CENTER FOR POLICY RESEARCH

THE MAXWELL SCHOOL

WORKING PAPER SERIES

Bayesian Spatial Bivariate Panel Probit Estimation

Badi H. Baltagi, Peter H. Egger, and Michaela Kesina

Paper No. 187
January 2016

ISSN: 1525-3066

426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
(315) 443-3114/email: ctrpol@syr.edu

http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

CENTER FOR POLICY RESEARCH –Spring 2016

Leonard M. Lopoo, Director
Associate Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director
Budget and Administration

John Yinger
Trustee Professor of Economics and PAIA
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi H. Baltagi.. Economics
Robert Bifulco..PAIA
Leonard Burman ....PAIA
Thomas Dennison ....PAIA
Alfonso Flores-Lagunes .... Economics
Sarah Hamersma ....PAIA
William C. Horrace .... Economics
Yilin Hou .... PAIA
Hugo Jales... Economics
Duke Kao.. Economics
Jeffrey Kubik.. Economics
Yoonseok Lee .... Economics
Amy Lutz.. Sociology
Yingyi Ma ....Sociology
Jerry Miner ....Economics
Cynthia Morrow .... PAIA
Jan Ondrich...Economics
John Palmer.. PAIA
David Popp .... PAIA
Stuart Rosenthal ....Economics
Rebecca Schewe ....Sociology
Amy Ellen Schwartz .... PAIA/Economics
Perry Singleton………………………….......Economics
Michael Wasylenko... ……………………….Economics
Peter Wilcoxen…...…PAIA

GRADUATE ASSOCIATES

Emily Cardon .... PAIA
Brianna Carrier .... PAIA
John T. Creedon... PAIA
Carlos Diaz ....Economics
Alex Falevich ....Economics
Wancong Fu ....Economics
Boqian Jiang ....Economics
Yusun Kim .... PAIA
Ling Li...Economics
Michelle Lofton .... PAIA
Judson Murchie .... PAIA
Brian Ohl... PAIA
Jindong Pang ....Economics
Malcolm Philogene .... PAIA
William Reed .... PAIA
Laura Rodriquez-Ortiz ....PAIA
Fabio Rueda De Vivero .... Economics
Max Ruppersburg ....PAIA
Iuliia Shybalkina..PAIA
Kelly Stevens ....PAIA
Mary Stottele...PAIA
Tian Tang..PAIA
Saied Toossi ....PAIA
Rebecca Wang .... Sociology
Nicole Watson...................................... Social Science
Katie Wood ....PAIA
Jinqi Ye .... Economics
Pengju Zhang ....PAIA
Xirui Zhang .... Economics

STAFF

Kelly Bogart.....…………...….Administrative Specialist
Kathleen Nasto.......................Administrative Assistant
Candi Patterson……………..……Computer Consultant
Mary Santy.…………………....Administrative Assistant
Katrina Wingle.....………..…….Administrative Assistant

Abstract

This paper formulates and analyzes Bayesian model variants for the analysis of systems of spatial panel data with binary dependent variables. The paper focuses on cases where latent variables of cross-sectional units in an equation of the system contemporaneously depend on the values of the same and, eventually, other latent variables of other cross-sectional units. Moreover, the paper discusses cases where time-invariant effects are exogenous versus endogenous. Such models may have numerous applications in industrial economics, public economics, or international economics. The paper illustrates that the performance of Bayesian estimation methods for such models is supportive of their use with even relatively small panel data sets.

JEL No. C11; C31; C35

Keywords: Spatial Econometrics; Panel Probit; Multivariate Probit

Acknowledgments: The authors Badi Baltagi (Economics Department and Center for Policy Research); Peter H. Egger (ETH Zurich, CEPR, CESifo, GEP) and Michaela Kesina (ETH Zurich) would like to thank an anonymous reviewer, Cheng Hsiao, and Jaya Krishnakumar for numerous helpful comments on earlier versions of the manuscript. Moreover we are grateful to the participants of the Advances in Econometrics Conference 2015 at Baton Rouge and the 26th (EC)2 Conference on *Theory and Practice of Spatial Econometrics*.

1 Introduction

Many fields in applied economics involve multinomial choice problems. Examples are the choice of sending children to public schools and to vote in favor of a school budget (see Greene, 1984) in public choice; The choice among various types of labor markets in labor economics (see Haque and Haque, 2009); The choice of different health care plans or treatments in health economics (see Jones, 2007); The choice of different types of preferential agreements (for trade and investment) in international economics (see Egger and Wamser, 2013). In treatment studies with a binary outcome such as graduating or not and a binary endogenous treatment such as private versus public schooling in education economics (see Altonji, Elder, and Taber, 2005). Historically, applications of such models use cross-section data, but recent applications include panel data, (see Johnson and Hensher, 1982; Börsch-Supan, Hajivassiliou, Kotlikoff, and Morris, 1992; Keane, 1997; Egger and Wamser, 2013; Mulkay, 2014, to mention a few). While a case for cross-sectional interdependence could be made – due to the presence of peer-group effects, social interaction, strategic interaction, spillovers, and general equilibrium effects – most applications ignore cross-sectional dependence. This paper proposes bivariate panel probit models which could be used in applied work in order to allow for equicorrelation due to the repeated observation of cross-sectional units over time as well as for cross-sectional dependence among the units within time.

The paper proposes a Bayesian bivariate probit model and analyzes its performance in small samples.[2] Monte Carlo simulation results are encouraging as parameter estimates can be obtained without much bias in small samples, and the root-mean-squared errors decline as the sample size increases, in particular, with the cross-sectional dimension. The paper illustrates how such models could readily be extended to the multivariate case with more than two equations. Also, the paper discusses the case where the explanatory

[2]In earlier research, – mostly cross-section – alternatives to Bayesian nonlinear probability model estimation had been proposed: see McMillen (1992) for expectation-maximization methods; see Beron and Vijverberg (2004) for simulated maximum likelihood methods; and Klier and McMillen (2008) for generalized methods of moments procedures.

variables are correlated with the time-invariant error components.

The remainder of the paper is organized as follows. The next section outlines a parsimonious model version. Section 3 describes the estimation of the model parameters of interest. Section 4 proposes extensions of the model allowing for a richer setting of cross-sectional dependence across equations. In particular, it outlines a multivariate model with more than two equations, and it discusses the case of a correlated random effects model. Section 5 summarizes the Monte Carlo simulation results for leading types of models addressed in the paper, and the last section concludes.

2 Econometric model

Let us denote the binary observable variables regarding the mth decision or equation for unit i at time t by $y_{m,it}$, where $m \in \{1, 2\}$ reflects the bivariate case. The total number of individual units and time periods be N and T, respectively. We observe this binary variable as

$$y_{m,it} = 1(y^*_{m,it} > 0), i = 1, 2, .., N \text{ and } t = 1, 2, .., T \tag{1}$$

$$y^*_{m,it} = \lambda_m \overline{y}^*_{m,it} + x_{m,it}\beta_m + \alpha_{m,i} + \nu_{m,it} \qquad \text{with} \qquad \overline{y}^*_{m,it} = \sum_{j=1}^{N} w_{ijt} y^*_{m,jt}, \tag{2}$$

where $y^*_{m,it}$ is a latent (i.e., unobserved) variable and denotes the net gains for i from choosing m at time t. $\lambda_m \overline{y}^*_{m,it}$ reflects a (global) *spillover effect* of other units on i.[3] w_{ijt} is a normalized weight describing the strength of the relationship between units i and j at time t. In the spatial panel econometrics literature, w_{ijt} is often assumed time-invariant. However, assuming that is not necessary. w_{ijt} is nonnegative if two distinct units i and j are neighbors and zero otherwise at time t; it is always zero for $i = j$.[4] Notice that the notion of neighborliness behind w_{ijt} is generic and can be related to space in a narrow sense or to other concepts (such as input-output relations, worker flows, information flows, etc.). λ_m denotes the spatial autocorrelation, contagion, interdependence, or

[3]The spillover effects are referred to as *global*, because the reduced form of the model involves an infinite number of cross-sectional effects and associated repercussions in the cross-sectional system.

