JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **32**: 422–439 (2017)
Published online 16 February 2016 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2505

TRANSITIONS AT DIFFERENT MOMENTS IN TIME: A SPATIAL PROBIT APPROACH

J.PAUL ELHORST,[a*] PIM HEIJNEN,[a] ANNA SAMARINA[a] AND JAN P. A. M. JACOBS[a,b,c,d]

[a] *University of Groningen, Netherlands*
[b] *University of Tasmania, Hobart, Tasmania*
[c] *CAMA, Australian National University, Canberra, Australia*
[d] *CIRANO, Montreal, Quebec, Canada*

SUMMARY

This paper adopts a spatial probit approach to explain interaction effects among cross-sectional units when the dependent variable takes the form of a binary response variable and transitions from state 0 to 1 occur at different moments in time. The model has two spatially lagged variables: one for units that are still in state 0 and one for units that had already transferred to state 1. The parameters are estimated on observations for those units that are still in state 0 at the start of the different time periods, whereas observations on units after they transferred to state 1 are discarded, just as in the literature on duration modeling. Furthermore, neighboring units that had not yet transferred may have a different impact from units that had already transferred. We illustrate our approach with an empirical study of the adoption of inflation targeting for a sample of 58 countries over the period 1985–2008.

Received 25 November 2013; Revised 15 August 2015

Supporting information may be found in the online version of this article.

1. INTRODUCTION

Spatial binary response models have seen increasing use in the spatial econometrics literature. This holds especially for the spatial probit model based on the normal distribution—the primary focus of this paper. Although this model may be used to explain interaction effects among cross-sectional units when the dependent variable takes the form of a binary response variable, one shortcoming is that it cannot be fruitfully used to explain the transition from one state to another when this transition for one cross-sectional unit takes place at a different moment in time than for another unit.

This paper proposes a spatial probit model with two spatially lagged variables: one for units that are still in state 0 and one for units that have already transferred to state 1. The parameters of this model will be estimated based on observations of those units that are still in state 0 at the start of the different time periods being considered; observations on units after they transferred to state 1 are removed. The dependent variable and the first spatial term are both specified in terms of unobserved choices, i.e. the propensity towards state 1, while the second spatial term is specified in terms of observed choices, i.e. the actual outcomes.

Our setting differs from LeSage *et al.* (2011), who investigate the decision of firms in New Orleans to reopen their stores dependent on the decision made by other firms 0–3 months, 0–6 months and 0–12 months in the aftermath of Hurricane Katrina. However, since they use their data in cross-section rather than splitting up the sample into different time periods, they implicitly assume that the transition to state 1 of all firms that reopened their stores took place at the same point in time. As a result, they

* Correspondence to: J.Paul Elhorst, Faculty of Economics and Business, University of Groningen, PO Box 800, Groningen 9700AV, The Netherlands. E-mail: j.p.elhorst@rug.nl

cannot answer the question why some firms reopened their stores earlier than others and which role the interaction among firms at different points in time played in this transition process.

Mukherjee and Singer (2008) analyze the decision of 78 countries to adopt a monetary policy strategy known as inflation targeting, dependent on the decision taken by other countries using time series cross-section data over the period 1987–2003. Due to the fact that the coefficient of this interaction term is positive, the probability that a country will transfer to state 1 increases if other countries have preceded. However, by just pooling cross-sectional data over time, they implicitly assume that the period that has expired since a neighboring country has taken a positive decision has no impact. In addition, they assume that neighboring countries that have not yet taken a positive decision have the same impact as countries that have already adopted inflation targeting.

These and related issues have been widely discussed in the literature on duration modeling (see Cameron and Trivedi, 2005, Ch. 17, for an excellent overview). Generally, duration models are used to explain the time that has passed to the moment when a particular unit transfers from state 0 to state 1. This literature has produced two results that are relevant for our study. Firstly, if the data are observed in discrete time intervals, one can use a discrete time transition model, since in each time interval two outcomes are possible: the transition takes place or it does not (Cameron and Trivedi, 2005, p. 602). A probit model based on the normal distribution function which restricts the coefficients of the regressors to be constant over time, except for the intercept, is then a straightforward and legitimate choice. Secondly, observations on units after they transferred to state 1 are generally removed from the sample. This is because explanatory variables that change over time may exhibit feedback and hence may not be strictly exogenous; once a unit has transferred to state 1, the explanatory variables may change as a result of this transition.

The standard probit model as suggested in Cameron and Trivedi (2005) for duration data is not appropriate for our setting since individual units are treated as independent entities in duration models. Interaction effects result in additional complications. In duration models the process that is observed may have begun at different points in time for different units in the sample. In our setting, not only is the time that has passed before units transfer to state 1 important, but also the time that has passed since the transfer of other units. Therefore, the transfer process can only be modeled adequately if the starting point of the observation period is the same for every unit in the sample.

The paper is structured as follows. Section 2 summarizes the literature on the basic spatial probit model, its extensions and estimation issues. A detailed description of our model and the results of a Monte Carlo simulation experiment are provided in Section 3. Section 4 illustrates our approach with an empirical study of the adoption of inflation targeting for a sample of 58 countries over the period 1985–2008. Section 5 concludes.

2. SPATIAL PROBIT MODELS: A REVIEW

2.1. The Basic Spatial Probit Model

The basic spatial probit model is a linear regression model with spatially correlated error terms ε_i for a cross-section of N observations ($i = 1, \ldots, N$). In vector notation, this model reads as

$$\boldsymbol{Y}^* = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} = \lambda \boldsymbol{W}\boldsymbol{\varepsilon} + \boldsymbol{v} \tag{1}$$

where $\boldsymbol{Y}^*$ is an $N \times 1$ vector consisting of one observation on the unobserved dependent variable y_i^* for every unit i ($i = 1, \ldots, N$) in the sample, and $\boldsymbol{X}$ is an $N \times K$ matrix of exogenous explanatory variables with parameters contained in a $K \times 1$ vector $\boldsymbol{\beta}$. $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)'$ and $\boldsymbol{v} = (v_1, \ldots, v_N)'$ represent the error terms of the model; $\boldsymbol{\varepsilon}$ reflects the spatially correlated error term with coefficient λ, while $\boldsymbol{v}$ follows a multivariate normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{I}$. We use $\boldsymbol{I}$ rather than $\sigma^2\boldsymbol{I}$ here since $\boldsymbol{\beta}$ and σ^2 cannot be separately identified. For this reason, σ^2 is set to 1. $\boldsymbol{W}$ is an

DOI: 10.1002/jae

$N \times N$ pre-specified non-negative spatial weights matrix describing whether or not the spatial units in the sample are neighbors of each other. Observed choices y_i are linked to the unobserved variable y_i^* by the rule: $y_i = 0$ if $y_i^* \leq 0$ and $y_i = 1$ if $y_i^* > 0$. Generally, the spatial error probit model is consistent with a situation where determinants of the binary response variable omitted from the model are spatially autocorrelated, and with a situation where unobserved shocks follow a spatial pattern.

The spatial error probit model in (1) can be rewritten as

$$\boldsymbol{Y}^* = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \lambda\boldsymbol{W})^{-1}\boldsymbol{v} \tag{2}$$

which implies that the covariance matrix of $\boldsymbol{\varepsilon}$ is $\boldsymbol{\Omega}_\lambda = [(\boldsymbol{I} - \lambda\boldsymbol{W})'(\boldsymbol{I} - \lambda\boldsymbol{W})]^{-1}$. The basic problem that needs to be solved in estimating this model is that the likelihood function cannot be written as the product of N one-dimensional normal probabilities as is the case with the standard (non-spatial) probit model. This is because the individual error terms ε_i $(i = 1, \ldots, N)$ are dependent on each other, as a result of which the likelihood function

$$L(\boldsymbol{\beta}, \lambda|\boldsymbol{Y}) = \int_{\boldsymbol{Y}^*} \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Omega}_\lambda|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon}'\boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\varepsilon}\right\} \mathrm{d}\boldsymbol{\varepsilon} \tag{3}$$

is an N-dimensional integral.

Another problem might be the inversion of the matrix $(\boldsymbol{I} - \lambda\boldsymbol{W})$ for large values of N when using a numerical algorithm to find the optimum of λ, since the number of steps which most practical algorithms require to determine the inverse of an $N \times N$ matrix is proportional to N^3. The spatial error probit model has mainly been used to present solutions to these methodological problems (see McMillen, 1992; Pinkse and Slade, 1998; LeSage, 2000; Beron and Vijverberg, 2004; Fleming, 2004; Klier and McMillen, 2008; Wang *et al.*, 2013), but it has rarely been used in empirical applications. One exception is Pinkse and Slade (1998), who use this model to investigate whether oil companies and their branded-services stations make contracts where either the company or the station operator sets the retail price and whether this contract-type decision depends on that of neighboring stations. They expect positive spatial error correlation if this decision is driven by price competition and negative spatial error correlation if driven by product differentiation.

