Parameter estimation in spatial econometric models with non-random missing data

Seya, Hajime
Tomari, Masashi
Uno, Shohei

(Citation)
Applied Economics Letters,28(6):440-446

(Issue Date)
2021-03-30

(Resource Type)
journal article

(Version)
Accepted Manuscript

Parameter estimation in spatial econometric models with non-random missing data

Hajime Seya[1], Masashi Tomari, Shohei Uno

Graduate School of Engineering Faculty of Engineering, Kobe University,

1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan,

E-mail: hseya@people.kobe-u.ac.jp

Abstract: This study examines the problem of parameter estimation in spatial econometric/social interaction models with non-random missing outcome data. First, we construct a sample selection model considering spatial lag (autoregressive) dependence. Then, we suggest a parameter estimation method for this model by slightly modifying the Bayesian Markov chain Monte Carlo algorithm proposed in an existing study. A simple illustration indicates that the proposed parameter estimation method performs well overall if the spatial autocorrelation is moderate (spatial parameter equals 0.5 or less), even under a relatively high missing data ratio (around 40%).

Keywords: sample selection, spatial lag model (SLM), spatial autocorrelation, social interaction, Bayesian Markov chain Monte Carlo (MCMC)

JEL: C40; C50; C13

Acknowledgement

This study was funded by JSPS KAKENHI Grant Numbers 17K14738 and 18H03628.

[1] Corresponding author

1. Introduction

Recently, there has been an increase in spatial/social network-related non-random missing data. For example, if, from a typical classroom, only excellent students reported their GPA scores in a survey on the effect of students' relationships on their GPA scores, the survey may then suffer from sample selection problems. This study addresses the problem of parameter estimation in spatial econometric/social interaction models with non-random missing outcome data due to sample selection[2].

Because applying the ordinary least squares (OLS) method to a regression model with non-random missing outcome data may result in biased parameter estimates, the sample selection model of Heckman (1976) is sometimes applied. Attempts to consider sample selection under the framework of Heckman (1976) in spatial econometric models were first made by McMillen (1995), followed by Flores-Lagunes and Schnier (2012). The latter study showed that consistent parameter estimates are obtainable through the generalized method of moments (GMM). However, there are apparent issues in its construction of an outcome models' spatial weight matrix using only observed samples: as noted in Hoshino (2019), an interaction network (or spatial weight) structure with only a subsample of the complete data obtained by discarding observations with missing items does not coincide with the "true" network structure. Consequently, such incomplete interaction networks (or spatial weights) may lead to biases in the parameter estimates. By contrast, Doğan and Taspinar (2018) (hereafter DT) proposed an estimation method for parameters integrating the spatial error model (SEM) and the sample selection model using the Bayesian Markov chain Monte Carlo (MCMC) methods. They adopted an imputing approach to create pseudo-complete data to consider the effects thorough an omitted network[3]. However, in DT, a parameter that measures the correlation between the outcome and the selection equation is always estimated close to zero and results in an inability to improve

[2] For the case of random missing data, a parameter estimation method was presented by Wang and Lee (2013). For spatial econometrics in general, see Arbia (2006). Our study focuses on spatial (auto)correlation among individuals, not among alternatives (De Grange et al., 2013). See Billé and Arbia for spatial discrete choice models (2019).

[3] Omori (2007) and Wiesenfarth and Kneib (2010) considered a sampler that imputes latent variables for the selection equation but uses only the observed responses from the outcome equation. By doing so, we can improve the mixing and convergence problem by integrating out missing values from a likelihood function. However, in spatial econometric/social interaction models, where *connections* or *networks* between samples are important, listwise deletion of missing data leads to the deletion of such connections or networks.

sample selection bias. In this study, a parameter estimation technique is presented in the form of a minor modification of the MCMC algorithm in DT in a sample selection model based on a spatial lag model (SLM). A simple numerical example follows, confirming the conditions in which this parameter estimation technique performs well.

2. Spatial econometric models and sample selection

The regression models frequently used in spatial econometrics are SLM[4], in which spatial autocorrelation is considered to be the autocorrelation between outcome variables, and SEM, in which spatial autocorrelation is considered to be the autocorrelation between error terms. DT formulated a model considering sample selection in an SEM as follows:

$$y_{1i}^* = \mathbf{x}'_{1i}\boldsymbol{\beta} + u_{1i}, \quad u_{1i} = \lambda_1 \textstyle\sum_{j=1}^{n} w_{ij}\, u_{1j} + \varepsilon_{1i}, \tag{1}$$

$$y_{2i}^* = \mathbf{x}'_{2i}\boldsymbol{\delta} + u_{2i}, \quad u_{2i} = \lambda_2 \textstyle\sum_{j=1}^{n} m_{ij}\, u_{2j} + \varepsilon_{2i}, \tag{2}$$

$$y_{2i} = \begin{cases} y_{2i}^* & if \quad y_{1i} = 1 \\ \text{missing} & if \quad y_{1i} = 0, \end{cases} \tag{3}$$

