Liesenfeld, Roman; Richard, Jean-François; Vogler, Jan

Working Paper

Analysis of discrete dependent variable models with spatial correlation

Economics Working Paper, No. 2013-01

Provided in Cooperation with:
Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Liesenfeld, Roman; Richard, Jean-François; Vogler, Jan (2013) : Analysis of discrete dependent variable models with spatial correlation, Economics Working Paper, No. 2013-01, Kiel University, Department of Economics, Kiel

This Version is available at:
http://hdl.handle.net/10419/68466

C | A | U
Christian-Albrechts-Universität zu Kiel
Department of Economics
Economics Working Paper
No 2013-01

analysis of discrete dependent variable models with spatial correlation

by Roman Liesenfeld, Jean-François Richard, and Jan Vogler

issn 2193-2476

Analysis of Discrete Dependent Variable Models with Spatial Correlation

Roman Liesenfeld*
Institute of Statistics and Econometrics, Christian-Albrechts-Universität Kiel, Germany

Jean-François Richard
Department of Economics, University of Pittsburgh, USA

Jan Vogler
Institute of Statistics and Econometrics, Christian-Albrechts-Universität Kiel, Germany

(January 3, 2013)

Abstract

In this paper we consider ML estimation for a broad class of parameter-driven models for discrete dependent variables with spatial correlation. Under this class of models, which includes spatial discrete choice models, spatial Tobit models and spatial count data models, the dependent variable is driven by a latent stochastic state variable which is specified as a linear spatial regression model. The likelihood is a high-dimensional integral whose dimension depends on the sample size. For its evaluation we propose to use efficient importance sampling (EIS). The specific spatial EIS implementation we develop exploits the sparsity of the precision (or covariance) matrix of the errors in the reduced-form state equation typically encountered in spatial settings, which keeps numerically accurate EIS likelihood evaluation computationally feasible even for large sample sizes. The proposed ML approach based upon spatial EIS is illustrated with estimation of a spatial probit for US presidential voting decisions and spatial count data models (Poisson and Negbin) for firm location choices.

JEL classification: C15; C21; C25; D22; R12.

Keywords: Count data models; Discrete choice models; Firm location choice; Importance sampling; Monte Carlo integration; Spatial econometrics.

*Corresponding address: Institut für Statistik und Ökonometrie, Christian-Albrechts-Universität zu Kiel, Olshausenstraße 40-60, D-24118 Kiel, Germany. Tel.: +49(0)4318803810; fax: +49(0)4318807605. *E-mail address:* liesenfeld@stat-econ.uni-kiel.de (R. Liesenfeld)

1. Introduction

Modeling spatial dependence in linear regression frameworks for continuous dependent variables has received much attention over the last three decades and specifications of cross-sectional spatial correlation in form of a linear model with a spatial lag or spatial error term has become a standard tool in econometrics. Excellent up-to-date overviews can be found in the monographs of Ariba (2009) and LeSage and Pace (2009) as well as in the handbook edited by Anselin, et al. (2010).

Spatial models for discrete dependent variables such as counts or binary and multinomial outcomes have received less attention in the spatial econometrics literature. This might be explained by the fact that the introduction of spatial dependence in traditional discrete dependent variable models greatly complicates estimation and specification testing (see, e.g., Anselin, 1999). However, research efforts focusing on spatial models for discrete variables and developing appropriate methods for their estimation have markedly expanded in recent years as they are sorely needed for empirical studies in social sciences analyzing phenomena involving discrete outcomes of interacting agents.

Early spatial applications for binary outcomes are found in Case (1992) and McMillen (1992) who consider probit models where the underlying latent model is specified as a spatially dependent process so that the likelihood takes the form of an analytically intractable high-dimensional integral whose dimension is equal to the sample size n. For estimation Case (1992) relies on Maximum-Likelihood (ML) based upon a normalized version of the latent model accounting for the heteroscedasticity induced by spatial dependence but ignoring the spatial correlation in the binary outcomes, while McMillen (1992) proposes an ML approach based on an Expectation-Maximization (EM) algorithm avoiding direct evaluations of the high-dimensional likelihood integral. More recent studies propose to estimate spatial probit models by Bayesian Markov-Chain Monte-Carlo (MCMC) techniques (see, e.g., Smith and LeSage, 2004, and Franzese et al., 2010) or by simulated ML based upon the GHK importance sampling procedure developed by Geweke (1991), Hajivassiliou (1990) and Keane (1994) (see, e.g., Beron and Vijverberg, 2004). As illustrated by the Monte-Carlo (MC) study in Beron and Vijverberg (2004) ML parameter estimates under the standard implementation of GHK are quite accurate for moderate sample sizes n, but become very time consuming for large ns as it involves the Cholesky-factorization of $n \times n$ precision matrices. In order to overcome the high computational

costs of standard GHK for spatial probits, Pace and LeSage (2011) propose a GHK implementation, which exploits the fact that in typical spatial applications (with spatial units having only a small number of direct neighbors) those covariance (or precision) matrices are sparse. ML-GHK as well as Bayesian MCMC are also applied for the estimation of spatial probit models for multiple outcomes, such as the multinomial probit model (see Bolduc, 1997) and the ordered probit specification (see LeSage and Pace, 2009).

Existing spatial models for discrete count data can broadly be specified as observation-driven or parameter-driven. An early observation-driven approach to model spatially dependent count data is the auto Poisson model proposed by Besag (1974), where the expected value of the count variable is specified as a function of neighboring counts. Direct extensions of this spatial count data approach, which is amenable to standard ML estimation, are the specification of Kaiser and Cressie (1997) based upon a truncated (Winzorized) Poisson distribution and the semi-parametric Negative-Binomial (Negbin) specification proposed by Basile et al. (2010). Observation-driven count data models where the expected value of a count variable is specified as a function of spatially lagged expectations are analyzed by Hays and Franzese (2009) and Lambert et al. (2010). In order to estimate those models the former study rely on a nonlinear Least Squares (LS) approach and the Generalized Method of Moments (GMM), while Lambert et al. (2010) use a two-step limited information ML (LIML) procedure. For an overview of further spatial observation-driven count data models – see Lambert et al. (2010).

An early parameter-driven approach for spatially dependent count variables is found in Clayton and Kaldor (1987) who propose a Poisson model where the expected values of the count variable depends on a latent stochastic and spatially correlated process accounting for cross-sectional spatial random effects. Further closely related count data models with spatial random effects are proposed by LeSage et al. (2007) and Gschlößl and Czado (2008). While parameter-driven approaches to model count data are typically more flexible to include different forms of spatial dependence and often easier to analyze w.r.t. the spatial effects than the class of observation-driven models, they are more difficult to estimate. In fact, as it is the case for the spatial probit models, ML estimation of parameter-driven count-data models requires high-dimensional numerical integration. In order to

circumvent this integration, Clayton and Kaldor (1987) apply an empirical Bayes estimator based upon quadratic local approximations of the log-likelihood function, while LeSage et al. (2007) and Gschlößl and Czado (2008) rely on a Bayesian MCMC posterior analysis.

In the present paper, we consider a simulated ML approach for a broad class of parameter-driven models for discrete dependent variables with spatial correlation. This class of models is characterized by discrete conditional densities for the dependent variable given a latent stochastic state variable which is specified as a spatial linear regression model allowing for different forms of spatial dependence (spatially lagged states, spatial autoregressive errors, spatial moving average errors, or spatial random effects). This class includes spatial probits for binary and multiple outcomes as well as parameter-driven count data models with spatial effects.

For the high-dimensional interdependent integration required for ML estimation of this class of models we propose to use the Efficient Importance Sampling (EIS) procedure developed by Richard and Zhang (2007). EIS is a high-dimensional MC integration technique, which is based on simple LS approximations, designed to maximize the numerical accuracy of MC likelihood estimation. The specific implementation of EIS we develop follows the idea of Pace and LeSage (2011) in that it explicitly exploits the sparsity of the precision (or covariance) matrix of the errors in the reduced-form state equation typically found in spatial applications. This keeps numerically accurate EIS likelihood evaluation computationally feasible even for large sample sizes ($n = 5000^{+}$). When applied to spatial probit models, the proposed EIS approach covers the spatial GHK implementation introduced by Beron and Vijverberg (2004) as a special case obtained by omitting the LS optimization step. However, the applicability of EIS is, in contrast to GHK, not restricted to truncated high-dimensional Gaussian integrals defining the likelihood of spatial probits, so that it can be used for a much broader class of spatial models compared to GHK. A further attractive feature of spatial EIS is that it is highly generic in that its basic structure does not depend on a specific discrete conditional density of the dependent variable or the specific form of spatial dependence. Hence, changes in the spatial model being analyzed typically require only minor adjustments of a baseline spatial EIS implementation.

In order to illustrate the spatial EIS approach, we consider ML estimation for spatial probit, Poisson and Negbin models. In particular, ML based on spatial EIS is illustrated through a set of MC

experiments analyzing its statistical and numerical properties for a large sample size. Furthermore, we use our procedure to estimate a spatial probit model for US presidential voting decisions and spatial count data models for firm investment location decisions.

The remainder of this paper is structured as follows. The class of parameter-driven discrete dependent variable models with spatial correlation is introduced in Section 2. Section 3 describes ML estimation based on spatial EIS and discusses its numerical and statistical performance. Empirical applications to a spatial probit model for US presidential voting decisions and spatial count data models for firm location decisions are presented in Section 4. Section 5 concludes.

2. Discrete Dependent Variable Models with Spatial Correlation

The class of models with spatial correlation we consider is characterized by a discrete conditional density $f(y_i|\lambda_i)$ for the dependent variable y_i observed for spatial unit i $(i = 1, ..., n)$ given a latent stochastic state variable λ_i. Conditional on the states $\lambda = (\lambda_1, ..., \lambda_n)'$ the dependent variables $y = (y_1, ..., y_n)'$ are assumed to be stochastically independent. The specification assumed for the state variables is a Gaussian linear spatial regression model designed to account for spatial correlation in y and to include observable exogenous variables. Here, we focus on the two most popular variants of the linear regression model with spatial correlation. First, the spatial autoregressive lagged dependent variable model (SAL), which includes spatially lagged dependent variables on the r.h.s. of the regression equation, and second, the spatial autoregressive error model (SAE), which permits spatial autocorrelation in the errors of the regression model (see, e.g., LeSage and Pace, 2009). The reduced forms of the SAL and SAE model for the vector of states λ are given by

$$\lambda = m + u, \qquad u \sim \mathrm{N}(0, H^{-1}), \tag{1}$$

with

$$\text{SAL:} \qquad m = (I - \rho W)^{-1} X\beta, \qquad H = (1/\sigma^2)(I - \rho W)'(I - \rho W), \tag{2}$$

$$\text{SAE:} \qquad m = X\beta, \qquad H = (1/\sigma^2)(I - \rho W)'(I - \rho W), \tag{3}$$

where $u = (u_1, ..., u_n)'$ is a vector of Gaussian (reduced form) errors with precision matrix H, the vector $m = (m_1, ..., m_n)'$ represents the conditional mean of λ given the exogenous variables in the $(n \times \ell)$ matrix X with an ℓ-dimensional vector of regression coefficients β, the variable I denotes the identity matrix, and $\sigma > 0$ is a scale factor. The $(n \times n)$ matrix W contains non-stochastic and pre-determined spatial weights w_{ij} specifying the distance or contiguity relation between the spatial units i and j, and the scalar ρ represents the spatial autoregressive parameter measuring the importance of spatial correlation. Specifically, under the SAL model ρ measures the strength of global spatial feedback and feed-forward effects of shocks between locations, while under the SAE specification ρ controls spatial correlation in the error process. In typical applications the spatial weights are non-negative numbers with $w_{ij} > 0$ for spatial units which are direct neighbors and $w_{ij} = 0$ for others (by convention the diagonal elements w_{ii} are set to zero). Since spatial units can be assumed to have only a small number of nearby neighbors, the spatial weight matrix W as well as $(I - \rho W)$ and the precision matrix H are sparse matrices containing a large portion of zeros. A sufficient condition for invertibility of $(I - \rho W)$ is that $\rho \in (1/\zeta_{\min}, 1/\zeta_{\max})$, where $\zeta_{\min}$ and $\zeta_{\max}$ are the smallest (on the real line) and largest eigenvalue of W, respectively (see, LeSage and Pace, 2009, Chap. 4.3.2.). Before we discuss some discrete conditional densities for $f(y_i|\lambda_i)$, we shall mention that the SAL and SAE specifications (1)-(3) are examples and that other members of the family of linear spatial regression models including a combination of SAL and SAE or spatial moving average processes, can be easily accommodated by our estimation approach. Such alternatives are discussed, e.g., in LeSage and Pace (2009, Chap. 2).