2.2. The Spatial Lag Probit Model

Another popular spatial probit model is the spatial lag probit model: a linear regression model with endogenous interaction effects among the unobserved dependent variable

$$\boldsymbol{Y}^* = \rho\boldsymbol{W}\boldsymbol{Y}^* + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v} \tag{4}$$

where ρ represents the spatial autoregressive coefficient. Endogenous interaction effects are typically considered as the formal specification for the equilibrium outcome of a spatial or social interaction process, in which the value of the dependent variable for one agent is jointly determined with that of neighboring agents. By rewriting the spatial lag probit model as

$$\boldsymbol{Y}^* = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{v} \equiv (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{5}$$

$$\boldsymbol{\varepsilon} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{v}$$

it can be seen that the covariance matrix of $\boldsymbol{\varepsilon}$ in this model is similar to that of the spatial error probit model, $\boldsymbol{\Omega}_\rho = [(\boldsymbol{I} - \rho\boldsymbol{W})'(\boldsymbol{I} - \rho\boldsymbol{W})]^{-1}$, the difference being that the parameter λ is replaced by ρ. To estimate this model, not only the integration of N-dimensional integral needs to be accounted for, but also the endogeneity of the variable $\boldsymbol{WY}^*$. Many studies have considered this model from a methodological viewpoint: McMillen (1992), LeSage (2000), Beron and Vijverberg (2004), Fleming (2004), Klier and McMillen (2008), LeSage and Pace (2009, Ch. 10), Franzese and Hays (2010), (Smirnov, 2010), Pace and LeSage (2011). In contrast to the spatial error probit model, it has been used in many empirical studies, among which are the two empirical applications of LeSage *et al.* (2011) and Mukherjee and Singer (2008) discussed in the Introduction.

2.3. Extensions of the Spatial Probit Model

The spatial probit model has been extended further in several ways. LeSage and Pace (2009) present a spatial probit model where more than two alternatives are observed that can be ordered. Bhat *et al.* (2010) deal with ordered-response models in general, among which are probit and logit. Wang and Kockelman (2009) construct a dynamic ordered spatial error probit model. Apart from spatial correlation, they add a (latent) dependent variable lagged in time to control for temporal dependence in the data. Instead of estimating a binary response variable, Pinkse *et al.* (2006) estimate transition probabilities: the probability of being in one state in period t conditional on having been in the same or another state in period $t-1$. Their model also allows for different coefficients and/or regressors depending on the current state. The spatial dependence structure takes the form of a group structure in which units in one group interact with each other but not with units in other groups. Pinkse and Slade (2007) consider ML estimation of a combined spatial lag spatial error model, although the spatial dependence is estimated ex post. Flores-Lagunes and Schnier (2012) extend the spatial error probit model to a so-called Tobit type II model: i.e. they first transform the dependent variable into a binary variable and explain this variable by a spatial error probit model; then they explain the magnitude of the dependent variable by a regular spatial error model for only those units that are in state 1. Some studies also deal with heteroskedasticity (McMillen, 1992; LeSage, 2000; Fleming, 2004),[1] without altering the basic multidimensional integration problem. It is perhaps for this reason that heteroskedasticity received less attention in later work.

2.4. Estimation

The expectation-maximization (EM) algorithm adapted by McMillen (1992) for the spatial probit model is one of the earliest attempts to deal with the multidimensional integration problem. The E-step takes the expectation of the log-likelihood function for the latent variable y_i^* conditional on its observed value y_i and the parameter vector. The initial parameter vector is obtained by estimating the spatial model as if the dependent variable is continuous, while subsequent values are obtained from the previous iteration. The M-step maximizes the likelihood function for the parameter vector conditional on the expected value of y_i obtained from the E-step, which boils down to estimating a regular spatial model for a continuous variable. These steps are then repeated until the parameter vector converges. This algorithm, however, has been severely criticized. First, there is a substantial computational burden in the repetitions of the algorithm (Fleming, 2004). Both the EM algorithm and the maximization of the regular spatial model in each M-step require an iterative two-stage procedure. Secondly, it does not produce an estimate of the variance–covariance matrix needed to determine the standard errors and t-values of the parameter estimates (LeSage, 2000; Fleming, 2004; Smirnov, 2010). It should be stressed that this is because of another important methodological shortcoming that has not been discussed in the literature before. Whereas the expectation of the latent variable y_i^* in the EM algorithm

[1] Generally, the diagonal elements of the variance–covariance matrix $\boldsymbol{\Omega}_\lambda = [(\boldsymbol{I}-\lambda\boldsymbol{W})'(\boldsymbol{I}-\lambda\boldsymbol{W})]^{-1}$ are not equal to each other. Some studies characterize this as heteroskedasticity too (Pinkse and Slade, 1998; Klier and McMillen, 2008). However, this type of heteroskedasticity is explicitly taken into account in the estimation of the spatial lag and the spatial error model.

DOI: 10.1002/jae

is determined conditional on the observed value y_i of the unit itself, it must be determined conditional on the observed values of *all* other units. Consequently, this algorithm produces inconsistent parameter estimates.

A similar type of problem applies to the Bayesian Markov chain Monte Carlo (MCMC) estimation procedure initially developed by LeSage (2000). This procedure is based on sequentially drawing model parameters from their conditional distributions. This process of sampling parameters continues until the distribution of draws converges to the targeted joint posterior distribution of the model parameters. Two different sampling schemes are used: the Gibbs sampler for model parameters that have standard conditional distributions ($\boldsymbol{\beta}$, $\boldsymbol{Y}^*$); and the Metropolis–Hastings sampler for the spatial parameter λ in the spatial error model or ρ in the spatial lag model, both of which have a non-standard distribution (LeSage and Pace, 2009, Ch. 5). The key problem is to sample $\boldsymbol{Y}^*$. In LeSage (2000), the individual elements of $\boldsymbol{Y}^*$ are obtained by sampling from a sequence of univariate truncated normal distributions. In later work, LeSage and Pace (2009, p. 285) point out that 'this *cannot* be done for the case of a truncated multivariate distribution' (emphasis in original). Draws for individual elements y_i^* should be based on the distribution of y_i^* conditional on all other $N-1$ elements $\left[y_1^*, \dots, y_{i-1}^*, y_{i+1}^*, \dots, y_N^*\right]$. Probably because James LeSage has made a Matlab routine of the (improved) Bayesian MCMC estimator of the spatial lag probit model available at his website (www.spatial-econometrics.com), it has been frequently used in empirical research (Bolduc *et al.*, 1997; Mukherjee and Singer, 2008; Wang and Kockelman, 2009; LeSage *et al.*, 2011). Another reason might be that Bayesian MCMC is faster than other estimation techniques (Franzese and Hays, 2010).

A third estimation method is the generalized method of moments (GMM), initially proposed by Pinkse and Slade (1998) for the estimation of a spatial error probit model.[2] To deal with the endogeneity of the spatially lagged dependent variables in the case of the spatial lag model, the variable $\boldsymbol{W}\boldsymbol{Y}^*$ is instrumented by $[\boldsymbol{X}\ \boldsymbol{W}\boldsymbol{X} \dots \boldsymbol{W}^g\boldsymbol{X}]$, where g is a pre-selected constant. Typically, one would take $g = 1$ or $g = 2$, dependent on the number of regressors and the type of model (see Kelejian *et al.*, 2004). To avoid repeated inversions of the matrix $(\boldsymbol{I} - \lambda\boldsymbol{W})$, they linearize the spatial parameters around the non-spatial parameter values that are obtained from a standard (non-spatial) probit or logit model. GMM studies do not specify the distribution function of the error terms, and therefore do not solve the multidimensional integration problem. They take into account that the diagonal elements of the covariance matrix are different from one unit to another by scaling the explanatory variables $\boldsymbol{X}_i$ of each unit i by σ_i, where σ_i represents the ith diagonal element of the covariance matrix of the error term $\boldsymbol{\Omega}_p = [(\boldsymbol{I} - p\boldsymbol{W})'(\boldsymbol{I} - p\boldsymbol{W})]^{-1}$ with $p = \lambda$ in the case of the spatial error model and $p = \rho$ in the case of the spatial lag model. However, they do no take into account that the off-diagonal elements of this matrix are non-zero too. Consequently, they overrule the basic notion underlying spatial econometric models in general and spatial discrete-response models in particular that units cannot be treated as independent entities. In other words, although these studies are right that the ML and Bayesian methods rely on the potentially inaccurate assumption of normally distributed error terms, they in turn ignore the spatial interaction effects among the error terms.