Here, eq. (1) is referred to as a selection equation, and eq. (2) is referred to as an outcome equation. y_{1i}^* and y_{2i}^* are continuous latent variables, $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ are the point i exogenous explanatory variables of $k_1 \times 1$ and $k_2 \times 1$ (including the respective constant terms), and $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are the corresponding estimation parameters of $k_1 \times 1$ and $k_2 \times 1$, respectively. Dummy variable y_{1i} (observation: 1, missing: 0) and latent variable y_1^* concerning the presence or absence of observation use indicator function $\mathrm{I}()$ and are linked by $y_{1i} = \mathrm{I}(y_{1i}^* > 0)$. In addition, y_{2i} is observed only when $(y_{1i} = 1)$ based upon eq. (3). w_{ij} and m_{ij} are spatial weight matrix components and can be the same. λ_1 and λ_2 are the spatial parameters.

In the case of an SLM, the selection equation and outcome equation may be formulated as follows:

$$y_{1i}^* = \rho_1 \textstyle\sum_{j=1}^{n} w_{ij}\, y_{1j}^* + \mathbf{x}'_{1i}\boldsymbol{\beta} + \varepsilon_{1i}, \tag{4}$$

$$y_{2i}^* = \rho_2 \textstyle\sum_{j=1}^{n} m_{ij}\, y_{2j}^* + \mathbf{x}'_{2i}\boldsymbol{\delta} + \varepsilon_{2i}. \tag{5}$$

---

[4] Alternatively, spatial autoregressive model.

Here, ρ_1 and ρ_2 are the spatial parameters. Negative spatial autocorrelation is present when ρ_1 (or ρ_2) < 0 , and positive spatial autocorrelation is present when ρ_1 (or ρ_2) > 0 . The assumptions for the error terms are described later. Eq. (4) and eq. (5) are not necessarily used as a pair. If price competition between gas stations is considered, the prices at one gas station and a nearby station are in a direct competition relationship; therefore, eq. (5) may be more desirable than eq. (2) as the outcome equation. However, we can still use eq. (1) or $y_{1i}^* = \mathbf{x}'_{1i}\boldsymbol{\beta} + \varepsilon_{1i}$ instead of eq. (4) for the selection equation if a spatial-lag type autocorrelation is difficult to assume.

The error terms ε_{1i} and ε_{2i} represent bivariate normal distribution and are assumed to be governed by $(\varepsilon_{1i}, \varepsilon_{2i})' \sim i.i.d. N(\mathbf{0}, \boldsymbol{\Sigma})$ ($\mathbf{0}$ is a 2×1 zero vector). Here, $\boldsymbol{\Sigma}$ is given by the following equation:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & \varrho\sigma_2 \\ \varrho\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{6}$$

Here, ϱ indicates the correlation coefficient of ε_{1i} and ε_{2i} . For convenience, $\boldsymbol{\Sigma}$ is often parameterized as follows (Omori, 2007):

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_{12}^2 + \xi^2 \end{pmatrix}. \tag{6}$$

When eq. (6) is used, even if an (constrained) inverse-Wishart distribution is not assumed for $\boldsymbol{\Sigma}$ in the MCMC estimation, σ_{12} and ξ^2 can be conveniently sampled on an individual basis[5].

Eq. (4) and eq. (5) can be written in matrix form as follows:

$$\mathbf{y}_1^* = \mathbf{S}^{-1}(\rho_1)\mathbf{X}_1\boldsymbol{\beta} + \mathbf{S}^{-1}(\rho_1)\boldsymbol{\varepsilon}_1, \tag{7}$$

$$\mathbf{y}_2^* = \mathbf{R}^{-1}(\rho_2)\mathbf{X}_2\boldsymbol{\delta} + \mathbf{R}^{-1}(\rho_2)\boldsymbol{\varepsilon}_2, \tag{8}$$

where $\mathbf{y}_1^*$ and $\mathbf{y}_2^*$ are vectors of $n \times 1$ respectively consisting of latent variables y_{1i}^* and y_{2i}^*, $\mathbf{X}_1$ and $\mathbf{X}_2$ are explanatory variable matrices of $n \times k_1$ and $n \times k_2$, and $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ are error term vectors of $n \times 1$ conferring elements thereof with ε_{1i} and ε_{2i}. In addition, $\mathbf{S}^{-1}(\rho_1) = (\mathbf{I} - \rho_1\mathbf{W})^{-1}$ and $\mathbf{R}^{-1}(\rho_2) = (\mathbf{I} - \rho_2\mathbf{M})^{-1}$, and $\mathbf{W}$ and $\mathbf{M}$ are (possibly row-standardized) spatial weight matrices of $n \times n$. The latent variable $\mathbf{y}_2^*$ and observed variables $\mathbf{y}_2$ are linked by the following equation:

---

[5] It may be a strong assumption that ε_{1i} and ε_{2i} are governed by two-dimensional normal distribution, and they are thought to form a flexible distribution with copulas as in, for example, Wojtys et al. (2016). However, emphasis is placed on widespread (or more standard) assumptions, and this study assumes two-dimensional normal distribution.