2.1. Spatial Probit Models

If the observed discrete dependent variable y_i reflects a binary choice outcome such that $y_i \in \{0, 1\}$, the latent state λ_i can be interpreted as a utility or profit difference, with $y_i = 1$ if $\lambda_i \geq 0$ and $y_i = 0$ for $\lambda_i < 0$. The corresponding conditional distribution of y_i given the signal λ_i is a degenerated distribution with a probability density function (pdf) which can be written as

$$f(y_i|\lambda_i) = \mathbb{1}(z_i \lambda_i < 0), \quad \text{with} \quad z_i = 1 - 2y_i, \tag{4}$$

where $\mathbb{1}(\cdot)$ denotes an indicator function. Equation (4) together with Equations (1)-(3) define spatial probit models where for identification the scaling parameter σ is set equal to 1. Such spatial probit models have attracted considerable attention in the spatial-econometric literature (see, Case, 1992 and McMillen, 1992, for early contributions and LeSage and Pace, 2009, for a textbook treatment). Applications of spatial probit models include analyses of the voting decisions of US counties in presidential elections (Smith and LeSage, 2004), the diffusion of legislative term-limits among US states and the World-War I entry decisions (Franzese et al., 2010). Various extensions of binary spatial probit models can be framed in our set-up, including a spatial ordered probit model as discussed in LeSage and Pace (2009, Chap. 10.2) and the multinomial probit model with spatial correlation as proposed by Bolduc et al. (1997).

2.2. Spatial Count Data Models

In situations where y_i represents a count variable with $y_i \in \{0, 1, 2, 3, ...\}$, we can assume for the conditional distribution of $y_i|\lambda_i$ a Poisson distribution with an appropriate link function b, whereby the mean of the conditional Poisson distribution for $y_i|\lambda_i$, denoted by θ_i, is expressed as $b(\theta_i) = \lambda_i$. Using a log-link function $b(\theta_i) = \ln(\theta_i)$, which ensures the positivity of the mean θ_i without parametric restrictions on the parameters of the state model (β, ρ, σ^2), the conditional pdf for y_i is given by

$$f(y_i|\lambda_i) = \frac{1}{y_i!} \exp\{y_i\lambda_i - \exp\{\lambda_i\}\}. \tag{5}$$

The Poisson models defined by Equations (1)-(3) and (5) adapt naturally the class of parameter-driven models for time-series of count-data with serial dependence introduced by Zeger (1988) to cross-sectional count-data processes with spatial dependence. As such the parameter-driven spatial count-data specifications (1)-(3) and (5) are closely related to the Poisson models with latent spatial random effects proposed by LeSage et al. (2007) and Gschößl and Czado (2008) and differ from the observation-driven spatial Poisson model of Lambert et al. (2010), which specifies the mean of the Poisson distribution θ_i as a *measurable* function of spatially lagged θ_is and the regressors. In fact, the spatial Poisson model proposed by Lambert et al. (2010) represents a special case of the parameter-driven SAL Poisson model defined by Equations (1), (2), and (5) since it obtains by specifying the

process of the states λ in Equation (1) without the vector of errors u, which is tantamount to assume in Equation (2) that the variance parameter tends to zero ($\sigma^2 \to 0$). The spatial Poisson models proposed in the abovementioned studies are applied to analyses of knowledge spill-overs measured by patent citation (LeSage et al., 2007), the number of claims for policyholders of insurance companies (Gschößl and Czado, 2008), and the location choice of firms (Lambert et al., 2010).

Under the Poisson distribution (5) the dispersion index (defined as the ratio between the variance and the mean) equals one. However, count data often exhibit substantial overdispersion with a dispersion index being significantly larger than 1, which cannot be captured by a conditional Poisson distribution even if marginalization of the conditional mean can generate by itself an over-dispersed unconditional distribution. Hence, it might be important to replace the Poisson by a conditional distribution which allows for overdispersion, such as the negative binomial (Negbin). For a log-link function the conditional pdf of a Negbin distribution for y_i is given by

$$f(y_i|\lambda_i) = \frac{\Gamma(y_i + s)}{\Gamma(s)\Gamma(y_i + 1)} \Big(\frac{1}{1 + \exp\{\lambda_i\}/s}\Big)^s \Big(\frac{\exp\{\lambda_i\}}{\exp\{\lambda_i\} + s}\Big)^{y_i}, \tag{6}$$

where $\Gamma(\cdot)$ denotes the Gamma function. The over-dispersion is a decreasing function of the parameter $s > 0$ and the Poisson distribution in Equation (5) obtains as $s \to \infty$.

2.3. Spatial Tobit Models

Tobit models are used in situations where a subset of observations for the dependent variables y_i are censored with, say, $y_i = \lambda_i$ if $\lambda_i > 0$ and $y_i = 0$ if $\lambda_i \leq 0$ such that $y_i \in \mathbb{R}_+$ (see, e.g., Winkelmann and Boes, 2006). This implies that for the uncensored observations with $y_i = \lambda_i$, the state variable is directly observed, while for the censored ones with $y_i = 0$ the state remains latent. Let $\{y_i, i = 1, ..., n'\}$ be the subset of censored observations, and $\{y_i, i = n' + 1, ..., n\}$ the set of uncensored observations, then the conditional distribution of y_i given the state λ_i for the censored y_i's is a degenerated distribution with a pdf which can be written as

$$f(y_i|\lambda_i) = \mathbb{1}(y_i = 0) \cdot \mathbb{1}(\lambda_i < 0), \qquad i = 1, ..., n'. \tag{7}$$

This implies that the likelihood contribution of the censored variables has the same form as the likelihood of the spatial probits of Section 2.1. The corresponding probit probabilities are obtained under the conditional normal distribution for the unobserved states $(\lambda_1, ..., \lambda_{n'})$ given the observed ones $(\lambda_{n'+1}, ..., \lambda_n)$ obtained from Equation (1), while the likelihood contribution of the uncensored observations is that of a linear Gaussian spatial regression model. Such a spatial Tobit model with a SAL specification for λ is applied by LeSage and Pace (2009, Chap. 10) to model interregional origin-destination commuting flows using a data set where the flow variable is zero for 15 percent of the observations.

3. EIS Likelihood Evaluation

In order to estimate the parameters for specifications from the class of spatial discrete dependent variable models introduced in Section 2 we propose to use ML based upon the EIS Monte-Carlo integration technique introduced by Richard and Zhang (2007).

The likelihood for the spatial discrete dependent variable models obtains by integrating the conditional joint pdf of the dependent variables y and the errors of the states u given the regressors X with respect to u. Factorizing the joint Gaussian density for the errors $f(u)$ given by Equations (1)-(3) 'back-recursively' yields

$$f(u) = \prod_{i=1}^{n} f(u_i|u_{(i+1)}), \tag{8}$$

where $u_{(i)} = (u_i, ..., u_n)'$ with $u_{(n+1)} = \emptyset$ and $u_{(1)} = u$. It follows that the likelihood integral to be evaluated can be written as

$$L(\psi) = \int_{\mathbb{R}^n} \prod_{i=1}^{n} \varphi_i(u_{(i)}) du, \tag{9}$$

with

$$\varphi_i(u_{(i)}) = f(y_i|u_i, X) f(u_i|u_{(i+1)}), \tag{10}$$

where ψ regroups the model parameters and $f(y_i|u_i, X)$ denotes the conditional density of y_i given $(u_i,\ X)$ obtained from the conditional density $f(y_i|\lambda_i)$ (see Equations 4-7). Under a probit and a

Poisson model, e.g., we have

$$f(y_i|u_i, X) = \begin{cases} \mathbb{1}(z_i u_i \leq -z_i m_i) & \text{(Probit)} \\ \frac{1}{y_i!} \exp\{y_i(m_i + u_i) - \exp\{m_i + u_i\}\} & \text{(Poisson).} \end{cases} \tag{11}$$

The EIS algorithm facilitates numerically accurate MC-evaluations of likelihood integrals of the form as given by Equations (9) and (10), and relies upon a sequence of auxiliary regressions used to obtain recursively an IS density for u that closely mimics the target density kernel $\prod_i \varphi_i$ which needs to be integrated.

Based upon the factorization of the likelihood integrand used in Equations (9) and (10) we propose an EIS implementation for the spatial models under consideration which differs from existing EIS algorithms in two critical ways. First, the fact that we use the factorization of the likelihood integrand obtained from the back-recursive decomposition of the density for u given in Equation (8) requires that the EIS-algorithm has to be implemented such that it constructs the IS density forward-recursively. In contrast, existing EIS implementations, as that in Liesenfeld and Richard (2010), are back-recursively as they would be based on a decomposition of the likelihood integrand using the reverse factorization of the joint density for u given by $f(u) = \prod_{i=1}^{n} f(u_i|u_{i-1}, ..., u_1)$. The reason for using a forward rather than a backward-recursive version of EIS is that for applications to spatial probits the former represents a direct generalization of the spatial GHK implementations proposed by Beron and Vijverberg (2004) and Pace and LeSage (2011), so that we can compare them directly to EIS.

The second difference of the EIS implementation we propose for the spatial models introduced in Section 2 relative to existing ones, is that it relies on a parametrization of the target density kernel obtained by using the precision matrix of the latent states (H) rather than its covariance matrix ($\Sigma = H^{-1}$) as used by typical EIS applications (see, e.g., Richard and Zhang, 2007, and Liesenfeld and Richard, 2010). This EIS adapted to work with the precision matrix takes advantage of the fact that for the spatial specifications under consideration the precision matrix is, in contrast to the covariance matrix, a sparse matrix (see Equations 2 and 3). Since matrix operations for sparse matrices require, in general, less operation counts and reduced memory requirements than for dense

matrices, the sparsity of the precision matrix can be exploited to reduce significantly the computing time relative to an EIS implementation based on the dense covariance matrix. Of course, if we were considering spatial specifications where it is the covariance matrix Σ which is sparse rather than the precision matrix H, we would need to use the standard EIS implementation based on the covariance matrix in order to ensure computational efficiency. An example for such a specification is the spatial moving average process, where the covariance of the reduced form error u has the form $\Sigma = \sigma^2(I-\rho W)(I-\rho W)'$. (For a discussion of the computational advantages of the GHK exploiting the sparsity of precision (or covariance) matrices, see Pace and LeSage, 2011).

In Section 3.1 we outline the basic principle of EIS, and in Section 3.2 we illustrate its implementation for the spatial probit and Poisson models. Once the algorithm is implemented for a probit or a Poisson model, it can be easily adapted to other specifications of the class of discrete spatial models as it would require only minor modifications of a baseline spatial EIS implementation.