Our paper adopts a maximum likelihood estimation method. Starting from McMillen (1992), Beron and Vijverberg (2004) developed a simulated maximum likelihood (SML) estimator for the spatial lag probit model. This simulation method is known as recursive-importance-sampling (RIS) and relies on Monte Carlo simulation of truncated multivariate normal distributions, as discussed by Vijverberg (1997). First, a lower-triangular Cholesky matrix of the variance–covariance matrix of the error terms is determined, and then the multidimensional integral in equation (3) is

[2] Klier and McMillen (2008) use the same technique to estimate a spatial lag logit model. Following these two studies, Diallo and Geniaux (2011) propose a GMM estimator for a logit model with both a spatially lagged dependent variable and a spatially autocorrelated error term. Flores-Lagunes and Schnier (2012) develop a GMM estimator for their so-called Tobit type II model. These studies criticize the Bayesian MCMC and ML estimation methods for relying on the potentially inaccurate assumption of normally distributed errors. Instead, they assume that the individual error terms υ_i are i.i.d. with mean zero and variance σ^2.

DOI: 10.1002/jae

evaluated. Beron and Vijverberg (2004) also point out that the RIS-normal simulator is identical to the Geweke–Hajivassiliou–Keane (GHK) simulator (Börsch-Supan and Hajivassiliou, 1993; Keane, 1993). The advantage of this estimation method is that it provides a feasible and efficient algorithm to approximate the N-dimensional truncated normal density function needed to maximize the log-likelihood function.

Franzese and Hays (2010) compare the performance of different estimation methods of the spatial lag probit model using Monte Carlo experiments and find that the RIS simulator produces more efficient estimates of the spatial parameter ρ than Bayesian MCMC.[3] However, due to the Cholesky factorization of the $N \times N$ covariance matrix, the RIS procedure turns out to be computationally intensive and time consuming, especially when N grows large. Fortunately, two recent studies developed estimation routines to speed up computation time of the SML estimator, beginning with Pace and LeSage (2011), who exploited the fact that often not more than 5% of the elements of the spatial weight matrix $\boldsymbol{W}$ is different from zero. By using sparse matrix algorithms that only store the non-zero elements, computation time can be reduced substantially. Following Pace and LeSage (2011), Liesenfeld *et al.* (2013) used sparse matrix algorithms, but instead of the GHK/RIS simulator they proposed efficient importance sampling (EIS), based on a procedure developed by Richard and Zhang (2007). They point out that EIS is a high-dimensional Monte Carlo integration technique, based on simple least-squares (LS) approximation, designed to maximize numerical accuracy of the SML estimator. Just as in Beron and Vijverberg (2004) and Pace and LeSage (2011), EIS is based on a recursive sequence of auxiliary importance sampling densities for v_i given v_{i+1}, but the difference is that it generalizes the GHK/RIS simulator for the spatial probit model by also imposing the LS optimization step. Full details can be found in Liesenfeld *et al.* (2013). Since the EIS-SML estimator is shown to outperform the GHK/RIS-SML estimator in a simple Monte Carlo simulation experiment, we will use it in our empirical analysis. To test the performance of this estimator in our setting, as well as the Bayesian MCMC estimator, we will carry out a Monte Carlo simulation experiment, which is introduced shortly.

3. TRANSITIONS AT DIFFERENT MOMENTS IN TIME

In the Introduction it was explained that the spatial probit model when pooling cross-sectional data over time cannot be fruitfully used to describe the transition from one state to another when this transition for one cross-sectional unit takes place at a different moment in time than for another unit. Following the literature on duration models, we propose an alternative specification in which observations on units after they transferred to state 1 are removed. This model takes the form

$$\boldsymbol{Y}_t^{0*} = \rho \boldsymbol{W}_t^{00} \boldsymbol{Y}_t^{0*} + \delta \boldsymbol{W}_t^{01} \boldsymbol{Y}_t^{1} + \boldsymbol{X}_t^{0} \boldsymbol{\beta} + \boldsymbol{v}_t^{0} \tag{6}$$

where $t = 1, \ldots, T$ is an index for the time dimension. The dependent variable only contains units that are still in state 0 at the start of every time period ($\boldsymbol{Y}_t^{0*}$). If N_t^0 denotes the number of observations that are not yet in state 1 at the start of time period t, the total number of observations to estimate the parameters of this model amounts to $\sum_{t=1}^{T} N_t^0$. Units that had not yet transferred may be affected by neighboring units that also had not yet transferred, and vice versa, and by neighboring units that had already transferred. The first variable on the right-hand side, $\boldsymbol{W}_t^{00} \boldsymbol{Y}_t^{0*}$, denotes the interaction effect with the first set of units. Since these units are also in state 0, this variable represents an endogenous interaction effect. The second variable on the right-hand side, $\boldsymbol{W}_t^{01} \boldsymbol{Y}_t^{1}$, denotes the interaction effect with the second set of units. Since it can be observed that these units are already in state 1, $\boldsymbol{Y}_t^1$ is specified as an observable variable. Furthermore, since observations on units in time periods after they transferred to state 1 are removed from the sample, units that had already transferred cannot be affected

[3] As an alternative to the SML, Wang *et al.* (2013) propose a partial maximum likelihood (PML) estimator of a bivariate spatial error probit model. The authors show through a simulation study that the PML estimator is more efficient than GMM.

by units that are still in state 0. Consequently, the right-hand-side variable $\boldsymbol{W}_t^{01}\boldsymbol{Y}_t^1$ may be treated as an exogenous explanatory variable. Finally, since it is reasonable to assume that neighboring units that are still in state 0 may have a different impact from neighboring units that had already transferred to state 1, we allow these two variables to have different coefficients ρ and δ. Hence the parameters in equation (6) can be estimated similarly to those of a standard spatial lag probit model, equation (4).

Equation (6) contains two submatrices extracted from the full $N \times N$ spatial weights matrix $\boldsymbol{W}$: $\boldsymbol{W}_t^{00}$ expressing spatial relations between the units that are in state 0 at the start of period t; and $\boldsymbol{W}_t^{01}$ describing spatial relations of the units in state 0 with the units in state 1 at the start of period t. Since the number of spatial units in state 0 and 1 may be different from one period to another, these submatrices are time dependent, indicated by the subscript t. The dimensions of these submatrices $\boldsymbol{W}_t^{00}$ and $\boldsymbol{W}_t^{01}$ are respectively $N_t^0 \times N_t^0$ and $N_t^0 \times N_t^1$, where $N_t = N_t^0 + N_t^1$ for all t.

Finally, it should be stressed that this specification is suitable only when state 1 is an absorbing state; once units have transferred to state 1, they do not return to state 0. Such a specification is appropriate for our empirical illustration where we analyze inflation targeting adoption by countries. Inflation targeting is treated as an absorbing state since countries do not leave it (see Section 4 for discussion). If instead units can exit state 1, the model needs to be further generalized.

3.1. Direct and Indirect Effects

It is well known that the point estimates of the parameter vector $\boldsymbol{\beta}$ in the probit model $\boldsymbol{Y}^* = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}$ and in the spatial lag model with a continuous dependent variable $\boldsymbol{Y} = \rho\boldsymbol{W}\boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}$ are not equal to their marginal effects (see Cameron and Trivedi, 2005, p. 466, and LeSage and Pace, 2009, pp. 293–297, respectively). LeSage *et al.* (2011) consider the marginal effects of the spatial probit model by combining these two models. When applied to our model set forth in equation (6), the matrix of partial derivatives of the expected value of $\boldsymbol{Y}$ with respect to the kth explanatory variable of $\boldsymbol{X}$ in unit 1 up to unit N (say x_{ik} for $i = 1, \ldots, N$, respectively) at a particular moment in time t takes the form

$$\left(\frac{\partial E(\boldsymbol{Y}_t)}{\partial x_{1k}} \cdots \frac{\partial E(\boldsymbol{Y}_t)}{\partial x_{Nk}} \right) = \begin{pmatrix} \frac{\partial E(y_{1t})}{\partial x_{1k}} & \cdots & \frac{\partial E(y_{1t})}{\partial x_{Nk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_{Nt})}{\partial x_{1k}} & \cdots & \frac{\partial E(y_{Nt})}{\partial x_{Nk}} \end{pmatrix} = \operatorname{diag}(\phi(\boldsymbol{\eta}))\left(\boldsymbol{I} - \rho\boldsymbol{W}_t^{00}\right)^{-1}\boldsymbol{I}_N\beta_k \qquad (7)$$