$$\mathbf{y_2} = \begin{cases} \mathbf{y}_2^* = \mathbf{R}^{-1}(\rho_2)\mathbf{X_2}\boldsymbol{\delta} + \mathbf{R}^{-1}(\rho_2)\boldsymbol{\varepsilon}_2 & if \quad y_{1i} = 1 \\ \text{missing} & if \quad y_{1i} = 0, \end{cases} \tag{9}$$

Here, if $\mathbf{y}^* = (\mathbf{y}_1^{*\prime}, \mathbf{y}_2^{*\prime})'$, $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0}_{n\times k_2} \\ \mathbf{0}_{n\times k_1} & \mathbf{X}_2 \end{pmatrix}$, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\delta}')'$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1'\mathbf{S}^{-1\prime}(\rho_1), \boldsymbol{\varepsilon}_2'\mathbf{R}^{-1\prime}(\rho_2))'$ and $\Xi(\rho_1, \rho_2) = \begin{pmatrix} \mathbf{S}^{-1}(\rho_1) & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \mathbf{R}^{-1}(\rho_2) \end{pmatrix}$, the selection equation and outcome equation can be compactly written as follows:

$$\mathbf{y}^* = \Xi(\rho_1, \rho_2)\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{10}$$

$$\boldsymbol{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Omega} = \Xi(\rho_1, \rho_2)(\boldsymbol{\Sigma} \otimes \mathbf{I})\Xi'(\rho_1, \rho_2), \tag{11}$$

where $\otimes$ denotes a Kronecker product. In the following, the SLM combined with the sample selection model is referred to as a spatial lag-sample selection model (SL-SSM).

3. Parameter estimation of the spatial lag-sample selection model

The following algorithm resembles Algorithm 2 in DT for the SEM, although different algorithms may be designed. DT indicate that there are no major differences between the estimates from the perspective of bias even if the other algorithm is used.

First, it is assumed that the prior distribution of each parameter is independent, and the following prior distribution is assumed: $\boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{\mu}_0, \mathbf{V}_0), \xi^2 \sim \mathrm{IG}\left(\frac{a_0}{2}, \frac{b_0}{2}\right), \sigma_{12} \sim \mathrm{N}(\gamma_0, G_0), \rho_1 \sim \mathrm{Unif}(-1,1)$, $\rho_2 \sim \mathrm{Unif}(-1,1)$, where $\mathrm{IG}\left(\frac{a_0}{2}, \frac{b_0}{2}\right)$ and $\mathrm{Unif}(-1,1)$ denote inverse gamma distributions with shape and scale parameters $a_0/2$ and $b_0/2$, and uniform distribution between –1 and 1, respectively. The posterior distribution may be written as $p(\boldsymbol{\theta}, \xi^2, \sigma_{12}, \rho_1, \rho_2, \mathbf{y}^*|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\xi^2)p(\sigma_{12})p(\rho_1)p(\rho_2)p(\mathbf{y}^*|\boldsymbol{\theta}, \xi^2, \sigma_{12}, \rho_1, \rho_2)p(\mathbf{y}|\boldsymbol{\theta}, \xi^2, \sigma_{12}, \rho_1, \rho_2, \mathbf{y}^*)$. The full conditional posterior distributions (or simply, full conditionals), except for $\mathbf{y}_1^*$ and $\mathbf{y}_2^*$, are given in appendix A.

The full conditionals for $\mathbf{y}_1^*$ and $\mathbf{y}_2^*$ are given as a truncated normal distribution (TN) (Geweke, 1991). When sample sets $N_0=\{i: y_{1i} = 0\}$ and $N_1=\{i: y_{1i} = 1\}$ are defined, $\mathbf{y}_2^*$ is sampled from the following:

$$\mathbf{y}_2^* \sim \mathrm{N}\left(\mathbf{R}^{-1}(\rho_2)\mathbf{X}_2\boldsymbol{\delta} + \sigma_{12}\mathbf{R}^{-1}(\rho_2)\mathbf{S}(\rho_1)(\mathbf{y}_1^* - \mathbf{S}^{-1}(\rho_1)\mathbf{X}_1\boldsymbol{\beta}), (\sigma_2^2 - \sigma_{12}^2)\mathbf{R}^{-1}(\rho_2)\mathbf{R}^{-1\prime}(\rho_2)\right). \tag{12}$$

For $i \in N_0$ and $i \in N_1$, we set $\mathbf{y}_2^* = \mathbf{y}_2$. DT, in the case of an SEM, sample $\mathbf{y}_1^*$ as follows:

$$\mathbf{y_1^*} \sim \mathrm{TN}_{(-\infty,0)}(\mathbf{X_1\beta}, \mathbf{S}^{-1}(\rho_1)\mathbf{S}^{-1\prime}(\rho_1)),\ for\ \ i \in N_0 \tag{13}$$

$$\mathbf{y_1^*} \sim \mathrm{TN}_{(0,\infty)}(\mathbf{X_1\beta}, \mathbf{S}^{-1}(\rho_1)\mathbf{S}^{-1\prime}(\rho_1)),\ for\ \ i \in N_1 \tag{14}$$

