Likelihood Evaluation of High-Dimensional Spatial Latent Gaussian Models with Non-Gaussian Response Variables

Roman Liesenfeld*
Institut für Ökonometrie and Statistik, Universität Köln, Germany

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

Jan Vogler
Institut für Ökonometrie and Statistik, Universität Köln, Germany

(February 25, 2015)

Abstract

We propose a generic algorithm for numerically accurate likelihood evaluation of a broad class of spatial models characterized by a high-dimensional latent Gaussian process and non-Gaussian response variables. The class of models under consideration includes specifications for discrete choices, event counts and limited dependent variables (truncation, censoring, and sample selection) among others. Our algorithm relies upon a novel implementation of Efficient Importance Sampling (EIS) specifically designed to exploit typical sparsity of high-dimensional spatial precision (or covariance) matrices. It is numerically very accurate and computationally feasible even for very high-dimensional latent processes. Thus Maximum Likelihood (ML) estimation of high-dimensional non-Gaussian spatial models, hitherto considered to be computationally prohibitive, becomes feasible. We illustrate our approach with ML 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 Ökonometrie and Statistik, Universität Köln, Universitätsstr. 22a, D-50937 Köln, Germany. Tel.: +49(0)221-470-2813; fax: +49(0)221-470-5074. *E-mail address:* liesenfeld@statistik.uni-koeln.de (R. Liesenfeld)

1. Introduction

Models that incorporate spatial dependence have received much attention over the last two decades. Foremost, there has been an explosive growth in the size of spatial databases across fields (including economics, social sciences, ecology, epidemiology and transportation, among others). Recent overviews including extensive lists of references can be found in the monographs of Ariba (2006) and LeSage and Pace (2009), as well as in the handbooks edited by Anselin, et al. (2010) and Fischer and Nijkamp (2014).

In this paper, we propose a generic Efficient Importance Sampling (EIS) procedure for accurate numerical evaluation of likelihood functions for a broad class of high-dimensional spatial 'latent Gaussian models', in the terminology of Rue et al. (2009). These are models in which the observable response variables are linked to a latent (state) Gaussian process that is spatially correlated. It follows that likelihood evaluation requires integration of a potentially high-dimensional interdependent state vector. When the response variables are themselves normally distributed with means that depend linearly on the state variables (and variances independent of the latter) integration of the state vector is carried out analytically.

However, response variables are frequently non-Gaussian as they represent discrete choices, event counts, censored and truncated data in which case likelihood evaluation requires numerical integration that becomes 'prohibitively difficult' (Wang et al., 2013), poses 'challenging problems' (Pace, 2014), and/or 'sometimes impossible computational burdens' (Wang, 2014) for high-dimensional state processes. As discussed further in Section 2 below, a variety of alternative inference techniques have been applied to specific sub-classes of spatial models such as probit models. See also Wang (2014) for a recent survey of methods that can be applied to spatial models for limited and censored dependent variables.

In this paper we propose a novel, generic and flexible procedure to compute likelihoods for a wide range of spatial dependent variable models for non-Gaussian responses. Our procedure is computationally feasible and numerically accurate even for very high-dimensional spatial applications. It consists of an original and powerful combination of two existing techniques. One is the EIS procedure that was initially proposed by Richard and Zhang (2007) and has since been successfully

applied to numerous time-series applications, including high-dimensional ones. See, e.g., Liesenfeld and Richard (2003, 2006, 2008), Bauwens and Galli (2009), Pastorello and Rossi (2010), Liesenfeld et al. (2010), Jung et al. (2011), Hafner and Manner (2012), Kleppe and Skaug (2012) or Scharth and Kohn (2013). The other technique regroups tools that allow for fast computations on large sparse matrices, such as the reverse Cuthill-McKee algorithm (Alan and Liu, 1981) or the approximate minimum degree ordering (Amestoy et al., 1996; Pace and Barry, 1997; LeSage and Pace, 2009; Pace and LeSage, 2011).

While combining EIS and sparse matrix algebra is conceptually fairly simple, it does require a novel implementation of EIS for spatial applications. EIS relies upon full sequential factorizations of the likelihood integral. It was developed specifically for time-series models with low-order Markovian (time-sequential) specifications for the latent state process, which directly translate into a likelihood factorization consisting of parsimoniously parameterized conditional densities. In this case EIS relies upon recursive sequence of operations on low-dimensional (covariance) matrices.

However, spatial models share a critical characteristic, that prevent using standard EIS. They do not have a sequential Markovian structure, so that any likelihood factorization inherently consist of a sequence of conditional densities that depend on large (auxiliary) parameter vectors in high-dimensional applications. In such cases, standard EIS becomes rapidly computationally prohibitive as sample size (n) increases. However, spatial units typically have small numbers of direct neighbors so that spatial precision matrices are generally sparse with large proportions of zero entries. Thus, the key to computationally feasible high-dimensional spatial EIS lies in sequential factorizations that operate on precision matrices (instead of covariance matrices) and, foremost, preserve sparsity through the entire sequence. In particular it requires an appropriate and automated (re)ordering of the data.

The likelihood factorization we propose resembles that used by Pace and LeSage (2011) to construct a fast GHK (Geweke, 1991; Hajivasiliou, 1990; Keane, 1994) importance sampling procedure for high-dimensional spatial probit models. Their approach relies upon a sparse Cholesky factorization of the precision matrix as needed to apply GHK. Our approach is different in that it relies upon a direct sparse sequential factorization of the likelihood function itself. By this we avoid computation-

ally costly matrix operations on the typically dense *inverse* of the high-dimensional Cholesky factor, which would be required for an EIS implementation using the likelihood factorization of Pace and LeSage. Nor is our approach restricted to probit applications. Last but not least, it is well known that the numerical accuracy of GHK rapidly deteriorates as the sample size n increases (see, e.g., Lee, 1997). In Section 4.1 below we illustrate the fact that our ML-EIS estimates are numerically significantly more accurate than their GHK counterparts.

The only EIS application to a model without a Markovian structure we are aware of is that by Liesenfeld and Richard (2010), where the authors analyze a low-dimensional ($n = 20$) probit model with a dense covariance matrix for the latent process. They do so by 'brute-force' Cholesky factorization of the covariance matrix resulting in an EIS implementation whose extension to high-dimensional spatial applications would be computationally prohibitive, requiring $O(n^3)$ operations. In sharp contrast, the combination of EIS and sparse matrix algebra we propose requires computing times that are $O(n^\delta)$ with $\delta \ll 3$, keeping it computationally feasible even for high dimensions ($n = 5000^+$). Foremost, it is generic in that it allows for considerable flexibility in the specification of the response process since adjustments for different response processes leave the core algorithm unchanged. As highlighted by a Monte-Carlo study for spatial probit and Poisson models in Section 4 below, our spatial EIS algorithm is very fast and remains numerically accurate even for very large n.

The remainder of the paper is organized as follows. The class of spatial models we consider is introduced in Section 2, with specific sub-classes and brief surveys of existing estimation methods. Our sparse spatial EIS algorithm is presented in Section 3, where we also analyze its performance by a Monte-Carlo study. Section 4 presents two empirical applications: a spatial probit model for US presidential voting decisions and spatial count data models for firm location decisions. Section 5 concludes.

2. Spatial Dependent Variable Models

2.1 Baseline model

Let $y = (y_1, \ldots, y_n)' \in \times_{i=1}^{n} S_i$ denote a vector of (non-Gaussian) observable response variables with support S_i, that are assumed to be mutually independent conditionally on a spatial Gaussian latent vector $\lambda = (\lambda_1, \ldots, \lambda_n)'$. Let X denote a $(n \times \ell)$ matrix of observable exogenous variables. The class of models we consider here is characterized by the following two assumptions:

$$\text{A1:} \qquad f(y|\lambda, X) \equiv \prod_{i=1}^{n} f(y_i|\lambda_i), \tag{1}$$

$$\text{A2:} \qquad \lambda|X \sim \mathrm{N}_n(m, H^{-1}), \tag{2}$$

where m denotes the conditional mean of λ given X and H its sparse precision matrix. $\mathrm{N}_n(\cdot, \cdot)$ represents the n-dimensional Gaussian distribution of the latent variable $\lambda|X$.

The spatial EIS algorithm we propose in Section 3 below allows for considerable flexibility in the specification of $f(y_i|\lambda_i)$, m, and H (subject to sparsity of H). Nevertheless, for illustration purposes, we shall pay special attention to some of the specifications commonly discussed in the spatial literature, whereby[1]

$$\text{A3:} \qquad H = (1/\sigma^2)(I_n - \rho W)'(I_n - \rho W), \quad \text{and} \tag{3}$$

$$\text{A4.1:} \qquad m = (I_n - \rho W)^{-1} X\beta, \quad \text{or} \tag{4}$$

$$\text{A4.2:} \qquad m = X\beta,$$

where $\sigma > 0$ denotes a scale factor and I_n the n-dimensional identity matrix. The $(n \times n)$ matrix W represents distance or contiguity relations across units, and the scalar ρ measures the overall intensity of spatial correlation. By convention, the diagonal elements of W are set equal to zero. In typical spatial settings, $w_{ij} > 0$ only for direct neighbors and $w_{ij} = 0$, otherwise. It follows that W and H

[1] The terminology in the spatial literature is somewhat ambiguous. Under A3 and A4.1, the model is referred to as 'Spatial Autoregressive' (SAR) or 'Spatial Autoregressive Lagged dependent' (SAL). Under A3 and A4.2 it is referred to as 'Spatial Autoregressive Error' (SAE) or 'Spatial Error Model' (SEM). See, e.g., Anselin (1999) or LeSage and Pace (2009, Chapter 2), where alternative specifications for m and H are also discussed.

are sparse matrices with increasingly large proportions of zeros as n increases. Conditions for the invertibility of $(I - \rho W)$ are discussed in LeSage and Pace (2009, Section 4.3). For matrices with real eigenvalues, a sufficient condition is that $\rho \in (1/\zeta_{\max}, 1/\zeta_{\min})$, where $\zeta_{\min}$ and $\zeta_{\max}$ denote the extreme eigenvalues of W. We note that the matrix W itself could be a (parametric) function of X. Since, however, this would not impact our EIS procedure, we ignore such extensions in our notation for H. Next, we briefly survey commonly used specifications for the conditional response density $f(y_i|\lambda_i)$ in Equation (1).