[4]In principle, the weights w_{ijt} could be specific to equation m.

spillover parameter for latent outcome of type m, and it is important to gauge the relative magnitude of spillovers. The $1 \times K$ vector of covariates $x_{m,it} = (x_{k,m,it})$ is indexed by m for reasons of parameter identification in multivariate probit models (see Keane, 1992; Munkin and Trivedi, 2008).

The time-varying idiosyncratic error is denoted by $\nu_{m,it}$ and the time-invariant random effect is denoted by $\alpha_{m,i}$. For these error components, we adopt the conventional assumptions that $E(\nu_{m,it}\nu_{m,jt}) = 0$ for all $i \neq j$, $E(\alpha_{m,i}\nu_{m,it}) = 0$ for all m, t, $E(\nu_{m,it}\nu_{m,is}) = 0$ for all $t \neq s$. More specifically, regarding the bivariate distributions of $(\alpha_{l,i}, \alpha_{m,i})$ and $(\nu_{l,it}, \nu_{m,it})$, we assume bivariate normality

$$\begin{pmatrix}\alpha_{1,i}\\ \alpha_{2,i}\end{pmatrix} \sim N\left(\begin{pmatrix}\overline{\alpha}_1\\ \overline{\alpha}_2\end{pmatrix}, \begin{pmatrix}\sigma_{\alpha,11} & \sigma_{\alpha,12}\\ \sigma_{\alpha,12} & \sigma_{\alpha,22}\end{pmatrix}\right), \quad \begin{pmatrix}\nu_{1,it}\\ \nu_{2,it}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}1 & \tau\\ \tau & 1\end{pmatrix}\right).$$

The variances of $\nu_{1,it}$ and $\nu_{2,it}$ are normalized to unity (see for instance Greene, 2003, for a treatment of the bivariate probit model without accounting for any form of spatial correlation) and τ denotes the tetrachoric correlation.

We shall impose the assumption that all elements of $x_{m,it}$ are *doubly exogenous* in the sense that $E[x_{m,it}\alpha_{m,i}] = 0$ and $E[x_{m,it}\nu_{m,it}] = 0$. [5]

As is common in spatial panel econometrics (see, e.g., Kapoor, Keleijian, and Prucha, 2007), the observations are stacked such that i is the fast index and t the slow index, which yields the following stacked model for equation m of the framework given in (1)–(2)

$$y_m = 1(y_m^* > 0), \tag{3}$$

$$y_m^* = \lambda_m \overline{y}_m^* + x_m\beta_m + \alpha_m + \nu_m \qquad \text{with} \qquad \overline{y}_m^* = W_{TN} y_m^*, \tag{4}$$

for $m \in \{1, 2\}$ where y_m, y_m^*, and ν_m are of dimension $TN \times 1$. The matrix x_m is of dimension $TN \times k_m$ and its parameter vector β_m is $k_m \times 1$. The spatial weights matrix $W_{TN} = diag(W_{tN})$ is of dimension $TN \times TN$ and contains zero diagonal elements. Its off-diagonal elements of W_{tN} are nonzero, reflecting the neighborliness between two cross-sectional units. Moreover, we assume the elements of W_{TN} to be normalized so that the

[5] Note that $x_{m,it}$ may contain time averages of some or all of the time-variant covariates. In the latter case, it is sufficient for the time-variant variables in $x_{m,it}$ to be singly-exogenous with only $E[x_{m,it}\nu_{m,it}] = 0$.

admissible parameter space of $\{\lambda_1, \lambda_2\}$ is known and less than unity in absolute value. For instance, a convenient normalization is dividing each element by the corresponding sum across elements in a row (see Anselin, 1988; and Kelejian and Prucha, 2010, for alternative normalizations).[6] The vector α_m is of dimension $N \times 1$.

Stacking both equations for $m \in \{1, 2\}$ below one another yields the following model for the latent dependent variables:

$$y^* = (\Lambda \otimes J_n) \odot W y^* + X\beta + A\alpha + \nu,$$

where $\otimes$ denotes the Kronecker product and $\odot$ the Hadamard product, ι_T is a vector of ones of dimension T, I_N is an identity matrix of dimension N, and J_n is a matrix of ones of dimension N.

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, W = \begin{pmatrix} W_{TN} & 0 \\ 0 & W_{TN} \end{pmatrix}, X = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}, A = \begin{pmatrix} \iota_T \otimes I_N & 0 \\ 0 & \iota_T \otimes I_N \end{pmatrix},$$

$y^* = (y_1^{*\prime}, y_2^{*\prime})'$, $\beta = (\beta_1', \beta_2')'$, $\alpha = (\alpha_1', \alpha_2')'$, and $\nu = (\nu_1', \nu_2')'$.

The reduced form is given by

$$y^* = S^{-1}(X\beta + A\alpha + \nu),$$

with $S = (I_{2TN} - (\Lambda \otimes J_{TN}) \odot W) = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}$ where $S_m = I_{TN} - \lambda_m W_{TN}$. Together with the normalization of W_{TN}, the admissible parameter space of Λ ensures invertibility of S.

3 Model estimation

3.1 Bayesian estimation procedure

Non-spatial bivariate panel data models can be estimated by maximum likelihood. The presence of the spatial lag $\overline{y}^*_{m,it}$ as a determinant of the latent variable $y^*_{m,it}$ leads to

[6] With a time-invariant, normalized, $N \times N$ spatial weights matrix, $W_{TN} = I_T \otimes W_N$, where I_T is an identity matrix of dimension T and $W_N = (w_{ij})$.

a reduced form of the model which is nonlinear in variables and parameters and to an N-dimensional integral in the likelihood function. However, ignoring relevant spatial lags in the system leads to inconsistent estimates of the parameters.

For implementation and estimation, we follow the generic Markov chain Monte Carlo (MCMC) simulation approach suggested by LeSage (2000) and LeSage and Pace (2009) who focused mainly on single-equation and cross-sectional models.

Bayesian MCMC simulation entails estimating the posterior distribution of all parameters by combining prior information on them with the likelihood for the respective model, and sampling each parameter sequentially from its conditional distribution. This approach involves both Gibbs and Metropolis Hastings sampling. Details on those are provided in the next subsections.

Building on the idea of Albert and Chib (1993) for non-spatial, cross-sectional, univariate probit models and on LeSage (2000) and LeSage and Pace (2009) for spatial, cross-sectional, univariate and multivariate probits, we introduce the latent variables as additional parameters. This provides for a considerable facilitation of the estimation procedure, as conditioning on latent variables yields simpler distributions which we can sample from.