3.1. Basic Principle of EIS

EIS MC integration for the likelihood factorized as in Equations (9) and (10) is based on a back-recursive sequence of auxiliary importance sampling densities for u_i given $u_{(i+1)}$ of the form

$$g_i(u_i|u_{(i+1)};a_i) = \frac{k_i(u_{(i)};a_i)}{\chi_i(u_{(i+1)};a_i)}, \quad \text{with} \quad \chi_i(u_{(i+1)};a_i) = \int_{\mathbb{R}^1} k_i(u_{(i)};a_i)du_i, \tag{12}$$

for $i = 1, ..., n$, where $\{k_i(u_{(i)};a_i), a_i \in A_i\}$ is a preassigned class of parametric density kernels with (analytically) available integrating factors in u_i denoted by $\chi_i(u_{(i+1)};a_i)$. The likelihood integral in Equation (9) is then transformed into its importance sampling representation given by

$$L(\psi) = \chi_n(a_n) \cdot \int_{\mathbb{R}^n} \prod_{i=1}^{n} \frac{\varphi_i(u_{(i)}) \cdot \chi_{i-1}(u_{(i)};a_{i-1})}{k_i(u_{(i)};a_i)} \prod_{i=1}^{n} g_i(u_i|u_{(i+1)};a_i)du, \tag{13}$$

where $\chi_0(\cdot) \equiv 1$, and the corresponding IS MC likelihood estimate obtains as

$$\bar{L}(\psi) = \chi_n(a_n) \cdot \frac{1}{S}\sum_{s=1}^{S} \tilde{w}^{(s)}, \qquad \text{with} \qquad \tilde{w}^{(s)} = \prod_{i=1}^{n} \frac{\varphi_i(\tilde{u}_{(i)}^{(s)}) \cdot \chi_{i-1}(\tilde{u}_{(i)}^{(s)};a_{i-1})}{k_i(\tilde{u}_{(i)}^{(s)};a_i)}, \tag{14}$$

where $\{\tilde{u}^{(s)}\}_{s=1}^{S}$ denotes S draws from the IS density $g(u;a) = \prod_{i=1}^{n} g_i(u_i|u_{(i+1)};a_i)$ with $a = (a_1,...,a_n) \in A = \times_{i=1}^{n} A_i$. The objective of EIS is to select a parametrization $\hat{a} \in A$ that minimizes in Equation (14) for each i the MC variation of the ratios $\varphi_i\chi_{i-1}/k_i$ defining the IS weights $\tilde{w}^{(s)}$. According to the EIS principle of Richard and Zhang (2007) a near optimal solution to this minimization problem for the back-recursive importance sampling representation of the likelihood integrand in Equation (13) obtain as solutions of the following forward-recursive sequence of least squares (LS) problems:

$$\hat{a}_i = \arg\min_{a_i \in A_i} \sum_{s=1}^{S} \left\{ \ln\left[\varphi_i\big(\tilde{u}_{(i)}^{(s)}\big) \cdot \chi_{i-1}\big(\tilde{u}_{(i)}^{(s)};\hat{a}_{i-1}\big)\right] - \ln k_i\big(\tilde{u}_{(i)}^{(s)};a_i\big) \right\}^2, \quad i = 1,...,n, \tag{15}$$

where $\{\tilde{u}^{(s)}\}_{s=1}^{S}$ denotes S trajectories drawn from an initial IS sampling density $g(u;a^{(0)})$. (Under the reverse factorization of the target density kernel mentioned above, EIS would construct instead a sequence of auxiliary sampling densities g_i for u_i given $(u_{i-1},...,u_1)$ which would require a back-recursive version of the sequence of EIS regressions in Equation (15), see Richard and Zhang, 2007.) As initial samplers we can use densities obtained from local approximations to $\varphi_i\chi_{i-1}$ or densities associated with density kernels for $u_{(i)}$ included in φ_i. In the probit case we will use the GHK sampling densities and for the Poisson application an appropriate second-order Taylor-series approximation (TSA). As discussed in Richard and Zhang (2007), the sequence of EIS regressions (15) is iterated over $\hat{a} = (\hat{a}_1,...,\hat{a}_n)'$ until a fixed-point solution is obtained. In order to achieve a fixed-point convergence for $\hat{a}$, the trajectories $\{\tilde{u}^{(s)}\}_{s=1}^{S}$ generated under different values for a must be obtained by transformation of a set of Common Random Numbers (CRNs). Note that $\hat{a}$ is an implicit function of the model parameters ψ. Therefore, complete reruns of the EIS algorithm is required for each new ψ value. The use of CRNs across those reruns ensures continuity of the MC likelihood approximation $\bar{L}(\psi)$ with respect to ψ.

3.2. Spatial EIS Implementation

The implementation of the sequential EIS scheme described in Section 3.1 to the likelihood of a particular model begins with selecting the class of parametric density kernels $k_i(u_{(i)};a_i)$ ought to

approximate $\varphi_i(u_{(i)}) \cdot \chi_{i-1}(u_{(i)}; a_{i-1})$ as a function in $u_{(i)}$. Since the conditional pdf $f(u_i|u_{(i+1)})$ included in φ_i represents a Gaussian kernel in $u_{(i)}$, a natural choice for k_i is the class of Gaussian kernels. If k_i is selected as a Gaussian kernel in $u_{(i)}$ it follows that its integrating constant χ_i w.r.t. u_i as given in Equation (12) itself includes a Gaussian kernel in $u_{(i+1)}$. This together with the fact that the family of Gaussian distributions is closed under multiplications suggests to specify k_i as the product of the Gaussian kernels included in $\varphi_i\chi_{i-1}$ and an additional Gaussian kernel designed to best approximate the non-Gaussian terms in $\varphi_i\chi_{i-1}$ via the EIS regression in Equation (15). In the following we will provide the functional forms of those Gaussian EIS kernels k_i and their integrating factors χ_i which can be used for the EIS implementation for spatial probit and Poisson models. The corresponding closed-form expressions for k_i, χ_i, and the resulting EIS densities g_i are obtained from standard Gaussian algebra, which essentially consists of combining, regrouping and integrating Gaussian kernels, and are derived in a sequence of lemmas provided in the Appendix. The MATLAB codes for the spatial EIS implementations are available at http://www.stat-econ.uni-kiel.de.

Spatial Probit Models: Under the spatial probit models given by Equations (1)-(3), and (4), the function φ_i defines according to Equations (10) and (11) a (truncated) Gaussian kernel in $u_{(i)}$, such that we can select a Gaussian kernel for k_i which includes φ_i. Due to the presence of the indicator function in φ_i, the resulting integral χ_i of k_i in u_i given $u_{(i+1)}$ takes the form of a product of a Gaussian kernel in $u_{(i+1)}$ and a Gaussian cumulative distribution function (cdf) in a linear combination of the elements in $u_{(i+1)}$. In order to recursively derive its actual expression, we use the following parametrization for χ_{i-1}:

$$\chi_{i-1}(u_{(i)}, a_{i-1}) = \chi^*_{i-1}(u_{(i)}, a_{i-1}) \cdot \Phi(\omega_i), \qquad i-1 = 1, ..., n, \tag{16}$$

with

$$\chi^*_{i-1}(u_{(i)}; a_{i-1}) = \exp -\frac{1}{2}(u'_{(i)} P^*_{i-1} u_{(i)} - 2u'_{(i)} q^*_{i-1} + r^*_{i-1}), \tag{17}$$

$$\omega_i = c_{i-1} + d'_{i-1} u_{(i)}, \tag{18}$$

where $\Phi(\cdot)$ denotes the standardized normal cdf and $(P^*_{i-1}, q^*_{i-1}, r^*_{i-1}, c_{i-1}, d_{i-1})$ are functions of the coefficients characterizing the EIS kernel k_i defined below. Their actual closed-form expressions are derived in Lemma 3 in the Appendix. It follows from Equation (16) that the product to be approximated by k_i has the form $\varphi_i \chi_{i-1} = \varphi_i \chi^*_{i-1} \Phi(\omega_i)$, where $\Phi(\omega_i)$ is the sole non-Gaussian term. Accordingly, we define k_i as

$$k_i(u_{(i)}; a_i) = \mathbb{1}(z_i u_i \leq -z_i m_i) \cdot f(u_i | u_{(i+1)}) \cdot \chi^*_{i-1}(u_{(i)}; a_{i-1}) \cdot k^*_i(\omega_i; a_i), \tag{19}$$

where k^*_i denotes a Gaussian kernel in the linear function ω_i of $u_{(i)}$ designed to approximate $\Phi(\omega_i)$. It is parameterized as

$$k^*_i(\omega_i, a_i) = \exp -\frac{1}{2}(\alpha_i \omega_i^2 - 2\beta_i \omega_i + \kappa_i), \tag{20}$$

where $a_i = (\alpha_i, \beta_i, \kappa_i)'$ defines the EIS auxiliary parameter. Note that the conditional Gaussian density $f(u_i | u_{(i+1)})$ included in the EIS kernel (19) defines a Gaussian density kernel for $u_{(i)}$ whose parameters depend on the precision matrix H_i of the normal distribution for $u_{(i)}$, which obtains from the $\mathrm{N}(0, H^{-1})$ distribution of $u = u_{(1)}$ assumed in Equation (1). Since $\chi_0 \equiv 1$, the first factor to be approximated by k_1 is $\varphi_1 \chi_0 = \varphi_1$, so that we can select the initial EIS kernel as $k_1(u_{(1)}; \cdot) = \mathbb{1}(z_1 u_1 \leq -z_1 m_1) \cdot f(u_1 | u_{(2)})$ which generates a perfect fit and obtains by setting in Equation (19) the corresponding parameters to $P^*_0 = 0$, $q^*_0 = 0$, $r^*_0 = 0$, and $a_1 = 0$.

The selection of the EIS density kernel k_i as given by Equations (19) and (20) implies that all Gaussian factors common to k_i and $\varphi_i \chi_{i-1}$ cancel out in the EIS regression (15). It follows that this regression for unit i reduces to the simple LS regression

$$\ln \Phi(\tilde{\omega}_i^{(s)}) = -\frac{1}{2}\alpha_i [\tilde{\omega}_i^{(s)}]^2 + \beta_i \tilde{\omega}_i^{(s)} - \frac{1}{2}\kappa_i + \xi_i^{(s)}, \quad s = 1, ..., S, \tag{21}$$

where $\tilde{\omega}_i^{(s)} = c_{i-1} + d'_{i-1} \tilde{u}_{(i)}^{(s)}$, and $\xi_i^{(s)}$ is the implicit regression error term. The functional form of k_i obtained by the combination of the three Gaussian kernels in Equation (19) is given by

$$k_i(u_{(i)}; a_i) = \mathbb{1}(z_i u_i \leq -z_i m_i) \cdot \exp -\frac{1}{2}(u'_{(i)} P_i u_{(i)} - 2u'_{(i)} q_i + r_i + \ln(2\pi)), \tag{22}$$

where (P_i, q_i, r_i) are appropriate functions of the EIS auxiliary parameter a_i, the coefficients $(P^*_{i-1}$, q^*_{i-1}, r^*_{i-1}, c_{i-1}, $d_{i-1})$ characterizing χ_{i-1}, and the precision matrix H_i of the Gaussian density for $u_{(i)}$ defining the moments of $f(u_i|u_{(i+1)})$ in Equation (19). Their closed-form expressions are derived in Lemma 4 in the Appendix. The corresponding EIS sampling density $g_i(u_i|u_{(i+1)}, a_i)$ obtains by dividing the kernel k_i in Equation (22) by its integral in u_i and is a truncated normal density given in Lemma 5.

In summary, the sequential computation of the χ_{i-1}-parameters in Equations (16)-(18) and of the k_i-parameters in Equation (22) define the following EIS recursion: In recursion step i the coefficients $(P^*_{i-1}$, q^*_{i-1}, $r^*_{i-1})$ are combined via Equation (19) with the EIS parameter $\hat{a}_i = (\hat{\alpha}_i, \hat{\beta}_i, \hat{\kappa}_i)'$ from the EIS regression (21) and the precision matrix H_i of $u_{(i)}$ to obtain the coefficients (P_i, q_i, r_i) characterizing the EIS sampler g_i as well as the coefficients $(P^*_i$, q^*_i, r^*_i, c_i, $d_i)$ required for the next EIS recursion step $i+1$. This sequence of recursion steps for $i = 1, ..., n$ is initialized by setting $P^*_0 = 0$, $q^*_0 = 0$, $r^*_0 = 0$ and $\hat{a}_1 = 0$ and is iterated over the EIS parameters $\{\hat{a}_i\}$ upon the EIS regression until a fixed-point solution for the optimal EIS parameters is obtained. At convergence, the EIS-likelihood estimate is computed as given in Equation (14) with IS weights which simplify to

$$\tilde{w}^{(s)} = \prod_{i=2}^{n} \frac{\Phi(\tilde{\omega}_i^{(s)})}{\exp -\frac{1}{2}(\hat{\alpha}_i[\tilde{\omega}_i^{(s)}]^2 - 2\hat{\beta}_i\tilde{\omega}_i^{(s)} + \hat{\kappa}_i)}, \tag{23}$$

where $\{\{\tilde{u}^{(s)}_{(i)}\}_{i=1}^n\}_{s=1}^S$ used to compute $\{\{\tilde{\omega}_i^{(s)}\}_{i=2}^n\}_{s=1}^S$ are S independent trajectories from the sequence of EIS densities $\{g_i(u_i|u_{(i+1)}, \hat{a}_i)\}_{i=1}^n$.

