 differs from case 2 in that the "missing" portion of the sample, referenced here by n_3, increases in the same order as the square root of the observed portion of the sample indexed here by $n_1 + n_2$.

2.3 Assumptions

In this section, we state our assumptions; their interpretations are given in the following section. Let

$$D_{n,1} = \left(X_{n,1}, \bar{J}_{n,1}\right) \quad \text{and} \quad D_{n,1+2} = \left(X_{n,1+2}, \bar{J}_{n,1+2}\right).$$

Our assumptions are given below.

Assumption 1 The elements of ε_n, namely $\{\varepsilon_{i,n}: 1 \leq i \leq n, n \geq 1\}$, are identically distributed. Further, $\{\varepsilon_{i,n}: 1 \leq i \leq n\}$ are for each n distributed (jointly) independently with $E(\varepsilon_{i,n}) = 0$, $E(\varepsilon_{i,n}^2) = 1$. Additionally, the innovations are assumed to possess a finite fourth moment.

Assumption 2 The elements of X_n, $H_{n,1}$, and $H_{n,1+2}$ are uniformly bound in absolute value; in addition, $D_{n,1}$ and $D_{n,1+2}$ have full column rank for n_1 and n_2 large enough.

In the limits below, no restrictions are placed on the magnitudes of n_2 and n_3, which may be related to n_1 and so may also increase beyond limit.

Assumption 3 We assume

(a) $\lim_{n_1 \to \infty} n_1^{-1} D_{n,1}' D_{n,1} = Q_{D_1 D_1}$,
(b) $\lim_{n_1 \to \infty} n_1^{-1} H_{n,1}' H_{n,1} = Q_{H_1 H_1}$,
(c) $\lim_{n_1 \to \infty} (n_1 + n_2)^{-1} D_{n,1+2}' D_{n,1+2} = Q_{D_{1+2} D_{1+2}}$,
(d) $\lim_{n_1 \to \infty} (n_1 + n_2)^{-1} H_{n,1+2}' H_{n,1+2} = Q_{H_{1+2} H_{1+2}}$,

where $Q_{D_1D_1}, Q_{D_{1+2}D_{1+2}}, Q_{H_1H_1}$ and $Q_{H_{1+2}H_{1+2}}$ are finite nonsingular matrices.

Assumption 4 (a) $|\lambda| < 1$ (b) The diagonal elements of W_n are zero, and R_n and $(I_n - \lambda W_n)$ are nonsingular (c). The row and column sums of R_n, W_n and $(I_n - \lambda W_n)^{-1}$ are uniformly bound in absolute value.

Assumption 5 We assume

(a) $p\lim_{n_1\to\infty} n_1^{-1}H'_{n,1}Z_{n,1} = Q_{H_1Z_1}$
(b) $\lim_{n_1\to\infty} n_1^{-1}H'_{n,1}R_{n,1}R'_{n,1}H_{n,1} = Q_{H_1R_1R_1H_1}$
(c) $p\lim_{n_1\to\infty}(n_1 + n_2)^{-1}H'_{n,1+2}Z_{n,1+2} = Q_{H_{1+2}Z_{1+2}}$
(d) $\lim_{n_1\to\infty}(n_1 + n_2)^{-1}H'_{n,1+2}R_{n,1+2}R'_{n,1+2}H_{n,1+2} = Q_{H_{1+2}R_{1+2}R_{1+2}H_{1+2}}$

where $Q_{H_1Z_1}, Q_{H_1R_1R_1H_1}, Q_{H_{1+2}Z_{1+2}}$, and $Q_{H_{1+2}R_{1+2}R_{1+2}H_{1+2}}$ are finite matrices with full column rank.

The next assumption is only maintained under Case 3.

Assumption 6 For $n_1\to\infty$ and $\frac{n_3}{(n_1+n_2)^{1/2}} \to c$, with $0 < c < \infty$, we assume

(a) $\lim_{n_1\to\infty}(n_1 + n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}J_{n,3} = F_{1+2}$
(b) $\lim_{n_1\to\infty}(n_1 + n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}(I_n - \lambda W_n)^{-1}_{n,3}X_n = L_{1+2}$
(c) $\lim_{n_1\to\infty}(n_1 + n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}(I_n - \lambda W_n)^{-1}_{n,3}W_nJ_n = S_{1+2}$

where F_{1+2}, L_{1+2}, and S_{1+2} are, respectively, $q_2 \times r$, $q_2 \times s$, and $q_2 \times r$ finite matrices, and where $(I_n - \lambda W_n)^{-1}_{n,3}$ is the $n_3 \times n$ matrix consisting of the last n_3 rows of $(I_n - \lambda W_n)^{-1}$.