where $\boldsymbol{\eta} = \left(\boldsymbol{I} - \rho\boldsymbol{W}_t^{00}\right)^{-1}\left(\delta\boldsymbol{W}_t^{01}\boldsymbol{Y}_t^1 + \boldsymbol{X}_t^0\boldsymbol{\beta}\right)$ denotes the vector of predicted values of $\boldsymbol{Y}_t^0$. The first matrix on the right-hand side of this equation is a diagonal matrix of order N whose elements ϕ_i represent the probability that the dependent variable takes its observed value, dependent on the observed values of the other units in the sample. For this reason, each observation has its own mean and variance. Define the matrix $\boldsymbol{\Pi}$ as $\boldsymbol{\Pi} = \boldsymbol{\eta}\boldsymbol{\eta}'$, π_{ij} as the (i, j)th element of $\boldsymbol{\Pi}$, $\boldsymbol{\Pi}_{-ii}$ as the $(N - 1) \times (N - 1)$ matrix that is obtained after removing both row and column i, and $\boldsymbol{\pi}_{-i}$ as the ith row vector and $\boldsymbol{\pi}_{i-}$ as the ith column vector removed from $\boldsymbol{\Pi}$. Then ϕ_i $(i = 1, \ldots, n)$ evaluates the normal probability density function for the observed value of y_i, which is either 0 or 1, with mean $\boldsymbol{\eta}_i + \boldsymbol{\pi}_{-i}\boldsymbol{\Pi}_{-ii}^{-1}(\boldsymbol{y}_{-i} - \boldsymbol{\eta}_{-i})$ and variance $\pi_{ii} - \boldsymbol{\pi}_{-i}\boldsymbol{\Pi}_{-ii}^{-1}\boldsymbol{\pi}_{i-}$.

The second matrix on the right-hand side of equation (7) is an $N \times N$ matrix whose diagonal elements represent the impact on the dependent variable of unit 1 up to N if the kth explanatory variable in the own unit changes, while its off-diagonal elements represent the impact on the dependent variable if the kth explanatory variable in another unit changes. LeSage and Pace (2009) define the *direct effect* as the average diagonal element of the full matrix expression on the right-hand side of equation (7), and the *indirect effect* as the average row or column sums of the off-diagonal elements of that matrix expression. Normally, the outcomes are independent from the time index, but in this case they are not since the spatial weight matrix changes over time. To obtain one summary indicator for

DOI: 10.1002/jae

the direct effect and one for the indirect effect of every explanatory variable in the model, we therefore propose to average the outcomes also over time.[4] Since $\boldsymbol{W}_t^{00}$ will be row-normalized, the sum of the direct and indirect effect of the kth explanatory variable, also known as the *total effect*, will take the form $\bar{\phi}(1-\rho)^{-1}\beta_k$, where $\bar{\phi}$ is the average of the ϕs.

The standard errors and t-values of the direct and indirect effects estimates are more difficult to determine, because they depend on β_k, ρ and the elements of the spatial weights matrix $\boldsymbol{W}_t^{00}$ in a complicated way. In order to draw inferences regarding the statistical significance of the direct and indirect effects, LeSage and Pace (2009, p. 39) suggest simulating the distribution of the direct and indirect effects using the variance–covariance matrix implied by the maximum likelihood estimates. If the full parameter vector $\boldsymbol{\theta} = (\rho, \delta, \boldsymbol{\beta}')'$ is drawn D times from $N(\hat{\boldsymbol{\theta}}, \text{asyvar}(\hat{\boldsymbol{\theta}}))$, the standard deviation of each summary indicator can be approximated by the standard deviation of the mean value over these D draws, and the significance by dividing each summary indicator by the corresponding estimated standard deviation.

3.2. Monte Carlo Simulation

This section describes the design and the results of a simulation experiment conducted to compare the performance of the efficient importance sampling simulated maximum likelihood (EIS-SML) and the Bayesian MCMC estimators when applied to the model proposed in equation (6).

The spatial arrangement of the units in the cross-sectional domain is based on the 'buckyball'. This is composed of $N = 60$ units distributed over the surface of a sphere in such a way that the distance from any unit to its first, second up to ninth nearest neighbors is the same for all units. The buckyball spatial weights matrix is a 60×60 symmetric matrix and has applications for physical objects, such as the seams in a soccer ball. Each unit has 59 neighbors, among which are 3 first nearest neighbors, 6 second, 8 third, 10 fourth, 10 fifth, 10 sixth, 8 seventh, 3 eighth and finally 1 ninth nearest neighbor. The off-diagonal elements of the spatial weights matrix are specified as the inverse distances to these neighbors, as a result of which each unit will still have neighbors even if many units already transferred to state 1.

To generate the dependent variable in the first period, we closely follow Beron and Vijverberg (2004) and Franzese and Hays (2010). In period 1 we have

$$\boldsymbol{Y}_1^{0*} = \left(\boldsymbol{I} - \rho \boldsymbol{W}_1^{00}\right)\left[\boldsymbol{X}_1^0 \boldsymbol{\beta} + \boldsymbol{v}_1^0\right] \tag{8}$$

We assume that all units are in state 0 at the start of this first period; this explains why the term $\delta \boldsymbol{W}_1^{01}\boldsymbol{Y}_1^1$ is lacking. We consider one exogenous explanatory variable $\boldsymbol{X}_1^0$ drawn from the standard normal distribution with coefficient $\beta = 1$. Similarly, the error terms are drawn from the standard normal distribution, since σ^2 is generally set to 1 for reasons of identification. Owing to this set-up, the probabilities that a unit in the first period stays in state 0 or transfers to state 1 are equal to each other. In subsequent periods ($t = 2, \ldots, T$ with $T = 5$), we generate the data by

$$\boldsymbol{Y}_t^{0*} = \left(\boldsymbol{I} - \rho \boldsymbol{W}_t^{00}\right)\left[\delta \boldsymbol{W}_t^{01}\boldsymbol{Y}_t + \left(\alpha \boldsymbol{X}_{t-1}^0 + (1-\alpha)\boldsymbol{X}_t^0\right)\boldsymbol{\beta} + \boldsymbol{v}_t^0\right] \tag{9}$$

The term $\delta \boldsymbol{W}_t^{01}\boldsymbol{Y}_t$ is added here, since the number of units in state 1 will be greater than zero in the second time period (except for some exceptional cases) and will further increase over time. The parameter α represents the degree of persistence in the explanatory variables $\boldsymbol{X}_t^0$ and is set to 0.95, since explanatory variables generally change only slowly over time. The coefficients ρ and δ are varied over the range -0.3 to 0.6 by increments of 0.3, producing a total of 16 parameter combinations. Given that the size of the matrix $\boldsymbol{W}_t^{00}$ changes over time, we row-normalize $\boldsymbol{W}_t^{00}$ for every period to

[4] A similar expression to equation (7) applies to the explanatory variable $\boldsymbol{W}_t^{01}\boldsymbol{Y}_t^1$ with coefficient δ.

DOI: 10.1002/jae

satisfy the regularity conditions for consistency and asymptotic normality of spatial parameter ρ. The rule $y_i = 0$ if $y_i^* \leq 0$ and $y_i = 1$ if $y_i^* > 0$ is used to generate $\boldsymbol{Y}$ from $\boldsymbol{Y}^*$.

Table I reports the bias in the parameters β, ρ and δ and their root mean squared errors (RMSE) based on 1000 replications for each of the experimental parameter combinations and three different estimators. The first is the ML estimator if $\boldsymbol{WY}^*$ were observable; it is used as a benchmark to compare the performance of the other two estimators. The second estimator is the Bayesian MCMC estimator made available by James LeSage at his website (www.spatial-econometrics.com) (routine sarp-g). The number of draws for burn-in within this routine is set to 1000 and the total number of draws to 5000. With this setting the computation time of the Bayesian MCMC estimator appeared to be comparable to that of the EIS-SML estimator based on Liesenfeld *et al.* (2013) (made available at www.stat-econ.uni-kiel.de), which is the last estimator that we consider.

When using the ML estimator, provided that $\boldsymbol{WY}^*$ is observable, the absolute value of the bias in the parameter β amounts to 0.004 averaged over all parameter combinations, which is negligible. By contrast, the parameter ρ is underestimated by 0.041 and the parameter δ by 0.072. The explanation for these biases is that the decision to remove observations from a particular unit i from the sample as soon as y_{it}^* takes a positive value for the first time in the data-generating process is quite strict, especially if this value is close to zero. The performance of this estimator slowly improves when increasing the sample size. We experimented with $T = 15, 25$ and $N = 400, 900$. In the case of T, the absolute value of the biases changed to 0.019 and 0.007 for β, 0.033 and 0.032 for ρ and 0.071 and 0.062 for δ, and in the case of N to 0.001 and 0.003 for β, 0.039 and 0.038 for ρ and 0.066 and 0.064 for δ. Biases of this size also occur in spatial probit models that do not remove observations. Using a similar set-up but then without the regressor $\delta \boldsymbol{W}_t^{01} \boldsymbol{Y}_t$, Franzese and Hays (2010, table 1) find that β is overestimated by 0.02 (true value 1) and ρ is underestimated by 0.18 (true value 0.5) for $N = 48$ and that β is underestimated by 0.02 and ρ by 0.08 for $N = 144$.