Here, TN $_{(a,b)}$ denotes normal distribution truncated on the interval (*a*,*b*). In the case of an SLM, it is given by:

$$\mathbf{y_1^*} \sim \mathrm{TN}_{(-\infty,0)}(\mathbf{S}^{-1}(\rho_1)\mathbf{X_1\beta}, \mathbf{S}^{-1}(\rho_1)\mathbf{S}^{-1\prime}(\rho_1)), for\ \ i \in N_0 \tag{15}$$

$$\mathbf{y_1^*} \sim \mathrm{TN}_{(0,\infty)}(\mathbf{S}^{-1}(\rho_1)\mathbf{X_1\beta}, \mathbf{S}^{-1}(\rho_1)\mathbf{S}^{-1\prime}(\rho_1)),\ for\ \ i \in N_1 \tag{16}$$

However, as shown in, for instance, Koop (2007), in the case of $i \in N_1$, we need to consider the correlation between error terms, so we can sample $\mathbf{y}_1^*$ as follows:

$$\mathbf{y_1^*} \sim \mathrm{TN}_{(0,\infty)}(\mathbf{S}^{-1}(\rho_1)\mathbf{X_1\beta} + \frac{\sigma_{12}}{\sigma_2^2}\mathbf{S}^{-1}(\rho_1)\mathbf{R}(\rho_2)(\mathbf{y_2^*} - \mathbf{R}^{-1}(\rho_2)\mathbf{X_2}\boldsymbol{\delta}), (1 - \frac{\sigma_{12}^2}{\sigma_2^2})\mathbf{S}^{-1}(\rho_1)\mathbf{S}^{-1\prime}(\rho_1)),$$

(17)

We expect that the bias in the estimate for σ_{12} can be improved using eq. (17) for $i \in N_1$ instead of eq. (16).

4. Numerical illustration

4.1. Experiment settings

Normally, the properties of an estimator are evaluated through Monte Carlo experiments, but MCMC requires a large computation load even for a one-time trial. Therefore, in this section, the validity of the proposed technique is ascertained through a numerical illustration.

We assume the following spatial-lag type data generating process (DGP).

$$\begin{pmatrix}\mathbf{y}_1^*\\ \mathbf{y}_2^*\end{pmatrix} = \begin{pmatrix}\mathbf{S}^{-1}(\rho_1) & \mathbf{0}_{N\times N}\\ \mathbf{0}_{N\times N} & \mathbf{R}^{-1}(\rho_2)\end{pmatrix}\begin{pmatrix}\mathbf{X}_1 & \mathbf{0}_{N\times K_2}\\ \mathbf{0}_{N\times K_1} & \mathbf{X}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\delta}\end{pmatrix} + \begin{pmatrix}\mathbf{S}^{-1}(\rho_1)\boldsymbol{\varepsilon}_1\\ \mathbf{R}^{-1}(\rho_2)\boldsymbol{\varepsilon}_2\end{pmatrix}.$$

The explanatory variables, when $\mathbf{X}_1 = (\mathbf{1}, \mathbf{X}_{1,1}, \mathbf{X}_{1,2})$ and $\mathbf{X}_2 = (\mathbf{1}, \mathbf{X}_{2,1}, \mathbf{X}_{2,2})$, were set to $\mathbf{X}_{1,1} \sim \mathrm{i.i.d.Unif}(0, \mathrm{upper}), \mathbf{X}_{1,2} \sim \mathrm{i.i.d.N}(0,3)$, $\mathbf{X}_{2,1} \sim \mathrm{i.i.d.N}(0,3)$, and $\mathbf{X}_{2,2} = \mathbf{X}_{1,2}$. Here, "upper" is modified as appropriate in order to adjust the missing ratio. The missing ratio here is the percentage of observed values in which $\mathbf{y}_1 = 0$—in other words, where $\mathbf{y}_2$ was not observed—and "upper" was adjusted so that the ratio was around 40% in the present verification. The regression coefficient true values were $\boldsymbol{\beta} = (2.0, -1.0, 1.0)'$ and $\boldsymbol{\delta} = (1.0,\ 0.5, -0.5)'$. A spatial weight matrix independently generated sample x and y coordinates from $\mathrm{i.i.d.Unif}\,(0,10)$, which was row normalized assuming values close to 4 to be of weight 1. Finally, spatial autocorrelation parameters were adjusted to when spatial autocorrelation is weak, moderate, and strong, and three respective patterns were set for $(\rho_1, \rho_2) = (0.2, 0.2)$; $(0.5, 0.5)$; and $(0.8, 0.8)$. In the above settings, the error term was assumed to be given in the following manner:

$$\begin{pmatrix} \varepsilon_{1,i} \\ \varepsilon_{2,i} \end{pmatrix} \sim \mathrm{i.i.d.N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}\right).$$

The correlation parameter σ_{12} is adjusted to when sample selection is weak, moderate, and strong, and three respective patterns were set for σ_{12} = 0.2, 0.5, 0.8. Samples were generated for sample size *n*=300 or *n*=700, and the nine scenarios under consideration are shown in Table 1.