2.4 A brief discussion of the assumptions

First note from (3), Assumptions 1 and part (b) of Assumption 4 that

$$Ey_n = (I_n - \lambda W_n)^{-1}[X_n\beta_1 + W_nJ_n\beta_2] \tag{13}$$

Assumption 2 and part (c) of Assumption 4 imply that the elements of Ey_n are uniformly bound in absolute value. Also note that Assumption 1 and part (c) of Assumption 4 imply that the row and column sums of the variance–covariance (henceforth, VC) matrix of y_n, say VC_{y_n}, are uniformly bound in absolute value, where

$$VC_{y_n} = (I_n - \lambda W_n)^{-1}R_nR'_n(I_n - \lambda W'_n)^{-1}. \tag{14}$$

Assumptions 1, 2, 3(a), and 4 are standard-type conditions that have been discussed elsewhere in the literature, see e.g., Kelejian and Prucha (1998, 2004). Note also that the condition $E(\varepsilon^2_{i,n}) = 1$ in Assumption 1 is not restrictive, since the only assumption maintained for R_n is that its row and column sums are uniformly bounded. Thus, the variance of the $\varepsilon_{i,n}$ is a scale factor, which can be incorporated into R_n. To see this, note that if $E(\varepsilon^2_{i,n}) = \sigma^2_\varepsilon$, the disturbance vector u_n in (3) could be redefined as:

$$u_n = R_n\varepsilon_n = [\sigma_\varepsilon R_n]\left[\frac{1}{\sigma_\varepsilon}\varepsilon_n\right] = R_n^*\varepsilon_n^*,$$

where the row and column sums of R_n^* are again uniformly bounded, and the elements of ε_n^* would have a variance of 1.0.

Consider part (c) of Assumption 3. The assumption requires that $(n_1+n_2)^{-1} X'_{n,1+2}X_{n,1+2}, (n_1+n_2)^{-1}X'_{n,1+2}W^o_{n,1+2}X_{n,1+2},$ and $(n_1+n_2)^{-1}X'_{n,1+2}W^{o\prime}_{n,1+2}W^o_{n,1+2} X_{n,1+2}$ converge. These limit assumptions are standard-type conditions. Given part (a) of Assumption 3, and if and if $n_2/n_1 \to 0$, then $Q_{D_{1+2}D_{1+2}} = Q_{D_1D_1}$. Similar comments apply to parts (b) and (d) of Assumption 3.

Now consider part (a) of Assumption 5, and note that

$$n_1^{-1}H'_{n,1}Z_{n,1} = n_1^{-1}H'_{n,1}(X_{n,1}, W_{n,1}J_{n,1+2}, W_{n,1}y_{n,1+2}). \tag{15}$$

Since assumptions concerning the limit of terms such as $n_1^{-1}H'_{n,1}X_{n,1}$ are standard, we will focus our attention on the other two components in (15), namely $\psi_{n,1} = n_1^{-1}H'_{n,1}W_{n,1}J_{n,1+2}$ and $\phi_{n,1} = n_1^{-1}H'_{n,1}W_{n,1}y_{n,1+2}$. Recalling that the columns of $J_{n,1+2}$ are a subset of those in $X_{n,1+2}$, it follows from Assumption 2 and part (c) of Assumption 4 that, regardless of the relative magnitudes of n_2 and n_1, the elements of $\psi_{n,1}$ are uniformly bound in absolute value. Hence, at this point, part (a) of Assumption 5 imposes the condition that the limits of the elements of $\psi_{n,1}$ exist. Now, consider $\phi_{n,1}$ and note

$$E(\phi_{n,1}) = n_1^{-1}H'_{n,1}W_{n,1}E[y_{n,1+2}] \tag{16}$$

where $E[y_{n,1+2}]$ represents the first $n_1 + n_2$ elements of Ey_n, which is given in (13). Again, Assumption 2 and part (c) of Assumption 4 imply that the elements of $E[y_{n,1+2}]$ are uniformly bound in absolute value. Now note, using evident notation, that

$$VC_{\phi_{n,1}} = n_1^{-2}H'_{n,1}\left[W_{n,1}VC_{y_{n,1+2,}}W'_{n,1}\right]H_{n,1} \tag{17}$$

where $VC_{y_{n,1+2,}}$ is the upper $n_1 + n_2 \times n_1 + n_2$ block of VC_{y_n}, which is given in (14). Assumption 2 and part (c) of Assumption 4 imply that the elements of $VC_{\phi_{n,1}}$ are $O(n^{-1})$ and therefore $VC_{\phi_{n,1}} \to 0$, as $n_1 \to \infty$. Hence, part (a) of Assumption 5 imposes the condition that the limits of the elements of $E\phi_{n,1}$ exist. Therefore, in essence, part (a) of Assumption 5 requires that $E(\phi_{n,1})$ converge to a limit, which is linearly independent of the limit of $n_1^{-1}H'_{n,1}X_{n,1}$. Similar arguments suggest that the other parts of Assumption 5 are reasonable.