2.2 Spatial probit models

If $y_i \in \{0, 1\}$ denotes a binary choice outcome, such that $y_i = 1$ if $\lambda_i \geq 0$ and $y_i = 0$ otherwise, then y_i is binomial with a conditional probability density function (pdf)

$$f(y_i|\lambda_i) = \mathbb{1}(z_i\lambda_i < 0), \quad \text{with} \quad z_i = 1 - 2y_i, \tag{5}$$

where $\mathbb{1}(\cdot)$ denotes an indicator function. The latent variable λ_i is typically interpreted as a utility difference and the scale factor σ is set equal to one for identification. Equations (1) to (5) characterize spatial probit models.

Spatial probit models and multinomial extensions thereof have received considerable attention in the literature. See, e.g., Case (1992) or McMillen (1992) for early contributions or LeSage and Pace (2009) for a textbook treatment. A few other references of interest (theory and/or applications) are McMillen (1995), Bolduc et al. (1997), Vijverberg (1997), Pinske and Slade (1998), Beron and Vijverberg (2004), Smith and LeSage (2004), Wang and Kockelman (2009), Franzese et al. (2010), Bhat (2011) and Elhorst et al. (2013).

Two recent papers deserve special mention in the context of the present paper. The first one is that by Wang et al. (2013), in which the authors propose a Partial ML Estimator (PMLE) obtained by regrouping observations into spatially correlated pairs and ignoring correlation across pairs[2]. Though partial, PMLE turns out to be statistically more efficient than Generalized Method of Moments

[2]A very closely related PMLE approach based on pairwise likelihood contribution is found in Heagerty and Lele (1998).

(GMM) counterparts used, e.g., by Pinske and Slade (1998). Naturally, we do expect additional efficiency gains by accounting for the complete spatial correlation structure. The second paper is that of Pace and LeSage (2011), in which the authors provide a fast implementation of GHK obtained by application of sparse matrix techniques to a likelihood factorization based on a Cholesky decomposition of the precision matrix H. Reliance upon sparse matrix techniques constitutes a very significant advance for high-dimensional applications. However, using Pace and LeSage's sparse likelihood factorization to implement EIS would require operations on the inverted Cholesky factor of H, which are computationally costly since the inverted Cholesky factor is a dense matrix. The procedure we propose below also relies upon sparse matrix operations but bypasses Cholesky to produce a direct sparse factorization of the likelihood immediately amenable to EIS (nor is it restricted to probit models).

2.3 Spatial count data models

The Poisson distribution is often used when y_i is a count variable with support $\mathbb{N} = \{0, 1, 2, \ldots\}$. Specifically, let the distribution of y_i given λ_i be a Poisson distribution whose mean $\theta_i > 0$ is a monotone increasing function of λ_i. If, in particular, $\theta_i = \exp(\lambda_i)$ then $\theta_i > 0$ without restrictions on the state parameters (β, ρ, σ^2) in Equations (3) and (4), and the conditional pdf for y_i is given by

$$f(y_i|\lambda_i) = \frac{1}{y_i!} \exp\{y_i\lambda_i - \exp(\lambda_i)\}. \tag{6}$$

The Poisson models defined by Equations (1)-(4) and (6) represent spatial counterparts to the class of 'parameter-driven' time-series models for serially correlated counts introduced by Zeger (1988). They also generalize a model used by Lambert et al. (2010) in which θ_i is a measurable function of spatially lagged θ_i's and X with $\sigma^2 \to 0$ in Equation (3). Examples of applications of spatial Poisson models are found, e.g., LeSage et al. (2007), Gschößl and Czado (2008), Lambert et al. (2010) and Buczkowska and de Lapparent (2014).

The dispersion index (ratio between the variance and the mean) associated with the conditional Poisson distribution in Equation (6) equals one. Marginalization with respect to λ_i generates overdispersion but may not suffice to capture the significant overdispersion frequently exhibited by count

data. In such cases, the Poisson can usefully be replaced by a distribution allowing for conditional overdispersion, such as the negative binomial (Negbin). With its mean set equal to $\exp(\lambda_i)$ as above, the conditional pdf of a Negbin with dispersion parameter $s > 0$ is given by

$$f(y_i|\lambda_i) = \frac{\Gamma(y_i + s)}{\Gamma(s)\Gamma(y_i + 1)} \Big(\frac{1}{1 + \exp\{\lambda_i\}/s}\Big)^s \Big(\frac{\exp\{\lambda_i\}}{\exp\{\lambda_i\} + s}\Big)^{y_i}, \tag{7}$$

where $\Gamma(\cdot)$ denotes the Gamma function. Its dispersion index is given by $1+\exp(\lambda_i)/s$ with a Poisson as limiting distribution when $s \to \infty$.

2.4 Censored Data

Recent years have witnessed an increasing number of applications involving spatially correlated (interval) censored data. Examples are infant mortality in Kneib (2005) and Banerjee et al. (2003), water quality detection in Toscas (2010) and uranium grade measurements in Militino and Ugarte (1999). A fairly generic and easily generalizable representation of such models is one whereby λ_i is measured as zero if $\lambda_i \leq 0$ and is measured with error if $\lambda_i > 0$. The corresponding pdf is given by

$$f(y_i|\lambda_i) = f_m(y_i|\lambda_i) \cdot \mathbb{1}(\lambda_i > 0) + \mathbb{1}(y_i = 0) \cdot \mathbb{1}(\lambda_i \leq 0). \tag{8}$$

A special case is the spatial tobit model which assumes that λ_i is measured without error when $\lambda_i > 0$, in which case $f_m(y_i|\lambda_i)$ is Dirac at $y_i = \lambda_i$ (see, e.g., Winkelmann and Boes, 2006). In such a case, the likelihood contribution of the censored observations has the same form as that of the spatial probit in Section 2.2 with a latent distribution that is Gaussian, conditional on the observed (uncensored) λ_i's. See LeSage and Pace (2009, Chap. 10) for an application to interregional origin-destination commuting flows using a data set where 15 percent of the observations are reported as zeros.

3. Spatial EIS

The EIS procedure introduced by Richard and Zhang (2007) has been designed for and successfully applied to likelihood evaluation for a wide range of high-dimensional time-series models for which there exists a natural ordering of the data and, foremost, where latent state processes are specified in the form of parsimonious low-order Markovian conditional densities. EIS takes full advantage of the 'natural' likelihood factorization based on those parsimonious conditional densities to construct a forward sequential efficient importance sampler obtained by means of a backward-recursive sequence of low-dimensional auxiliary regressions and matrix operations.

The situation is quite different for the spatial models in Section 2. There is neither a natural ordering of the n observations nor a corresponding likelihood factorization that would consist of a sequence of parsimoniously parameterized conditional densities. Actually, sequential factorizations of the likelihood function (whether based upon H or H^{-1}) imply sequences of conditional densities with increasing numbers of auxiliary parameters. In this case, the EIS auxiliary regressions are not inherently parsimonious and EIS computations will need to operate on a sequence of large-dimensional parameter matrices. Thus, overall computing time of a 'brute-force' EIS implementation would be $O(n^3)$, as in Liesenfeld and Richard (2010) where the authors analyze low-dimensional ($n = 20$) correlated probit models. Our key contribution in the present paper is that of proposing a novel implementation of EIS that takes full advantage of the sparsity of H to produce a recursive sequence of EIS regressions with a fixed small number of regression parameters irrespective of the sample size n and overall computing time $O(n^\delta)$ with $\delta \ll 3$ ($\delta \simeq 1$ for Poisson and Negbin models and $\delta \simeq 1.5$ for probits).

The necessary modifications of the original EIS principle while conceptually fairly simple are far from trivial in their identification and implementations. They are presented in the next three sections: In Section 3.1 we outline the basic EIS principle; in Section 3.2 we present a likelihood factorization based on a sequential partitioning of the sparse precision matrix H and discuss how to preserve sparse matrix operations throughout the entire EIS sequence; in Section 3.3 we present the corresponding spatial EIS algorithm and apply it in Section 3.4 to spatial probit and in Section 3.5 to Poisson and Negbin models.

3.1 EIS principle

The presentation which follows applies to an arbitrary ordering of the spatial observations (ordering is discussed in Section 3.2 below). For the ease of notation we introduce the error term u, defined as

$$u = \lambda - m \sim \mathrm{N}_n(0, H^{-1}), \tag{9}$$

and delete explicit reference to the matrix X. The joint density $f(u)$ is 'back-recursively' decomposed as

$$f(u) = \prod_{i=1}^{n} f(u_i|u_{(i+1)}), \quad \text{with} \quad u_{(i)} = (u_i, ..., u_n)', \; u_{(n+1)} = \emptyset, \tag{10}$$

and the likelihood is factorized accordingly as

$$L(\psi) = \int_{\mathbb{R}^n} \prod_{i=1}^{n} \varphi_i(u_{(i)}) du, \quad \text{with} \quad \varphi_i(u_{(i)}) = f(y_i|u_i) f(u_i|u_{(i+1)}), \tag{11}$$

where ψ regroups the model parameters and $f(y_i|u_i)$ obtains from $f(y_i|\lambda_i)$ through the linear transformation $u_i = \lambda_i - m_i$.

Parametric importance sampling (IS) densities for u are partitioned conformably with $f(u)$ into

$$g(u; a) = \prod_{i=1}^{n} g(u_i|u_{(i+1)}; a_i), \tag{12}$$

where $a_i \in A_i$ and $a = (a_1, ..., a_n) \in A = \times_{i=1}^{n} A_i$. The conditional IS densities in Equation (12) obtain as a normalized version of a parametric density kernel $k_i(u_{(i)}; a_i)$ with known analytical integrating factor w.r.t. u_i, say

$$g_i(u_i|u_{(i+1)}; a_i) = \frac{k_i(u_{(i)}; a_i)}{\chi_i(u_{(i+1)}; a_i)}, \quad \text{with} \quad \chi_i(u_{(i+1)}; a_i) = \int_{\mathbb{R}^1} k_i(u_{(i)}; a_i) du_i, \tag{13}$$

where $\chi_n(u_{(n+1)}; a_n) \equiv \chi_n(a_n)$.