For modelling the time-invariant, unobserved heterogeneity across cross-sectional units through $\alpha = (\alpha_1', \alpha_2')'$, we assume a hierarchical structure, whereby all $\alpha_i = (\alpha_{1i}, \alpha_{2i})'$ are based on a distribution, which has some parameters in common, which we refer to as *hyperparameters*. These hyperparameters – namely mean μ_α and variance V_α – are drawn in a separate step and used when drawing α_i.

All parameters to be estimated we subsume in the parameter vector $\theta = \{\beta, \lambda_1, \lambda_2, \tau, y^*, \alpha, \mu_\alpha, V_\alpha\}$.

Using $y = (y_1', y_2')'$, the joint posterior distribution is given by

$$
\begin{aligned}
& p(\theta|y, X, W) \\
\propto\ & p(y|y^*, X, W) p(y^*|\beta, \lambda_1, \lambda_2, \tau, \alpha, \mu_\alpha, V_\alpha, X, W) \\
& p(\beta) p(\lambda_1) p(\lambda_2) p(\tau) p(\alpha|\mu_\alpha, V_\alpha) p(\mu_\alpha) p(V_\alpha).
\end{aligned}
$$

where the first term in the second line relates the observed dependent variables to their latent counterparts, the second term in the second line denotes the likelihood, and the third line contains the priors. Details on these components will be given in the following paragraphs. Since the expressions above turn out to be intractable, we calculate the conditional distributions for all model parameters given the data and the other parameters, $\theta_\ell|\theta_{-\theta_\ell}$, which are given in detail in Subsection 3.2.

Likelihood

The likelihood is stated in terms of the latent variables y_m^*. The joint distribution of (y_1^*, y_2^*) is given by

$$\begin{pmatrix} y_1^* \\ y_2^* \end{pmatrix} \sim N\left(S^{-1}(X\beta + A\alpha), \begin{pmatrix} (S_1'S_1)^{-1} & \tau(S_2'S_1)^{-1} \\ \tau(S_1'S_2)^{-1} & (S_2'S_2)^{-1} \end{pmatrix}\right).$$

This yields the likelihood

$$p(y_1^*, y_2^*|\theta, X) = \frac{|S_1||S_2|}{(2\pi)^{TN}|\Sigma|^{TN/2}} exp\left[\left(-\frac{1}{2} tr\left(R\Sigma^{-1}\right)\right)\right],$$

where $\Sigma = \begin{pmatrix} 1 & \tau \\ \tau & 1 \end{pmatrix}$ and $R = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$ is a 2×2 matrix containing the elements $r_{uv} = (S_u y_u^* - (x_u\beta_u + (\iota_T \otimes \alpha_u)))'(S_v y_v^* - (x_v\beta_v + (\iota_T \otimes \alpha_v)))$.

Priors

The prior distributions are assumed to be

$$\beta \sim N(\underline{\beta}, \underline{V}) \quad \text{where} \quad \underline{\beta} = 0_{2k\times 1} \quad \text{and} \quad \underline{V} = I_{2k} \cdot 1e^{12} \tag{5}$$

$$\lambda_m \sim U(-1,1) \quad \tau \sim U(-1,1). \tag{6}$$

Notice that λ_m and τ are parameters, which are bounded theoretically in absolute value. For instance, with a row-normalized matrix W_{TN} (and, hence W) and the model proposed in this section, both λ_m and τ need to be smaller than unity in absolute value.

Modelling the unobserved heterogeneity with a hierarchical prior, we draw hyperparameters from distributions using the following priors:

$$\mu_\alpha \sim N(\underline{\mu}_{\mu\alpha}, \underline{V}_{\mu\alpha}) \quad \text{where} \quad \underline{\mu}_{\mu\alpha} = 0_{2\times 1} \quad \text{and} \quad \underline{V}_{\mu\alpha} = I_2 \tag{7}$$

$$V_\alpha^{-1} \sim \mathcal{W}(\underline{V}_{V\alpha}^{-1}, \underline{v}_{V\alpha}) \quad \text{where} \quad \underline{V}_{V\alpha}^{-1} = I_2 \quad \text{and} \quad \underline{v}_{V\alpha} = 2 \tag{8}$$

where $\mathcal{W}$ denotes the Wishart distribution. The choice of the prior parameters leads to relatively uninformative priors reflecting a large degree of uncertainty about them. Intuitively, in calculating the posterior distribution less weight is placed on the priors and more on the data as a consequence.

3.2 Conditional distributions

We calculate the conditional distribution of each of the parameters given all the other parameters of the model.

Conditional distribution of y_1^* and y_2^*

The posterior distributions for the latent variables are calculated using the joint distribution of (y_1^*, y_2^*). The conditional distribution of y_1^* given the other parameters is given by

$$y_1^*|\theta_{-y_1^*} \sim N\left(S_1^{-1}\{x_1\beta_1 + \iota_T \otimes \alpha_1 + \tau(S_2 y_2^* - x_2\beta_2 - \iota_T \otimes \alpha_2)\}, (1-\tau^2)(S_1'S_1)^{-1} \right).$$

The conditional distribution of y_2^* given the other parameters is given by

$$y_2^*|\theta_{-y_2^*} \sim N\left(S_2^{-1}\{x_2\beta_2 + \iota_T \otimes \alpha_2 + \tau(S_1 y_1^* - x_1\beta_1 - \iota_T \otimes \alpha_1)\}, (1-\tau^2)(S_2'S_2)^{-1} \right).$$

y_m^* is truncated multivariate normal. Thus, we apply the method by Geweke (1991). When drawing y_m^*, we account for the observed binary y_m, taking draws from a right-truncated normal if y_m is 0 and from a left-truncated normal if y_m is 1.

Conditional distribution of $\boldsymbol{\beta}$

The conditional distribution of $\beta = (\beta_1', \beta_2')'$ given the other parameters is

$$\beta|\theta_{-\beta} \sim N(\overline{\beta}, \overline{V}_\beta), \tag{9}$$

where

$$\begin{aligned} \overline{\beta} &= \overline{V}_\beta \left(X' \left(\Sigma^{-1} \otimes I_{TN} \right) (Sy^* - A\alpha) + \underline{V}^{-1} \underline{\beta} \right) \\ \overline{V}_\beta &= \left(X' \left(\Sigma^{-1} \otimes I_{TN} \right) X + \underline{V}^{-1} \right)^{-1} \end{aligned}$$

We apply Gibbs sampling to draw values for β.

Conditional distribution of $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$

The conditional distribution of λ_m for $m \in \{1, 2\}$ is given by

$$\lambda_m|\theta_{-\lambda_m} \propto |S_m| exp\left[-\frac{1}{2} trace\left(R\Sigma^{-1}\right)\right] \tag{10}$$

This conditional distribution is of an unknown form. Thus we apply Metropolis-Hastings rather than Gibbs sampling for drawing it. We follow LeSage and Pace (2009) and draw a proposal candidate λ_m^c using $\lambda_m^c = \lambda_m + c_{\lambda_m} \cdot N(0, 1)$, where λ_m denotes the previous value and c_{λ_m} a tuning parameter. When taking draws we only use candidates lying in the admissible parameter space between -1 and 1. Using λ_m, λ_m^c, and (10), we calculate an acceptance probability to decide whether using the new candidate value or keeping the previous one. To ensure an acceptance probability between 40% and 60% we adapt the tuning parameter c_{λ_m}.[7]

Conditional distribution of τ

The conditional distribution of τ is given by

$$\tau|\theta_{-\tau} \propto \frac{1}{(1-\tau^2)^{NT/2}} exp\left[-\frac{1}{2} trace\left(R\Sigma^{-1}\right)\right] \tag{11}$$

Akin to λ_m, the conditional posterior distribution of τ takes an unknown form and we apply Metropolis-Hastings for drawing it. We apply the same approach as for drawing λ_m and draw new values using $\tau^c = \tau + c_\tau \cdot N(0, 1)$. Since τ lies in the interval between -1 and $+1$, we only accept those candidate values τ^c which lie in this interval. Both τ and τ^c are evaluated using (11) to calculate an acceptance probability and the tuning parameter c_τ is adapted to ensure an acceptance probability between 40% and 60%.