Finally, consider Assumption 6, and recall that this assumption relates to the case in which $n_1 \to \infty$ and $\frac{n_3}{(n_1+n_2)^{1/2}} \to c$, where c is a finite non-zero constant. In reference to part (a) of Assumption 6, note that Assumption 2 and part (c) of Assumption 4 imply that the elements of the $q_2 \times r$ matrix $H'_{n,1+2}W^u_{n,1+2}J_{n,3}$ are $O(n_3)$ or equivalently $O((n_1+n_2)^{1/2})$. Therefore, the elements of $(n_1+n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}J_{n,3}$ are $O(1)$. Similarly, in parts (b) and (c) of Assumption 6, the elements of the $q_2 \times s$ and $q_2 \times r$ matrices, namely $(n_1+n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}(I_n - \lambda W_n)^{-1}_{n,3}X_n$ and

$(n_1+n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}(I_n-\lambda W_n)^{-1}_{n,3}W_nJ_n$, respectively, are $O(1)$. Given these results, the assumed limits in Assumption 6 are reasonable.

3 Suggested estimators

In this section, we suggest two estimators. One is based on the complete subset of the sample in which the dependent variable is $y_{n,1}$. In this case, the sample size is n_1, and all observations on the regressors determining $y_{n,1}$ are available. The other estimator is based on a larger subset of the sample, which involves missing data. In this case, the sample size is $n_1 + n_2$, and the dependent variable is $y_{n,1+2}$.

3.1 The estimator based on $y_{n,1}$

Consider the model in (9), and let $P_{n,H_1} = H_{n,1}\left(H'_{n,1}H_{n,1}\right)^{-1}H_{n,1}$ and $\hat{Z}_{n,1} = P_{n,H_1}Z_{n,1}$. Our suggested estimator of γ, which is based on the complete portion of the sample is

$$\hat{\gamma}_1 = \left(\hat{Z}'_{n,1}\hat{Z}_{n,1}\right)^{-1}\hat{Z}'_{n,1}y_{n,1} \tag{18}$$

The proof of Theorem 1 is given in the "Appendix".

Theorem 1 *Given* (9), *and Assumptions* 1, 2, 3(a), (b), 4, *and* 5(a), (b)

$$n_1^{1/2}(\hat{\gamma}_1-\gamma)\xrightarrow{D}N\left(0, Q^{-1}_{\hat{Z}_1\hat{Z}_1}\Omega_1 Q^{-1}_{\hat{Z}_1\hat{Z}_1}\right) \tag{19}$$

as $n_1\to\infty$ *where*

$$Q_{\hat{Z}_1\hat{Z}_1} = Q'_{H_1Z_1}Q^{-1}_{H_1H_1}Q_{H_1Z_1}$$

and

$$\Omega_1 = Q'_{H_1Z_1}Q^{-1}_{H_1H_1}Q_{H_1R_1R_1H_1}Q^{-1}_{H_1H_1}Q_{H_1Z_1.}$$

Small sample inferences could be based on

$$\hat{\gamma}_1 \sim N\left(\gamma, \hat{VC}_{\hat{\gamma}_1}\right)$$

where

$$\hat{VC}_{\hat{\gamma}_1} = n_1 G'_{n,1}\hat{Q}_{H_1R_1R_1H_1}G_{n,1} \tag{20}$$

with

$$G_{n,1} = \left(H'_{n,1}H_{n,1}\right)^{-1}H'_{n,1}Z_{n,1}\left(\hat{Z}'_{n,1}\hat{Z}_{n,1}\right)^{-1},$$

and where $\hat{Q}_{H_1R_1H_1R_1}$ is a consistent estimator of $Q_{H_1R_1H_1R_1}$. Note, given further assumptions especially relating to distance measures, $Q_{H_1R_1H_1R_1}$ can be estimated non-parametrically as a SHAC estimator as considered in Kelejian and Prucha (2007), because $R_{n,1}R'_{n,1}$ is the $n_1 \times n_1$ VC matrix of $u_{n,1}$, and $u_{n,1}$ can be estimated via (9).[5] Of course, in the special case where the disturbances are not spatially correlated and thus $Q_{H_1R_1R_1H_1} = \sigma^2 Q_{H_1H_1}$, we can employ the estimator $\hat{Q}_{H_1R_1R_1H_1} = \left(n_1^{-1}\widehat{u}'_{n,1}\widehat{u}_{n,1}\right)\left(n^{-1}H'_{n,1}H_{n,1}\right)$, where $\widehat{u}_{n,1}$ denotes the estimated disturbances via (9).