For any particular $a \in A$, the corresponding IS representation of the likelihood $L(\psi)$ in Equation

(11) is given by

$$L(\psi) = \chi_n(a_n) \cdot \int_{\mathbb{R}^n} \prod_{i=1}^{n} \frac{\varphi_i(u_{(i)}) \cdot \chi_{i-1}(u_{(i)}; a_{i-1})}{k_i(u_{(i)}; a_i)} \prod_{i=1}^{n} g_i(u_i | u_{(i+1)}; a_i) du, \tag{14}$$

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

$$\bar{L}(\psi) = \chi_n(a_n) \cdot \frac{1}{S} \sum_{s=1}^{S} \tilde{w}^{(s)}(a), \qquad \text{with} \qquad \tilde{w}^{(s)}(a) = \prod_{i=1}^{n} \frac{\varphi_i(\tilde{u}_{(i)}^{(s)}) \cdot \chi_{i-1}(\tilde{u}_{(i)}^{(s)}; a_{i-1})}{k_i(\tilde{u}_{(i)}^{(s)}; a_i)}, \tag{15}$$

where $\{\tilde{u}^{(s)}\}_{s=1}^{S}$ denotes S i.i.d. draws from the IS density $g(u; a)$. The objective of EIS is to select a parameter $\hat{a} \in A$ that minimizes the MC variance of the IS ratio $\tilde{w}^{(s)}(a)$. A near optimal solution to this minimization problem obtains as solutions of the following 'forward-recursive' sequence of auxiliary least squares (LS) problems:

$$\hat{a}_i = \arg \min_{a_i \in A_i} \sum_{s=1}^{S} \left\{ \ln \left[\varphi_i\big(\tilde{u}_{(i)}^{(s)}\big) \cdot \chi_{i-1}\big(\tilde{u}_{(i)}^{(s)}; \hat{a}_{i-1}\big) \right] - \ln k_i\big(\tilde{u}_{(i)}^{(s)}; a_i\big) \right\}^2, \quad i = 1, ..., n, \tag{16}$$

where $\{\tilde{u}^{(s)}\}_{s=1}^{S}$ denotes S independent trajectories drawn from $g(u; a)$ itself. Thus, $\hat{a}$ obtains as a fixed point solution to a sequence $\{\hat{a}^{[0]}, \hat{a}^{[1]}, \ldots\}$ in which $\hat{a}^{[j]}$ results from Equation (16) under trajectories drawn from $g(u; \hat{a}^{[j-1]})$. Convergence is typically very fast when trajectories generated for $\{\hat{a}^{[j]}\}_j$ all obtain by transformations of a canonical set $\{\tilde{z}^{(s)}\}_{s=1}^{S}$ of Common Random Numbers (CRNs). Note that $\hat{a}$ is an implicit function of the model parameters ψ. Thus, the EIS algorithm has to be rerun for each new value of ψ. The use of CRNs offers the additional advantage that the IS MC likelihood estimates $\bar{L}(\psi; \hat{a})$ are continuous in ψ and generally amenable to numerical differentiation.

It should be noted that, as discussed in Koopman et al. (2014), using the same set of CRNs for EIS optimization in Equation (16) and likelihood estimation in Equation (15) biases the latter. Note, however, that the subsequent log transformation to obtain the log-likelihood also generates a bias that is an increasing function of the MC variance of the likelihood estimate (see, e.g., Gourieroux and Monfort, 1996). As illustrated in DeJong et al. (2013), this 'log bias' can be significantly larger than the 'CRN bias', so that MC-variance reduction is by far the most critical component of EIS. Moreover, the CRN bias can trivially be eliminated at little costs by using a new set of random draws

in Equation (15).

A critical difference between spatial and Markovian time-series implementations of EIS lies in the fact that the dimension of $u_{(i)}$ in the likelihood factor φ_i is $n-i+1$ (as opposed to a low order and fixed dimension for time series). Thus, for a Gaussian kernel k_i, for example, the EIS parameter a_i to be obtained from the auxiliary EIS regressions (16) could potentially include $O([n-i+1]^2)$ parameters for $i=1,\ldots,n$, implying a prohibitive $O(n^3)$ dimension for the auxiliary parameter space A. Moreover, the construction of the sequence of EIS densities in Equation (13) will require operations on a sequence of $O([n-i+1]^2)$ auxiliary parameter matrices. However, as we shall discuss in the next sections, a likelihood factorization based upon a sequential partitioning of the sparse precision matrix together with careful selection of the EIS kernels $\{k_i(u_{(i)};a_i)\}$ enables us to take full advantage of the fact that the latent process in Equation (2) is Gaussian and, accordingly, to reduce the set of parameters in regression (16) to a small and fixed number, irrespectively of i and n. It also allows us to use fast matrix functions for all high-dimensional matrix operations needed to construct the corresponding EIS densities. Such matrix operations for sparse matrices generally require significantly less operation counts and reduced memory requirements than the corresponding operations on dense matrices (see, e.g., Pace and LeSage, 2011), and are available in software packages like GAUSS and MATLAB.

In conclusion of this presentation of the EIS principle, we ought to mention that we could conceivably rely upon a forward recursive factorization of $f(u)$ instead of the one used in Equation (10). However, there would be no gain in doing so in the absence of a parsimonious Markovian representation. Actually, in the context of the probit model as discussed in Section 3.4, the back-recursive factorization of $f(u)$ we use yields an EIS implementation which is directly comparable to the spatial GHK implementation of Beron and Vijverberg (2004) and Pace and LeSage (2011), allowing for a direct comparison of relative numerical efficiency.

3.2 Sparse likelihood factorization

In this section we first present the algebra of back-recursive factorization of $f(u)$ in Equation (10) in terms of the precision matrices H which is assumed to be sparse. Then we discuss how to preserve

sparsity through the recursive sequence in order to obtain a 'fully' sparse likelihood factorization. Lemma 1 regroups standard results for the sequential factorization of Gaussian densities and introduces the kernel notation we shall use for the conditional densities $\{f(u_i|u_{(i+1)})\}$.

Lemma 1. *Let $u_{(i)} \sim N_{n-i+1}(0, H_i^{-1})$ for $i = 1, \ldots, n$, with $H_1 = H$ for $u_{(1)} = u$. Let partition H_i conformably with $u_{(i)} = (u_i, u'_{(i+1)})'$ into*

$$H_i = \begin{pmatrix} h_{11}^i & h_{12}^i \\ h_{21}^i & H_{22}^i \end{pmatrix}. \tag{17}$$

Then $u_{(i+1)} \sim N_{n-i}(0, H_{i+1}^{-1})$ with

$$H_{i+1} = H_{22}^i - h_{21}^i h_{12}^i / h_{11}^i, \tag{18}$$

and $u_i|u_{(i+1)} \sim N_1\left(-[h_{12}^i/h_{11}^i]u_{(i+1)}, 1/h_{11}^i\right)$, whose density can be written as the following Gaussian density kernel for $u_{(i)}$:

$$f(u_i|u_{(i+1)}) = \frac{1}{\sqrt{2\pi}}\sqrt{h_{11}^i}\exp\left\{-\frac{1}{2}u'_{(i)}H_i^* u_{(i)}\right\}, \tag{19}$$

$$\textit{where} \qquad H_i^* = \begin{pmatrix} h_{11}^i & h_{12}^i \\ h_{21}^i & h_{21}^i h_{12}^i / h_{11}^i \end{pmatrix}. \tag{20}$$

Note that for H_{i+1} and H_i^* to be sparse matrices, H_i must be sparse. Furthermore, the computation of H_{i+1} and H_i^* essentially operates on the first row (and column) of the $(n-i+1 \times n-i+1)$ matrix H_i. Thus, as long as we can maximize the sparsity of first row h_{12}^i of the H_i matrices when they are large (corresponding to low values of i), the number of floating point operations to compute the entire sequences $\{H_{i+1}\}$ and $\{H_i^*\}$ will be greatly reduced by the use of fast sparse matrix functions. In order to achieve this and before factorizing $f(u)$ according to Lemma 1, we reorder the n spatial units relying upon a Symmetric Approximate Minimum Degree (SAMD) permutation that concentrates the non-zero elements of H in its lower right corner (see, Amestoy et al. 1996)[3].

[3]A symmetric approximate minimum degree permutation is also used by Pace and LeSage (2011) for their fast

The effect of such a permutation is illustrated in Figure 1 (right panel) where we plot the non-zero elements of a typical H matrix as shown before permutation in the left panel. The specific matrix we use contains 99.7% zero elements and is constructed in the same way as those we use for our MC study (Section 3.4) and spatial applications (Section 4).

As we shall see in the next section, the combination of sparse factorization of $f(u)$ in Lemma 1 and SAMD permutation allows for fast sparse matrix operations not only for likelihood factorization but also for the complete EIS implementation. This is one of the critical conditions which ensure that EIS remains computationally feasible even for very large n's.

We conclude this section by mentioning that, while it is the precision matrix H in Equation (3) that is typically sparse, there exist occasional specifications for which it is $\Sigma = H^{-1}$ that is sparse. One example is provided by the spatial moving average process in Anselin (1999) where Σ is defined as

$$\Sigma = \sigma^2(I - \rho W)(I - \rho W)'. \tag{21}$$

In such cases the factorization of $f(u)$ in Lemma 1 would be reformulated in terms of a recursive factorization of the sparse Σ matrix.

3.3 EIS spatial kernels and auxiliary regressions

In this section we propose a generic procedure to construct the auxiliary EIS kernels k_i in Equation (13) ought to approximate $\varphi_i\chi_{i-1}$. It applies to the broad class of (high-dimensional) spatial models introduced in Section 2.1. It only requires that, as function of u_i given y_i, $f(y_i|u_i)$ be of the form[4]

$$f(y_i|u_i) \equiv \mathbb{1}(u_i \in D_i) \cdot f_i(u_i), \tag{22}$$

where D_i denotes the support of $u_i = \lambda_i - m_i$ given y_i. As shown in Equation (19), $f(u_i|u_{(i+1)})$ can be rewritten as a Gaussian kernel in $u_{(i)}$. Thus, it is natural to select for k_i a truncated Gaussian

GHK implementation. However, there it is used in order to maximize the sparsity of the Cholesky factor of H.

[4]Our approach trivially covers the more general case where $f(y_i|u_i) = \sum_j \mathbb{1}(u_i \in D_{ij}) \cdot f_{ij}(u_i)$ as, for example, in Equation (8).

kernel in $u_{(i)}$, say

$$k_i(u_{(i)}; a_i) = \mathbb{1}(u_i \in D_i) \exp\left\{-\frac{1}{2}\left(u'_{(i)} P_i u_{(i)} - 2u'_{(i)} q_i + r_i + \ln(2\pi)\right)\right\}, \tag{23}$$

with EIS auxiliary parameters $a_i = (P_i, q_i, r_i)$. We note that the dimension of a_i is $(n-i+1)^2/2 + (n-i+1) + 1$. It follows that a brute force approach to the auxiliary EIS regressions in Equation (16) would require $O(n^3)$ operations, which would be computationally prohibitive for large n's.