Conditional distribution of α

The conditional distribution of the $2N \times 1$ vector $\alpha = (\alpha_1', \alpha_2')'$ is

$$\alpha|\theta_{-\alpha} \sim N(\overline{\alpha}, \overline{V}_\alpha),$$

[7]For more details we refer the reader to LeSage and Pace (2009), p. 136/137.

where

$$\begin{aligned} \overline{\alpha} &= \overline{V}_{\alpha} \left(A' \left(\Sigma^{-1} \otimes I_{TN} \right) (Sy^* - X\beta) + (V_{\alpha}^{-1} \otimes I_N)(\mu_{\alpha} \otimes \iota_N) \right) \\ \overline{V}_{\alpha} &= \left((T\Sigma^{-1} + V_{\alpha}^{-1}) \otimes I_N \right)^{-1} \end{aligned}$$

which are based on the hyperparameters μ_{α} and V_{α}, which are drawn as

$$\mu_{\alpha} | \theta_{-\alpha} \sim N(\overline{\mu}_{\mu\alpha}, \overline{V}_{\mu\alpha}),$$

using

$$\begin{aligned} \overline{\mu}_{\mu\alpha} &= \overline{V}_{\mu\alpha} \left((V_{\alpha}^{-1} \otimes \iota_N')\alpha + \underline{V}_{\mu\alpha}^{-1} \underline{\mu}_{\mu\alpha} \right) \\ \overline{V}_{\mu\alpha} &= \left(N V_{\alpha}^{-1} + \underline{V}_{\mu\alpha}^{-1} \right)^{-1} \end{aligned}$$

and

$$V_{\alpha}^{-1} | \theta_{-\alpha} \sim \mathcal{W}(\overline{V}_{V\alpha}, \overline{v}_{V\alpha}),$$

with

$$\begin{aligned} \overline{v}_{V\alpha} &= \underline{v} + N \\ \overline{V}_{V\alpha} &= (H + \underline{V}_{V\alpha})^{-1} \end{aligned}$$

and the 2×2 matrix $H = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}$ containing the elements $h_{uv} = (\alpha_u - \mu_{\alpha u}' \iota_N)'(\alpha_v - \mu_{\alpha v}' \iota_N)$, where ι_N is an $N \times 1$ vector of ones. All of these parameters have known distributions. Specifically, we apply Gibbs sampling, drawing the hyperparameters μ_{α} and V_{α} first and then using those in drawing the elements of α.

Interpretation of results

Clearly, the point estimates of the parameters are a key ingredient for a quantitative assessment of the results. As with standard probit models, marginal effects of changes in explanatory variables cannot be read off the parameters but need to be evaluated at a certain point, typically the sample mean of the data. The computation of marginal effects in standard (nonspatial) probit models is outlined, e.g., in Greene (2003). With an index of the probit model whose reduced form is itself nonlinear in parameters, this issue is exacerbated. As with standard spatial models, direct, indirect, and total effects can be distinguished. What is of interest in probit models are the direct and total effects (with the indirect effects being defined as the difference between the latter and the former) on the probability that the outcome of interest is unity. This issue is exhaustively discussed for univariate probit models in LeSage, Pace, Lam, Campanella, and Liu (2011) and in Lacombe and LeSage (2015). The computation of marginal effects in bivariate probit models involves a straightforward combination of the approach outlined in Greene (1996) for non-spatial bivariate probits and in LeSage, Pace, Lam, Campanella, and Liu (2011) and in Lacombe and LeSage (2015) for spatial univariate probits.

4 Extensions

In this section, we consider three extensions. First, we introduce a richer framework of interdependence than the one introduced in Sections 2-3. This may be a useful extension, if the researcher believes that spillovers across individuals are not only related to a specific but to all latent outcomes. Second, we briefly discuss the case of more than two equations, which may be generally referred to as multinomial spatial probit estimation. Such a case may emerge, for instance, if researchers analyze problems with many discrete decisions (e.g., market entry with multi-product firms; market-entry with multi-national firms; etc.). Third, we discuss the case of estimation with correlated random effects, where some of the explanatory variables may be correlated with the unobserved individual-specific characteristics. The latter is often considered to be more plausible than the case of so-called double-exogeneity as assumed before, where the explanatory variables are uncorrelated with both the time-invariant and the time-variant characteristics of the cross-sectional units.

4.1 A richer structural latent-variable framework

Model

In what follows, we use the same notation as before as far as this is possible. In applications, the consideration of within-equation spatial dependence dominates. This might even be the case with systems of equations with binary dependent variables. However, in the latter case, economic theory or intuition of the researcher might support a more general set-up, where cross-sectional spillovers are associated not only with the latent variable pertaining to the same equation as the binary outcome but also ones pertaining to other binary outcomes in the system.

The model is given by

$$y_{m,it} = 1(y^*_{m,it} > 0), \tag{12}$$

$$y^*_{m,it} = \sum_{l=1}^{2} \lambda_{ml} \overline{y}^*_{ml,it} + x_{m,it}\beta_m + \alpha_{m,i} + \nu_{m,it} \text{ with } \overline{y}^*_{ml,it} = \sum_{j=1}^{N} w_{ml,ijt} y^*_{l,jt}, \tag{13}$$

The observations are stacked such that i is the fast index and t the slow index, which yields the following stacked model for equation m of the model given in (12) –(13)

$$y_m = 1(y^*_m > 0), \tag{14}$$

$$y^*_m = \sum_{l=1}^{2} \lambda_{ml} \overline{y}^*_{ml} + x_m\beta_m + \iota_T \otimes \alpha_m + \nu_m \text{ with } \overline{y}^*_{ml} = W_{ml,TN} y^*_l. \tag{15}$$

The only thing that is now needed is a redefinition of Λ, W, and S: $\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}$, $W = \begin{pmatrix} W_{11,TN} & W_{12,TN} \\ W_{21,TN} & W_{22,TN} \end{pmatrix}$, and $S = (I_{2TN} - (\Lambda \otimes J_{TN}) \odot W) = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$. The $2TN \times 2TN$ matrix W consists of 4 $TN \times TN$ spatial weights matrices $W_{ij,TN}$ for $i, j \in \{1, 2\}$. Of course, one can also assume the same $W_{ij,TN}$ for all i and j.