Remark 1 Note that the result in (20) relates to the model in (9). The dependent variable in that model is $y_{n,1}$, which is an $n_1 \times 1$ vector; on the other hand, the endogenous spatial lag in that model, namely $\bar{y}_{n,1} = W_{n,1}y_{n,1+2}$, involves the $(n_1 + n_2) \times 1$ vector $y_{n,1+2}$. The result in (20) holds for all possible values of n_2 including $n_2/n_1 \to \infty$ – i.e., the incomplete portion of the sample can be infinitely large relative to the complete portion. A crucial assumption underlying this result is part (c) of Assumption 4, namely that the row and column sums of the weighting matrix W_n are uniformly bound in absolute value.

3.2 The estimator based on $y_{n,(1,2)}$ and $\frac{n_3}{(n_1+n_2)^{1/2}} \to c$

If n_2 is "large", a researcher might be tempted to estimate the parameters of the model in terms of a sample that is larger than just the first n_1 observations. Given the structure of the model, in such a case, the researcher might estimate the parameters of the model in terms of (11) instead of (9). In this case, however, it is clear from (11) that missing data problems arise if $W_{n,23} \neq 0$. In this section, we give results that correspond to two cases. In the first case, the missing data are of no consequence in the sense that they are asymptotically negligible. In this case, $c = 0$. In the second case, the missing data are of consequence. In this case, c is a finite non-zero constant.

3.2.1 The case $c = 0$

Consider the model in (11), and let $P_{n,H_{1+2}} = H_{n,1+2}\left(H'_{n,1+2}H_{n,1+2}\right)^{-1}H'_{n,1+2}$ and $\hat{Z}_{n,1+2} = P_{n,H_{1+2}}Z_{n,1+2}$. Our suggested estimator of γ is the two-stage least squares estimator

$$\hat{\gamma}_2 = \left(\hat{Z}'_{n,1+2}\hat{Z}_{n,1+2}\right)^{-1}\hat{Z}'_{n,1+2}y_{n,1+2} \tag{21}$$

The proof of Theorem 2 is given in the "Appendix".

Theorem 2 *Given Assumptions* 1, 2, 3(c), (d), 4 *and* 5(c), (d)

$$(n_1 + n_2)^{1/2}(\hat{\gamma}_2 - \gamma) \xrightarrow{D} N\left(0, Q^{-1}_{\hat{Z}_{1+2}\hat{Z}_{1+2}}\Omega_2 Q^{-1}_{\hat{Z}_{1+2}\hat{Z}_{1+2}}\right) \tag{22}$$

[5] The proof of the consistency of the SHAC estimator in Kelejian and Prucha (2007) is based on the assumption that the disturbance vector and the innovation vector are of the same dimension. The VC matrix appearing in (22) relates to $u_{n,1}$ which is $n_1 \times 1$. However, from (1) it is clear that the innovation vector defining $u_{n,1}$ has dimension $n_1 + n_2 + n_3 \times 1$. This difference requires tedious, but straight forward, adjustments of the formal proof of consistency of the SHAC estimator given in Kelejian and Prucha (2007). Essentially, the reason for this is part (c) of Assumption 4.

as $n_1 \to \infty$ *where*

$$\Omega_2 = Q'_{H_{1+2}Z_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} Q_{H_{1+2}R_{1+2}R_{1+2}H_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} Q_{H_{1+2}Z_{1+2}}$$

and

$$Q_{\hat{Z}_{1+2}\hat{Z}_{1+2}} = \left[Q'_{H_{1+2}Z_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} Q_{H_{1+2}Z_{1+2}} \right]$$

Small sample inference can be based on

$$\hat{\gamma}_2 \sim N\left(\gamma, \hat{VC}_{\hat{\gamma}_2}\right), \tag{23}$$

where

$$\hat{V}_{\hat{\gamma}_2} = (n_1 + n_2) G'_{n,1+2} \hat{Q}_{H_{1+2}R_{1+2}R_{1+2}H_{1+2}} G_{n,1+2} \tag{24}$$

with

$$G_{n,1+2} = (H'_{n,1+2} H_{n,1+2})^{-1} H'_{n,1+2} Z_{n,1+2} \left(\hat{Z}'_{n,1+2} \hat{Z}_{n,1+2} \right)^{-1},$$

and where $\hat{Q}_{H_{1+2}R_{1+2}R_{1+2}H_{1+2}}$ is a consistent estimator of $Q_{H1+2}R_{1+2}R_{1+2}H_{1+2}$.

Remark 2 Observe that the difference between the disturbance vector and the residual vector in (11) is given by

$$\xi_n - u_{n,1+2} = W^u_{n,1+2} J_{n,3} \beta_2 + \lambda W^u_{n,1+2} y_{n,3}. \tag{25}$$

We conjecture that given additional assumptions, we should again be able to estimate the matrix $Q_{H_{1+2}R_{1+2}R_{1+2}H_{1+2}}$ non-paramterically via the SHAC estimation approach put forward in Kelejian and Prucha (2007).