Our approach combines two critical components in order to reduce the computational burden to a manageable $O(n^\delta)$ with $\delta \ll 3$. One is the use of the sparse likelihood factorization introduced in Section 3.2. The other one is careful construction of the EIS kernel k_i in order to dramatically reduce the number of parameters in the auxiliary EIS regressions (16). In Lemma 2, we derive the integrating factor χ_i of the kernel k_i in Equation (23). In Lemma 3, we then present the recursive derivation of the auxiliary parameters $\{a_i\}_{i=1}^n$. In order to do so, we rely upon the partitioning of P_i and q_i conformably with $u'_{(i)} = (u_i, u'_{(i+1)})$, say

$$P_i = \begin{pmatrix} p_{11}^i & p_{12}^i \\ p_{21}^i & P_{22}^i \end{pmatrix}, \qquad q_i = \begin{pmatrix} q_1^i \\ q_2^i \end{pmatrix}. \tag{24}$$

Lemma 2. *The integrating factor χ_i of the truncated Gaussian kernel k_i in Equation (23) takes the form of a product of two kernels in $u_{(i+1)}$:*

$$\chi_i(u_{(i+1)}; a_i) = \chi_i^1(u_{(i+1)}; a_i) \cdot \chi_i^2(u_{(i+1)}; a_i), \tag{25}$$

where χ_i^1 is a Gaussian kernel in $u_{(i+1)}$ of the form:

$$\chi_i^1(u_{(i+1)}; a_i) = \exp\left\{-\frac{1}{2}\left(u'_{(i+1)} P_i^* u_{(i+1)} - 2u'_{(i+1)} q_i^* + r_i^*\right)\right\}, \tag{26}$$

with

$$P_i^* = P_{22}^i - \frac{p_{21}^i p_{12}^i}{p_{11}^i}, \qquad q_i^* = q_2^i - \frac{p_{21}^i q_1^i}{p_{11}^i}, \qquad r_i^* = r_i - \frac{(q_1^i)^2}{p_{11}^i} + \ln p_{11}^i, \tag{27}$$

and χ_i^2 *is a non-Gaussian kernel that depends only on a scalar linear combination of* $u_{(i+1)}$

$$\chi_i^2(u_{(i+1)}; a_i) \equiv \chi_i^2(\nu_{i+1}; a_i), \tag{28}$$

with

$$\nu_{i+1} = \alpha_i + \beta_i' u_{(i+1)}, \qquad \alpha_i = \frac{q_1^i}{p_{11}^i}, \qquad \beta_i' = -\frac{p_{12}^i}{p_{11}^i}. \tag{29}$$

Proof: We rely upon the partitioning introduced in Equation (24) in order to partition k_i in Equation (23) into a truncated Gaussian density in u_i given ν_{i+1} and a Gaussian kernel in $u_{(i+1)}$. Standard Gaussian algebra produces the partitioning

$$k_i(u_{(i)}; a_i) = \left[\chi_i^1(u_{(i+1)}; a_i)\right] \left[\mathbb{1}(u_i \in D_i) f_N(u_i|\nu_{i+1}, 1/p_{11}^i)\right], \tag{30}$$

where $f_N(\cdot|\mu, \varpi^2)$ denotes the density of a Gaussian random variable with mean μ and variance ϖ^2. It immediately follows that:

$$\chi_i^2(\nu_{i+1}; a_i) = \int \mathbb{1}(u_i \in D_i) f_N(u_i|\nu_{i+1}, 1/p_{11}^i) du_i, \tag{31}$$

which, as a function of $u_{(i+1)}$, only depends on ν_{i+1}. □

Our next step is the actual construction of the kernel k_i in Equation (23). Following Equations (11), (22), (25) and (28), the product $\varphi_i \chi_{i-1}$ to be approximated by k_i is given by

$$\varphi_i(u_{(i)})\chi_{i-1}(u_{(i)}; a_{i-1}) = \mathbb{1}(u_i \in D_i) \left[f(u_i|u_{(i+1)})\chi_{i-1}^1(u_{(i)}; a_{i-1})\right] \left[f_i(u_i)\chi_{i-1}^2(\nu_i; a_{i-1})\right], \tag{32}$$

where the product $f_i \chi_{i-1}^2$ is the sole non-Gaussian kernel. Thus, in order to reduce the dimensionality of the auxiliary EIS regressions (16), we define k_i as the (truncated) product of the two Gaussian kernels in $\varphi_i \chi_{i-1}$ and of an EIS Gaussian kernel approximation to the product $f_i \chi_{i-1}^2$ (or just f_1 for $i = 1$, since $\chi_0 \equiv 1$).

Thus, for $i > 1$, the kernel k_i in Equation (23) obtains as the following product of three Gaussian

kernels:

$$k_i(u_{(i)}; a_i) = \mathbb{1}(u_i \in D_i) \left[f(u_i|u_{(i+1)}) \chi_{i-1}^1(u_{(i)}; a_{i-1}) \right] k_i^2(u_i, \nu_i; a_i^*), \tag{33}$$

where k_i^2 denotes a bivariate EIS Gaussian kernel approximation of the product $f_i \chi_{i-1}^2$ in Equation (32), which can be written as

$$k_i^2(u_i, \nu_i; a_i^*) = k_i^2(\omega_i; a_i^*) = \exp\left\{ -\frac{1}{2} \left(\omega_i' B_i \omega_i - 2\omega_i' c_i + d_i \right) \right\}, \tag{34}$$

with

$$\omega_i = \begin{pmatrix} u_i \\ \nu_i \end{pmatrix} = \gamma_i + \Delta_i u_{(i)}, \quad \gamma_i = \begin{pmatrix} 0 \\ \alpha_{i-1} \end{pmatrix}, \quad \Delta_i = \begin{pmatrix} e_i' \\ \beta_{i-1}' \end{pmatrix}, \quad e_i' = (1, 0, \ldots, 0), \tag{35}$$

and the EIS regression parameters $a_i^* = (B_i, c_i, d_i)$.

For $i = 1$, we have $\chi_0 \equiv 1$ in which case k_i^2 only needs to approximate $f_1(u_1)$ and can be written as

$$k_1^2(u_1, \nu_1; a_1^*) = k_1^2(u_1; a_1^*) = \exp\left\{ -\frac{1}{2} \left(b_1 u_1^2 - 2c_1 u_1 + d_1 \right) \right\}, \tag{36}$$

with EIS regression parameter $a_1^* = (b_1, c_1, d_1)$. As we shall see below, similar univariate simplifications apply to the spatial probit model, for which $f_i(u_i) \equiv 1$, and to spatial count data models for which $D_i \equiv \mathbb{R}$ so that $\chi_i^2 = 1$.

With these definitions in place, it is now a simple matter to derive the recursion providing the EIS parameters $\{a_i\}_{i=1}^n$ since it amounts to combining the three Gaussian kernels in Equation (33). For the ease of later reference, we regroup these recursion formulas in the following lemma.

Lemma 3. *With k_i as defined in Equation (33), the EIS auxiliary parameters $\{a_i\}_{i=1}^n$ in Equation (23) obtain through the following recursion:*

For $i = 1$ (k_i with only two Gaussian kernels)

$$P_1 = H_1^* + b_1 e_1 e_1', \qquad q_1 = e_1 c_1, \qquad r_1 = d_1 - \ln h_{11}^1. \tag{37}$$

For $i > 1$ (k_i with three Gaussian kernels)

$$P_i = H_i^* + P_{i-1}^* + \Delta_i' B_i \Delta_i, \tag{38}$$

$$q_i = q_{i-1}^* + \Delta_i'(c_i - B_i\gamma_i), \tag{39}$$

$$r_i = r_{i-1}^* + d_i + \gamma_i' B_i \gamma_i - 2\gamma_i' c_i - \ln h_{11}^i, \tag{40}$$

where $(P_{i-1}^*, q_{i-1}^*, r_{i-1}^*)$ and (γ_i, Δ_i), as defined respectively in Equations (27) and (35) are all functions of a_{i-1}.

We now have all the components in place to present our simplified spatial EIS auxiliary regressions, the EIS densities, as well as the corresponding EIS weights. Foremost, we note that after elimination of the common factors in $\varphi_i \chi_{i-1}$ and k_i, as given respectively by Equations (32) and (33), the auxiliary regressions (16) simplify into

$$\hat{a}_i^* = \arg\min_{a_i^*} \sum_{s=1}^{S} \left\{ \ln \left[f_i\left(\tilde{u}_i^{(s)}\right) \cdot \chi_{i-1}^2\left(\tilde{\nu}_i^{(s)}; \hat{a}_{i-1}\right) \right] - \ln k_i^2\left(\tilde{u}_i^{(s)}, \tilde{\nu}_i^{(s)}; a_i^*\right) \right\}^2, \tag{41}$$

where, as in Equation (16), $\{\tilde{u}^{(s)}\}_{s=1}^S$ denotes S i.i.d. trajectories drawn from $g(u; a)$. By construction, the latter factorizes into a sequence of the following conditional densities for u_i given $u_{(i+1)}$:

$$g_i(u_i | u_{(i+1)}, a_i) = \frac{\mathbb{1}(u_i \in D_i) \cdot f_N(u_i | \nu_{i+1}, 1/p_{11}^i)}{\chi_i^2(\nu_{i+1}; a_i)}. \tag{42}$$

Accounting for observation $i = 1$, the corresponding EIS weights in Equation (15) are given by

$$\tilde{w}^{(s)}(\hat{a}) = \frac{f_1(\tilde{u}_1^{(s)})}{k_1^2(\tilde{u}_1^{(s)}; \hat{a}_1^*)} \prod_{i=2}^{n} \frac{f_i(\tilde{u}_i^{(s)}) \chi_{i-1}^2(\tilde{\nu}_i^{(s)}; \hat{a}_{i-1})}{k_i^2(\tilde{\omega}_i^{(s)}; \hat{a}_i^*)}. \tag{43}$$

All in all, the sequential construction of the kernels k_i, their integrating factors χ_i and the EIS densities g_i amounts to combine in step i the EIS regression parameters $\hat{a}_i^*$ from Equation (41) with H_i^* in Lemma 1 and the parameters $(P_{i-1}^*, q_{i-1}^*, r_{i-1}^*, \alpha_{i-1}, \beta_{i-1})$ in order to obtain according to Lemma 3 the optimal EIS parameters $\hat{a}_i = (P_i, q_i, r_i)$ as well as the parameters $(P_i^*, q_i^*, r_i^*, \alpha_i, \beta_i)$ in Lemma 2 for the next step $i + 1$.