As mentioned above, the spatial GHK implementation of Beron and Vijverberg (2004) and Pace and LeSage (2011) represents a special case of our proposed EIS procedure. It uses as IS density kernels

$$k_i(u_{(i)}; \cdot) = \varphi_i(u_{(i)}) = \mathbb{1}(z_i u_i \leq -z_i m_i) \cdot f(u_i|u_{(i+1)}), \tag{24}$$

with an integrating factor, which has the form

$$\chi_i(u_{(i+1)}; \cdot) = \Phi(c^*_i + d^{*\prime}_i u_{(i+1)}), \quad i = 1, ..., n-1, \qquad \chi_n(\cdot) = \Phi(c^*_n), \tag{25}$$

where c_i^* and d_i^* are functions of z_i, m_i, and of the mean and variance associated with the Gaussian density $f(u_i|u_{(i+1)})$. This Gaussian IS density kernel used by GHK and its integrating factor obtain by setting in the EIS recursion $\hat{a}_i = 0$ for $i = 1, ..., n$, from which follows that the GHK-procedure is numerically less efficient than EIS (see, also Liesenfeld and Richard, 2010). In the EIS application below, we use this GHK sampler as the initial sampling densities $g_i(u_i|u_{(i+1)}, a_i^{(0)})$ with $a_i^{(0)} = 0$ to initialize the fixed-point iterations over the EIS parameters $\hat{a} = \{\hat{a}_i\}$. It follows from Equations (24), and (25) that the IS weights used to obtain the GHK likelihood estimate according to Equation (14) are given by

$$\tilde{w}^{(s)} = \prod_{i=1}^{n-1} \Phi(c_i^* + d_i^{*\prime} \tilde{u}_{(i+1)}^{(s)}). \tag{26}$$

The relevance of the sparsity of the precision matrix $H = H_1$ for the vector of errors $u = u_{(1)}$, mentioned above, is through the fact that this sparsity translates into the complete sequence of precision matrices $\{H_i\}_{i=1}^n$ of the normal distributions for the $u_{(i)}$s obtained from the $N(0, H^{-1})$-distribution for u, which enter the sequence of matrix operations required for the EIS recursion. By exploiting this sparsity using sparse matrix functions available in software packages such as GAUSS or MATLAB, the computational time of running the EIS recursion for a large number of recursion steps ($n = 1000^+$) are dramatically lower than for the EIS recursion that obtains when using the dense covariance matrix $\Sigma = H^{-1}$. Furthermore, note that the computation of the matrices P_i and P_{i-1}^* in Equations (17) and (22) critically depends on the first row of the $(n - i + 1 \times n - i + 1)$-dimensional precision matrix H_i (see Lemmas 3 and 4). Hence, we can greatly reduce the total number of floating-point operations during the sparse matrix operations required for the EIS recursion by maximizing the sparsity of the first row of H_i, when H_i is a large matrix which is the case for low values of the index i. For this purpose we reorder, before running the EIS recursion, the rows and columns of H by using a symmetric approximate minimum degree permutation which concentrates the non-zero elements of H in its lower right corner (see Amestoy et al., 1996).

Spatial Poisson Models: The EIS implementation for the spatial Poisson models given by Equations (1)-(3), and (5) requires only minor modifications of that for the probit models. According to Equations (10) and (11), the function φ_i consists of the product of the Gaussian kernel in $u_{(i)}$ defined by $f(u_i|u_{(i+1)})$ and the non-Gaussian term given by the Poisson density $f(y_i|u_i, X)$. This

suggests to select k_i as a Gaussian kernel to include $f(u_i|u_{(i+1)})$ and a Gaussian approximation to $f(y_i|u_i, X)$ as a function in u_i. It follows that its integrating constant χ_i is itself a Gaussian kernel in $u_{(i+1)}$. Since there is no indicator function in φ_i and in the selected k_i, which was the case under the probit models, the functional form of χ_{i-1} for the Poisson models simplifies to

$$\chi_{i-1}(u_{(i)}, a_{i-1}) = \exp -\frac{1}{2}(u'_{(i)}P^*_{i-1}u_{(i)} - 2u'_{(i)}q^*_{i-1} + r^*_{i-1}), \tag{27}$$

where the parameters $(P^*_{i-1}, q^*_{i-1}, r^*_{i-1})$ have the same form as under the probit models – see Remark 1 to Lemma 3 of the Appendix. Since $\chi_0 \equiv 1$, the initial values are $P^*_0 = 0$, $q^*_0 = 0$, $r^*_0 = 0$. Accordingly, we can define the EIS density kernel as

$$k_i(u_{(i)}; a_i) = f(u_i|u_{(i+1)}) \cdot \chi_{i-1}(u_{(i)}; a_{i-1}) \cdot k^*_i(\lambda_i; a_i), \tag{28}$$

where k^*_i denotes a Gaussian kernel in the linear function $\lambda_i = m_i + u_i$ of u_i designed to approximate the sole non-Gaussian term in $\varphi_i\chi_{i-1}$ given by $f(y_i|u_i, X)$. It is parameterized as

$$k^*_i(\lambda_i, a_i) = \exp -\frac{1}{2}(\alpha_i\lambda_i^2 - 2\beta_i\lambda_i + \kappa_i). \tag{29}$$

Under this selection for k_i, the EIS regression as given in Equation (15) reduces to the LS regression

$$[-\ln(y_i!) + y_i\tilde{\lambda}^{(s)}_i - \exp\tilde{\lambda}^{(s)}_i] = -\frac{1}{2}\alpha_i[\tilde{\lambda}^{(s)}_i]^2 + \beta_i\tilde{\lambda}^{(s)}_i - \frac{1}{2}\kappa_i + \xi^{(s)}_i, \quad s = 1, ..., S. \tag{30}$$

The functional form of the Gaussian kernel k_i which obtains from Equations (27)-(29) is given by

$$k_i(u_{(i)}; a_i) = \exp -\frac{1}{2}(u'_{(i)}P_iu_{(i)} - 2u'_{(i)}q_i + r_i + \ln(2\pi)), \tag{31}$$

where the corresponding closed-form expressions for the parameters (P_i, q_i, r_i) are given in Lemma 4. The EIS density g_i for $u_i|u_{(i+1)}$ associated with the kernel in Equation (31) is an (untruncated) Gaussian density – see Remark 2 to Lemma 5.

All in all, the EIS-recursion for Poisson specifications is essentially the same as that for probit

models except for the fact that the functional form of the parameters (P_i, q_i, r_i) and the terms entering the EIS regressions need to be adjusted to changes of the functional form of $\varphi_i\chi_{i-1}$ when moving from the Probit to the Poisson specification. In order to initialize the fixed-point iteration over the EIS parameter $\hat{a}$, we use for $g_i(u_i|u_{(i+1)}, a_i^{(0)})$ the Gaussian density obtained by using in Equations (28)-(31) for $\ln k_i^*(\lambda_i, \cdot)$ a second-order TSA of $\ln f(y_i|u_i, X)$ around $u_i = 0$. The IS weights used to compute the EIS likelihood estimate for the Poisson models according to Equation (14) are given by

$$\tilde{w}^{(s)} = \prod_{i=1}^{n} \frac{\exp(y_i\tilde{\lambda}_i^{(s)} - \exp\tilde{\lambda}_i^{(s)})/y_i!}{\exp -\frac{1}{2}(\hat{\alpha}_i[\tilde{\lambda}_i^{(s)}]^2 - 2\hat{\beta}_i\tilde{\lambda}_i^{(s)} + \hat{\kappa}_i)}, \tag{32}$$

where $\{\{\tilde{u}_{(i)}^{(s)}\}_{i=1}^n\}_{s=1}^S$ used to compute $\{\{\lambda_i^{(s)}\}_{i=1}^n\}_{s=1}^S$ according to $\tilde{\lambda}_i^{(s)} = m_i + \tilde{u}_i^{(s)}$ are S independent trajectories from the sequence of EIS densities $\{g_i(u_i|u_{(i+1)}, \hat{a}_i)\}_{i=1}^n$.

If we replace the conditional Poisson density in Equation (5) by the conditional Negbin density in Equation (6), we only need to modify accordingly the dependent variable in the EIS regression (15), which requires to change only a few lines in the program code.

Since the joint EIS sampling density $g(u; \hat{a}) = \prod_{i=1}^n g_i(u_i|u_{(i+1)}, \hat{a}_i)$ for the spatial count-data models consists of untruncated conditional Gaussian densities g_i, the implementation of the spatial EIS for those models can be computationally simplified. Rather than constructing the components g_i of the joint EIS density $g(u; \hat{a})$ sequentially unit by unit i via the combination of the coefficients matrices $(P_{i-1}^*, q_{i-1}^*, r_{i-1}^*)$ of the integrating constants χ_{i-1} with the EIS parameter $\hat{a}_i = (\hat{\alpha}_i, \hat{\beta}_i, \hat{\kappa}_i)'$ from the EIS regression (30) and the precision matrix H_i of $u_{(i)}$, we can construct the joint EIS density $g(u; \hat{a})$ in a single step after running the sequence of independent EIS regressions. In fact, we can rewrite the joint EIS sampling density obtained from the sequential EIS-recursion according to Equation (12) together with Equations (28)-(29) as

$$g(u; a) = \frac{f(u) \cdot k^*(\lambda, a)}{\chi_n(a_n)}, \quad \text{with} \quad k^*(\lambda, a) = \prod_{i=1}^{n} k_i^*(\lambda_i, a_i), \tag{33}$$

where k_i^* is given by Equation (29), $f(u)$ represents the Gaussian density (8), and $\chi_n(a_n)$ is the integrating factor of the EIS density for the last unit $g_n(u_n; a_n) = k(u_n; a_n)/\chi_n(a_n)$. Hence, the joint Gaussian EIS density obtains directly in a single step from the Gaussian density kernel for u

given by $f(u)k^*(\lambda, a)$ with an integrating factor which obtains analytically from the corresponding multivariate Gaussian integral in u so that $\chi_n(a_n) = \int_{\mathbb{R}^n} f(u)k^*(\lambda, a)du$. This computational simplification avoiding the sequential combination of large matrices when constructing the sequence of the conditional EIS densities g_i increases the computational speed of the spatial EIS by the factor 45. Note that this simplification of the spatial EIS implementation is not feasible for the probit models since the conditional EIS densities g_i are truncated Normals which implies that the integrating factor for the corresponding joint EIS density $g(u; a)$ is an n-dimensional Gaussian cdf.

3.3. Monte Carlo Study

We proceed by presenting a MC study designed to explore the sampling distribution and numerical accuracy of the ML estimator based on EIS for spatial probit and Poisson models. In our design we consider for the probit as well as the Poisson model both, the SAL and the SAE specification given by Equations (2) and (3), where the regression function $(X\beta)$ for unit i is specified as

$$x_i'\beta = \beta_0 + \beta_1 x_i. \tag{34}$$

The regressors x_i are assumed to be i.i.d. uniform random variables on the interval $(-3, 4)$ for the probit, and on the interval $(0, 1)$ for the Poisson models. Following LeSage and Pace (2009, Chap. 4.11), we construct the spatial weight matrix W by simulating for each spatial unit i a pair of coordinates from a uniform-$(0, 1)$ distribution. The points associated with those coordinates are then transformed into a spatial weight matrix W assigning six neighbors to each unit by using a Delaunay triangulation carried out with the function *fasymneighbors2.m* in Kelly Pace's spatial statistics toolbox for MATLAB 2.0 (http://www.spatial-statistics.com). Finally, the spatial weight matrix is row-standardized.