3.2.2 The case $c \neq 0$

In this case, the large sample distribution $\hat{\gamma}_2$ is the same as in (22) except that the mean of that large sample distribution is not zero. In particular, the proof of Theorem 3 is in the "Appendix".

Theorem 3 *Given that* $\frac{n_3}{(n_1+n_2)^{1/2}} \to c \neq 0$, *and Assumptions* 1, 2, 3(c), (d), 4, 5(c), (d), *and Assumption 6*

$$\begin{aligned}
&(n_1 + n_2)^{1/2} (\hat{\gamma}_2 - \gamma) \xrightarrow{D} N\left(\mu, Q^{-1}_{\hat{Z}_{1+2}\hat{Z}_{1+2}} \Omega_2 Q^{-1}_{\hat{Z}_{1+2}\hat{Z}_{1+2}}\right) \\
&\mu = Q^{-1}_{\hat{Z}_{1+2}\hat{Z}_{1+2}} \Big[Q'_{Z_{1+2}H_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} F_{1+2}\beta_2 \\
&\qquad + \lambda Q'_{Z_{1+2}H_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} (L_{1+2}\beta_1 + S_{1+2}\beta_2) \Big]
\end{aligned} \tag{26}$$

Remark 3 Note that (26) implies that $\hat{\gamma}_2$ is consistent. However, the mean of the large sample distribution involves limits that involve $X_{n,3}$, which, if $X_{n,3}$ is not observed, suggests that the result in (26) is of limited use for making inferences concerning γ. Of course, if $(n_1 + n_2)$ is "large", small sample inferences can be based on (23), which would be the approximation taken if $\frac{\mu}{n_1+n_2} \simeq 0$.

4 Summary and suggestions for further research

In this paper, we suggest estimators for models that have a spatial lag in the dependent variable, a non-parametrically specified disturbance term, and missing observations. The specification of the disturbance term is such that it allows for both spatial correlation and heteroskedasticity. We consider various configurations of the missing observations. In one case, the missing observations are asymptotically negligible; in another case, they are not asymptotically negligible. One of our suggested estimators is based on a small but complete portion of the sample; our other suggested estimator is based on a larger sample that includes both the complete portion of the sample, as well as part of the incomplete portion. We give formal large sample results for our suggested estimators and suggest small distribution approximations that can be used for purposes of inference. Our large sample results account for the various configurations of the missing observations.

Among other things, a Monte Carlo study that focuses on the small sample properties of our suggested estimators would be of interest. Our results suggest that such a study should consider various configurations of the missing observations, as well as various specifications of the disturbance term. Specifically, small sample results relating to our estimators for cases in which the complete portion of the sample is large, but yet small relative to the incomplete portion should be of particular interest. It should also be of interest to consider various specifications of the disturbance term involving spatial correlation, etc. and corresponding hypothesis tests relating to the regression parameters based on our small sample approximations.

Acknowledgments We gratefully acknowledge financial support from the National Institute of Health through the SBIR grants R43 AG027622 and R44 AG027622. We would like to thank two referees for insightful comments on an earlier draft.

Appendix

Proof of Theorem 1

It follows from (18) and from (9) that

$$\begin{aligned} n_1^{1/2}(\hat{\gamma}_1 - \gamma) &= n_1\left(\hat{Z}'_{n,1}\hat{Z}_{n,1}\right)^{-1} n^{-1/2}\hat{Z}'_{n,1}u_{n,1} \\ &= \left(n_1^{-1}\hat{Z}'_{n,1}\hat{Z}_{n,1}\right)^{-1} n^{-1}Z'_{n,1}H_{n,1}\left(n_1^{-1}H'_{n,1}H_{n,1}\right)^{-1} n^{-1/2}H'_{n,1}R_{n,1}\varepsilon_n \end{aligned} \tag{27}$$

Observe that $n_1^{-1}\hat{Z}'_{n,1}\hat{Z}_{n,1} = \left[n_1^{-1}Z'_{n,1}H_{n,1}\right]\left[n_1^{-1}H'_{n,1}H_{n,1}\right]^{-1}\left[n_1^{-1}H'_{n,1}Z_{n,1}\right]$. Assumptions 3(a), (b) and 5(a) imply

$$\begin{aligned} n_1^{-1}\hat{Z}'_{n,1}\hat{Z}_{n,1} &\xrightarrow{P} Q'_{H_1Z_1}Q^{-1}_{H_1H_1}Q_{H_1Z_1} \equiv Q_{\hat{Z}_1\hat{Z}_1} \\ n_1^{-1}H'_{n,1}Z_{n,1} &\xrightarrow{P} Q_{H_1Z_1}, \\ n_1^{-1}H'_{n,1}H_{n,1} &\rightarrow Q_{H_1H_1}, \end{aligned} \tag{28}$$

where $Q_{\hat{Z}_1\hat{Z}_1}$ is a positive definite matrix.