Most importantly, the results we presented in Lemmas 1 to 3 provide the key to a computationally feasible EIS implementation for high-dimensional spatial models. There are two reasons for this. Foremost, the sequential computation of the high dimensional $(n-i+1)\times(n-i+1)$ matrices (P_i, P_i^*) and $(n-i+1)$ vectors (q_i, q_i^*, β_i) depend on the first row of H_i^* and, therefore, fully preserve their sparsity under the SAMD permutation. This follows from the fact that the SAMD permutation maximizes the sparsity of the first row of H_i^* that determines via P_i in Equation (38) the parameters P_i^*, q_i^*, β_i in Equations (27) and (29) which in turn are needed to compute for the next step P_{i+1} and q_{i+1}. Thus, the recursive computation of all the high-dimensional auxiliary parameter vectors required for the spatial EIS can be computed using computationally fast sparse matrix functions. Moreover, as shown in Equations (34) and (36), the dimension of the auxiliary regressions (41) is 6 for the more general case or 3 in the special cases ($i=1$, $f_i(u_i)\equiv 1$ or $D_i\equiv\mathbb{R}$), irrespective of n and i. These two key components of our sparse spatial EIS implementation produce computing times that are $O(n^\delta)$ with $\delta \ll 3$, instead of $O(n^3)$ that would obtain under brute force EIS.

We mentioned earlier that the EIS auxiliary regressions in Equation (41) have to be iterated since the trajectories $\{\tilde{u}^{(s)}\}$ are to be drawn from $g(u;\hat{a})$ itself. In practice, this requires constructing an initial $\hat{a}^{[0]}$ from initial values for $\{\hat{a}_i^*\}_{i=1}^n$ and then for iteration $j=1,\dots J$, using draws from $g(u;\hat{a}^{[j-1]})$ to compute a new $\hat{a}^{[j]}$. Actually, only the first few iterations produce significant MC variance reductions. Thus, we typically preset J at a low fixed number, say from 3 to 5. As for $\hat{a}^{[0]}$, it can generally obtained by means of a local Taylor-series approximation to the target $f_i\chi_i^2$ to produce initial values for $\hat{a}_i^*$, or even by setting $\hat{a}_i^*=0$.

In conclusion of this generic presentation, we provide next a pseudo-code summary of the spatial EIS algorithm. Full MATLAB codes are available at http://www.wisostat.uni-koeln.de/index.php?id=27287.

Spatial EIS algorithm

(i) *Use CRNs to generate S new i.i.d. trajectories $\{\tilde{u}^{(s)}\}$ from $g(u|\hat{a}^{[j-1]})$, where $\hat{a}^{[j-1]} = (\hat{a}_1^{[j-1]}, \ldots, \hat{a}_n^{[j-1]})$.*

(ii) *(initialization): set $P_0^* = 0$, $q_0^* = 0$, $r_0^* = 0$, $\alpha_0 = 0$, $\beta_0 = 0$.*

(iii) *For $i = 1$*

- *run the EIS regression as given by Equations (36) and (41) to obtain $\hat{a}_1^*$;*
- *use Equation (37) to compute $\hat{a}_1^{[j]} = (P_1, q_1, r_1)$;*
- *use Equations (27) and (29) to compute $(P_1^*, q_1^*, r_1^*, \alpha_1, \beta_1)$ and pass them to step $i = 2$.*

For $i : 2 \to n$

- *run the EIS regression as given by Equations (34) and (41) to obtain $\hat{a}_i^*$;*
- *use Equations (38) to (40) to compute $\hat{a}_i^{[j]} = (P_i, q_i, r_i)$;*
- *use Equations (27) and (29) to compute $(P_i^*, q_i^*, r_i^*, \alpha_i, \beta_i)$ and pass them to step $i + 1$.*

(iv) *If $j \geq J$, then stop and set $\hat{a} = \hat{a}_{[J]}$, else, $j \to j + 1$ and rerun.*

After iteration J, the EIS likelihood estimate is computed from Equations (15) and (43).

3.4 Spatial probit model

Under the spatial probit models, the two factors in the measurement density in Equation (22) are given by

$$\mathbb{1}(u_i \in D_i) = \mathbb{1}(z_i u_i \leq -z_i m_i), \qquad f_i(u_i) \equiv 1. \tag{44}$$

It follows that the integrating factor in Equation (31) is given by

$$\chi_i^2(\nu_{i+1}; a_i) = \Phi(-z_i \sqrt{p_{11}^i}[m_i + \nu_{i+1}]), \tag{45}$$

where Φ denotes the cdf of a standard normal. For ease of comparison with GHK it proves convenient to redefine accordingly α_i and β_i in Equation (29) as

$$\alpha_i = -z_i\sqrt{p_{11}^i}(m_i + \frac{q_1^i}{p_{11}^i}), \qquad \beta_i' = z_i\frac{p_{12}^i}{\sqrt{p_{11}^i}}, \tag{46}$$

so that $\chi_i^2(\nu_{i+1}, a_i) = \Phi(\nu_{i+1})$. Moreover, since $f_i(u_i) \equiv 1$, $\ln k_i^2$ in Equation (34) only needs to approximate $\ln\Phi(\nu_i)$. This implies the following simplifications of the spatial EIS algorithm. For $i = 1$ with $\chi_0^2 \equiv 1$ and $\ln f_1 \equiv 0$, we have $\hat{a}_1^* = 0$ in Equation (36) and $k_1 \equiv \varphi_1\chi_0$ (yielding a perfect fit). For $i > 1$, the auxiliary kernel k_i^2 in Equation (34) only depends on ν_i and Equations (34), (35) and (37)-(40) simplify accordingly into

$$k_i^2(u_i, \nu_i; a_i^*) = k_i^2(\nu_i; a_i^*) = \exp\left\{-\frac{1}{2}\left(b_i\nu_i^2 - 2c_i\nu_i + d_i\right)\right\}, \tag{47}$$

$$\gamma_i = \begin{pmatrix} 0 \\ \alpha_{i-1} \end{pmatrix}, \quad \Delta_i = \begin{pmatrix} 0 \\ \beta_{i-1}' \end{pmatrix}, \tag{48}$$

$$P_i = H_i^* + P_{i-1}^* + b_i\beta_{i-1}\beta_{i-1}', \qquad P_1 = H_1^*, \tag{49}$$

$$q_i = q_{i-1}^* + (c_i - b_i\alpha_{i-1})\beta_{i-1}, \qquad q_1 = 0, \tag{50}$$

$$r_i = r_{i-1}^* + d_i + b_i\alpha_{i-1}^2 - 2\alpha_{i-1}c_i - \ln h_{11}^i, \qquad r_1 = -\ln h_{11}^1, \tag{51}$$

and the EIS weights in Equation (43) are then given by

$$\tilde{w}^{(s)}(\hat{a}) = \prod_{i=2}^{n}\frac{\Phi(\tilde{\nu}_i^{(s)})}{\exp\{-\frac{1}{2}(\hat{b}_i[\tilde{\nu}_i^{(s)}]^2 - 2\hat{c}_i\tilde{\nu}_i^{(s)} + \hat{d}_i)\}}. \tag{52}$$

As mentioned above, our EIS procedure covers GHK as a special case. Specifically, the IS density kernels under GHK are defined as

$$k_i(u_{(i)}; \cdot) = \varphi_i(u_{(i)}) = \mathbb{1}(z_iu_i \leq -z_im_i) \cdot f(u_i|u_{(i+1)}), \tag{53}$$

with integrating factor $\chi_i(\nu_{i+1}, \cdot) = \Phi(\nu_{i+1})$. This is equivalent to setting $k_i^2 \equiv 1$ with $a_i^* = (b_i, c_i, d_i) = 0$ in Equations (47)-(52). In particular, the IS weights in Equation (52) simplify into

$$\tilde{w}^{(s)} = \prod_{i=2}^{n} \Phi(\tilde{\nu}_i^{(s)}), \tag{54}$$

where $\{\tilde{\nu}_i^{(s)}\}$ are drawn from the sequential GHK-IS truncated Gaussian densities associated with the kernels in Equation (53). Since $a_i^* = 0$ is generally EIS-suboptimal, it immediately follows that GHK is numerically less efficient than EIS. See Liesenfeld and Richard (2010) for a low-dimensional application (for which sparse matrix operations are not required). In Section 3.6 below, we provide a high-dimensional comparison of the relative numerical efficiencies of GHK and EIS, which is trivial since we initialize EIS with the GHK sampler by setting $\hat{a}_i^* = 0$ to obtain the initial EIS auxiliary parameters $\hat{a}^{[0]}$. We note that our GHK implementation is directly comparable to those of Beron and Vijverberg (2004) and Pace and LeSage (2011), though the latter relies upon sparse Cholesky factorizations of H, instead of the EIS procedure described above.

3.5 Spatial Poisson model

Under the Possion model the measurement density as given by Equation (22) consists of

$$\mathbb{1}(u_i \in D_i) \equiv 1, \qquad f_i(u_i) = f(y_i|u_i) = \frac{1}{y_i!} \exp\{y_i(m_i + u_i) - \exp(m_i + u_i)\}. \tag{55}$$

Therefore, $\chi_i^2 \equiv 1$ in Equation (28). It follows from Equation (41) that for $i = 1, \ldots, n$ we only have to EIS approximate $\ln f(y_i|u_i)$ by a univariate (log)kernel $\ln k_i^2(u_i; a_i^*)$. Thus Equations (34), (35), and (38)-(40) simplify accordingly into

$$k_i^2(u_i, \nu_i; a_i^*) = k_i^2(u_i; a_i^*) = \exp\left\{-\frac{1}{2}\left(b_i u_i^2 - 2c_i u_i + d_i\right)\right\}, \tag{56}$$

$$\gamma_i = 0, \;\; \Delta_i = \begin{pmatrix} e_i' \\ 0 \end{pmatrix}, \tag{57}$$

$$P_i = H_i^* + P_{i-1}^* + b_i e_i e_i', \qquad q_i = q_{i-1}^* + c_i e_i, \qquad r_i = r_{i-1}^* + d_i - \ln h_{11}^i. \tag{58}$$

The corresponding EIS weights in Equation (43) are then given by

$$\tilde{w}^{(s)}(\hat{a}) = \prod_{i=1}^{n} \frac{\exp\{y_i(m_i + \tilde{u}_i^{(s)}) - \exp(m_i + \tilde{u}_i^{(s)})\}/y_i!}{\exp\{-\frac{1}{2}(\hat{b}_i[\tilde{u}_i^{(s)}]^2 - 2\hat{c}_i\tilde{u}_i^{(s)} + \hat{d}_i)\}}. \tag{59}$$

Since the EIS regressions only require approximating $\ln f(y_i|u_i)$ by $\ln k_i^2(u_i; a_i^*)$, it immediately follows that we can directly construct the joint EIS density from Equations (12), (13), (33) and (56) as

$$g(u; a) = \frac{f(u)\prod_{i=1}^{n} k_i^2(u_i; a_i^*)}{\chi_n(a_n)}, \tag{60}$$

where

$$\chi_n(a_n) = \int_{\mathbb{R}^n} f(u)\prod_{i=1}^{n} k_i^2(u_i; a_i^*)du \tag{61}$$

obtains directly using standard Gaussian algebra since the integrand in Equation (61) represents an untruncated Gaussian kernel in u. This construction of the joint EIS density in a single step eliminates sequential combination of large matrices and reduces accordingly the computing times (by a factor of 45 in the application presented in Section 4 below).