Define $\widetilde{S} = S^{-1} = \begin{pmatrix} \widetilde{S}_{11} & \widetilde{S}_{12} \\ \widetilde{S}_{21} & \widetilde{S}_{22} \end{pmatrix}$. Then stacking both equations for $m \in \{1, 2\}$ yields

$$y^* = (\Lambda \otimes J_n) \odot W y^* + X\beta + A\alpha + \nu \tag{16}$$

and its reduced form is given by

$$y^* = \widetilde{S}(X\beta + A\alpha + \nu). \tag{17}$$

Joint distribution of (y_1^*, y_2^*) and the likelihood

Based on (17), the joint distribution of (y_1^*, y_2^*) is given by

$$\begin{pmatrix} y_1^* \\ y_2^* \end{pmatrix} \sim N\left(\begin{pmatrix} \widetilde{S}_{11}(x_1\beta_1 + \iota_T \otimes \alpha_1) + \widetilde{S}_{12}(x_2\beta_2 + \iota_T \otimes \alpha_2) \\ \widetilde{S}_{21}(x_1\beta_1 + \iota_T \otimes \alpha_1) + \widetilde{S}_{22}(x_2\beta_2 + \iota_T \otimes \alpha_2) \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\right) \tag{18}$$

with $\begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} \widetilde{S}_{11} & \widetilde{S}_{12} \\ \widetilde{S}_{21} & \widetilde{S}_{22} \end{pmatrix} (\Sigma \otimes I_N) \begin{pmatrix} \widetilde{S}'_{11} & \widetilde{S}'_{21} \\ \widetilde{S}'_{12} & \widetilde{S}'_{22} \end{pmatrix}$.

This yields the likelihood

$$p(y^*|\theta, X) = \frac{1}{2\pi^{TN}|\Sigma|^{TN/2}} |S_{22}||S_{11} - S_{12}S_{22}^{-1}S_{21}| exp\left[\left(-\frac{1}{2} trace\left(R\Sigma^{-1}\right)\right)\right],$$

where $\Sigma = \begin{pmatrix} 1 & \tau \\ \tau & 1 \end{pmatrix}$ and $R = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$ is a 2×2 matrix containing the elements $r_{uv} = r'_u \cdot r_v$ for $u, v \in \{1, 2\}$ with

$$\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} S_{11}y_1^* + S_{12}y_2^* - (x_1\beta_1 + \iota_T \otimes \alpha_1) \\ S_{21}y_1^* + S_{22}y_2^* - (x_2\beta_2 + \iota_T \otimes \alpha_2) \end{pmatrix}.$$

Priors

We use the same uninformative priors given by (5), (6), (7) and (8). In line with the previous subsection we assume a uniform uninformative prior for all λ_{ij} for $i, j \in \{1, 2\}$.

Conditional distribution of y_1^* and y_2^*

Using the joint distribution of (y_1^*, y_2^*) in (18) the conditional distribution of y_1^* given the other parameters is given by

$$\begin{aligned} y_1^*|\theta_{-y_1^*} \sim\; & N\Big(\widetilde{S}_{11}(x_1\beta_1 + \iota_T \otimes \alpha_1) + \widetilde{S}_{12}(x_2\beta_2 + \iota_T \otimes \alpha_2) + \Omega_{12}\Omega_{22}^{-1}[y_2^* - \widetilde{S}_{21}(x_1\beta_1 + \iota_T \otimes \alpha_1) \\ & -\widetilde{S}_{22}(x_2\beta_2 + \iota_T \otimes \alpha_2)], \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\Big). \end{aligned}$$

The conditional distribution of y_2^* given the other parameters is given by

$$\begin{aligned} y_2^*|\theta_{-y_2^*} \sim\; & N\Big(\widetilde{S}_{21}(x_1\beta_1 + \iota_T \otimes \alpha_1) + \widetilde{S}_{22}(x_2\beta_2 + \iota_T \otimes \alpha_2) + \Omega_{21}\Omega_{11}^{-1}[y_1^* - \widetilde{S}_{11}(x_1\beta_1 + \iota_T \otimes \alpha_1) - \\ & \widetilde{S}_{12}(x_2\beta_2 + \iota_T \otimes \alpha_2)], \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}\Big). \end{aligned}$$

Conditional distribution of $\boldsymbol{\lambda}_{gh}$

The conditional distribution of λ_{11}, λ_{12}, and λ_{21} is given by

$$\lambda_{uv}|\theta_{-\lambda_{uv}} \propto |S_{uu} - S_{uv}S_{vv}^{-1}S_{vu}|exp\left[-\frac{1}{2}trace\left(R\Sigma^{-1}\right)\right] \tag{19}$$

for $uv \in \{11, 12, 21\}$. Since this distribution takes an unknown form, we apply a Metropolis-Hastings procedure where we draw the new candidate values λ'_{11}, λ'_{12}, and λ'_{21}, and for λ_{22} we use

$$\lambda_{22}|\theta_{-\lambda_{22}} \propto |S_{22}||S_{11} - S_{12}S_{22}^{-1}S_{21}|exp\left[-\frac{1}{2}trace\left(R\Sigma^{-1}\right)\right] \tag{20}$$

We apply the Metropolis-Hastings procedure as in Subsection 3.2. However due to the more complex equation system we need to take stability conditions into account. E.g., when drawing the new candidate values, we only accept those where $|\lambda_{h1}| \leq 1 - |\lambda_{h2}|$ and $|\lambda_{h2}| \leq 1 - |\lambda_{h1}|$for $h \in \{1, 2\}$.

Conditional distribution of β, α, and τ

For β, α, and τ, the conditional distributions are the same as in section 3.2. The only difference is that we now use the S defined in section 4.1.

4.2 Multinomial spatial probit estimation with more than two equations

The proposed procedure can be extended to more than two decisions. Suppose one has M decisions, which corresponds to M equations. The dimensionality of the observed variable y and its latent counterpart y^* are $MTN \times 1$. The matrix of covariates is then of dimension $MTN \times \sum_{m=1}^{M} k_m$ where k_m denotes the dimensionality of the covariate matrix in the respective equation. The spatial weights matrix W and the matrices S and A are of dimension $MTN \times MTN$. The unobserved heterogeneity α is $MN \times 1$. The matrix of the spatial autocorrelation parameters Λ and the matrices R and Σ are then of dimension $M \times M$.[8] The likelihood and the conditional distributions of β, α, and λ are mathematically equivalent to and only of a different dimensionality than in the bivariate case. Each latent variable, y_m^* for the m-th equation, is to be drawn conditional on the ones in all the other equations. The relatively biggest difference is with respect to the estimation of the off-diagonal elements of Σ which have to be drawn based on a Wishart distribution subject to normalization constraints (see the discussion in Koop, 2003).

4.3 Endogenous time-invariant effects

In Bayesian econometrics it is common to assume that all covariates and also the unobserved individual-specific effect are purely random variables.
However, this is not the case in many empirical applications, where it is likely that some of the covariates of an individual are correlated with its unobserved individual-specific characteristics: in wage equations, education is correlated with individuals' unmeasurable ability, and affects discrete labor-market choices of individuals; total factor productivity is correlated with unobservable managerial or entrepreneurial talent and organization and affects discrete (market-entry or scope) decisions by firms; regional observable attributes are correlated with unobservable amenities and their hedonic valuation by mobile residents; etc. Ignoring potential correlation between unobserved heterogeneity and covariates might lead to an omitted variables bias in coefficients of interest on observable variables.

One prominent way to account for a potential correlation between unobserved heterogeneity and the covariates is proposed by Mundlak (1978). He proposed to include the averages of time-varying covariates as additional regressors into the regression equation to approximate the unobserved heterogeneity. In a second extension we follow his suggestion. By and large, this leaves our approach described in section 3.2 unchanged. The only difference is that the matrix of covariates now consists of $[X, \overline{X}]$ where $\overline{X}$ contains the time averages of the columns of X that pertain to time-variant explanatory

[8] In this case, Σ is symmetric with unitary diagonal and contains $M(M-1)/2$ unknown off-diagonal elements.

variables.