Table 1: Verification scenarios

4.2. Results of the experiments

Whether each parameter can be estimated close to true values is confirmed for each scenario given in Table 1. In the Bayesian MCMC estimation, samples were taken from each full conditional using 10,000 as the number of iterations (including a burn-in period of 2000)[6]. Hyperparameters were set as $\boldsymbol{\mu}_0 = \mathbf{0}, \mathbf{V}_0 = 10^3 \times \mathbf{I}, a_0 = b_0 = 1, \gamma_0 = 0,$ and $G_0 = 10^3$. The estimation results are shown in Tables 2–10. The name of the columns, "Mean," "Std. dev.," "B.C.I. 5%," and "B.C.I. 95%," denote the mean, standard deviation, and Bayesian credible intervals of 5 % and 95%, respectively (see also Table 1).

Note that scenario A is based on our algorithm given in eq. (17) and scenario A' is based on the SLM version of DT's algorithm given in eq. (16). By comparing the results under scenarios A and A', we can see that it is possible to improve the accuracy in reproducing σ_{12} by using eq. (17). Therefore, for the other scenarios B to H, we only use our algorithm (eq. (17)).

In the case of n = 300, the bias in the spatial autocorrelation parameter ρ_2 is found to not be minor. However, when $\rho_1, \rho_2 = 0.2$, the biases in the other parameters are not large. We find that as spatial autocorrelation increases, the biases in the coefficient and variance parameters increase. When n = 700, biases are relatively smaller, as expected, which also applies to the spatial parameters. However, when ρ_1 and $\rho_2 = 0.8$, similar to cases of n = 300, relatively large biases are found in the regression coefficients. "The results for scenarios G and H suggest that compared to the effect of high positive spatial autocorrelation (ρ_1 and $\rho_2 = 0.8$), the effect of the strong sample selection (σ_{12}=0.8) on the parameter bias (Mean – True value) is minor, both for the coefficients and variance σ_{22} estimates. It is also notable that focusing on the estimate for σ_{12}, when spatial autocorrelation is 0.5 or less, fairly accurate reproduction is attained for both n = 300 and 700 in terms of parameter bias (see Tables 2, 4, 6, and 7)."

These results suggest the potential of this technique's application when the spatial autocorrelation is minor to moderate (spatial parameter equals 0.5 or less), even under a relatively high missing data ratio (around 40%).

---

[6] MCMC convergence was confirmed by Geweke (1992)'s convergence diagnostic. Because our model structure was fairly simple, the shape of the posterior distributions of each parameter was obtained as unimodal distributions.

Table 2: Scenario A (n = 300, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.2$)

Table 3: Scenario A' (n = 300, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.2$)

Table 4: Scenario B (n = 300, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.5$)

Table 5: Scenario C (n = 300, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.8$)

Table 6: Scenario D (n = 700, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.2$)

Table 7: Scenario E (n = 700, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.5$)

Table 8: Scenario F (n = 700, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.8$)

Table 9: Scenario G (n = 700, σ_{12} = 0.2, $\rho_1, \rho_2 = 0.2$)

Table 10: Scenario H (n = 700, σ_{12} = 0.8, $\rho_1, \rho_2 = 0.2$)

References

1. Arbia, G. (2006) *Spatial Econometrics*, Springer, Berlin.
2. Billé, A.G. and Arbia, G. (2019) Spatial limited dependent variable models: A review focused on specification, estimation, and health economics applications, *Journal of Economic Surveys*, 33 (5), 1531–1554.
3. De Grange, L.; Boyce, D.; González, F. and Ortúzar, J. D. (2013) Integration of spatial correlation into a combined travel model with hierarchical levels, *Spatial Economic Analysis*, 8, 71–91.
4. Doğan, O. and Taşpinar, S. (2018) Bayesian inference in spatial sample selection models, *Oxford Bulletin of Economics and Statistics*, 80 (1), 90–121.
5. Flores-Lagunes, A. and Schnier, K.E. (2012) Estimation of sample selection models with spatial dependence, *Journal of Applied Econometrics*, 27 (2), 173–204.
6. Koop, G., Poirier, D.J. and Tobias, J.L. (2007) *Bayesian Econometric Methods*, Cambridge University Press, Cambridge.
7. Geweke, J. (1991) Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities, in E.M. Keramidas and S.M. Kaufman (eds.), *Computing Science and Statistics*, 571–578.
8. Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments, in J.M. Bernardo, J.O. Berger, A.P. Dawid and. A.F.M. Smith (eds.), *Bayesian statistics 4*, 641–649.
9. Hoshino, T. (2019) Two-step estimation of incomplete information social interaction models with sample selection, *Journal of Business & Economic Statistics*, 37 (4), 598–612.
10. Heckman, J. (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models, *Annals of Economic and Social Measurement*, 5, 475–492.
11. LeSage, J.P. and Pace, R.K. (2009) *Introduction to Spatial Econometrics*, CRC Press, Boca Raton.
12. McMillen, D.P. (1995) Selection bias in spatial econometric models, *Journal of Regional Science*, 35 (3), 417–436.
13. Omori, Y. (2007) Efficient Gibbs sampler for Bayesian analysis of a sample selection model, *Statistics & Probability Letters*, 77 (12), 1300–1311.
14. Wang, W. and Lee, L.F. (2013) Estimation of spatial autoregressive models with randomly missing data in the dependent variable, *The Econometrics Journal*, 16 (1), 73–102.
15. Wiesenfarth, M. and Kneib, T. (2010) Bayesian geoadditive sample selection models, *Journal of the Royal Statistical Society: Series C*, 59 (3), 381–404.
16. Wojtys, M., Marra, G. and Radice, R. (2016) Copula regression spline sample selection models: The R Package SemiParSampleSel, *Journal of Statistical Software*, 71 (6).