The parameter values for the four model specifications are selected as follows. For the SAL and SAE probit model we fix the regression parameters at $(\beta_0, \beta_1) = (-1.5, 3)$ and for the SAL and SAE Poisson model at $(\beta_0, \beta_1, \sigma) = (-0.25, 0.8, 0.3)$, where σ is the scaling parameter in the precision matrix H (which is set equal to one for the probit models to ensure identification). For all four specifications we vary the degree of spatial correlation by taking $\rho = 0.75$ or $\rho = 0.85$, and

considered different sample sizes ranging from a moderate size of $n = 100$ to a fairly large size of $n = 5000$. Here we will represent only the results for $n = 5000$ which represents the most challenging situation for EIS.

Based on those data-generating processes (DGP) we construct the sampling distribution of the ML-EIS estimator, distinguishing between the statistical and the numerical sampling distribution (see, Richard and Zhang, 2007). The (conventional) statistical sampling distribution of the ML-EIS estimator is obtained by repeating the ML-EIS estimation for 50 different data sets using a single set of CRNs for EIS. In contrast, the numerical properties of the ML-EIS estimates as MC approximations to the true but infeasible ML estimates are analyzed by repeating the ML-EIS estimation for the first of the simulated data sets using 50 different sets of CRNs. For the probit specifications we compare the statistical and numerical properties of the ML-EIS estimator to those of the ML estimator based on the corresponding spatial GHK procedure described above. For the Possion models, however, no such direct ML based competitor is readily available.

The ML-EIS estimator is implemented using a simulation sample size of $S = 20$ and three fixed-point EIS iterations. One likelihood EIS evaluation for the probit models with a sample size of $n = 5000$ takes about 45 s on a Intel i7 Core computer with 2.67 GHz for a code written in MATLAB and that for the Poisson models using the simplified implementation of EIS only 1 s. Since the GHK, which we consider for the probit models, does not include a LS optimization step like that used by EIS, it is for a given simulation sample size S typically numerically less accurate but computationally faster than the EIS procedure (see, Liesenfeld and Richard, 2010). In order to make ML-GHK directly comparable to ML-EIS, we select for GHK a simulation sample size such that the computing time for a likelihood evaluation is nearly the same as for EIS with $S = 20$. This amounts to be $S = 500$. For all ML-EIS and ML-GHK estimation we use the BFGS optimizer.

Spatial probit models: The results of our MC experiments for the spatial probit models are summarized in Table 1. Reported are the statistical means, standard deviations and root mean squared errors (RMSE) around the true parameter values obtained from 50 simulated data sets all with one set of EIS CRNs. Also reported are the numerical means, standard deviations and RMSEs around the 'true' ML estimates obtained for the first simulated data set under 50 different sets of

CRNs. The 'true' ML values are computed as the ML-EIS estimates using a simulation sample size of $S = 1000$. The results for the statistical sampling distribution reported on the left hand side of Table 1 indicate that the ML-EIS estimator for the spatial probit models is virtually unbiased with estimates which are well centered around the true parameter values for both the SAL and the SAE specification and for all parameters. As for the numerical properties of the ML-EIS estimates as MC approximations to the true ML values, the results on the right hand side of Table 1 reveal that the numerical mean ML-EIS estimates are very close to the true ML values under all four DGPs. The associated MC numerical standard deviations are in all cases substantially smaller than the corresponding conventional statistical standard deviations of the ML-EIS estimator indicating a high numerical accuracy. These results reveal a performance of ML-EIS which is remarkably good, considering the fairly small simulation sample size of $S = 20$ used to estimate a 5000-dimensional likelihood integral.

Next, we note that the statistical sampling distribution of the ML-GHK estimator with $S = 500$ also reported in Table 1 shows significant biases with RMSEs which are for all parameters and all DGPs substantially larger than those associated with ML-EIS. The largest RMSEs of the ML-GHK estimator are observed for the parameters of the SAE specification with a strong spatial dependence ($\rho = 0.85$). We also note that the ML-GHK estimates exhibit fairly large numerical standard deviations and severe biases relative to the true ML estimates. At the same time, the numerical biases are in close accordance to their corresponding statistical biases, which suggest that the statistical biases observed for the ML-GHK estimates for the parameters are driven by numerical biases of the ML-GHK estimates as MC approximations to the true ML values.

A reason for this fairly poor performance of GHK for the spatial probit models with a sample size as large as $n = 5000$ is that the GHK IS density as defined in Equations (24) and (25) represents a poor global approximation to the very high dimensional target density kernel $\prod_{i=1}^{n} \varphi_i(u_{(i)})$. In order to verify this conjecture we analyzed for the GHK and EIS likelihood estimator their respective normalized IS weights given by $\bar{w}^{(s)} = \tilde{w}^{(s)} / \sum_{\tau=1}^{S} \tilde{w}^{(\tau)}$, where $\tilde{w}^{(s)}$ is given by Equations (23) and (26), respectively. The optimal IS density with a perfect global fit to the target density kernel would produce IS weights given by $\bar{w}^{(s)} = 1/S \ \forall s$. Figure 1 plots the histogram of the normalized IS

weights for the GHK and EIS procedure both implemented with $S = 500$ and applied to the SAE probit specification with $\rho = 0.85$ where we set the parameters in ψ equal to their true values. The histogram of the normalized GHK IS weights shows an extremely skewed distribution, indicating that only two of the 500 simulated u's carry weights which are effectively different from zero (one with a weight of 0.39 and the other with a weight of 0.61), while the remaining weights are essentially zero. This corroborates our conjecture that the GHK IS density poorly approximates the target density kernel. In contrast, the histogram of the EIS weights indicates a much better behaved EIS density. We also note that our GHK results are consistent with the findings of Pace and LeSage (2011) who report a similar distribution of the GHK IS weights for a spatial probit model applied to large data sets ($n = 100,000$). As a remedy they propose to replace the standard GHK likelihood estimator, which is according to Equations (14) and (26) given by $\bar{L} = \chi_n \cdot \frac{1}{S} \sum_{s=1}^{S} \prod_{i=1}^{n-1} \Phi(c_i^* + d_i^{*\prime} \tilde{u}_{(i+1)}^{(s)})$, with a likelihood MC approximation of the form $\hat{L} = \chi_n \cdot \prod_{i=1}^{n-1} [\frac{1}{S} \sum_{s=1}^{S} \Phi(c_i^* + d_i^{*\prime} \tilde{u}_{(i+1)}^{(s)})]$. However, the reliability of this MC likelihood approximation critically depends on the correlation structure of the sequence $\{\Phi(c_i^* + d_i^{*\prime} u_{(i+1)})\}_{i=1}^{n-1}$ under the GHK-IS density. In fact, if the elements in that sequence are mutually uncorrelated, $\hat{L}$ would produce unbiased likelihood estimates. However, this condition is hard to verify theoretically and would need to be checked empirically for each application.

Spatial Poisson models: The MC results for the spatial SAL and SAE Poisson model are summarized in Table 2. They show that for all four DGPs and for all parameters the statistical distribution of the ML-EIS estimates with a simulation sample size as low as $S = 20$ is well centered around the true parameter values. As to the numerical accuracy, the results obtained from the repeated parameter estimation using different sets of CRNs for one simulated data set indicate a high numerical accuracy of ML-EIS with numerical standard deviations which are significantly smaller than their statistical counterparts and (numerical) means which are very close to the true ML values.

All in all, the simulation results suggest that the ML-EIS estimator based on a simulation sample size of $S = 20$ is a reliable estimator for the parameters of spatial probit and Poisson models under an SAL and an SAE assumption for a sample size as large as $n = 5000$.

4. Empirical Applications

We now turn to simple empirical applications, where we use ML-EIS to estimate a spatial probit model for US presidential voting decisions (Section 4.1) and spatial count data models for firm investment location decisions (Section 4.2).

4.1 Spatial Probit for the 1996 Presidential Election

Following LeSage and Pace (2009, Chap. 10) we use an SAL probit specification as given by Equations (1), (2), and (4) to model the voting decisions from the 1996 US presidential election in each of the $n = 3110$ US counties, where $y_i = 1$ if in county i the Democratic candidate Clinton won the majority of votes and $y_i = 0$ for counties won by the Republican candidate Dole. Spatial dependence in the election outcomes at the county level may well be expected since voters located at similar regions in the US may tend to exhibit a similar voting behavior, so that the observed outcome of the election in one county is similar to the voting behavior observed in nearby counties. The spatial weight matrix W is constructed as for our MC study using the counties' geographical coordinates (latitude and longitude) which are transformed via a Delaunay triangulation to assign to each county six neighbors.

As explanatory variables in the vector x_i we use the log of the urban population and the following four education variables expressed as a proportion of the county population with degrees: the population with some years at college, the population with associate degrees, the population with college degrees, and the population with graduate or professional degrees. The data are taken from James LeSage's spatial econometric toolbox (http://www.spatial-econometrics.com).

The ML-EIS parameter estimates based upon a simulation sample size of $S = 20$ and 3 EIS fixed-point iterations are reported in Table 3 alongside with the ML-GHK estimates with $S = 500$. Also reported are the estimates of the average marginal effects which are computed as (see, e.g., Beron and Vijverberg, 2004)

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \text{prob}(y_i = 1|X, W)}{\partial x_i} = \frac{1}{n}\sum_{i=1}^{n}\phi(h_i^{\frac{1}{2}}[A_{i1}x_1'\beta + \cdots + A_{in}x_n'\beta]) \cdot h_i^{\frac{1}{2}} A_{ii}\beta, \tag{35}$$

where $\phi(\cdot)$ denotes the standardized normal density function, A_{ij} is the element (i,j) of the matrix $A = (I - \rho W)^{-1}$, and h_i is the precision of the marginal distribution of the error u_i obtained from the joint $\mathrm{N}(0, H^{-1})$-distribution for the vector u (see Equation 1). The marginal effects as given by Equation (35) jointly account for the direct impact of a change in $x_i'\beta$ on λ_i and its indirect impact caused by the spatial interdependence and captured by the matrix A.

The numerical standard deviations reported in Table 3 obtained from 50 ML-estimations under different CRNs confirm the result of the MC study that ML-EIS is numerically much more accurate than ML-GHK. The ML-EIS estimate for ρ indicates significant spatial dependence in the counties' voting behavior which is in line with the Bayesian MCMC results reported by LeSage and Pace (2009, Table 10.3) for the voting behavior in the 2000 US presidential election. We also note that the ML-EIS estimate for ρ is substantially larger than its ML-GHK counterpart. This appears to be consistent with the downward bias of the ML-GHK estimates of the spatial correlation parameter found in our MC study. Next, we see that the ML-EIS estimates for the β parameters and the associated marginal effects are generally smaller in absolute values than the corresponding ML-GHK estimates and lead to different conclusions with respect to the statistical significance at the 1% level for the impact of the variable *college degree* compared to the ML-GHK estimates. The only two variables for which ML-EIS indicates a significant effect are the variable *some college* with a negative impact on Clinton winning and the variable *graduate/professional degree* with a positive effect, which seems to suggest that the higher the education level of a county population the higher the probability of voting for Clinton. Finally, observe that ML-EIS estimation yields a maximized log-likelihood value which is substantially larger than that obtained by using ML-GHK. This reflects the fact that the GHK MC approximation of the (log)likelihood values for probit models tend to be downward biased (see, also Liesenfeld and Richard, 2010).