Assumptions 2 and 4(c) imply that the elements of $H'_{n,1}R_{n,1}$ are uniformly bound. Given this, and Assumptions 1 and 5(b), and the central limit theorem in Kelejian and Prucha (1998), it follows that

$$n^{-1/2}H'_{n,1}R_{n,1}\varepsilon_n \xrightarrow{D} N(0, Q_{H_1R_1R_1H_1}) \tag{29}$$

The proof of Theorem 1 follows from (27)–(29).

Proof of Theorem 2

Recall that under Theorem 2, we have $c = 0$. It follows from (11) and (21) that

$$(n_1+n_2)^{1/2}(\hat{\gamma}_2-\gamma) = \left[(n_1+n_2)\left(\hat{Z}'_{n,1+2}\hat{Z}_{n,1+2}\right)^{-1}\right]\left[(n_1+n_2)^{-1/2}\hat{Z}'_{n,1+2}\xi_n\right] \tag{30}$$

where it should be recalled that $\xi_n = [W^u_{n,1+2}J_{n,3}\beta_2 + \lambda W^u_{n,1+2}y_{n,3} + u_{n,1+2}]$ and $u_{n,1+2} = R_{n,1+2}\,\varepsilon_n$. Consider the first bracketed term in (30). Assumptions 3(d) and 5(c) imply that

$$\begin{aligned} (n_1+n_2)^{-1}\hat{Z}'_{n,1+2}\hat{Z}_{n,1+2} &\xrightarrow{P} Q'_{H_{1+2}Z_{1+2}}Q^{-1}_{H_{1+2},H_{1+2}}Q_{H_{1+2}Z_{1+2}} \equiv Q_{\hat{Z}_{1+2}\hat{Z}_{1+2}}, \\ (n_1+n_2)^{-1}H'_{n,1+2}Z_{n,1+2} &\xrightarrow{P} Q_{H_{1+2}Z_{1+2}}, \\ (n_1+n_2)^{-1}H'_{n,1+2}H_{n,1+2} &\rightarrow Q_{H_{1+2}H_{1+2}}, \end{aligned} \tag{31}$$

where $Q_{\hat{Z}_{1+2}\hat{Z}_{1+2}}$ is a finite and nonsingular matrix. Now, consider the second bracketed term in (30). In light of (11),

$$\begin{aligned} (n_1+n_2)^{-1/2}\hat{Z}'_{n,1+2}\xi_n &= (n_1+n_2)^{-1/2}\hat{Z}'_{n,1+2}\left[W^u_{n,1+2}J_{n,3}\beta_2 + \lambda W^u_{n,1+2}y_{n,3} + u_{n,1+2}\right] \\ &= (n_1+n_2)^{-1/2}\hat{Z}'_{n,1+2}u_{n,1+2} + \Delta_n, \\ \Delta_n &= (n_1+n_2)^{-1/2}\hat{Z}'_{n,1+2}\left[W^u_{n,1+2}J_{n,3}\beta_2 + \lambda W^u_{n,1+2}y_{n,3}\right]. \end{aligned} \tag{32}$$

We show below that the term Δ_n in (32) limits to zero in probability so that

$$(n_1+n_2)^{-1/2}\hat{Z}'_{n,1+2}\xi_n - (n_1+n_2)^{-1/2}\hat{Z}'_{n,1+2}u_{n,1+2} \xrightarrow{P} 0 \tag{33}$$

and so, from (30) to (33)

$$(n_1+n_2)^{1/2}(\hat{\gamma}_2-\gamma) - Q^{-1}_{\hat{Z}_{1+2}\hat{Z}_{1+2}}(n_1+n_2)^{-1/2}\hat{Z}'_{n,1+2}u_{n,1+2} \xrightarrow{P} 0. \tag{34}$$

Consequently, the two terms on the l.h.s. of (34) have the same limiting distribution. In determining this distribution, note from Assumptions 3(d) and 5(c) that

$$(n_1+n_2)^{-1/2}\hat{Z}'_{n,1+2}u_{n,1+2} - Q'_{H_{1+2}Z_{1+2}}Q^{-1}_{H_{1+2}H_{1+2}}(n_1+n_2)^{-1/2}H'_{n,1+2}R_{n,1+2}\varepsilon_n \xrightarrow{P} 0, \quad (35)$$

provided the last term on the r.h.s. converges in distribution. To see this, observe that Assumptions 2 and 4(c) imply that the elements of the $q_2 \times n_1 + n_2 + n_3$ matrix $H'_{n,1+2}R_{n,1+2}$ are uniformly bound. Therefore, Assumptions 1, 5(d), and the central limit theorem in Kelejian and Prucha (1998) imply that

$$(n_1+n_2)^{-1/2}H'_{n,1+2}R_{n,1+2}\varepsilon_n \xrightarrow{D} N(0, Q_{H_{1+2}R_{1+2}R_{1+2}H_{1+2}}) \quad (36)$$

Theorem 2 follows from (34) to (36).