Finally, the overall RMSEs of the three parameters amount to 0.106 for β, 0.278 for ρ and 0.213 for δ. Since the latter outcomes are almost constant over the different parameter combinations, they have not been reported for each single combination to save space.

Table I. Biases of ML and Bayesian spatial probit estimators with respect to ρ, δ and β

True values		ML $\boldsymbol{WY}^*$			Bayesian MCMC probit			EIS-SML probit			
ρ	δ	β	ρ	δ	β	ρ	δ	β	ρ	δ	$(1-\rho)^{-1}\beta$
−0.3	−0.3	0.000	−0.061	−0.070	0.051	−0.049	−0.089	0.050	0.003	−0.028	0.041
	0	−0.001	−0.033	−0.053	0.052	−0.039	−0.120	0.047	0.020	−0.017	0.068
	0.3	−0.010	−0.008	−0.042	0.051	−0.016	−0.138	0.021	−0.007	−0.039	0.017
	0.6	−0.009	0.007	−0.073	0.043	0.056	−0.123	0.026	0.024	−0.050	0.029
0	−0.3	−0.010	−0.071	−0.079	0.040	−0.188	−0.220	0.029	−0.132	−0.115	−0.051
	0	−0.008	−0.050	−0.083	0.038	−0.143	−0.253	0.036	−0.070	−0.120	−0.031
	0.3	−0.009	−0.033	−0.072	0.019	−0.129	−0.297	0.015	−0.074	−0.144	−0.117
	0.6	−0.011	0.001	−0.077	0.023	0.013	−0.274	0.005	0.034	−0.119	0.043
0.3	−0.3	0.001	−0.108	−0.090	0.043	−0.421	−0.361	0.044	−0.322	−0.199	−0.126
	0	0.000	−0.058	−0.085	0.042	−0.316	−0.386	0.033	−0.129	−0.158	0.053
	0.3	−0.001	−0.029	−0.097	0.056	−0.114	−0.312	0.046	−0.044	−0.150	−0.024
	0.6	−0.002	0.010	−0.095	0.036	0.005	−0.338	0.027	0.004	−0.154	0.069
0.6	−0.3	0.000	−0.026	−0.033	−0.059	−0.321	−0.376	0.086	0.029	−0.021	0.075
	0	−0.004	−0.083	−0.056	0.010	−0.327	−0.361	0.070	−0.076	−0.079	−0.005
	0.3	−0.001	−0.055	−0.071	0.082	−0.284	−0.389	0.069	−0.034	−0.103	0.027
	0.6	−0.003	−0.021	−0.078	0.063	−0.043	−0.440	0.053	0.005	−0.140	0.066
Average bias		0.004	0.041	0.072	0.044	0.154	0.280	0.041	0.063	0.102	0.053
RMSE		0.106	0.278	0.213	0.213	0.438	0.261	0.215	0.442	0.321	0.458

Note: $N = 60$, $T = 5$ and $\boldsymbol{W}$ specified as an inverse distance matrix whose size diminishes over time due to removing observations that transferred to state 1. Average bias in absolute value.

DOI: 10.1002/jae

Just as in Franzese and Hays (2010), we find that the performance of the Bayesian MCMC estimator is relatively poor. Another problem is that in almost half of the cases the estimator did not converge; instead it produced the outcome $\rho = 0.5$. This might be fixed by increasing the number of draws, though at the expense of more computation time. Here we calculate the biases based on the number of solutions that did converge.

When using the ML probit estimator the biases increase in magnitude to 0.041 for β, 0.063 for ρ and 0.102 for δ. Liesenfeld *et al.* (2013), who do not remove any observations, find a bias in β that may run up to 0.037 (true value 3.0) and to 0.002 for ρ. They, however, only consider true values for ρ of 0.75 and 0.85. When considering the case $\rho = \delta = 0.6$ in Table I, the EIS-SML also produces a bias which appears to be small: namely 0.005. The stability of the reported biases over the different parameter combinations suggests that a bias correction procedure might apply; Lee and Yu (2010) show that bias correction procedures for static and dynamic panel data models may lead to significant improvements, which is an interesting topic for further research. Furthermore, it should be noted that the positive bias in the parameter estimate of β is partly compensated by the negative bias in the parameter estimate of ρ when determining the direct and indirect effect estimates using equation (7), since it is based on the expression $(1-\rho)^{-1}\beta$. The overall bias of β and ρ within this expression falls to 0.053 (calculated as the mean absolute value of the biases). Finally, the RMSEs of the three parameters amount to 0.215 for β, 0.442 for ρ and 0.321 for δ; almost twice the RMSEs found for the ML estimator if $\boldsymbol{W}\boldsymbol{Y}^*$ would be observable, except for δ (factor 1.5).

The conclusions of this Monte Carlo experiment are twofold. First, not one single estimator is free of biases, including the ML estimator if $\boldsymbol{W}\boldsymbol{Y}^*$ were observable. This means that researchers need to be careful when drawing any conclusions based on the parameter estimates. In this respect, they would better focus on the direct and indirect effect estimates which depend on more than one parameter and which have the empirical property to partly compensate each other's biases. Second, the EIS-SML estimation procedure developed by Liesenfeld *et al.* (2013) outperforms the Bayesian MCMC estimator in terms of bias, RMSE and computation time.

4. ILLUSTRATION

To illustrate our model, we analyze the transition of countries from one type of monetary policy strategy to another. Specifically, we focus on the adoption of inflation targeting (thereafter, IT), a monetary policy strategy that involves the public announcement of medium-term targets for inflation and a strong commitment to price stability as a final monetary policy objective (Mishkin and Schmidt-Hebbel, 2001). The decision of countries to adopt IT is influenced by their own characteristics as well as choices of other countries that decide either to adopt IT or use an alternative monetary strategy. To explain the interdependence between countries in IT adoption, we explore the literature on international policy diffusion (e.g. Simmons and Elkins, 2004). Policy diffusion means that policy choices of one country lead to similar policy choices of other countries. We identify two mechanisms of IT diffusion: competition and information.

The competition mechanism implies that economic policy elsewhere can alter the payoffs associated with choosing or maintaining a particular policy through economic competition (Simmons and Elkins, 2004). In this context, IT diffusion could be motivated by competition of central banks for having a strong domestic currency (White, 2003). The implementation of IT by one central bank can reduce inflation and strengthen the value of domestic currency, thereby making it more attractive internationally. On observing this outcome, central banks in other countries may also decide to adopt IT to strengthen their own domestic currencies. Conversely, central banks that resist adoption of IT may face reputational consequences.

According to the second diffusion mechanism, policy innovations in some countries provide valuable information about the effects and construction of such polices (Simmons and Elkins, 2004).

DOI: 10.1002/jae

Based on this information, other countries decide which policy is suitable for them. The informational mechanism explains the diffusion of IT between countries through policy learning in that countries gather information and 'learn' from the experience of other countries. Central banks observe each other's monetary strategy choices and share relevant information about their requirements, structure and effects. They may use this information in their decision whether to adopt IT. The important channel through which central banks can 'learn' about IT is network proximity. Countries that are similar in terms of their economic and institutional background also tend to have more information about each other's economic policies and monetary strategies. Consequently, similar countries tend to adopt similar monetary strategies due to better information access and peer effects. Furthermore, countries that decide to adopt IT in the current period could motivate their peers to follow this choice. In addition, the impact of countries that have experience of implementing IT might be different from the impact of countries that have not yet adopted IT.

Hence we include two interaction effects: one for countries that have not yet adopted IT and one for countries that have already adopted IT. In our analysis, we assume that in each time period (year) a country can be in one of two possible states: state 1 corresponds to the adoption of IT, while state 0 corresponds to an alternative monetary strategy. Considering that none of the inflation targeters has so far been willing to abandon this monetary strategy, IT can be treated as an absorbing state.[5]

4.1. Data Description

Our panel dataset is based on the data of Samarina and de Haan (2014) and consists of 58 countries over the period 1985–2008.[6] By coincidence, in our sample 29 countries (17 OECD and 12 non-OECD countries) adopted IT (IT group) and 29 other countries did not adopt IT (non-IT group) during the analyzed period. To reduce the risk of selection bias, we include both OECD and non-OECD countries in the non-IT group. The OECD part of this group consists of 13 remaining OECD non-inflation targeters. Following the approach of Rose (2007) and Lin and Ye (2009), the non-OECD part covers 16 emerging and developing countries that have a population at least as large as the population of the smallest non-OECD inflation targeter, and/or a level of GDP per capita that is at least as high as that of the poorest non-OECD inflation targeter.[7] Table II provides the list of countries in our dataset with the official adoption dates based on central banks' announcements following the 'half-year-rule': if a country adopts IT in the second half of year t, the adoption year is $(t + 1)$; otherwise the adoption year is t.