5 Monte Carlo simulation study

To illustrate the performance of the bivariate panel probit model with triangular data, we perform Monte Carlo experiments for a spatial bivariate model structure along the aforementioned lines.

Design for the basic model:
In particular, we assume two variables $x_{1,it} \sim N(0,1)$ and $x_{2,it} \sim N(0,1)$ where both enter equations 1 and 2. Their true parameter values are for the first equation $\beta_1 = (\beta_{11}, \beta_{12}) = (-2, 1.25)$ and for the second equation $\beta_2 = (\beta_{21}, \beta_{22}) = (-1, 0.5)$.
We assume a time-invariant, row-normalized, 5-before-5-behind neighborhood structure regarding W_N so that $W = diag_T(W_N)$, and $w_{ii} = 0$ and all non-diagonal elements w_{ij} are either zero (non-neighbors) or 0.1 (neighbors).

Moreover, we specify the bivariate normality about the 2×1 vector α_i as $\alpha_i \sim N\left(\begin{pmatrix} 0.5 \\ 0.25 \end{pmatrix}\begin{pmatrix} 1 & 0.4 \\ 0.4 & 1.25 \end{pmatrix}\right)$.

We consider four alternative sets of parameters $(\lambda_1, \lambda_2, \tau)$ with

$$(\lambda_1, \lambda_2, \tau) = \begin{cases} (0.4, 0.6, 0.5) \\ (0.2, 0.3, 0.5) \\ (0.4, 0.6, 0.8) \\ (0.4, 0.6, 0) \end{cases}$$

In general we consider two configurations each for N and T with $N \in \{100; 500\}$ and $T \in \{5; 7\}$.

For each of the four parameter and four panel configurations, we draw 500 $2NT \times 1$ vectors of residuals ν. For each one of these 8,000 experiments we do an MCMC simulation with a chain of 30,000 elements of which 4,000 are burn-ins and only every 10th of the remaining elements is used (i.e., a thinning ratio of one-tenth is applied).

Simulation results for the basic model:
We summarize the corresponding simulation results in Tables 1 and 2. In Table 1, we report on $\hat{\theta} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\beta}_{22}, \hat{\tau})$ and moments of the elements of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ for the first parameter configuration and alternative sample-size configurations. In Table 2, we focus on the configuration of $N = 100$ and $T = 5$ for the remaining three considered true parameter configurations.

−− Tables 1 and 2 about here −−

The results in Table 1 suggest that the parameter biases are relatively small, even in the case of $\{N = 100; T = 5\}$. Obviously, the biases decline as the sample size grows in the T- and, in particular, the N-dimensions. With $\{N = 500; T = 5\}$ the biases of most parameters are down to a range of about five to ten percent of the true values only, across the board. These biases are about twice as high with $\{N = 100; T = 5\}$. However, when comparing these results with non-spatial models, we would support the use of spatial panel data probits even with small to moderate data-sets at hand.

The underlying correlations between the true and the predicted latent variables are: 0.8699 for $(y_1^*, \hat{y}_1^*)$ and 0.7760 $(y_2^*, \hat{y}_2^*)$ for $\{N, T\} = \{100, 5\}$; 0.8812 for $(y_1^*, \hat{y}_1^*)$ and 0.8710 $(y_2^*, \hat{y}_2^*)$ for $\{N, T\} = \{100, 7\}$; and 0.7875 for $(y_1^*, \hat{y}_1^*)$ and 0.7760 $(y_2^*, \hat{y}_2^*)$ for $\{N, T\} = \{500, 5\}$. These numbers indicate that there is enough noise in the data-generating process so that the small bias figures point to a relatively good performance of the proposed estimation routines.

Design for a framework for within- and across-equation spatial correlation:
For an analysis of the richer model, we assume a framework as outlined in Section 4.1, where y_1^* as well as y_2^* affect both latent outcomes y_1^* and y_2^*. For this, we assume the same spatial weights matrix W_N for all terms. The corresponding spatial autocorrelation parameters are: $\{\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}\} = \{0.4, 0.2, 0.1, 0.3\}$.

In this design, we assume two variables $x_{11,it} \sim N(0, 1)$ and $x_{12,it} \sim N(0, 1)$, which enter the first equation, $x_{21,it} \sim N(0, 1)$ and $x_{22,it} \sim N(0, 1)$, which enter the second equation. Their true parameter values are, as above, $\beta_1 = (\beta_{11}, \beta_{12}) = (-2, 1.25)$ and $\beta_2 = (\beta_{21}, \beta_{22}) = (-1, 0.5)$ for the first and the second equation, respectively. α, and ν are drawn in the same way as in the benchmark design.

Simulation results for the within- and across-equation spatial-correlation model:
Table 3 summarizes the Monte Carlo simulation results for this richer design.

$--$ Table 3 about here $--$

The findings in the table suggest that the richer design does not involve systematically larger biases or root mean squared errors across the parameters of interest. Hence, the findings are assuring that even more complex designs with spillovers across different latent variables in the system between cross-sectional units can be analyzed even with relatively small samples at hand.

Design for the correlated random effects model:
In an extension, we let α be correlated with X at different intensities. In this set-up, we consider the vectors α_1 and α_2 to be correlated with x_2, maintaining the assumption that x_1 is exogenous. Specifically, we decompose x_2 into its between (bar) and within (tilde) parts (where between and within refer to cross-sectional units i), $x_2 = \overline{x}_2 + \widetilde{x}_2$ and assume that $\alpha_m = \alpha_m^* + c_\alpha \cdot \overline{x}_2$ for $m = \{1, 2\}$, considering several alternative degrees of endogeneity with $c_\alpha = \{1; 2; 4\}$. α_m^* is drawn in the same way as α_m in the basic design. In the generated data sets for $N = 100$, this yields an average correlation between x_2 and $\{\alpha_1, \alpha_2\}$ of about $\{0.438; 0.387\}$ with c_α=1, of about $\{0.701; 0.654\}$ with c_α=2, and of about $\{0.892; 0.869\}$ with c_α=4, across all Monte Carlo runs, respectively.

Clearly, with this setting of correlated random effects, the parameters on x_2, $\{\beta_{12}; \beta_{22}\}$, will be biased unless $\overline{x}_2$ is included as a control function as suggested by Mundlak (1978), Chamberlain (1982), and Wooldridge (1995). The corresponding results for specifications where the control function (whose parameters we suppress) is included in X are summarized in Table 4 for the sample-size configuration $\{N = 100; T = 5\}$. The true parameter values are $\beta_1 = (\beta_{11}, \beta_{12}) = (-2, 1.25)$, $\beta_2 = (\beta_{21}, \beta_{22}) = (-1, 0.5)$, and $(\lambda_1, \lambda_2, \tau) = (0.4, 0.6, 0.5)$ in this case.

Simulation results for the correlated random effects model:

We summarize the Monte Carlo simulation results for the correlated random effects model and the three configurations $c_\alpha = \{1; 2; 4\}$ in Table 4.

$--$ Table 4 about here $--$

The results in Table 4 suggest that the proposed approach works well in small samples even with endogenous cross-sectional effects when conditioning on individual-specific variable means. We have seen that the correlations between x_2 and α are relatively strong even in the case of $c_\alpha = 1$. In that case, the biases of the parameters amount to less than ten percent on average. The root mean-squared error (RMSE) is relatively highest for the coefficients on the endogenous variable, $\{\beta_{12}; \beta_{22}\}$, and it amounts to less than one-fifth for each of those. Clearly, both the biases and the RMSEs tend to be somewhat larger with a higher degree of endogeneity (a larger value of c_α). However, as said before, the degree of correlation studied here is rather strong, which is supportive of the proposed procedure.