Proof that $\mathbf{\Delta}_n \xrightarrow{P} \mathbf{0}$: We now show that indeed $\mathbf{\Delta}_n$ defined in (32) converges in probability to zero as claimed. Recalling that $\hat{Z}_{n,1+2} = P_{n,H_{1+2}}Z_{n,1+2}$, Assumptions 3(d) and 5(c) imply

$$\begin{aligned} &\Delta_n - \Delta_n^* \xrightarrow{P} 0 \\ &\quad \Delta_n^* = \Delta_{n1}^* + \Delta_{n2}^* \\ &\quad \Delta_{n1}^* = Q'_{Z_{1+2}H_{1+2}}Q^{-1}_{H_{1+2}H_{1+2}}(n_1+n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}J_{n,3}\beta_2 \\ &\quad \Delta_{n2}^* = \lambda Q'_{Z_{1+2}H_{1+2}}Q^{-1}_{H_{1+2}H_{1+2}}(n_1+n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}y_{n,3} \end{aligned} \quad (37)$$

Recalling that the columns of J_n are a subset of those in X_n, Assumptions 2 and 4(c) imply that the elements of $q_2 \times r$ nonstochastic matrix, $H'_{n,1+2}W^u_{n,1+2}J_{n,3}$ are $O(n_3)$. Therefore, if $\frac{n_3}{(n_1+n_2)^{1/2}} \to 0$, it follows that

$$(n_1+n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}J_{n,3} \to 0 \quad (38)$$

and so

$$\Delta_{n1}^* \to 0. \quad (39)$$

Now consider Δ_{n2}^*. From (13),

$$\begin{aligned} E\Delta_{n2}^* &= \lambda Q'_{Z_{1+2}H_{1+2}}Q^{-1}_{H_{1+2}H_{1+2}}(n_1+n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}E(y_{n,3}) \\ E(y_{n,3}) &= (I_n - \lambda W_n)^{-1}_{n,3}[X_n\beta_1 + W_nJ_n\beta_2] \end{aligned} \quad (40)$$

where it should be recalled that $(I_n - \lambda W_n)^{-1}_{n,3}$ is the last n_3 rows of $(I_n - \lambda W_n)^{-1}$, and $E(y_{n,3})$ is an $n_3 \times 1$ vector. Assumptions 2 and 4(c) imply that the elements of $E(y_{n,3})$ are uniformly bound in absolute value; therefore, the elements of the $q_2 \times 1$ vector $(n_1+n_2)^{-1/2}H'_{n,1+2}W^u_{n,1+2}E(y_{n,3})$ are $O(\frac{n_3}{(n_1+n_2)^{1/2}})$ and hence,

$$\lim_{n_1\to\infty} E\Delta_{n2}^* = 0 \quad (41)$$

if $\frac{n_3}{(n_1+n_2)^{1/2}} \to 0$. Now consider the variance–covariance matrix of Δ_{n2}^*, say $VC_{\Delta_{n2}^*}$. In (37), let $M'_n = Q'_{Z_{1+2}H_{1+2}}Q^{-1}_{H_{1+2}H_{1+2}}H'_{n,1+2}W^u_{n,1+2}$, and note that Assumptions 2 and part (c) of 4 imply that the elements of the $s + r + 1 \times n_3$ matrix M'_n are uniformly bound in absolute value. It follows from (37) that

$$VC_{\Delta_{n2}^*} = \lambda^2(n_1+n_2)^{-1}M'_nVC_{y_{n,3}}M_n \quad (42)$$

where $VC_{y_{n,3}}$ is the lower $n_3 \times n_3$ block of VC_{y_n} given in (15). Assumption 4(c) implies that the row and column sums of $VC_{y_{n,3}}$ are uniformly bound in absolute value. Therefore, the elements of $M_n' VC_{y_{n,3}} M_n$ are $O(n_3)$ and so

$$VC_{\Delta_{n2}^*} \to 0 \tag{43}$$

if $\frac{n_3}{(n_1+n_2)^{1/2}} \to 0$. The results in (41) and (43) imply, via Chebyshev's inequality, that

$$\Delta_{n2}^* \xrightarrow{P} 0 \tag{44}$$

It follows from (37), (39), and (44) that $\Delta_n \xrightarrow{P} 0$. □

Proof of Theorem 3

Recall that under Theorem 3, we have $c > 0$. In light of (30)–(32), it should be clear that the large sample distribution of $(n_1+n_2)^{1/2}(\hat{\gamma}_2 - \gamma)$ differs when $c = 0$ when compared to when $c \neq 0$ only because if $c \neq 0$ the probability limit of Δ_n is not zero. In this case, the large sample distribution of $(n_1+n_2)^{1/2}(\hat{\gamma}_2 - \gamma)$ has a non-zero mean, which is the probability limit of $Q^{-1}_{\hat{Z}_{1+2}\hat{Z}_{1+2}} \Delta_n$ but is otherwise the same as in the case in which $c = 0$.