Replacing the conditional Poisson density in Equation (6) by the conditional Negbin density in Equation (7), only requires modifying accordingly the dependent variable $\ln f(y_i|u_i)$ in the EIS auxiliary regressions.

3.6 Censored Data

Under censoring as defined by Equation (8), the pdf for censored data is of the generic form given by Equation (22) where D_i can represent an arbitrary censoring interval for observation i. It follows that k_i^2 in Equation (34) effectively depends on both u_i and ν_i so that the auxiliary EIS parameter a_i^* is now 6-dimensional. Thus, our generic EIS algorithm as described in Section 3.3 immediately applies to a wide range of spatial models for censored data.

4. Monte Carlo Study

In this section, we conduct a MC analysis of the statistical and numerical properties of ML-EIS estimates for spatial probit and Poisson models. We consider both the SAL (A4.1) and the SAE (A4.2) specifications introduced in Equation (4), where the regression function $X\beta$ for unit i is specified as

$$x_i'\beta = \beta_0 + \beta_1 x_i. \tag{62}$$

The regressors x_i are assumed to be i.i.d. uniform random variables on the interval $(-3, 4)$ for the probit models, 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 a pair of coordinates from a uniform-$(0, 1)^2$ distribution for each spatial unit i. 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 probit models we set the regression parameters at $(\beta_0, \beta_1) = (-1.5, 3)$ and for the Poisson models at $(\beta_0, \beta_1, \sigma) = (-0.25, 0.8, 0.3)$, where σ is the scale parameter in Equation (3) (for the probit models it is set equal to one to ensure identification). For all four specifications we vary the degree of spatial correlation by taking $\rho = 0.75$ or $\rho = 0.85$. We consider different sample sizes ranging from a moderate $n = 100$ to a fairly large $n = 5000$. We only report the results for $n = 5000$, which represents a compromise between our interest in large data sets and the need to conduct an extended MC analysis. We also ran a few tries with $n = 50,000$ in order to confirm the feasibility of our approach for very large sample sizes and to verify that computing times are indeed $O(n^\delta)$ with $\delta \ll 3$.

Using the Data-Generating Processes (DGPs) we just described we construct two different MC sampling distributions for the ML-EIS estimators. One is the conventional (finite-sample) *statistical* distribution (used for likelihood-based inference) and the other is the *numerical* distribution used

to assess numerical accuracy (see, e.g., Richard and Zhang, 2007). The statistical 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 a given data set using 50 different sets of CRNs. Clearly, reliable statistical inference requires that numerical biases and standard deviations be negligible relative to statistical standard deviations.

As explained in Section 3.4, we use GHK to initialize our EIS procedure for spatial probit models. This enables us to trivially compare the statistical and numerical properties of ML-GHK and ML-EIS estimates for these models. Unfortunately, no such direct ML competitor is readily available for spatial count models.

ML-EIS estimates are computed using simulation sample size $S = 20$ and $J = 3$ EIS iterations. For the probit models with $n = 5000$, a single likelihood EIS evaluation using a MATLAB code takes of the order of 45 s on a Intel i7 Core computer with 2.67 GHz. The corresponding computing time for the Poisson models is only 1 s, as we rely upon the joint EIS implementation described in Equation (60). For the probit models GHK is much faster than EIS (by an approximate factor of 25), since it does not require (iterated) auxiliary regressions, but is also significantly less accurate. Thus, we compare EIS based on $S = 20$ with GHK using $S = 500$ in order to approximately equate computing times. For all ML estimations we use the BFGS optimizer.

4.1 Spatial probit models

The main results of our four MC experiments for the spatial probit model are reported in Table 1. The DGP parameter values are listed in column 1. Statistical means, standard deviations and Root Mean Squared Errors (RMSEs) for ML-GHK and ML-EIS are reported in columns 2 and 3, respectively. 'True' ML estimates, defined as ML-EIS estimates with $S = 1000$ are listed in column 5. Numerical means, standard deviations and RMSEs for ML-GHK and ML-EIS are listed in column 6 and 7, respectively.

Our main finding is that the ML-EIS estimates are statistically virtually unbiased and numerically

highly accurate. The ratios between statistical and numerical standard errors range roughly from 20 to 100. Thus, statistical inference based on ML-EIS is largely unaffected by numerical errors and fully reliable.

In sharp contrast, ML-GHK estimates are statistically significantly biased with RMSEs 2 to 23 times larger than those of ML-EIS. Foremost, ML-GHK estimates are numerically considerably less accurate than their ML-EIS counterparts with RMSEs ratios ranging from 90 to 384. Clearly, statistical results obtained by ML-GHK are subject to large numerical errors and are essentially unreliable. In order to match EIS-ML numerical accuracy, GHK would require a prohibitively large simulation sample size S.

The explanation for such poor numerical performance is that the GHK-IS density $g(u;.)$ obtained from the kernels in Equation (53) represents a very poor *global* approximation to the high-dimensional target $\prod_i \varphi_i(u_{(i)})$. In order to verify this conjecture we present in Figure 2 histograms of the normalized IS weights $\bar{\omega}^{(s)} = \tilde{\omega}^{(s)} / \sum_{s=1}^{S} \tilde{\omega}^{(s)}$ for both EIS and GHK, as defined in Equations (52) and (54), respectively. A perfect IS fit would imply that $\bar{\omega}^{(s)} = 1/S, \forall s$. The histogram of the normalized GHK-IS weights exhibits extreme skewness with only 2 out of 500 weights that are effectively larger than zero (0.39 and 0.61). The corresponding histogram for the 500 EIS weights is clearly much better behaved. Our results are consistent with the findings of Pace and LeSage (2011) who also report an extremely skewed distribution of the GHK-IS weights for $n = 100,000$ and $S = 30$. As a remedy, they propose to replace the GHK likelihood estimate $\bar{L} = \chi_n \cdot \frac{1}{S} \sum_{s=1}^{S} \prod_{i=2}^{n} \Phi(\tilde{\nu}_i^{(s)})$, as defined by Equations (15) and (54), by the alternative estimate $\hat{L} = \chi_n \cdot \prod_{i=2}^{n} [\frac{1}{S} \sum_{s=1}^{S} \Phi(\tilde{\nu}_i^{(s)})]$ which would remain unbiased if the terms of the sequence $\{\Phi(\tilde{\nu}_i^{(s)})\}_{i=2}^{n}$ were mutually uncorrelated. This condition, however, would need to be checked empirically for each application.

The results of the four MC experiments obtained by using Page and LeSage's modified MC likelihood estimate $\hat{L}$ using $S = 500$ (GHK^{PL}) are reported in columns 4 and 8 of Table 1. They reveal that ML-GHK^{PL} performs better than the standard ML-GHK in terms of statistical and numerical properties but remains dominated by ML-EIS with statistical RMSEs ratios ranging from 1 to 1.5 and numerical RMSEs ratios between 4 and 53.

4.2 Spatial Poisson models

The MC results for the Spatial SAL and SAE Poisson model are reported in Table 2. They show that for all four DGPs and for all parameters, the ML-EIS estimators are virtually unbiased with S as low as 20. Numerical accuracy is high with means that are close to the 'true' ML values and standard deviations that are significantly smaller than their statistical counterparts.

We conclude that likelihood based inference results obtained by spatial EIS for probit and Poisson models with $n = 5,000$ are numerically very accurate even with MC sample sizes as low as 20. As for the dependence on n, we have run experiments with larger n's (up to 50,000) and have found computing times of the order of $O(n^{1.5})$ for the probit models and virtually $O(n)$ for the Poisson models under the simplified implementation according to Equation (60).

5. Empirical Applications

In this section we present two empirical applications relying upon ML-EIS. One uses a spatial probit model for voting decisions at the county level in the 1996 US presidential election, and the other a spatial count model (Poisson and Negbin) for firm location decisions in US manufacturing. Both models rely upon the SAL specification A4.1 in Equation (4).

5.1 Spatial probit for the 1996 US presidential election

We use a SAL probit to model the voting decisions of the $n = 3,110$ US counties in the 1996 US presidential election. The dependent variable y_i equals one if the Democratic Candidate Bill Clinton won the majority in county i and zero if the Republican candidate Bob Dole won. The SAL specification is the same as that used for the MC study in Section 4, expect that x_i in Equation (62) is now a vector which includes the log urban population and the following four educational level variables (expressed as a proportion of the county population): some years at college, associate degrees, college degrees and graduate or professional degrees. The data are taken from LeSage's spatial econometric toolbox (http://www.spatial-econometrics.com).

Spatial dependence reflects the fact that voters from neighboring counties tend to exhibit similar

voting patterns. As was the case for our MC study, we construct W using the geographical coordinates (latitude and longitude) of the counties and transforming them by a Delaunay triangulation in order to assign to each county its six 'closest' neighbors. As in Section 4.1, the ML-EIS parameter estimates rely upon simulation sample size $S = 20$ and $J = 3$ EIS iterations, while ML-GHK uses $S = 500$.

Following, e.g., Beron and Vijverberg (2004), we also compute average marginal effects defined as:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \text{prob}(y_i = 1|X, W)}{\partial x_i} = \frac{1}{n}\sum_{i=1}^{n}\phi(h_i^{\frac{1}{2}}[v_{i1}x_1'\beta + \cdots + v_{in}x_n'\beta]) \cdot h_i^{\frac{1}{2}}v_{ii}\beta, \tag{63}$$

where $\phi(\cdot)$ denotes the standardized normal density function, v_{ij} is the element (i, j) of the matrix $V = (I - \rho W)^{-1}$, and h_i is the precision of the marginal distribution of the error u_i. These marginal effects account for both the direct impact on λ_i of a change in $x_i'\beta$ and its indirect impact through the spatial interdependence reflected by the matrix V.