While we illustrated that a consideration of correlated random effects is possible in Table 4, it is the purpose of Table 5 to document the consequences of disregarding correlated random effects when they are present.

$--$ Table 5 about here $--$

As in Table 4, we focus on the case of $\{N, T\} = \{100, 5\}$, and, for the sake of brevity, we summarize the results for the case of $c_\alpha = 4$, where the correlated-random-effects assumption is relatively important. A comparison of the respective rows in the table indicates that both the bias and the RMSE on $\{\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}\}$ are much higher in Table 5 than in Table 4. Hence, the merits of considering a correlated-random-effects version of the model in practice are obvious.

6 Conclusion

This paper analyzes a Bayesian estimation procedure for bivariate and, eventually, higher-variate panel probit models with spatial interdependence in the dependent variable. Such models could be interesting to use for an array of empirical problems where contagion or spillovers in a broad sense are important, the choices are not mutually exclusive, and there is time variation in those choices. Examples are discrete preferential policy choices of countries (e.g., with respect to trade agreement and/or investment agreement membership), discrete global-market-participation decisions of firms as exporters and/or multinational firms, discrete market-entry decisions of firms in a set of markets (such as countries and/or products), discrete consumption decisions of households with regard to certain products, discrete portfolio-acquisition decisions of investors, etc. All of these choices are ones where earlier empirical work had identified independently the existence of contagious effects and the interdependence between those choices. The approach presented in this paper is capable of treating the features of contagion or spillovers and cross-issue correlation simultaneously.

For estimation, the paper proposes a Bayesian spatial bivariate panel probit model. An advantage of this estimation procedure relative to standard maximum-likelihood estimation is that it can be used with large, interdependent cross-sections of data that are repeatedly observed over relatively short time periods. Our Monte Carlo simulation study suggests that the procedure works well even in small to moderately large samples.

7 References

Albert, J.H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88(422), 669–679.

Altonji, J., T. Elder, and C. Taber (2005). Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. *Journal of Political Economy* **113**(1), 151–184.

Anselin, L. (1988). *Spatial econometrics methods and models.* Dordrecht, The Netherlands: Kluwer.

Beron, K.J. and Vijverberg, W.P.M. (2004). Probit in a spatial context: A Monte Carlo analysis. In: Anselin, L., Florax, R.J.G.M., and Rey, S.J. (Eds). Advances in spatial econometrics. Methodology, tools and applications, Berlin, Germany: Springer.

Börsch-Supan, A., Hajivassiliou, V., Kotlikoff, L.J., and Morris, J.N. (1992). Health, children, and elderly living arrangements: a multi-period-multinomial probit model with unobserved heterogeneity and autocorrelated errors. In: Wise, D. (ed.), *Topics in the Economics of Aging.* Chicago: Chicago University Press, 79–108.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics* **18**(1), 5–46.

Egger, P.H. and Wamser, G. (2013). Multiple Faces of Preferential Market Access: Their Causes and Consequences, *Economic Policy* **28**(73), 145–187.

Geweke, J. (1991). Efficient simulation from the multivariate normal and Student t-distributions subject to linear constraints. In: *Computer Sciences and Statistics Proceedings of the 23d Symposium on the Interface*, 571–578.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., and Smith, A.F.M. (eds.), *Bayesian Statistics.* Oxford, UK: Clarendon Press.

Greene, W.H. (1984). Estimation of the correlation coefficient in a bivariate probit model using the method of moments. *Economics Letters* **16**(3-4), 285-291.

Greene, W.H. (1996). Marginal Effects in the Bivariate Probit Model. *Working Paper.*

Greene, W.H. (2003). *Econometric Analysis.* Upper Saddle River, NJ, USA: Prentice-Hall.

Haque, R. and Haque, M. (2009). *Gender, Ethnicity and Employment.* Heidelberg: Physica Verlag.

Jones, A.M. (2007). *Applied Econometrics for Health Economists*, 2nd ed. Radcliffe Publishing.

Johnson, L. and Hensher, D. (1982). Application of multinomial probit to a two-period panel data set. *Transportation Research A* **16A**(5-6), 457-464.

Kapoor, M., Kelejian, H.H., and Prucha, I.R. (2007). Panel Data Models with Spatially Correlated Error Components. *Journal of Econometrics* **140**(1), 97–130.

Keane, M.P. (1992). A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics* **10**(2), 193–200.

Keane, M.P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business and Economic Statistics* **15**(3), 310–327.

Kelejian, H.H. and Prucha, I.R. (2010). Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances. *Journal of Econometrics* **157**(1), 53–67.

Klier, T. and McMillen, D.P. (2008). Clustering of auto supplier plants in the United States: Generalized method of moments spatial logit for large samples. *Journal of Business and Economic Statistics* **26**(4), 460–471.

Koop, G. (2003). *Bayesian Econometrics.* Chichester, England: John Wiley & Sons Ltd.

Lacombe, D.J. and LeSage, J.P. (2015). Use and interpretation of spatial autoregressive probit models. *Annals of Regional Science*, forthcoming.

LeSage, J.P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis* **32**(1), 19–35.

LeSage, J.P. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. Boca Raton, FL, USA: CRC Press/Taylor & Francis.

LeSage, J.P., Pace, R.K., Lam, N., Campanella, R. and Liu, X. (2011). New Orleans business recovery in the aftermath of Hurricane Katrina. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **174**, 1007–1027.

McMillen, D.P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science* **32**(3), 335–348.

Mulkay, B. (2014). Bivariate probit estimation for panel data: a two-step Gauss-Hermite Quadrature approach with an application to product and process innovations for France. Unpublished manuscript, Université Montpellier.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica* **46**(1), 69–85.

Munkin, M.K. and Trivedi, P.K. (2008). Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics* **143**(2), 334–348.

Raftery, A.E. and Lewis, S.M.(1992). How many iterations in the Gibbs sampler? In: Bernardo, J.M., Berger, J.O., Dawid, A.P., and Smith, A.F.M. (eds.), *Bayesian Statistics*. Oxford, UK: Clarendon Press.

Wooldridge, J.A. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics* **68**(1), 115–132.

Tables

Table 1: Results for various cases of $\{N, T\}$

	β_{11}	β_{12}	β_{21}	β_{22}	λ_1	λ_2	τ	α_1	α_2
True	−2	1.25	−1	0.5	0.4	0.6	0.5	0.5	0.25
					$N = 100$, $T = 5$				
Mean	−2.168	1.329	−1.082	0.517	0.422	0.559	0.440	0.391	0.262
Bias	−0.168	0.079	−0.082	0.017	0.022	−0.041	−0.060	−0.109	0.012
RMSE	0.342	0.219	0.162	0.111	0.068	0.094	0.140	0.145	0.079
I-statistic	5.096	1.963	1.954	1.095	2.393	1.993	3.143	1.176	1.086
GT p-value	0.462	0.467	0.530	0.485	0.502	0.511	0.451	0.479	0.504
					$N = 100$, $T = 7$				
Mean	−2.084	1.305	−1.107	0.557	0.409	0.617	0.479	0.380	0.180
Bias	−0.084	0.055	−0.107	0.057	0.009	0.017	−0.021	−0.120	−0.070
RMSE	0.205	0.147	0.161	0.111	0.057	0.054	0.106	0.141	0.087
I-statistic	4.014	1.989	1.796	1.114	2.073	1.845	3.261	1.129	1.046
GT p-value	0.427	0.440	0.502	0.519	0.478	0.506	0.480	0.441	0.497
					$N = 500$, $T = 5$				
Mean	−2.096	1.292	−1.083	0.527	0.407	0.639	0.496	0.531	0.255
Bias	−0.096	0.042	−0.083	0.027	0.007	0.039	−0.004	0.031	0.005
RMSE	0.158	0.092	0.105	0.057	0.031	0.048	0.057	0.057	0.028
I-statistic	4.140	2.430	1.731	1.139	2.285	1.758	3.376	1.416	1.068
GT p-value	0.456	0.460	0.494	0.501	0.484	0.469	0.465	0.474	0.511

The I-statistic is calculated following Raftery and Lewis (1992). GT p-value denotes the p-value of the Geweke (1992) test.