Appendix A

The full conditionals, except for $\mathbf{y}_1^*$ and $\mathbf{y}_2^*$, are given as follows:

$$\boldsymbol{\theta}|\rho_1,\rho_2,\boldsymbol{\Sigma},\mathbf{y}^*,\mathbf{y} \sim \mathrm{N}(\boldsymbol{\mu}_1,\mathbf{V}_1), \tag{A-1}$$

$$\xi^2|\boldsymbol{\theta},\sigma_{12},\rho_1,\rho_2,\mathbf{y}^*,\mathbf{y} \sim \mathrm{IG}\left(\frac{a_1}{2},\frac{b_1}{2}\right), \tag{A-2}$$

$$\sigma_{12}|\boldsymbol{\theta},\xi^2,\rho_1,\rho_2,\mathbf{y}^*,\mathbf{y} \sim \mathrm{N}(\gamma_1,G_1), \tag{A-3}$$

where $\mathbf{V}_1 = [\mathbf{V}_0^{-1} + \{\Xi(\rho_1,\rho_2)\mathbf{X}\}'^{\boldsymbol{\Omega}^{-1}}\Xi(\rho_1,\rho_2)\mathbf{X}]^{-1}, \boldsymbol{\mu}_1 = \mathbf{V}_1(\mathbf{V}_0^{-1}\boldsymbol{\mu}_0 + \{\Xi(\rho_1,\rho_2)\mathbf{X}\}'\boldsymbol{\Omega}^{-1}\mathbf{y}^*)$,

$$a_1 = a_0 + n,\ b_1 = b_0 + \sigma_{12}^2\mathbf{e}_1{}'\mathbf{e}_1 - 2\sigma_{12}\mathbf{e}_2{}'\mathbf{e}_1 + \mathbf{e}_2{}'\mathbf{e}_2,$$

$$\mathbf{e}_1 = \mathbf{S}^{-1}(\rho_1)\mathbf{y}_1^* - \mathbf{X}_1\boldsymbol{\beta};\ \ \mathbf{e}_2 = \mathbf{R}^{-1}(\rho_2)\mathbf{y}_2^* - \mathbf{X}_2\boldsymbol{\delta},$$

$$G_1 = \ (G_0^{-1} + \xi^{-2}\mathbf{e}_1'\mathbf{e}_1)^{-1};\ \ \gamma_1 = G_1(G_0^{-1}\gamma_0 + \xi^{-2}\mathbf{e}_1{}'\mathbf{e}_2).$$

For spatial parameters ρ_1 and ρ_2, we have

$$p(\rho_1|\boldsymbol{\theta},\rho_2,\boldsymbol{\Sigma},\mathbf{y}^*,\mathbf{y}) \propto |\mathbf{S}(\rho_1)| \times \exp\left\{-\frac{1}{2}\left(\mathbf{e_1}'\mathbf{e_1} + \frac{\sigma_{12}^2}{\xi^2}\mathbf{e_1}'\mathbf{e_1} - \frac{2\sigma_{12}}{\xi^2}\mathbf{e_2}'\mathbf{e_1}\right)\right\}, \tag{A-4}$$

$$p(\boldsymbol{\rho_2}|\boldsymbol{\theta},\rho_1,\boldsymbol{\Sigma},\mathbf{y}^*,\mathbf{y}) \propto |\mathbf{R}(\rho_2)| \times \exp\left\{-\frac{1}{2}\left(\frac{1}{\xi^2}\mathbf{e_2}'\mathbf{e_2} - \frac{2\sigma_{12}}{\xi^2}\mathbf{e_2}'\mathbf{e_1}\right)\right\}. \tag{A-5}$$

The full conditional of ρ_1 and ρ_2 is not a standard distribution and Gibbs sampling cannot be used; thus, sampling is conducted using, for instance, the random walk Metropolis algorithm.