4.2 Spatial Count Data Models for Firm Location Choice

We now present an empirical application where we use the parameter driven spatial SAL count data specifications as defined in Section 2.2 to model firm location choices in the US manufacturing industry at the county level ($n = 3078$) in the lower 48 United States. A reason to expect spatial

dependence in the firm location choices is that site selection in a given region may simultaneously be determined by firm birth events in neighboring regions due to localization economies, and modeling such dependence would provide detailed information about regional linkages supporting local industry clustering and regional economic development (see Lambert et al., 2010). Our application is based upon the same dataset as that used in the study of Lambert et al. (2010), who propose to model firm location choices by an observation-driven spatial Poisson specification, which obtains from our SAL Poisson model given by Equations (1), (2), and (5) as $\sigma^2 \to 0$ (see, Section 2.2). In order to estimate their model they use a two-step limited information ML (LIML) approach.

The dependent variable y_i for county i is defined as the cumulative number of new single-unit start-up firms in the manufacturing industry observed between 2000 and 2004. Following Lambert et al. (2010), we construct the weight matrix W based on the Delaunay triangulation algorithm and assign to each county eight neighbors. The set of explanatory variables consists of location factors of the counties related to agglomeration economies, market structure, labor market, infrastructure, and the fiscal policy regime. The agglomeration variables are the manufacturing share of employment (*Msemp*), total establishment density (*Tfdens*), percentage of manufacturing establishments with less than 10 (*Pelt10*), and more than 100 employees (*Pemt100*). The market structure variables are median household income (*Mhhi*), population (*Pop*), and the share of workers in creative occupations (*Cclass*). Properties of the regional labor markets are measured by the average wage per job (*Awage*), net flows of wages per commuter (*Netflow*), Unemployment rate (*Uer*), percentage of adults with associate degree (*Pedas*). The variables characterizing the regional infrastructure are the public road density (*Proad*), interstate highway miles (*Interst*), public expenditures on highways per capita (*Hwypc*), the percentage of farmland to total county area (*Avland*). The fiscal policy variables are a tax business climate index (*Bci*), per capita government expenditures on education (*Educpc*). Also included in the set of regressors are dummy variables identifying counties as belonging to metropolitan (*Metro*) or micropolitan (*Micro*) areas.

The ML-EIS estimation results for the SAL Poisson model obtained for a simulation sample size $S = 20$ and 3 fixed point iterations are summarized on the left-hand side of Table 4. The ML-EIS parameter estimates for the impact of the counties' characteristics in x_i obtained under the SAL

Poisson specification are in line with the LIML estimates of Lambert et al. under the observation-driven Poisson model for most of the characteristics and confirm their conclusion that counties with agglomeration economies, labor availability, low costs for labor, availability of skilled labor, a business-friendly infrastructure and fiscal policy where more likely to attract new start-ups. Only for the variables *Pelt10*, *Uer*, *Awage*, *Netflow*, *Educpc*, *Proad* and *Hwypc* our estimate leads to different conclusions with respect to statistical significance compared to Lambert et al.'s estimate. The ML-EIS estimate of the spatial correlation parameter ρ is significantly larger than zero suggesting that start-ups in neighboring counties are important and its estimate of 0.51 is substantially larger than its estimate of 0.18 reported by Lambert et al.. Next, we note that the ML-EIS estimate of the variance parameter for the innovations of the state variable σ is statistically significantly larger than zero, which is the benchmark value expected under Lambert et al.'s observation-driven Poisson model. Hence, there is strong evidence in favor of the parameter-driven SAL model and against its observation-driven counterpart.

Since count data often exhibit over-dispersion, which cannot be fully captured by a conditional Poisson distribution, we also fitted the corresponding SAL Negbin model to the firm investment data, which allows for over-dispersion in the conditional distribution for $y_i|\lambda_i$. As explained in Section 3.2, the replacement of the Poisson density in Equation (5) by the Negbin density (6) only requires minor modifications of the ML-EIS algorithm implemented for the Poisson model. The ML-EIS results of the SAL Negbin model are reported on the right-hand side of Table 4. They show that the substitution of the Negbin for the Poisson assumption increases the maximized log-likelihood value by 395 which indicates a much better fit. Also the estimate of the Negbin parameter s of 2.90 reveals a significant deviation from the Poisson distribution which obtains as $s \to \infty$. Nevertheless, the estimates for the impact of counties' characteristics obtained under the Poisson and Negbin specification are typically quite similar. Only for the effect of the variables *Tfdens* and *Proad* the estimation under the Negbin model leads to changes with respect to statistical significance compared to the estimation under the Poisson model. Finally, we note that the move from the Poisson to the Negbin assumption has decreased the estimates for the ρ as well as the σ parameter, implying less spatial correlation and less variation in the innovation of the latent state variable λ_i. These differences suggest that

a significant part of the variation in the firm birth events across locations, which was attributed to shocks with global spatial feedback and feed-forward effects between locations under the Poisson model, is interpreted under the Negbin model as aspatial variation and attributed to conditional over-dispersion ($s < \infty$).

5. Conclusions

We have developed a generic simulation-based ML approach for parameter estimation in a broad class of parameter-driven models for discrete dependent variables with spatial correlation. Under this class of models the dependent variable is driven by a latent stochastic state variable which is assumed to follow a linear spatial regression model. The estimation approach is based on the Efficient Importance Sampling (EIS) technique for MC approximations of the likelihood, which is adapted to spatial settings. Computational efficiency of the spatial EIS implementation even for large sample sizes is due to the ability to exploit the sparsity of the precision (or covariance) matrix of the spatially correlated errors of the reduced form representation for the linear spatial regression model typically found in spatial cross-sectional applications with spatial units having only a small number of direct neighbors.

The spatial EIS is illustrated with ML estimation of spatial binary probit, Poisson and Negbin models with spatially lagged states (SAL) and spatial autoregressive errors (SAE). In a set of MC experiments we have shown that our proposed ML-EIS approach is numerically very accurate even for sample sizes as large as $n = 5000$ and produces reliable parameter estimates. With empirical applications to a spatial probit model for US presidential voting decisions and spatial count data models for firm investment location decisions we have illustrated the power of this approach.

Even if we have limited ourselves to illustrate spatial EIS with binary probits, Poisson and Negbin models with spatial SAL and SAE specifications, it is due to its very generic structure easily applicable to other spatial specifications for discrete and limited dependent variables including spatial ordered and multinomial probits, spatial Tobit models, and spatial binomial models as well as other forms of spatial dependence. This would require only minor adjustments in the spatial EIS implementations we used for the binary probit or the Poisson model. All in all, we believe that the spatial EIS

procedure provides a useful tool for the analysis of spatial models beyond the class of spatial linear Gaussian models for continuous dependent variables.

Acknowledgements

The authors thank Jason Brown for providing the firm location choice data set used in this paper. R. Liesenfeld and J. Vogler acknowledge support by the Deutsche Forschungsgemeinschaft (grant LI 901/3-1)

Appendix: Technical Details

This Appendix derives in a sequence of lemmas the recursive closed-form expressions for the EIS density kernels k_i defined in Section 3.2, their integrating factors χ_i and the resulting EIS sampling densities g_i.

Let $u \in \mathbb{R}^n$ be jointly normally distributed with mean vector $\mu = P^{-1}q$ and covariance matrix $\Sigma = P^{-1}$. Its density is denoted by

$$f(u) = f_N^n(u \mid P^{-1}q\ ,\ P^{-1}). \tag{A-1}$$

Let u, μ, q, Σ and P be partitioned conformably with one another into

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \tag{A-2}$$

with $u_i \in \mathbb{R}^{n_i}$, $i = 1, 2$.

Lemma 1. *The conditional density for u_1 given u_2 and the marginal density for u_2 obtained from the multivariate normal distribution in Equation (A-1) are given by*

$$f(u_1|u_2) = f_N^{n_1}(u_1 \mid P_{11}^{-1}[q_1 - P_{12}u_2]\ ,\ P_{11}^{-1}) \tag{A-3}$$

$$f(u_2) = f_N^{n_2}(u_2 \mid P_{22.1}^{-1}q_{2.1}\ ,\ P_{22.1}^{-1}) \tag{A-4}$$

with

$$P_{22.1} = P_{22} - P_{21}P_{11}^{-1}P_{12}, \quad q_{2.1} = q_2 - P_{21}P_{11}^{-1}q_1. \tag{A-5}$$

Proof. The partitioned inverse of P can be written as

$$\Sigma = P^{-1} = \begin{pmatrix} P_{11}^{-1} + P_{11}^{-1}P_{12}P_{22.1}^{-1}P_{21}P_{11}^{-1} & -P_{11}^{-1}P_{12}P_{22.1}^{-1} \\ -P_{22.1}^{-1}P_{21}P_{11}^{-1} & P_{22.1}^{-1} \end{pmatrix}. \tag{A-6}$$

It follows immediately that

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = P_{11}^{-1}, \qquad \Sigma_{12}\Sigma_{22}^{-1} = -P_{11}^{-1}P_{12}, \qquad \Sigma_{22} = P_{22.1}^{-1}, \tag{A-7}$$

$$\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(u_2 - \mu_2) = P_{11}^{-1}[q_1 - P_{12}u_2], \qquad \mu_2 = P_{22.1}^{-1}q_{2.1}. \qquad \square \tag{A-8}$$

Lemma 2. *Let $n_1 = 1$, $z \in \{-1, 1\}$, and $m \in \mathbb{R}^1$. Then*

$$\int_{\mathbb{R}^1} \mathbb{1}(zu_1 \leq -zm) f(u) du_1 = f_N^{n_2}(u_2 \mid P_{22.1}^{-1} q_{2.1} \ , \ P_{22.1}^{-1}) \cdot \Phi(c + d' u_2), \tag{A-9}$$

where $\Phi(\cdot)$ denotes the standardized normal cdf and

$$c = -z\sqrt{P_{11}}(m + \frac{q_1}{P_{11}}), \qquad d' = z\frac{P_{12}}{\sqrt{P_{11}}}. \tag{A-10}$$

Proof. Equation (A-9) follows from the partitioning in Lemma 1. Integration of $f(u_1|u_2)$ is carried out under the standardizing transformation

$$x = (zu_1 - zP_{11}^{-1}[q_1 - P_{12}u_2])/\sqrt{z^2 P_{11}^{-1}}, \tag{A-11}$$

where $z^2 = 1$. $\square$

Lemma 3. *Let P_i and q_i in the (truncated) Gaussian density kernel as given by Equation (22) be partitioned conformably with $u_{(i)} = (u_i, u'_{(i+1)})'$ into*

$$q_i = \begin{pmatrix} q_1^i \\ q_2^i \end{pmatrix}, \quad P_i = \begin{pmatrix} P_{11}^i & P_{12}^i \\ P_{21}^i & P_{22}^i \end{pmatrix}. \tag{A-12}$$

The integral of $k_i(u_{(i)}, .)$ in u_i is given by

$$\chi_i(u_{(i+1)}, .) = \chi_i^*(u_{(i+1)}, .) \cdot \Phi(\omega_{i+1}), \tag{A-13}$$

where

$$\chi_i^*(u_{(i+1)}, .) = \exp -\frac{1}{2}\{u'_{(i+1)} P_i^* u_{(i+1)} - 2u'_{(i+1)} q_i^* + r_i^*\}, \tag{A-14}$$

$$P_i^* = P_{22}^i - \frac{P_{21}^i P_{12}^i}{P_{11}^i}, \qquad q_i^* = q_2^i - \frac{P_{21}^i q_1^i}{P_{11}^i}, \qquad r_i^* = r_i - \frac{(q_1^i)^2}{P_{11}^i} + \ln|P_{11}^i|, \tag{A-15}$$

$$\omega_{i+1} = c_i + d_i' u_{(i+1)}, \qquad c_i = -z_i\sqrt{P_{11}^i}(m_i + \frac{q_1^i}{P_{11}^i}), \qquad d_i' = z_i\frac{P_{12}^i}{\sqrt{P_{11}^i}}. \tag{A-16}$$

Proof. We first rewrite k_i in Equation (22) as

$$k_i(u_{(i)},.) = \mathbb{1}(z_i u_i \leq -z_i m_i) \cdot f_N^{n-i+1}(u_{(i)} \mid P_i^{-1} q_i \ , \ P_i^{-1}) \tag{A-17}$$

$$\times \exp -\frac{1}{2}\{r_i - q_i' P_i^{-1} q_i + \ln|P_i| - (n-i)\ln(2\pi)\}.$$