In line with the theoretical notion that similar countries in terms of their economic and institutional background adopt similar monetary strategies, we use institutional proximity between countries to construct the spatial weight matrix. An important measure of institutional proximity is common legal tradition. Countries with similar legal origins are more strongly connected with each other and more inclined to follow similar policy choices. Let J_i denote the total number of countries that have the same origin of legal system as country i. We adopt a legal similarity weight matrix in which the weights are equal to $1/J_i$ if countries i and j have the same origin of the legal system, and 0 otherwise. The data on legal systems are based on La Porta *et al.* (1999), who distinguish English, French, German,

[5] Three EU members (Finland, Spain and Slovakia) abandoned IT when they joined the EMU, but this decision was caused by institutional commitment to adopt the euro and to unify countries' monetary policy conduct with the ECB (Samarina and Sturm, 2014). Furthermore, although these countries gave up IT formally, their monetary strategy under the ECB framework is similar to implicit IT (Rose, 2007).

[6] IT was adopted for the first time in 1990 in New Zealand. The study period begins in 1985 to allow for a pre-adoption period. Although available, data after 2008 are not used, since IT lost is popularity during the financial crisis; Table 2 shows that the last adoption took place in 2007.

[7] We include only those countries for which the data are available in the analyzed period.

Table II. Country sample

Country	Adoption year	Country	Adoption year
Inflation targeting countries (29)			
Australia	1993	Norway	2001
Brazil	1999	Peru	2002
Canada	1991	Philippines	2002
Chile	1991	Poland	1999
Colombia	2000	Romania	2006
Czech Republic	1998	Slovakia	2005
Finland	1993	South Korea	1998
Ghana	2007	South Africa	2000
Guatemala	2005	Spain	1995
Hungary	2001	Sweden	1993
Iceland	2001	Switzerland	2000
Indonesia	2005	Thailand	2000
Israel	1992	Turkey	2006
Mexico	2001	UK	1993
New Zealand	1990		
Non-inflation targeting countries (29)			
Argentina	Denmark	Ireland	Netherlands
Austria	Egypt	Italy	Pakistan
Belgium	Estonia	Japan	Panama
Bolivia	France	Latvia	Portugal
Bulgaria	Germany	Lithuania	Singapore
China	Greece	Luxembourg	USA
Costa Rica	India	Malaysia	Venezuela
Cyprus			

Source: Samarina and de Haan (2014).

Scandinavian and socialist legal origins. All estimation results reported below are conditional on this choice of the spatial weight matrix.[8]

The matrix $\boldsymbol{X}$ includes six exogenous explanatory variables, considered to be the relevant factors driving countries' motivation to adopt IT: inflation, output growth, exchange rate regime, government debt, financial development and central bank instrument independence. For details see Samarina and de Haan (2014). Table III describes the explanatory variables and their data sources.

Inflation is an important factor in the decision to adopt IT. Countries more often choose this strategy to maintain low inflation rather than to fight high inflation. Thus we expect that low inflation increases the probability to adopt IT. Next, we include output growth to control for the macroeconomic performance of countries. This variable is expected to have a negative effect on the probability of IT adoption—a high-growth country may be reluctant to focus on inflation targets as this could lead to lower economic growth.

To avoid the risk of prioritizing exchange rate stability at the expense of higher inflation, countries are advised to have flexible exchange rates when they adopt IT.[9] The exchange rate regime indicator is based on the de facto coarse classification of Reinhart and Rogoff (2004), with higher values implying more flexible exchange rates. We expect that countries with more flexible exchange rate regimes are more likely to adopt IT. In the presence of large public debt, a central bank may be forced to generate high inflation to reduce the real value of debt. This increases the risk of missing the inflation

[8] Samarina (2014, Ch. 5) considers two other spatial weights matrices: ten-nearest neighbors and common language. She finds that especially the coefficient estimates of the two spatial lags and their significance levels are sensitive to the choice of weights matrix. However, the log-likelihood of the model based on the common legal origin matrix appears to be higher than those of these alternative spatial weights matrices. By using legal tradition as the main principle in the construction of the spatial weight matrix, each country in the sample has at least one other country as 'neighbor'. Geographical measures, which are and also turn out to be of less importance here, have the side effect that many countries end up as 'islands'.

[9] Although having a flexible exchange rate regime is desirable for IT adoption, this is not a necessary precondition. Some emerging and developing countries initially adopted a soft version of IT while still using crawling exchange rate bands. Thus in some cases inflation targets can coexist with exchange rate pegs or bands.

DOI: 10.1002/jae

Table III. Explanatory variables and data sources

Variable	Description and data sources
Inflation	Annual CPI inflation rate transformed as $\frac{\pi_t/100}{1+\pi_t/100}$. Sources: IFS IMF; Datastream
Output growth	Annual GDP growth rates (%). Source: IFS IMF
Exchange rate regime	Indicator, from 1 (hard peg) to 4 (freely floating). In several cases takes value 5 (freely falling). Sources: Reinhart and Rogoff (2004); Ilzetzki *et al.* (2011)
Government debt	Central government debt as % of GDP. Sources: Datastream; OECD Stat; Jaimovich and Panizza (2010)
Financial development	Domestic credit provided by the financial sector/GDP. Source: WDI World Bank
Central bank instrument independence	Dummy variable: 1, central bank is instrument independent; 0, otherwise. Sources: Cukierman *et al.* (2002); Arnone *et al.* (2007); central banks laws

target. Thus low government debt is expected to increase the probability of adopting IT. Samarina and de Haan (2014) find that countries with less developed financial systems are more likely to adopt IT, possibly because higher financial development, measured by credit-to-GDP, makes it harder to control inflation. We expect a similar result. Central bank instrument independence implies that a central bank is independent from government in choosing instruments to achieve its goals. Higher instrument independence gives more freedom for central banks to pursue their policy objectives; hence it is expected to increase the probability of IT adoption.

Our dataset is not complete; the percentage of missing observations on different explanatory variables ranges from 1% to 8% of all observations. An imputation technique is used for filling in missing observations.[10] Finally, the explanatory variables do not highly correlate with each other (not reported here).

4.2. Estimation Results

Table IV reports the coefficient estimates and their t-statistics for two specifications of the spatial probit model. Column (1) of Table IV contains results for the standard spatial lag probit model when pooling the cross-sectional data over time. This model can be obtained from equation (4) by adding a subscript t, which runs from 1 to T, to the variables and the error terms of that equation. This model is similar to the one employed in Mukherjee and Singer (2008) for their analysis of IT adoption. We find that the coefficient estimate ρ of the endogenous interaction effects is positive and significant, while Mukherjee and Singer (2008) report a positive but insignificant result. One explanation is that we use data over a longer time period: 1985–2008 versus 1987–2003. The findings for two regressors used in their study and ours—exchange rate regime and central bank independence—are comparable, while the result for inflation is different, both in terms of the sign and significance of the estimate.

Column (2) contains results for our spatial probit model with two spatially lagged variables and a full set of regressors. The results show that the estimate of the spatial lag reflecting countries that had not yet make the transition to inflation targeting (ρ) is positive and significant at the 5% level, whereas the estimate of the spatial lag reflecting countries that had already transferred to inflation targeting (δ) is insignificant. Compared to the standard spatial probit model, the most important change is that the coefficient of the first spatial lag halves from 0.464 to 0.231. According to Table I, this coefficient might be downward biased, but the potential magnitude of this bias is too small (<0.05) to explain this significant change. Apparently, applying the standard spatial probit leads to overestimation of the spatial parameter. Since the estimate of the second spatial lag is insignificant, this can only be explained by removing observations from the sample after they have transferred to IT—the common approach in the duration literature.

[10] We apply the expectation-maximization (EM) algorithm for missing values imputation, which is proposed by Dempster *et al.* (1977) and described, for example, in Schafer (1997).