$$\begin{pmatrix}\rho_1{}^{new}\\ \rho_2{}^{new}\end{pmatrix} = \begin{pmatrix}\rho_1{}^{t-1}\\ \rho_2{}^{t-1}\end{pmatrix} + \begin{pmatrix}\eta_{\rho_{1,t}} & 0\\ 0 & \eta_{\rho_{2,t}}\end{pmatrix} \times \begin{pmatrix}\tau_{\rho_{1,t}}\\ \tau_{\rho_{2,t}}\end{pmatrix};\quad \begin{pmatrix}\tau_{\rho_{1,t}}\\ \tau_{\rho_{2,t}}\end{pmatrix} \sim \mathrm{N}(\mathbf{0}_{2\times1},\mathbf{I}_2).$$

Tuning parameter η is adjusted during the burn-in period so that the adoption rate is from 40% to 60% (LeSage and Pace, 2009). The adoption rate is given as follows:

$$\alpha(\rho_1{}^{t-1},\rho_1{}^{new}) = \min\left(\frac{p(\rho_1{}^{new}|\mathbf{rest},\mathbf{y}^*,\mathbf{y})}{p(\rho_1{}^{t-1}|\mathbf{rest},\mathbf{y}^*,\mathbf{y})},1\right), \tag{A-6}$$

$$\alpha(\rho_2{}^{t-1}\rho_2{}^{new}) = \min\left(\frac{p(\rho_2{}^{new}|\mathbf{rest},\mathbf{y}^*,\mathbf{y})}{p(\rho_2{}^{t-1}|\mathbf{rest},\mathbf{y}^*,\mathbf{y})},1\right), \tag{A-7}$$

where $\mathbf{rest}$ denotes a vector of parameters other than ρ_1 or ρ_2. Candidates $\rho_1{}^{new}$ and $\rho_2{}^{new}$ are adopted with probability $\alpha(\rho_1{}^{t-1},\rho_1{}^{new})$ and $\alpha(\rho_2{}^{t-1}\rho_2{}^{new})$.

Tables

Table 1: Verification scenarios

Scenario	A	A'	B	C	D	E	F	G	H
n	300	300	300	300	700	700	700	700	700
ρ_1, ρ_2	0.2	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.2
σ_{12}	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.2	0.8
Upper	3.0	3.0	3.2	3.5	3.0	3.2	3.5	3.0	3.0
Missing rate	0.377	0.377	0.363	0.343	0.419	0.413	0.431	0.416	0.410
Results	Table 2	Table 3	Table 4	Table 5	Table 6	Table 7	Table8	Table 9	Table10

Table 2: Scenario A (n = 300, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.2$)

	True value	Mean	Std.dev.	B.C.I. 5%	B.C.I. 95%
β_1	2.0	1.626	0.214	1.277	1.979
β_2	–1.0	–0.800	0.117	–0.993	–0.608
β_3	1.0	0.799	0.092	0.649	0.956
δ_1	1.0	1.253	0.153	1.010	1.511
δ_2	0.5	0.559	0.041	0.493	0.626
δ_3	–0.5	–0.483	0.062	–0.586	–0.382
σ_{12}	0.5	0.424	0.190	0.092	0.710
σ_{22}	1.0	0.985	0.116	0.811	1.190
ρ_1	0.2	0.194	0.089	0.044	0.334
ρ_2	0.2	0.012	0.081	–0.124	0.145

Table 3: Scenario A' (n = 300, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.2$)

	True value	Mean	Std.dev.	B.C.I. 5%	B.C.I. 95%
β_1	2.0	1.600	0.215	1.252	1.959
β_2	–1.0	–0.788	0.118	–0.985	–0.595
β_3	1.0	0.772	0.088	0.628	0.921
δ_1	1.0	1.416	0.138	1.195	1.648
δ_2	0.5	0.566	0.042	0.495	0.636
δ_3	–0.5	–0.547	0.053	–0.634	–0.460
σ_{12}	0.5	0.061	0.091	–0.087	0.214
σ_{22}	1.0	0.943	0.100	0.790	1.119
ρ_1	0.2	0.198	0.088	0.052	0.344
ρ_2	0.2	0.016	0.073	–0.110	0.136

Table 4: Scenario B (n = 300, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.5$)

	True value	Mean	Std.dev.	B.C.I. 5%	B.C.I. 95%
β_1	2.0	1.544	0.216	1.201	1.913
β_2	–1.0	–0.745	0.119	–0.948	–0.556
β_3	1.0	0.765	0.093	0.619	0.926
δ_1	1.0	1.548	0.164	1.284	1.824
δ_2	0.5	0.594	0.046	0.517	0.669
δ_3	–0.5	–0.470	0.067	–0.579	–0.358
σ_{12}	0.5	0.437	0.210	0.091	0.773
σ_{22}	1.0	1.123	0.135	0.928	1.369
ρ_1	0.5	0.432	0.083	0.284	0.556
ρ_2	0.5	0.223	0.068	0.111	0.331