Next, we apply Lemma 2 to obtain

$$\chi_i(u_{(i+1)},.) = f_N^{n-i}(u_{(i+1)} \mid (P_i^*)^{-1} q_i^* \ , \ (P_i^*)^{-1}) \cdot \Phi(c_i + d_i' u_{(i+1)}) \tag{A-18}$$

$$\times \exp -\frac{1}{2}\{r_i - q_i' P_i^{-1} q_i + \ln|P_i| - (n-i)\ln(2\pi)\},$$

wherefrom Equations (A-13)-(A-16) follow with

$$r_i^* = r_i - q_i' P_i^{-1} q_i + (q_i^*)'(P_i^*)^{-1} q_i^* + \ln|P_i| - \ln|P_i^*|, \tag{A-19}$$

which simplifies into the expression for r_i^* as given in Equation (A-15) by application of the partitioned inverse expression for P_i^{-1} as given by Equation (A-6) and by using the partitioned determinant for P_i, given by $|P_i| = |P_{11}^i| \cdot |P_{22}^i - P_{21}^i (P_{22}^i)^{-1} P_{12}^i|$. $\square$

Remark 1. From Lemmas 2 and 3, it follows immediately, that the density kernel k_i in Equation (22) without the indicator function $\mathbb{1}(\cdot)$ as given in Equation (31) has an integral in u_i, which is given by

$$\chi_i(u_{(i+1)},.) = \chi_i^*(u_{(i+1)},.) = \exp -\frac{1}{2}\{u_{(i+1)}' P_i^* u_{(i+1)} - 2u_{(i+1)}' q_i^* + r_i^*\}. \tag{A-20}$$

Lemma 4. *Let the covariance and precision matrix of the joint normal distribution for $u_{(i)} = (u_i, ..., u_n)'$ obtained from $u = (u_1, ..., u_n) \sim N(0, \Sigma)$ be denoted by Σ_i and H_i, respectively, and let H_i be partitioned conformably with $u_{(i)} = (u_i, u_{(i+1)}')'$ into*

$$H_i = \begin{pmatrix} H_{11}^i & H_{12}^i \\ H_{21}^i & H_{22}^i \end{pmatrix}. \tag{A-21}$$

Then the parameters (P_i, q_i, r_i) of the density kernel of the EIS sampling density $k_i(u_{(i)}, a_i)$ in Equation (22)

(probit case) are given by

$$P_i = H_i^* + P_{i-1}^* + \alpha_i d_{i-1} d_{i-1}', \tag{A-22}$$

$$q_i = q_{i-1}^* + (\beta_i - \alpha_i c_{i-1}) d_{i-1}, \tag{A-23}$$

$$r_i = r_{i-1}^* + \kappa_i + \alpha_i c_{i-1}^2 - 2\beta_i c_{i-1} - \ln|H_{11}^i|, \tag{A-24}$$

where

$$H_i^* = \begin{pmatrix} H_{11}^i & H_{12}^i \\ H_{21}^i & H_{21}^i H_{12}^i / H_{11}^i \end{pmatrix}, \qquad H_{i+1} = H_{22}^i - H_{21}^i H_{12}^i / H_{11}^i, \tag{A-25}$$

with $H_1 = \Sigma^{-1}$.

The parameters (P_i, q_i, r_i) *of the density kernel of the EIS sampling density* $k_i(u_{(i)}, a_i)$ *in Equation (31) (Poisson case) obtain as*

$$P_i = H_i^* + P_{i-1}^* + \alpha_i e_{(i)} e_{(i)}', \tag{A-26}$$

$$q_i = q_{i-1}^* + (\beta_i - \alpha_i m_i) e_{(i)}, \tag{A-27}$$

$$r_i = r_{i-1}^* + \kappa_i + \alpha_i m_i^2 - 2\beta_i m_i - \ln|H_{11}^i|, \tag{A-28}$$

where $e_{(i)} = (1, 0, ..., 0)'$.

Proof. It follows from Lemma 1 (with $q = \mu = 0$) that the conditional pdf of $u_i | u_{(i+1)}$ associated with $u_{(i)} \sim \mathrm{N}(0, \Sigma_i)$ is given by

$$f(u_i | u_{(i+1)}) = f_N^1\left(u_i \mid -\frac{H_{12}^i}{H_{11}^i} u_{(i+1)}\ ,\ \frac{1}{H_{11}^i}\right), \tag{A-29}$$

and that the precision matrix H_{i+1} for the distribution of $u_{(i+1)}$ obtains by the recursion given in Equation (A-25). The quadratic term in the exponent of the pdf (A-29) is

$$-\frac{1}{2} H_{11}^i \left(u_i + \frac{H_{12}^i}{H_{11}^i} u_{(i+1)}\right)^2 = -\frac{1}{2} u_{(i)}' H_i^* u_{(i)}, \tag{A-30}$$

wherefrom H_i^* as given in Equation (A-25) follows. Whence, the conditional pdf $f(u_i | u_{(i+1)})$ represents a density kernel of a joint normal distribution for $u_{(i)}$ given by

$$f(u_i | u_{(i+1)}) = \exp -\frac{1}{2}\{u_{(i)}' H_i^* u_{(i)} + \ln(2\pi) - \ln|H_{11}^i|\}. \tag{A-31}$$

Combining $f(u_i | u_{(i+1)})$ as given by Equation (A-31) with the remaining factors defining k_i according to Equa-

tions (19) and (28) yield the parameters as given in Equations (A-22)-(A-24) and (A-26)-(A-28), respectively. □

Lemma 5. *Let P_i and q_i in the (truncated) Gaussian density kernel k_i given by Equation (22) be partitioned as in Lemma 3. Then the density for $u_i|u_{(i+1)}$ obtained from the density kernel k_i is*

$$g_i(u_i|u_{(i+1)};.) = \frac{\mathbb{1}(z_i u_i \leq -z_i m_i) \cdot f_N^1(u_i|\mu_i^*, \sigma_i^{*2})}{\Phi(c_i + d_i' u_{(i+1)})}, \tag{A-32}$$

where

$$\mu_i^* = (q_1^i - P_{12}^i u_{(i+1)})/P_{11}^i, \qquad \sigma_i^{*2} = 1/P_{11}^i, \tag{A-33}$$

and (c_i, d_i) are given in Equation (A-16).

Proof. We first rewrite k_i in Equation (22) as

$$\begin{aligned} k_i(u_{(i)},.) = \mathbb{1}(z_i u_i \leq -z_i m_i) \cdot f_N^{n-i+1}(u_{(i)} \mid P_i^{-1} q_i \ , \ P_i^{-1}) \\ \times \exp -\frac{1}{2}\{r_i - q_i' P_i^{-1} q_i + \ln|P_i| - (n-i)\ln(2\pi)\}. \end{aligned} \tag{A-34}$$

Next, we apply Lemma 1 to $f_N^{n-i+1}(u_{(i)} \mid P_i^{-1} q_i \ , \ P_i^{-1})$, where P_i and q_i are partitioned as in Lemma 3 to obtain

$$\begin{aligned} k_i(u_{(i)},.) = \mathbb{1}(z_i u_i \leq -z_i m_i) \cdot f_N^1(u_i \mid (q_1^i - P_{12}^i u_{(i+1)})/P_{11}^i \ , \ 1/P_{11}^i) \\ \times \quad f_N^{n-i}(u_{(i+1)} \mid (P_i^*)^{-1} q_i^* \ , \ (P_i^*)^{-1}) \\ \times \quad \exp -\frac{1}{2}\{r_i - q_i' P_i^{-1} q_i + \ln|P_i| - (n-i)\ln(2\pi)\}, \end{aligned} \tag{A-35}$$

where P_i^* and q_i^* are given in Equation (A-15) of Lemma 3. Integrating the r.h.s. of Equation (A-35) in u_i and dividing it by the resulting integrating factor χ_i yields the truncated Gaussian density for $u_i|u_{(i+1)}$ given in Equation (A-32). □

Remark 2. From Lemma 5, it follows immediately, that the density for $u_i|u_{(i+1)}$ associated with the EIS-kernel k_i in Equation (22) without an indicator function as given by Equation (31) is $g_i(u_i|u_{(i+1)};\cdot) = f_N^1(u_i|\mu_i^*, \sigma_i^{*2})$.

References

Amestoy, P.R., Davis, T.A., Duff, I.S., 1996. An approximate minimum degree ordering algorithm. SIAM Journal on Matrix Analysis and Applications 17, 886-905.

Anselin, L., 1999. Spatial Econometrics. University of Texas at Dallas School of Social Sciences: Bruton Center.

Anselin, L., Florax, R.J.G.M., Rey, S.J., 2010. Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer.

Ariba, G., 2009. Spatial Econometrics. Springer.

Basile, R., Benfratello L., Castellani, D., 2010. Location determinants of greenfield foreign investments in the enlarged europe: evidence from a spatial autoregressive negative binomial additive model. Working paper No. 10, Department of Economics and Public Finance, University of Rome.

Beron, K.J., Vijverberg, W.P.M., 2004. Probit a spatial context: a Monte-Carlo analysis. In Anselin, L., Florax, R., Rey, S.J., Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer, 169–195.

Besag, J.B., 1974. Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society, Series B (Methodological) 36 No. 2, 192–236.

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.

Case, A.C., 1992. Neighborhood influence and technological change. Regional Science and Urban Economics 22, 491–508.

Clayton, D., Kaldor, J., 1987. Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. Biometrics, Vol. 43, No. 3, 671–681.

Franzese, R.J., Hays, J.C., Schaeffer, L.M., 2010. Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. Working paper, University of Michigan, Ann Arbor.

Geweke, J., 1991. Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities. University of Minnesota Department of Economics; Published in: Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface, 571–578.

Gschlößl, S., Czado, C., 2008. Does a Gibbs sampler approach to spatial Poisson regression models outperform a single site MH sampler? Computational Statistics and Data Analyis 52, 4184–4202.

Hajivassiliou, V., 1990. Smooth simulation estimation of panel data LDV models. Mimeo Yale University.

Hays, C.H., Franzese, R.J., 2009. A Comparison of the Small-Sample Properties of Several Estimators for Spatial-Lag Count Models, incomplete Working Paper.

Kaiser, M.S., Cressie, N., 1997. Modeling Poisson variables with positive spatial dependence. Statistic & Probability Letters 35, 423–432.

Keane, M., 1994. A computationally practical simulation estimator for panel data. Econometrica 62, 95–116.

Lambert, D. M., Brown, J.P., Florax, R.J.G.M., 2010. A two-step estimator for spatial lag model of counts: theory, small sample performance and application. Regional Science and Urban Economics 40, 241–252.

LeSage, J.P., Fischer, M.M., Scherngell, T., 2007. Knowledge spillovers across Europe: evidence from a Poisson spatial interaction model with spatial effects. Papers in Regional Science 86, 393–422.

LeSage, J.P., Pace, R.K., 2009. Introduction to Spatial Econometrics. CRC Press, Taylor and Francis Group.

Liesenfeld, R., Richard, J.-F., 2010. Efficient estimation of probit models with correlated errors. Journal of Econometrics 156, 367 – 376.

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

Pace, R.K., LeSage, J.P., 2011. Fast simulated maximum likelihood estimation of the spatial probit model capable of handling large samples. Working paper, Louisiana State University, Baton Rouge.

Richard, J.-F., Zhang, W., 2007. Efficient high-dimensional importance sampling. Journal of Econometrics 141, 1385–1411.

Smith, T.E., LeSage, J.P., 2004. A Bayesian probit model with spatial dependencies. In LeSage, J.P., Pace, R.K., Advances in Econometrics, Vol. 18, Spatial and Spatiotemporal Econometrics. Elsevier, 127–160.

Winkelmann, R., Boes, S., 2006. Analysis of Microdata. Springer.