The results are reported in Table 3. Asymptotic (statistical) standard deviations are derived from a numerical approximation of the Hessian. Numerical standard deviations are based upon 50 ML-EIS estimations using different CRNs. They confirm our earlier finding that ML-EIS is numerically far more accurate than ML-GHK. Our ML-EIS estimate of ρ is highly significant and implies strong spatial dependence across neighboring counties, in line with the Bayesian MCMC results obtained by LeSage and Pace (2009, Table 10.3) for the 2000 US presidential election. Our results also confirm the downward bias of the ML-GHK estimates of ρ, as already observed in our MC study. The ML-EIS estimates of the β parameters and the corresponding marginal effects are systematically lower in absolute value than their ML-GHK counterparts. Under ML-EIS only two variables have a significant impact: negative for *some college* and positive for *graduate/professional degree*. Thus, higher education levels play in favor of Clinton. Finally, ML-EIS produces a significantly higher value than ML-GHK for the maximized log likelihood. This confirms a negative bias we had previously observed when comparing GHK with 'brute force' EIS for $n = 20$ in Liesenfeld and Richard (2010).

5.2 Spatial count model for US firms location choices

In this section we apply the spatial count models (Poisson and Negbin) introduced in Section 2.3 to US firm births (start-ups) at the county level ($n = 3,078$) during the period 2000-2004. We rely upon a data set from Lambert et al. (2010, Section 6), hereafter LBF, who applied a two-step Limited-Information ML (LIML) procedure to estimate an 'observation-driven' version of the SAL-Poisson model introduced in Section 2.3, with $\sigma \to 0$ in Equation (3). Instead we apply Full-Information ML (FIML) to the 'parameter-driven' version ($\sigma > 0$) of that model. Spatial correlation is expected to reflect such factors as industry clustering and regional economic development policies. We refer the reader to LBF for an insightful discussion of the covariates used in the SAL regression.

The dependent variable y_i is defined as the cumulative number of new single-unit start-ups from 2000 to 2004. Following LBF we construct the weight matrix W using the Delaunay triangulation algorithm to assign eight neighbors to each county. 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.

Our SAL-Poisson ML-EIS estimates obtained with a MC sample size $S = 20$ and $J = 3$ EIS iterations are reported in the left panel of Table 4. They are mostly in line with those reported in LBF's Table 6, though generally with significantly smaller statistical standard deviations, as expected

from using FIML instead of LIML. We observe significant sign changes for only three covariates *Pelt10*, *Awage* and *Netflow*. Three other parameters have the same sign but are, nevertheless, significantly different (based on the FIML standard deviations): *Pemt100*, *Pop* and *Pedas*. Overall our results confirm LBF's conclusion that counties with agglomeration economies, labor availability, low costs for labor, availability of skilled labor, a business-friendly infrastructure and fiscal policy were more likely to attract new start-ups. However, there are two key differences with LBF's results. One is that we find a significantly higher estimate of the spatial correlation parameter ρ: 0.514 (0.018) instead of 0.181 (0.062). The other is that our estimate of σ is 0.697 (0.013) and implies a strong rejection of LBF's specification where σ is set equal to zero.

Since counts frequently exhibit over-dispersion, we also estimated a SAL-Negbin model version that, as explained above, only requires minor modifications of the baseline spatial-EIS algorithm. The ML-EIS Negbin results are reported in the right panel of Table 4. The estimates of the regression coefficients are largely in agreement with those obtained under the Poisson specification, with only one statistically significant sign difference for *Mhhi*. The dispersion parameter estimate of s equals 2.902 (0.204) and is statistically significantly smaller than infinity, implying rejection of the Poisson model, as further confirmed by a large log-likelihood difference of 395. The Negbin estimate of σ is significantly lower (0.261 instead of 0.697) but this was to be expected from the substitution of an over-dispersed Negbin for the Poisson. Last but not least, we also find a significantly lower estimate of ρ (0.321 instead of 0.514). These differences suggest that a significant part of the variation in the firm-birth events across counties, that was attributed to shocks with global spatial feedback and feed-forward transmission between locations under the Poisson model, is interpreted as conditional over-dispersion together with lesser spatial transmission under the Negbin model. Numerical accuracy is good for the Poisson model and excellent for the Negbin specification. While we could easily improve the numerical accuracy of the Poisson estimates by increasing S, we decided there was little reason for doing so, since the Poisson model is strongly rejected against the Negbin model.

6. Conclusions

We have proposed a generic EIS-based procedure for numerically accurate likelihood evaluation of a broad class of high-dimensional spatial latent Gaussian models with non-Gaussian response variables (discrete choices, counts, truncation, censoring and sample selection). Our algorithm consists of an original and powerful combination of sparse-matrix techniques and EIS. Its two key features are: (i) a novel forward-recursive implementation of EIS that is specifically constructed in such a way that it preserves the sparsity of large (auxiliary) precision matrices throughout the entire recursive sequence so that all high-dimensional EIS matrix operations can be performed using computationally fast sparse matrix functions; and (ii) a selection of auxiliary importance sampling kernels that produces low-dimensional auxiliary EIS regressions, irrespectively of the data size n. The combination of these two features produces an algorithm that remains computationally feasible for large sample sizes n, with computing times of $O(n^{\delta})$ with $\delta \ll 3$ ($\delta \simeq 1.5$ for probit applications and $\delta \simeq 1$ for Poisson and Negbin models with n as large as 5,000, or even 50,000 in test trials).

Two empirical applications with sample sizes of the order of 3,000 have illustrated further the full potential of our spatial EIS procedure for accurate likelihood based inference. While in the paper we restricted our attention to probit, Poisson and Negbin models, the generic structure of our algorithm, as presented in Section 3, indicates that it can easily be applied to a much broader class of spatial models including but not limited to spatial ordered or multinomial probits, spatial tobit models, spatial truncated and sample selection models. Most importantly, applications to such models would only require minor adjustments of the baseline spatial EIS implementation presented in Section 3 (essentially adapting the dependent variable in the auxiliary EIS regressions and the corresponding (E)IS weights).

All in all, our procedure allows for numerically accurate likelihood based inference in a broad class of high-dimensional spatial latent Gaussian models, a task hitherto considered to be computationally prohibitive.

Acknowledgements

The authors thank Jason Brown for providing the firm location choice data set used in this paper and Albrecht Mengel for providing access to the grid computing facilities of the Institute of Statistics and Econometrics at University of Kiel. R. Liesenfeld and J. Vogler acknowledge support by the Deutsche Forschungsgemeinschaft (grant LI 901/3-1). For the helpful comments and suggestions they provided on earlier versions of the paper, we thank seminar and conference participants at the University of Kiel, University of Cologne, University of Université catholique de Louvain (CORE), 2013 Spatial Statistics conference (Ohio), the 2013 Econometric Society European Meeting (Gothenburg), the 2013 International Conference on Computational and Financial Econometrics (London), the 2014 World Conference of the Spatial Econometrics Association (Zürich), the 2014 International Conference on Computational Statistics (Geneva), the 2014 ERSA Congress (Saint Petersburg), the Statistische Woche 2014 (Hannover). A former version of this paper circulated under the title 'Analysis of discrete dependent variable models with spatial correlation'.

References

Alan, G., Liu, J.W.H., 1981. Computer solution of large sparse positive definite systems. Prentice-Hall.

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., 2006. Spatial Econometrics. Springer.

Banerjee, S., Wall, M.M., Carlin, B.P., 2003. Frailty modeling for spatially correlated survival data, with applications to infant mortality in Minnesota. Biostatistics 4, 123–142.

Bauwens, L., Galli, F., 2009. Efficient importance sampling for ML estimation of SCD models. Computational Statistics and Data Analysis 53, 1974–1992.

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

Bhat, C.R., 2011. The maximum approximated composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transportation Research Part B: Methodological 45, 923–939.

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.

Buczkowska, S., de Lapparent, M., 2014. Location choices of newly created establishments: Spatial patterns at the aggregate level. Regional Science and Urban Economics 48, 68–81.

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

DeJong, D.N., Liesenfeld, R., Moura, G.V., Richard, J.-F., Dharmarajan, H., 2013. Efficient likelihood evaluation of state-space representations. Reviews of Economic Studies 80, 538–567.

Elhorst, P., Heijnen, P., Samarina, A., Jacobs, J., 2013. State transfers at different moments in time: A spatial probit approach. Working paper 13006-EEF, University of Groningen.

Fischer, M.M., Nijkamp, P., 2014. Handbook of Regional Science. Springer-Verlag, Berlin.

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.

Gourieroux, C., Monfort, A., 1996. Simulation-based Econometric Methods. Oxford University Press.

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 Analysis 52, 4184–4202.

Hafner, C.M., Manner, H., 2012. Dynamic stochastic copula models: estimation, inference and applications. Journal of Applied Econometrics 27, 269–295.

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

Heagerty, P.J., Lele, S.R., 1998. A composite likelihood approach to binary spatial data. Journal of the American Statistical Association 93, 1099–1111.

Jung, R.C., Liesenfeld, R., Richard, J.-F., 2011. Dynamic factor models for multivariate count data: An application to stock-market trading activity. Journal of Business and Economic Statistics 29, 73–85.

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

Kleppe, T.S., Skaug, H.J., 2012. Fitting general stochastic volatility models using Laplace accelerated sequential importance sampling. Computational Statistics and Data Analysis 56, 3105–3119.

Kneib, T., 2005. Geoadditive hazard regression for interval censored survival times. Working Paper 447, Sonderforschungsbereich 386, Ludwig-Maximilians-Universität München.

Koopman, S.J., Lucas, A., Scharth, M., 2014. Numerically accelerated importance sampling for nonlinear non-Gaussian state space models. Journal of Business and Economic Statistic, forthcoming.

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.

Lee, L.-F., 1997. Simulated maximum likelihood estimation of dynamic discrete choice models – some Monte Carlo results. Journal of Econometrics 82, 1–35.

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., Moura, G.V., Richard, J.-F., 2010. Determinants and dynamics of current account reversals: An empirical analysis. Oxford Bulletin of Economics and Statistics 72, 486–517.

Liesenfeld, R., Richard, J.-F., 2003. Univariate and multivariate stochastic volatility models: Estimation and diagnostics. Journal of Empirical Finance 10, 505–531.

Liesenfeld, R., Richard, J.-F., 2006. Classical and Bayesian analysis of univariate and multivariate stochastic volatility models. Econometric Reviews 25, 335–360.