Table 5: Scenario C (n = 300, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.8$)

	True value	Mean	Std.dev.	B.C.I. 5%	B.C.I. 95%
β_1	2.0	1.502	0.240	1.120	1.900
β_2	–1.0	–0.713	0.125	–0.922	–0.511
β_3	1.0	0.652	0.096	0.499	0.819
δ_1	1.0	2.239	0.275	1.804	2.701
δ_2	0.5	0.676	0.066	0.569	0.787
δ_3	–0.5	–0.524	0.084	–0.659	–0.386
σ_{12}	0.5	0.254	0.221	–0.112	0.613
σ_{22}	1.0	2.180	0.299	1.729	2.707
ρ_1	0.8	0.730	0.037	0.666	0.787
ρ_2	0.8	0.569	0.049	0.486	0.649

Table 6: Scenario D (n = 700, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.2$)

	True value	Mean	Std.dev.	B.C.I. 5%	B.C.I. 95%
β_1	2.0	2.050	0.179	1.766	2.354
β_2	–1.0	–1.049	0.096	–1.208	–0.897
β_3	1.0	1.126	0.080	0.995	1.259
δ_1	1.0	1.061	0.092	0.912	1.216
δ_2	0.5	0.498	0.028	0.453	0.543
δ_3	–0.5	–0.469	0.044	–0.541	–0.398
σ_{12}	0.5	0.540	0.122	0.327	0.726
σ_{22}	1.0	1.001	0.077	0.882	1.133
ρ_1	0.2	0.195	0.052	0.107	0.280
ρ_2	0.2	0.145	0.045	0.070	0.219

Table 7: Scenario E (n = 700, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.5$)

	True value	Mean	Std.dev.	B.C.I. 5%	B.C.I. 95%
β_1	2.0	1.885	0.173	1.610	2.183
β_2	–1.0	–0.967	0.094	–1.124	–0.820
β_3	1.0	1.033	0.081	0.903	1.173
δ_1	1.0	1.241	0.094	1.091	1.402
δ_2	0.5	0.522	0.031	0.470	0.573
δ_3	–0.5	–0.498	0.045	–0.572	–0.424
σ_{12}	0.5	0.441	0.131	0.214	0.647
σ_{22}	1.0	1.173	0.092	1.031	1.330
ρ_1	0.5	0.515	0.033	0.460	0.569
ρ_2	0.5	0.391	0.041	0.317	0.453

Table 8: Scenario F (n = 700, σ_{12} = 0.5, $\rho_1, \rho_2 = 0.8$)

	True value	Mean	Std.dev.	B.C.I. 5%	B.C.I. 95%
β_1	2.0	1.543	0.164	1.278	1.816
β_2	–1.0	–0.781	0.086	–0.924	–0.641
β_3	1.0	0.846	0.073	0.726	0.967
δ_1	1.0	1.597	0.142	1.363	1.828
δ_2	0.5	0.517	0.046	0.443	0.592
δ_3	–0.5	–0.558	0.070	–0.672	–0.442
σ_{12}	0.5	0.158	0.151	–0.097	0.398
σ_{22}	1.0	2.212	0.249	1.840	2.656
ρ_1	0.8	0.746	0.019	0.714	0.776
ρ_2	0.8	0.694	0.024	0.655	0.734

Table9: Scenario G (n = 700, σ_{12} = 0.2, $\rho_1, \rho_2 = 0.2$)

	True value	Mean	Std.dev.	B.C.I. 5%	B.C.I. 95%
β_1	2.0	2.136	0.189	1.840	2.456
β_2	–1.0	–1.090	0.101	–1.260	–0.930
β_3	1.0	1.156	0.083	1.027	1.296
δ_1	1.0	1.081	0.097	0.922	1.239
δ_2	0.5	0.516	0.028	0.469	0.563
δ_3	–0.5	–0.474	0.045	–0.548	–0.400
σ_{12}	0.2	0.347	0.144	0.111	0.586
σ_{22}	1.0	0.996	0.074	0.881	1.123
ρ_1	0.2	0.202	0.047	0.124	0.281
ρ_2	0.2	0.133	0.048	0.055	0.214

Table 10: Scenario H (n = 700, σ_{12} = 0.8, $\rho_1, \rho_2 = 0.2$)

	True value	Mean	Std.dev.	B.C.I. 5%	B.C.I. 95%
β_1	2.0	1.987	0.177	1.712	2.288
β_2	–1.0	–0.994	0.093	–1.156	–0.849
β_3	1.0	1.073	0.081	0.944	1.212
δ_1	1.0	1.069	0.085	0.929	1.209
δ_2	0.5	0.495	0.027	0.451	0.540
δ_3	–0.5	–0.482	0.038	–0.545	–0.420
σ_{12}	0.8	0.756	0.084	0.610	0.885
σ_{22}	1.0	1.015	0.078	0.892	1.149
ρ_1	0.2	0.193	0.046	0.116	0.267
ρ_2	0.2	0.148	0.050	0.066	0.231