Zeger, S.L., 1988. A regression model for time series of counts. Biometrika 75, 621–629.

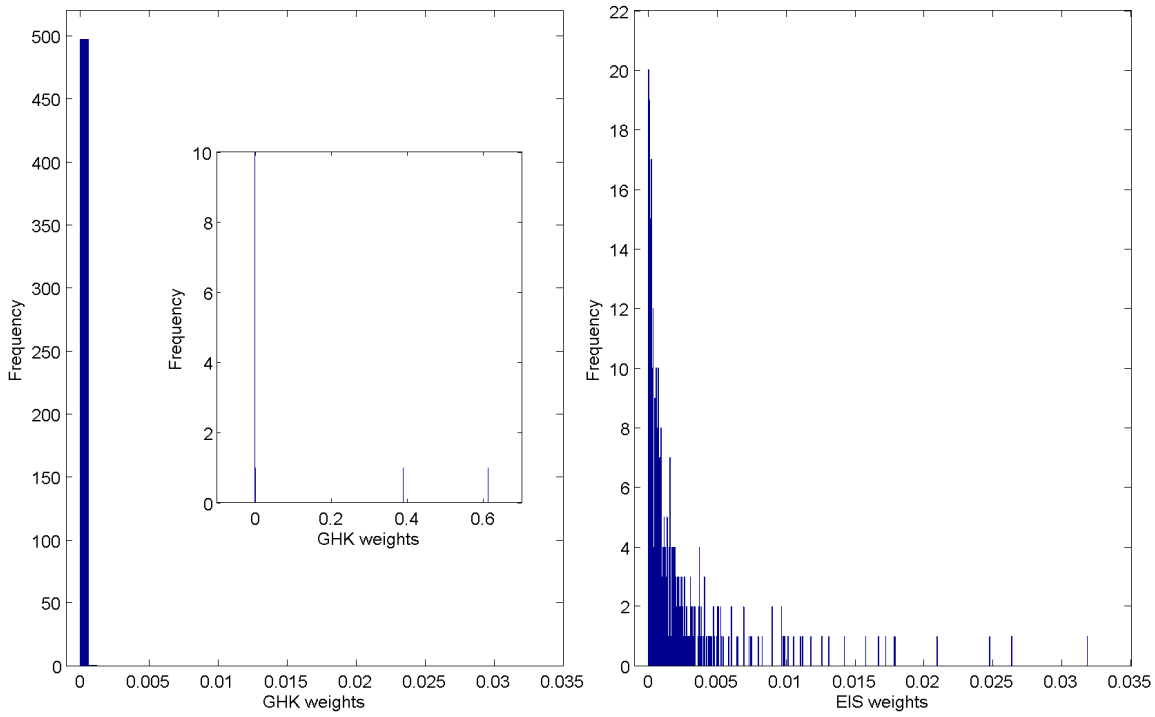

Figure 1. Histogram of the normalized IS weights $\tilde{w}^{(s)}/\sum_{\tau=1}^{S}\tilde{w}^{(\tau)}$ *of the GHK (plotted for different scales) and EIS procedure with* $S = 500$ *for the likelihood of the SAE probit model with* $\rho = 0.85$*. The model parameters* ψ *are set to their true values.*

Table 1. ML-EIS and ML-GHK for the SAL and SAE Probit Model

Parameter	True	Statistical properties		True ML	Numerical properties	
		GHK	EIS		GHK	EIS
SAL-probit model						
ρ	.750	.740	.750	.748	.739	.748
		(.006)	(.005)		(.003)	(.0001)
		[.012]	[.005]		[.009]	[.0001]
β_1	−1.500	−1.354	−1.498	−1.502	−1.371	−1.502
		(.061)	(.050)		(.025)	(.0006)
		[.158]	[.050]		[.134]	[.0007]
β_2	3.000	2.726	3.007	3.039	2.770	3.038
		(.121)	(.108)		(.049)	(.001)
		[.300]	[.108]		[.273]	[.001]
ρ	.850	.839	.850	0.851	.840	.851
		(.005)	(.003)		(.003)	(.0001)
		[.012]	[.003]		[.011]	[.0001]
β_1	−1.500	−1.243	−1.516	−1.489	−1.220	−1.488
		(.076)	(.070)		(.034)	(.002)
		[.268]	[.072]		[.272]	[.002]
β_2	3.000	2.488	3.037	2.989	2.444	2.986
		(.145)	(.138)		(.067)	(.003)
		[.532]	[.142]		[.549]	[.004]
SAE-probit model						
ρ	.750	.380	.748	.747	.364	.748
		(.033)	(.028)		(.024)	(.001)
		[.372]	[.028]		[.384]	[.001]
β_1	−1.500	−1.108	−1.480	−1.381	−1.027	−1.381
		(.068)	(.111)		(.011)	(.003)
		[.398]	[.113]		[.354]	[.003]
β_2	3.000	2.246	2.987	2.879	2.133	2.880
		(.096)	(.158)		(.015)	(.006)
		[.760]	[.159]		[.746]	[.006]
ρ	.850	.482	0.848	.843	.513	.843
		(.033)	(.016)		(.021)	(.0009)
		[.369]	[.016]		[.331]	[.0009]
β_1	−1.500	−.894	−1.483	−1.444	−.898	−1.443
		(.076)	(.140)		(.018)	(.004)
		[.610]	[.141]		[.546]	[.004]
β_2	3.000	1.829	3.013	2.845	1.759	2.844
		(.067)	(.176)		(.019)	(.007)
		[1.173]	[.177]		[1.086]	[.008]

NOTE: The reported numbers for ML-EIS are mean, standard deviation (in parentheses) and RMSE (in brackets). The simulation sample size for EIS is $S = 20$ and for GHK $S = 500$. The true ML values are the ML-EIS estimates based on $S = 1000$.

Table 2. ML-EIS for the SAL and SAE Poisson Model

Parameter	True	Statistical properties	True ML	Numerical properties
SAL-Poisson model				
ρ	.750	.750	.756	.756
		(.019)		(.0007)
		[.019]		[.0008]
β_1	−.250	−.249	−.254	−.254
		(.017)		(.0001)
		[.017]		[.0001]
β_2	.800	.800	.801	.800
		(.032)		(.0006)
		[.032]		[.0007]
σ	.300	.294	.286	.285
		(.015)		(.0009)
		[.016]		[.0015]
ρ	.850	.854	.841	.842
		(.018)		(.002)
		[.018]		[.002]
β_1	−.250	−.241	−.188	−.186
		(.022)		(.003)
		[.024]		[.004]
β_2	.800	.786	.737	.735
		(.034)		(.003)
		[.036]		[.004]
σ	.300	.291	.285	.284
		(.014)		(.003)
		[.016]		[.003]
SAE-Poisson model				
ρ	.750	.747	.741	.741
		(.031)		(.002)
		[.031]		[.002]
β_1	−.250	−.244	−.237	−.236
		(.038)		(.0006)
		[.038]		[.0007]
β_2	.800	.794	.760	.760
		(.048)		(.0002)
		[.048]		[.0002]
σ	.300	.301	.304	.303
		(.026)		(.002)
		[.026]		[.002]
ρ	.850	.855	.849	.849
		(.024)		(.002)
		[.025]		[.002]
β_1	−.250	−.223	−.276	−0.272
		(.065)		(.006)
		[.070]		[.008]
β_2	.800	.794	.830	.830
		(.043)		(.0005)
		[.043]		[.0005]
σ	.300	.289	.304	.302
		(.033)		(.003)
		[.035]		[.004]

NOTE: The reported numbers for ML-EIS are mean, standard deviation (in parentheses) and RMSE (in brackets). The simulation sample size for EIS is $S = 20$. The true ML values are the ML-EIS estimates based on $S = 1000$.

Table 3. ML-EIS and ML-GHK results for the SAL Probit Model for the 1996 U.S. Presidential Election

Variable	Parameters		Marg. Eff.	
	GHK	EIS	GHK	EIS
Constant	.634*	.597*		
	(.120)	(.122)		
	[.033]	[.0001]		
Urban population	8.445	4.894	3.131	1.745
	(5.494)	(5.635)	(2.058)	(2.030)
	[1.386]	[.0003]	[.512]	[.0002]
Some college	−3.192*	−2.792*	−1.183*	−.996*
	(.470)	(.479)	(.167)	(.165)
	[.122]	[.0005]	[.044]	[.0002]
Associate degree	1.118	.901	.414	.321
	(.893)	(.909)	(.331)	(.323)
	[.241]	[.0005]	[.089]	[.0002]
College degree	−2.056*	−1.885	−.762*	−.672
	(.783)	(.805)	(.299)	(.295)
	[.171]	[.0006]	[.063]	[.0002]
Graduate/professional degree	5.185*	4.610*	1.922*	1.644*
	(1.371)	(1.405)	(.507)	(.450)
	[.317]	[.0009]	[.115]	[.0004]
Spatial lag (ρ)	.508*	.633*		
	(.024)	(.025)		
	[.009]	[.0001]		
Log-likelihood	−1947.3	−1910.2		

NOTE: The reported numbers are mean ML-EIS and ML-GHK estimates for the parameters and the marginal effects, the asymptotic (statistical) standard deviation (in parentheses) and the numerical standard deviation (in brackets). The asymptotic standard deviation for the ML parameter estimates are obtained from a numerical approximation to the Hessian and those for the marginal effects are obtained as MC approximation using 2000 draws from the asymptotic distribution of the ML estimators. The simulation sample size for EIS is $S = 20$ and for GHK $S = 500$; $*$ statistically significant at the 1% level.

Table 4. ML-EIS results for the SAL Poisson and Negbin Model for Firm Location Choice

Variable	Poisson			Negbin		
Constant	−1.113*	(.130)	[.044]	−.855*	(0.139)	[.0002]
Msemp	.032*	(.002)	[.0005]	.042*	(0.002)	[.00003]
Pelt10	.006*	(.0009)	[.0002]	.005*	(0.001)	[.00001]
Pemt100	−.014*	(.002)	[.0004]	−.015*	(0.002)	[.00002]
Tfdens	−.003	(.011)	[.004]	−.056*	(0.018)	[.0001]
Mhhi	−.250	(.278)	[.134]	.517	(0.369)	[.005]
Pop	.006*	(.0002)	[.0001]	.013*	(0.001)	[.00002]
Cclass	.059*	(.003)	[.001]	.084*	(0.004)	[.00001]
Uer	.030*	(.006)	[.003]	.047*	(0.010)	[.0001]
Pedas	.030*	(.006)	[.002]	.035*	(0.007)	[.00002]
Awage	−.025*	(.003)	[.001]	−.035*	(0.005)	[.00003]
Netflow	−.004*	(.0005)	[.0002]	−.016*	(0.001)	[.00001]
Proad	.020	(.010)	[.004]	.053*	(0.012)	[.00003]
Interst	.003*	(.0004)	[.0002]	.004*	(0.0006)	[.000003]
Avland	−.003*	(.0005)	[.0002]	−.004*	(0.0006)	[.00001]
Bci	.095*	(.009)	[.004]	.028	(0.011)	[.00002]
Educpc	.006*	(.001)	[.0005]	.004*	(0.001)	[.00001]
Hwypc	−.044*	(.007)	[.003]	−.032*	(0.008)	[.00006]
Metro	1.025*	(.037)	[.009]	.823*	(0.040)	[.0002]
Micro	.683*	(.038)	[.006]	.541*	(0.036)	[.00003]
Spatial lag (ρ)	.514*	(.018)	[.008]	.321*	(0.024)	[.0006]
σ	.697*	(.013)	[.002]	.261*	(0.042)	[.002]
s				2.902*	(0.204)	[.008]
Log likelihood	-10,720			-10,325		

NOTE: The reported numbers are mean ML-EIS estimates for the parameters, the asymptotic (statistical) standard deviation (in parentheses) and the numerical standard deviation (in brackets). The asymptotic standard deviation for the ML parameter estimates are obtained from a numerical approximation to the Hessian. The simulation sample size for EIS is $S = 20$; $*$ statistically significant at the 1% level.