DOI: 10.1002/jae

Table IV. Estimation results: spatial lag probit (1985–2008)

Variable	(1) Standard probit	(2) Spatial probit
ρ	0.464***	0.231**
	(4.95)	(1.97)
δ		0.601
		(1.39)
Inflation	−8.293***	−4.673***
	(−40.24)	(−3.16)
Output growth	−0.006	−0.048*
	(−0.01)	(−1.80)
Exchange rate regime	0.496***	0.369***
	(33.18)	(4.49)
Government debt	−0.006	−0.003
	(−0.13)	(−0.91)
Financial development	−0.370***	−0.669***
	(−224.61)	(−3.26)
Central bank instrument independence	0.707***	0.583***
	(8.00)	(3.14)
Constant	−0.865***	−1.610***
	(−19.91)	(−4.31)
Observations	1392	1127
Log-likelihood	−499.41	−110.86

Note: The table reports coefficient estimates and their t-values (in parentheses). $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Column (1) shows the results for a standard spatial lag probit model and column (2) for our spatial probit model.

The estimation of our spatial probit model may suffer from temporal dependency problems, as in discrete time duration models. This means that the probability of a country to adopt IT in year t depends on the duration of the non-IT period, i.e. the time that has passed from the start of the sample period until the IT adoption date. Ignoring temporal dependence may lead to inefficiency and inaccurate statistical inference. To correct for temporal dependence, we followed the approach of Beck *et al.* (1998) and generated a set of time dummies that mark each non-IT duration period. Including these time dummies yields comparable outcomes to the results reported in column (2) of Table IV, as shown in Appendix A (supporting information). The reason is that the time dummies appeared to be jointly insignificant.

We examine how the estimates of explanatory variables change in the two model specifications. Comparing the results in column (1) with column (2) in Table IV, we notice that the coefficients of inflation, exchange rate regime and central bank instrument independence become much smaller (in absolute value) in our spatial probit, where we exclude observations for the post-adoption period. The factors driving IT adoption may become endogenous after adoption; i.e. once countries adopt IT, their economic characteristics and institutions change as a result of using this strategy. Note that ρ also became smaller and thus that we have a different situation as in the Monte Carlo simulation experiment, where a negative bias in ρ was compensated by positive bias in β. Hence we may conclude that post-adoption observations overestimate the response parameters if not excluded.

The findings in column (2) suggest that countries with lower inflation and output growth and more flexible exchange rate regimes are more likely to adopt IT. This is in line with the literature and our theoretical expectations. Financial system development has a negative significant effect, implying that countries with less-developed financial systems are more likely to choose IT. Finally, the estimate of central bank instrument independence is significant with a positive sign; indeed, IT is more likely to be adopted when central banks have autonomy in choosing instruments to achieve their objectives.

DOI: 10.1002/jae

Table V. Marginal effects of our spatial probit model

Variable	Direct effects on $y*$	Indirect effects on y^*	Direct effects on y	Indirect effects on y ($\times 10^{-2}$)
Countries that already	0.606	0.555	0.053	0.021
adopted IT ($\boldsymbol{Z}_t$)	(1.35)	(1.34)	(1.24)	(0.93)
Inflation	−4.708***	−4.250***	−0.476**	−0.205
	(−3.25)	(−3.14)	(−2.48)	(−1.39)
Output growth	−0.048*	−0.044*	−0.005*	−0.002
	(−1.93)	(−1.90)	(−1.68)	(−1.13)
Exchange rate regime	0.374***	0.337***	0.039**	0.017
	(4.77)	(4.67)	(2.55)	(1.36)
Government debt	−0.003	−0.002	−0.000	−0.000
	(−0.77)	(−0.77)	(−0.66)	(−0.50)
Financial development	−0.651***	−0.587***	−0.066**	−0.028
	(−3.27)	(−3.17)	(−2.54)	(−1.43)
Central bank instrument	0.571***	0.513***	0.061**	0.028
independence	(3.29)	(3.35)	(2.04)	(1.25)

Note: Direct and indirect effects with t-values (in parentheses) are derived from the parameter estimates of model (2) in Table IV. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

4.3. Direct and Indirect Effects

Table V shows the direct effects of changes in explanatory variables on y^* and y, as well as the indirect effects of explanatory variables on institutionally proximate countries obtained from the coefficient estimates of our spatial probit model. For two reasons the direct effects are different from the coefficient estimates reported in Table IV. The first reason is the feedback effect that arises as a result of impacts passing through neighboring countries and back to the country where the change in one of the explanatory variables originated from. The second reason is the probability that a country adopts IT. An example may illustrate this. The coefficient estimate of the interaction effect with countries that already adopted IT is 0.601 (see Table IV). The corresponding direct effect on the unobserved variable y^*—the willingness to adopt IT—is 0.606 (see Table V). This implies that the feedback effect is equal to $0.606 - 0.601 = 0.005$, which corresponds to 1.0% of the direct effect on y^*. This is a common finding in standard spatial econometric models (Elhorst, 2014, Sections 2.9, 3.6 and 4.7). Generally, feedback effects are only a fraction (<10%) of the corresponding direct effects.

To obtain the impact on the observed value y, the impact on the vector of unobserved variables in each country also has to be multiplied by the diagonal matrix whose diagonal elements contain the probability to adopt IT $\mathrm{diag}(\phi(\boldsymbol{\eta}))$, as spelled out in equation (7). These probabilities are relatively small in this study. When leaving the pre-adoption period aside, the average value of $\mathrm{diag}(\phi(\boldsymbol{\eta}))$ calculated over the period 1990–2008, representing the average probability to adopt IT in a particular year, amounts to 0.077. Owing to this low value, the direct effect of countries that have already adopted IT on countries that have not yet adopted IT drops to 0.053. Similar figures apply to other explanatory variables and indicate that the probability of adopting IT has a much greater impact on the direct effect of the explanatory variables relative to their coefficient estimates than the feedback effects. Furthermore, since not only the willingness to adopt IT but also the probability of adopting IT is due to standard errors in the parameter estimates, also the significance levels of the direct effects estimates fall.[11] Nevertheless, every variable that appears to have a significant direct effect on the willingness to adopt IT (y^*) also has such effect on the actual decision (y).

Just like the feedback effects, the indirect effects appear to be very small. The explanation is again multiplication by the probability of adopting IT. If we calculate the ratio between the indirect effect and

[11] The willingness to adopt depends on both ρ and the β coefficient of a particular explanatory variable, while the probability of adopting depends on the β coefficients of all explanatory variables. Owing to this nonlinear relationship, the direct effects on y are also not just the direct effects on y^* multiplied by 0.077.

DOI: 10.1002/jae

the direct effect on the unobserved variable y^*—the willingness to adopt IT—we obtain an average value of 90.52% over all explanatory variables in the model. For example, the ratio of the indirect to the direct effect of the interaction effect with countries that have already adopted IT is $0.555/0.606 = 0.923$. However, when accounting for the probability of adopting IT, the indirect effects eventually drop to less than 1%. It is for this reason that we introduced a scaling factor of 10^{-2} in the last column of Table V. For example, we obtain $0.021(*10^{-2})/0.054 = 0.40\%$ for the interaction effect with countries that already adopted IT. In other words, although we have empirical evidence that neighboring countries reconsider their willingness to adopt IT due to changing circumstances in other countries, this evidence is hard to observe in practice since eventually neither of the indirect effects on the observed choices appears to be significant.

5. CONCLUSION

The standard spatial probit model can be employed to describe interaction effects among cross-sectional units when the dependent variable takes the form of a binary response variable. Unfortunately, it cannot adequately deal with transitions from one state to another when these transitions take place at different moments in time for different cross-sectional units.

This paper proposes a spatial probit model with two spatially lagged variables: one for units that had not yet transferred to the other state, and one for units that had already transferred. Observations on units that made the transfer from one state to the other are removed after the transfer. The model is estimated by maximum likelihood methods, using the Efficient Importance Sampling (EIS) algorithm developed by Liesenfeld *et al.* (2013) to evaluate the truncated multidimensional normal distribution. The results of our Monte Carlo simulation experiment show that this estimator outperforms its Bayesian MCMC counterpart, but is still not free of biases. These biases diminish when considering direct and indirect effects. This is because these effects depend on several parameters and have the empirical property that biases in individual parameters partly cancel each other out. Nevertheless, estimating a spatial probit model without biases in the parameters if transitions occur at different moments in time remains a challenge.