Liesenfeld, R., Richard, J.-F., 2008. Improving MCMC using efficient importance sampling. Computational Statistics and Data Analysis 53, 272–288.

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.

McMillen, D.P., 1995. Spatial effects in probit models: A Monte Carlo investigation. In Anselin, L., Florax, R. (eds.), New Directions in Spatial Econometrics. Springer-Verlag, Berlin, 169–195.

Militino, A.F., Ugarte, M.D., 1999. Analyzing censored spatial data. Mathematical Geology 31, 551–561.

Pace, R.K., 2014. Maximum likelihood estimation. In Fischer, M.M., Nijkamp, P. (eds.), Handbook of Regional Science. Springer-Verlag, Berlin, 1553–1569.

Pace, R.K., Barry, R., 1997. Quick computation of regressions with a spatially autoregressive dependent variable. Geographical Analysis 29, 232–247.

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.

Pastorello, S., Rossi, E., 2010. Efficient importance sampling maximum likelihood estimation of stochastic differential equations. Computational Statistics and Data Analysis 54, 2753–2762.

Pinske, J., Slade, M.E., 1998. Contracting in space: An application of spatial statistics to discrete choice models. Journal of Econometrics 85, 125–154.

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

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society, Series B, 319–392.

Scharth, M., Kohn, R., 2013. Particle efficient importance sampling. Working paper, University of New South Wales, Australian School of Business.

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

Toscas, P.J., 2010. Spatial modelling of left censored water quality data. Environmetrics 21, 632–644.

Vijverberg, W.P.M., 1997. Monte Carlo evaluation of multivariate normal probabilities. Journal of Econometrics 76, 281–307.

Wang, H., Iglesias, E.M., Wooldridge, J.M., 2013. Partial maximum likelihood estimation of spatial probit models. Journal of Econometrics 172, 77–89.

Wang, X., 2014. Limited and censored dependent variable models. In Fischer, M.M., Nijkamp, P. (eds.), Handbook of Regional Science. Springer-Verlag, Berlin, 1619–1635.

Wang, X., Kockelman, K., 2009. Bayesian inference for ordered response data with a dynamic spatial-ordered probit model. Journal of Regional Science 49, 877–913.

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.

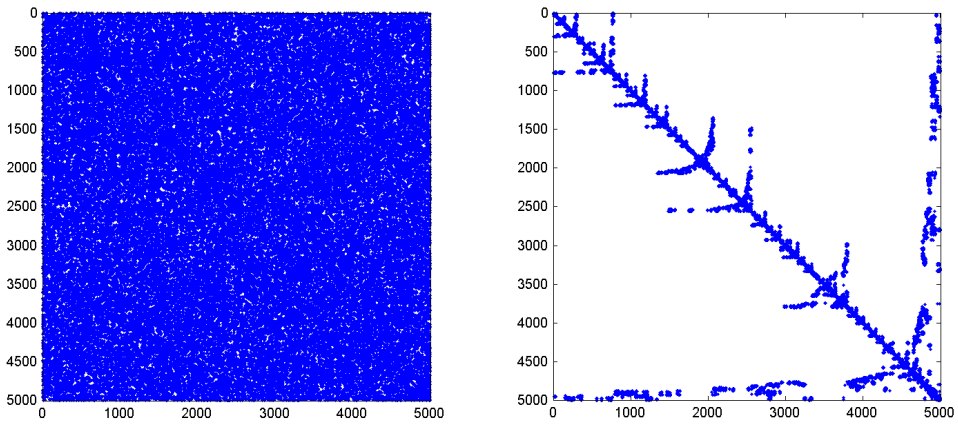

Figure 1. Left panel: (5000×5000) *spatial precision matrix* H *as defined in Equations (3) and using a spatial weight matrix* W *assigning six neighbors to each spatial unit* i*; blue dots indicate the non-zero elements; right panel:* H *after reordering the spatial units using the symmetric minimum degree permutation (see Amestoy et al., 1996).*

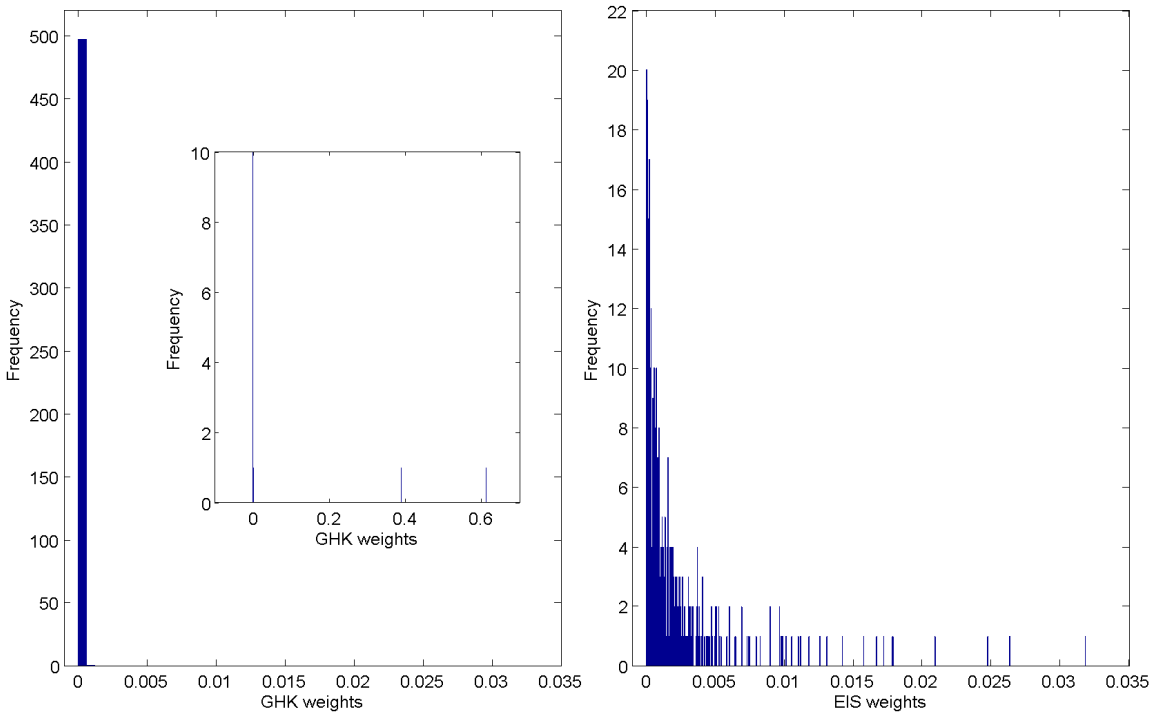

Figure 2. Histogram of the normalized IS weights $\tilde{w}^{(s)}/\sum_{\tau=1}^{S}\tilde{w}^{(\tau)}$ *of the GHK (plotted for different scales) and EIS procedure with* $S = 500$ *for the likelihood of the SAE probit model with* $\rho = 0.85$*. The model parameters* ψ *are set to their true values.*

Table 1. ML-EIS, ML-GHK and ML-GHK^{PL} for the SAL and SAE Probit Model

Parameter	True	Statistical properties			True ML	Numerical properties		
		GHK	EIS	GHK$^{\mathrm{PL}}$		GHK	EIS	GHK$^{\mathrm{PL}}$
SAL-probit model								
ρ	.750	.740	.750	.750	.748	.739	.748	.748
		(.006)	(.005)	(.005)		(.003)	(.0001)	(.0002)
		[.012]	[.005]	[.005]		[.009]	[.0001]	[.0006]
β_1	−1.500	−1.354	−1.498	−1.482	−1.502	−1.371	−1.502	−1.496
		(.061)	(.050)	(.055)		(.025)	(.0006)	(.002)
		[.158]	[.050]	[.057]		[.134]	[.0007]	[.007]
β_2	3.000	2.726	3.007	2.975	3.039	2.770	3.038	3.020
		(.121)	(.108)	(.114)		(.049)	(.001)	(.004)
		[.300]	[.108]	[.117]		[.273]	[.001]	[.019]
ρ	.850	.839	.850	.851	0.851	.840	.851	.853
		(.005)	(.003)	(.004)		(.003)	(.0001)	(.0002)
		[.012]	[.003]	[.004]		[.011]	[.0001]	[.002]
β_1	−1.500	−1.243	−1.516	−1.475	−1.489	−1.220	−1.488	−1.454
		(.076)	(.070)	(.072)		(.034)	(.002)	(.004)
		[.268]	[.072]	[.076]		[.272]	[.002]	[.036]
β_2	3.000	2.488	3.037	2.954	2.989	2.444	2.986	2.909
		(.145)	(.138)	(.140)		(.067)	(.003)	(.008)
		[.532]	[.142]	[.147]		[.549]	[.004]	[.080]
SAE-probit model								
ρ	.750	.380	.748	.771	.747	.364	.748	.800
		(.033)	(.028)	(.037)		(.024)	(.001)	(.005)
		[.372]	[.028]	[.042]		[.384]	[.001]	[.053]
β_1	−1.500	−1.108	−1.480	−1.520	−1.381	−1.027	−1.381	−1.504
		(.068)	(.111)	(.129)		(.011)	(.003)	(.020)
		[.398]	[.113]	[.130]		[.354]	[.003]	[.125]
β_2	3.000	2.246	2.987	3.078	2.879	2.133	2.880	3.107
		(.096)	(.158)	(.216)		(.015)	(.006)	(.038)
		[.760]	[.159]	[.229]		[.746]	[.006]	[.231]
ρ	.850	.482	0.848	.849	.843	.513	.843	.854
		(.033)	(.016)	(.019)		(.021)	(.0009)	(.002)
		[.369]	[.016]	[.019]		[.331]	[.0009]	[.012]
β_1	−1.500	−.894	−1.483	−1.434	−1.444	−.898	−1.443	−1.476
		(.076)	(.140)	(.148)		(.018)	(.004)	(.012)
		[.610]	[.141]	[.162]		[.546]	[.004]	[.034]
β_2	3.000	1.829	3.013	2.937	2.845	1.759	2.844	2.821
		(.067)	(.176)	(.180)		(.019)	(.007)	(.019)
		[1.173]	[.177]	[.191]		[1.086]	[.008]	[.031]

NOTE: The reported numbers for ML-EIS, ML-GHK and ML-GHK$^{\mathrm{PL}}$ are mean, standard deviation (in parentheses) and RMSE (in brackets). The simulation sample size for EIS is $S = 20$ and for GHK and GHK$^{\mathrm{PL}}$ $S = 500$. 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.