Table 2: Results for alternative values of $\{\lambda_1, \lambda_2, \tau\}$ and $N = 100$ and $T = 5$

	β_{11}	β_{12}	β_{21}	β_{22}	λ_1	λ_2	τ	α_1	α_2
True	−2	1.25	−1	0.5	0.2	0.3	0.5	0.5	0.25
Mean	−2.089	1.296	−1.025	0.497	0.215	0.189	0.439	0.400	0.265
Bias	−0.089	0.046	−0.025	−0.003	−0.085	−0.111	−0.061	−0.100	0.015
RMSE	0.264	0.179	0.117	0.093	0.127	0.172	0.144	0.148	0.097
I-statistic	4.759	1.879	1.605	1.075	2.734	2.292	3.187	1.185	1.120
GT p-value	0.439	0.449	0.505	0.519	0.465	0.492	0.466	0.477	0.479
True	−2	1.25	−1	0.5	0.4	0.6	0.8	0.5	0.25
Mean	−2.258	1.398	−1.142	0.558	0.443	0.591	0.671	0.394	0.253
Bias	−0.258	0.148	−0.142	0.058	0.043	−0.009	−0.129	−0.106	0.003
RMSE	0.424	0.275	0.223	0.141	0.078	0.079	0.151	0.150	0.076
I-statistic	7.471	2.446	2.658	1.189	3.189	2.669	4.285	1.232	1.104
GT p-value	0.461	0.465	0.500	0.515	0.475	0.499	0.472	0.490	0.478
True	−2	1.25	−1	0.5	0.4	0.6	0	0.5	0.25
Mean	−2.150	1.336	−1.059	0.510	0.423	0.549	0.008	0.389	0.261
Bias	−0.150	0.086	−0.059	0.010	0.023	−0.051	0.008	−0.111	0.011
RMSE	0.324	0.219	0.142	0.106	0.069	0.092	0.159	0.150	0.080
I-statistic	4.752	1.799	1.645	1.075	2.036	1.732	3.149	1.137	1.083
GT p-value	0.457	0.464	0.486	0.482	0.479	0.514	0.484	0.463	0.522

The I-statistic is calculated following Raftery and Lewis (1992). GT p-value denotes the p-value of the Geweke (1992) test.

Table 3: Results for latent variable extension, $N = 100$ and $T = 5$

	β_{11}	β_{12}	β_{21}	β_{22}	λ_{11}	λ_{12}	λ_{21}	λ_{22}	τ	α_1	α_2
True	-2	1.25	-1	0.5	0.4	0.2	0.1	0.3	0.5	0.5	0.25
Mean	-2.095	1.364	-1.050	0.538	0.388	0.270	0.114	0.173	0.408	0.561	0.290
Bias	-0.095	0.114	-0.050	0.038	-0.012	0.070	0.014	-0.127	-0.092	0.061	0.040
RMSE	0.336	0.260	0.137	0.111	0.082	0.170	0.082	0.191	0.174	0.158	0.133
I-statistic	6.369	2.117	1.893	1.104	4.173	3.093	2.745	4.437	3.850	1.910	1.689
GT p-value	0.526	0.539	0.535	0.509	0.509	0.507	0.525	0.493	0.457	0.492	0.501

The I-statistic is calculated following Raftery and Lewis (1992). GT p-value denotes the p-value of the Geweke (1992) test.

Table 4: Results for CRE-variant in CRE world; different c_α; $N = 100$ and $T = 5$

	β_{11}	β_{12}	β_{21}	β_{22}	λ_1	λ_2	τ	α_1	α_2
True	-2	1.25	-1	0.5	0.4	0.6	0.5	0.5	0.25
					$c_\alpha = 1$				
Mean	-1.901	1.167	-1.022	0.517	0.404	0.571	0.429	0.369	0.264
Bias	0.099	-0.083	-0.022	0.017	0.004	-0.029	-0.071	-0.131	0.014
RMSE	0.277	0.204	0.141	0.120	0.068	0.081	0.143	0.158	0.077
I-statistic	4.493	1.598	1.892	1.086	2.366	2.084	3.219	1.146	1.090
GT p-value	0.459	0.458	0.491	0.508	0.501	0.490	0.470	0.469	0.502
					$c_\alpha = 2$				
Mean	-1.814	1.107	-1.061	0.543	0.384	0.575	0.429	0.367	0.275
Bias	0.186	-0.143	-0.061	0.043	-0.016	-0.025	-0.071	-0.133	0.02
RMSE	0.320	0.221	0.172	0.135	0.078	0.078	0.153	0.161	0.084
I-statistic	4.743	1.612	2.262	1.126	2.638	2.456	3.470	1.164	1.108
GT p-value	0.463	0.465	0.505	0.495	0.488	0.495	0.464	0.479	0.489
					$c_\alpha = 4$				
Mean	-1.885	1.166	-1.160	0.594	0.373	0.616	0.440	0.399	0.302
Bias	0.115	-0.084	-0.160	0.094	-0.027	0.016	-0.060	-0.101	0.052
RMSE	0.359	0.220	0.266	0.188	0.090	0.077	0.160	0.148	0.104
I-statistic	6.385	1.974	2.915	1.229	3.964	3.469	4.960	1.280	1.183
GT p-value	0.454	0.464	0.478	0.469	0.514	0.485	0.460	0.473	0.461

The I-statistic is calculated following Raftery and Lewis (1992). GT p-value denotes the p-value of the Geweke (1992) test.

Table 5: Results for non-CRE-variant in CRE world for $c_\alpha = 4$; $N = 100$ and $T = 5$

	β_{11}	β_{12}	β_{21}	β_{22}	λ_1	λ_2	τ	α_1	α_2
True	-2	1.25	-1	0.5	0.4	0.6	0.5	0.5	0.25
					$c_\alpha = 4$				
Mean	-2.234	1.616	-1.238	0.768	0.383	0.620	0.474	0.451	0.308
Bias	-0.234	0.366	-0.238	0.268	-0.017	0.020	-0.026	-0.049	0.058
RMSE	0.459	0.457	0.347	0.324	0.088	0.080	0.180	0.135	0.106
I-statistic	8.865	2.697	3.459	1.371	4.469	3.971	5.673	1.416	1.212
GT p-value	0.482	0.493	0.490	0.523	0.518	0.515	0.504	0.476	0.479

The I-statistic is calculated following Raftery and Lewis (1992). GT p-value denotes the p-value of the Geweke (1992) test.