We illustrate our approach with a study of the adoption of IT for a sample of 58 countries over the period 1985–2008. We investigate whether countries that had not yet adopted IT interact with other countries, thereby making a distinction between countries that also had not yet adopted IT and countries that had. We find that the first interaction effect is positive and significant when using the standard spatial probit approach, but that it halves when using the spatial probit approach, i.e. when removing post-adoption observations, representing the common approach in the duration literature. Nevertheless, this endogenous interaction effect remains significant. Similarly, the magnitudes and significance levels of the explanatory variables fall when excluding post-adoption observations, among which are inflation, output growth, exchange rate regime, financial development and central bank instrument dependence. The second interaction effect is not found to be significant. Despite these changes, almost all variables produce significant direct and indirect effects on the unobserved dependent variable, as well as significant direct effects on the observed dependent variable, respectively the willingness and actual decision to adopt IT. By contrast, none of these variables produce significant indirect effects on the observed dependent variable. The main explanation for this finding is that the probability of transferring from state 0 to state 1 in this study is relatively small (approximately 8%). It is to be expected that more indirect effects will remain significant in an application where this probability is higher.

Our spatial probit model has various applications in economics, business and political studies. The approach can also be used to explain contagion of financial crises when countries enter crises at different times. Studies on the introduction of new brands and firms' entry decisions might also include transitions at different periods. Additionally, our spatial probit can be applied to analyze land use con-

DOI: 10.1002/jae

version models, where a landowner's decision to convert undeveloped land to farmland depends on current as well as past decisions of his/her neighbors.

ACKNOWLEDGEMENTS

This paper is the result of our reading group on spatial binary response models. We thank Paul Bekker and Tom Wansbeek for their valuable input in discussing the articles; Badi Baltagi, Chandra Bhat, Jakob de Haan, Peter Egger and James LeSage for helpful discussions and comments; the participants of the 59th Annual North American Meetings of the Regional Science Association International, Ottawa, the KOF Research Seminar, Zurich, the Scottish Economic Society Annual Conference, Perth, 12th International Workshop Spatial Econometrics and Statistics, Orléans, France, the Netherlands Econometric Study Group, Amsterdam, and the 9th Triennial Choice Symposium, Noordwijk, the Netherlands; and three anonymous referees for comments and suggestions.

REFERENCES

Arnone M, Laurens BJ, Segalotto J-F, Sommer M. 2007. Central bank autonomy: lessons from global trends. IMF Working Paper 07/88. International Monetary Fund.

Beck N, Katz JN, Tucker R. 1998. Taking time seriously: time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science* **42**: 1260–1288.

Beron KJ, Vijverberg WP. 2004. Probit in a spatial context: a Monte Carlo analysis. In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Anselin L, Florax R, Sergio J (eds). Springer: Berlin; 169–195.

Bhat CR, Sener IN, Eluru N. 2010. A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B* **44**: 903–921.

Bolduc D, Fortin B, Gordon S. 1997. Multinomial probit estimation of spatially interdependent choices: an empirical comparison of two new techniques. *International Regional Science Review* **20**: 77–101.

Börsch-Supan A, Hajivassiliou VA. 1993. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics* **58**: 347–368.

Cameron CA, Trivedi PK. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press: Cambridge, UK.

Cukierman A, Miller GP, Neyapti B. 2002. Central bank reform, liberalization and inflation in transition economies: an international perspective. *Journal of Monetary Economics* **49**: 237–264.

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**: 1–38.

Diallo A, Geniaux G. 2011. Spatial logit for large samples with local spatial lag and regional spatial random effects using linearized GMM: an application to land use models, *5th World Conference of the Spatial Econometrics Association (SEA)*: Toulouse.

Elhorst JP. 2014. *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. Springer: Berlin.

Fleming MM. 2004. Techniques for estimating spatially dependent discrete choice models. In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Anselin L, Florax R, Sergio J (eds). Springer: Berlin; 145–167.

Flores-Lagunes A, Schnier KE. 2012. Estimation of sample selection models with spatial dependence. *Journal of Applied Econometrics* **27**: 173–204.

Franzese RJ, Jr, Hays JC. 2010. Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. Presented at the Annual Meeting of the Public Choice Society, 6 March 2009.

Ilzetzki E, Reinhart CM, Rogoff KS. 2011. The country chronologies and background material to exchange rate arrangements into the 21st century: will the anchor currency hold? Working paper. London School of Economics, University of Maryland/Harvard University/NBER.

Jaimovich D, Panizza U. 2010. Public debt around the world: a new data set of central government debt. *Applied Economics Letters* **17**: 19–24.

Keane MP. 1993. Simulation estimation for panel data models with limited dependent variables. In *Handbook of Statistics 11: Econometrics*, Maddala GS, Rao CR, Vinod HD (eds). North-Holland: Amsterdam; 545–571.

Kelejian HH, Prucha IR, Yuzefovich Y. 2004. Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample results. In *Spatial and Spatiotemporal Econometrics*, LeSage JP, Pace KR (eds). Elsevier: Amsterdam; 163–198.
Klier T, McMillen DP. 2008. Clustering of auto supplier plants in the United States: generalized method of moments spatial logit for large samples. *Journal of Business Economics and Statistics* **26**: 460–471.
La Porta R, de Silanes FL, Shleifer A, Vishny R. 1999. The quality of government. *Journal of Law, Economics and Organization* **15**: 222–279.
Lee L-f, Yu J. 2010. Some recent developments in spatial panel data models. *Regional Science and Urban Economics* **40**: 255–271.
LeSage JP. 2000. Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis* **32**: 19–35.
LeSage JP, Pace RK. 2009. *Introduction to Spatial Econometrics*. Chapman & Hall/CRC: Boca Raton, FL.
LeSage JP, Pace RK, Lam N, Campanella R, Liu X. 2011. New Orleans business recovery in the aftermath of Hurricane Katrina. *Journal of the Royal Statistical Society A* **174**: 1007–1027.
Liesenfeld R, Richard J-F, Vogler J. 2013. Analysis of discrete dependent variable models with spatial correlation. Available at SSRN: http://ssrn.com/abstract=2196041.
Lin S, Ye H. 2009. Does inflation targeting make a difference in developing countries? *Journal of Development Economics* **89**: 118–123.
McMillen DP. 1992. Probit with spatial autocorrelation. *Journal of Regional Science* **32**: 335–348.
Mishkin FS, Schmidt-Hebbel K. 2001. One decade of inflation targeting in the world: what do we know and what do we need to know? NBER Working Paper 8397. National Bureau of Economic Research.
Mukherjee B, Singer DA. 2008. Monetary institutions, partisanship, and inflation targeting. *International Organization* **62**: 323–358.
Pace RK, LeSage JP. 2011. Fast simulated maximum likelihood estimation of the spatial probit model capable of handling large samples. Available at SSRN: http://ssrn.com/abstract=1966039.
Pinkse J, Slade M. 1998. Contracting in space: an application of spatial statistics to discrete-choice models. *Journal of Econometrics* **85**: 125–154.
Pinkse J, Slade M. 2007. Semi-structural models of advertising competition. *Journal of Applied Econometrics* **22**: 1227–1246.
Pinkse J, Slade M, Shen L. 2006. Dynamic spatial discrete choice using one-step GMM: an application to mine operating decisions. *Spatial Economic Analysis* **1**: 53–97.
Reinhart CM, Rogoff KS. 2004. The modern history of exchange rate arrangements: a reinterpretation. *Quarterly Journal of Economics* **119**: 1–48.
Richard J-F, Zhang W. 2007. Efficient high-dimensional importance sampling. *Journal of Econometrics* **141**: 1385–1411.
Rose AK. 2007. A stable international monetary system emerges: inflation targeting is Bretton Woods, reversed. *Journal of International Money and Finance* **26**: 663–681.
Samarina A. 2014. Monetary policy strategies: abandonment, adoption, and performance. PhD Thesis. University of Groningen, Netherlands.
Samarina A, de Haan J. 2014. Right on target: exploring the factors leading to inflation targeting adoption. *Contemporary Economic Policy* **32**: 372–389.
Samarina A, Sturm J-E. 2014. Factors leading to inflation targeting: the impact of adoption. *Applied Economics Letters* **21**: 918–923.
Schafer JL. 1997. *Analysis of Incomplete Multivariate Data*, Monographs on Statistics and Applied Probability No. 72. Chapman & Hall/CRC: London.
Simmons BA, Elkins Z. 2004. The globalization of liberalization: policy diffusion in the international political economy. *American Political Science Review* **98**: 171–189.
Smirnov OA. 2010. Modeling spatial discrete choice. *Regional Science and Urban Economics* **40**: 292–298.
Vijverberg WP. 1997. Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics* **76**: 281–307.
Wang H, Iglesias E, Wooldridge J. 2013. Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics* **172**: 77–89.
Wang X, Kockelman KM. 2009. Bayesian inference for ordered response data with a dynamic spatial-ordered probit model. *Journal of Regional Science* **49**: 877–913.
White LH. 2003. Currency competition and consumer-driven unification. *Cato Journal* **23**: 139–145.

DOI: 10.1002/jae