To determine this probability limit, first note that the limits in (37) do not depend upon the assumption that $\frac{n_3}{(n_1+n_2)^{1/2}} \to 0$. Hence, we determine the probability limit of Δ_n in terms of those of Δ_{n1}^* and Δ_{n2}^*.

Consider Δ_{n1}^*, and note by Assumption 6(a) that $(n_1+n_2)^{-1/2} H'_{n,1+2} W^u_{n,1+2} J_{n,3} \beta_2 \to F_{1+2}\beta_2$. Therefore, in light of (37),

$$\Delta_{n1}^* \to Q'_{Z_{1+2}H_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} F_{1+2}\beta_2. \tag{45}$$

Now consider Δ_{n2}^*. Assumption 6(b),(c) imply that

$$\begin{aligned} E\Delta_{n2}^* &= \lambda Q'_{Z_{1+2}H_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} (n_1+n_2)^{-1/2} H'_{n,1+2} W^u_{n,1+2} \\ &\quad \times (I_n - \lambda W_n)^{-1}_{n,3} [X_n\beta_1 + W_n J_n \beta_2] \\ &\to \lambda Q'_{Z_{1+2}H_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} [L_{1+2}\beta_1 + S_{1+2}\beta_2], \end{aligned} \tag{46}$$

which is a finite $(s+r) \times 1$ vector. The variance of Δ_{n2}^* was computed in the proof of Theorem 2. The expression for the variance of Δ_{n2}^* is given in (42) and was shown to converge to zero since $\frac{n_3}{(n_1+n_2)} \to 0$. Hence by Chebychev's inequality,

$$\Delta_{n2}^* \xrightarrow{P} \lambda Q'_{Z_{1+2}H_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} [L_{1+2}\beta_1 + S_{1+2}\beta_2] \tag{47}$$

Pulling results together, it now follows from (37), (45), and (47) that

$$\begin{aligned} Q^{-1}_{\hat{Z}_{1+2}\hat{Z}_{1+2}} \Delta_n \xrightarrow{P} Q^{-1}_{\hat{Z}_{1+2}\hat{Z}_{1+2}} \Big[& Q'_{Z_{1+2}H_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} F_{1+2}\beta_2 \\ & + \lambda Q'_{Z_{1+2}H_{1+2}} Q^{-1}_{H_{1+2}H_{1+2}} (L_{1+2}\beta_1 + S_{1+2}\beta_2) \Big] \end{aligned} \tag{48}$$

Theorem 3 follows from (30)–(32), (34), (37), (45), (47), and (48). □

References

Afifi A, Elashoff R (1966) Missing observations in multivariate statistics I. Review of the literature. J Am Stat Assoc 61(315):595–604

Afifi A, Elashoff R (1967) Missing observations in multivariate statistics II. Point estimation in simple linear regression. J Am Stat Assoc 62(317):10–29

Afifi A, Elashoff R (1969) Missing observations in multivariate statistics III. Large sample analysis of simple linear regression. J Am Stat Assoc 64(325):337–358

Anderson T (1957) Maximum likelihood estimates for a multivariate distribution when some observations are missing. J Am Stat Assoc 52(278):200–203

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Boston

Cressie NAC (1993) Statistics of spatial data. Wiley, New York

Friedman M (1962) The interpolation of time series by related series. J Am Stat Assoc 57(300):729–757

Haining R (2003) Spatial data analysis. Cambridge University Press, Cambridge

Haitovsky Y (1968) Missing data in regression analysis. J R Stat Soc Ser B 30(1):67–82

LeSage J, Pace K (2004) Models for spatially dependent missing data. J Real Estate Finan Econ 29(2):233–254

Kelejian H (1969) Missing observations in multivariate regression: efficiency of a first order method. J Am Stat Assoc 64(328):1609–1616

Kelejian HH, Prucha IR (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J Real Estate Finan Econ 17(1):99–121

Kelejian H, Prucha I (2004) Estimation of simultaneous systems of spatially interrelated cross sectional equations. J Econom 118:27–50

Kelejian H, Prucha I (2007) HAC estimation in a spatial framework. J Econom 140:31–154

Kelejian H, Murrell P, Shepotylo O (2008) Spatial spillovers in the development of institutions. Manuscript, University of Maryland

Kmenta J (1986) Elements of econometrics, 2nd edn. Macmiilan, New York

Little J, Rubin D (2002) Statistical analysis with missing data, 2nd edn. Wily, New York

Maddala GS (1977) Econometrics. McGraw-Hill, New York

Müller W (2007) Collecting spatial data, 3rd edn. Springer, Berlin