A Bayesian Probit Model with Spatial Dependencies

Tony E. Smith
Department of Systems Engineering
University of Pennsylvania
Philadephia, PA 19104
email: tesmith@ssc.upenn.edu

James P. LeSage
Department of Economics
University of Toledo
Toledo, Ohio 43606
phone: (419) 530–4754
e-mail: jlesage@spatial-econometrics.com

April 8, 2002

Abstract

A Bayesian probit model with individual effects that exhibit spatial dependencies is set forth. Since probit models are often used to explain variation in individual choices, these models may well exhibit spatial interaction effects due to the varying spatial location of the decision makers. That is, individuals located at similar points in space may tend to exhibit similar choice behavior. The model proposed here allows for a parameter vector of spatial interaction effects that takes the form of a spatial autoregression. This model extends the class of Bayesian spatial logit/probit models presented in LeSage (2000) and relies on a hierachical construct that we estimate via Markov Chain Monte Carlo methods. We illustrate the model by applying it to the 1996 presidential election results for 3,110 US counties.

1 Introduction

Probit models with spatial dependencies were first studied by McMillen (1992), where an EM algorithm was developed to produce consistent (maximum likelihood) estimates for these models. As noted by McMillen, such estimation procedures tend to rely on asymptotic properties, and hence require large sample sizes for validity. An alternative hierarchical Bayesian approach to non-spatial probit models was introduced by Albert and Chib (1993) which is more computationally demanding, but provides a flexible framework for modeling with small sample sizes. LeSage (2000) first proposed extending Albert and Chib's approach to models involving spatial dependencies, and this work extends the class of models that can be analyzed in this framework. Our extension relies on an error structure that involves an additive error specification first introduced by Besag, et al. (1991) and subsequently employed by many authors (as for example in Gelman, et al. 1998). As will be shown, this approach allows both spatial dependencies and general spatial heteroscedasticity to be treated simultaneously.

The paper begins by motivating the basic probit model in terms of an explicit choice-theoretic context involving individual behavioral units. Since probit models are often used to explain variation in individual choices, these models may well exhibit spatial interaction effects due to the varying spatial location of the decision makers. That is, individuals located at similar points in space may tend to exhibit similar choice behavior. A key element in this context is the spatial grouping of individuals by region. Here we assume that individuals within each region are homogeneous, suggesting that all spatial dependencies and heteroscedastic effects occur at the regional level. We also show that the case of spatial dependencies between individuals can be handled by treating individuals as separate 'regions'.

The model proposed here is set forth in section 2. This model allows for a parameter vector of spatial interaction effects that takes the form of a spatial autoregression. It extends the class of Bayesian spatial logit/probit models presented in LeSage (2000) and relies on a hierachical construct also presented in this section. We estimate the model using Markov Chain Monte Carlo (MCMC) methods to derive estimates by simulating draws from the complete set of conditional distributions for parameters in the model. Section 3 sets forth these conditional distributions for our model.

Section 4 provides illustrations based on generated data sets as well as an application of the method to the voting decisions from 3,110 US counties in the 1996 presidential election. Conclusions are contained in section 5.

2 A Spatial Probit Model

For the sake of concreteness, we motivate the model in terms of an explicit choice-model formulation detailed in Amemiya (1985, section 9.2) in section 2.1. Section 2.2 sets forth the Bayesian hierarchical structure that we use.

2.1 Choices involving spatial agents

Suppose there exists data on the observed choices for a set of individuals distributed within a system of spatial regions (or zones), $i = 1, \ldots, m$. In particular, suppose that the relevant choice context involves two (mutually exclusive and collectively exhaustive) alternatives, which we label '0', and '1'. Examples might be a voting decision or specific type of purchase decision. The observed choice for each individual $k = 1, \ldots, n_i$ in region i is treated as the realization of a random *choice variable*, Y_{ik}, where $Y_{ik} = 1$ if individual k chooses alternative 1 and $Y_{ik} = 0$ otherwise. In addition, it is postulated that choices are based on *utility maximizing* behavior, where k's *utility* for each of these alternatives is assumed to be of the form:

$$\begin{aligned} U_{ik0} &= \gamma' \omega_{ik0} + \alpha_0' s_{ik} + \theta_{i0} + \varepsilon_{ik0} \\ U_{ik1} &= \gamma' \omega_{ik1} + \alpha_1' s_{ik} + \theta_{i1} + \varepsilon_{ik1} \end{aligned} \qquad (1)$$

Here ω_{ika} is a ω-dimensional vector of *observed attributes* of *alternative* $a (= 0, 1)$ taken to be relevant for k (possibly differing in value among individuals), and s_{ik} is an $s-$dimensional vector of *observed attributes of individual* k. It is convenient to assume that k's region of occupancy is always included as an observed attribute of k. To formalize this, we let $\delta_i(k) = 1$ if k is in region i and $\delta_i(k) = 0$ otherwise, and henceforth assume that $s_{ikj} = \delta_j(k)$ for $j = 1, \ldots, m$ (so that by assumption $s \geq m$). The terms $\theta_{ia} + \varepsilon_{ika}$, represent the contribution to utility of all other relevant *unobserved* properties of both i, k and a. These are separated into a *regional effect*, θ_{ia}, representing the unobserved utility components of alternative a common to all individuals in region i, and an *individualistic effect*, ε_{ika}, representing all other unobserved components. In particular, the individualistic components $(\varepsilon_{ika} : k = 1, \ldots, n_i)$ are taken to be *conditionally independent* given θ_{ia}, so that all unobserved dependencies between individual utilities for a within region i are assumed to be captured by θ_{ia}. If we let the *utility difference* for individual k be denoted by

$$
\begin{aligned}
z_{ik} &= U_{ik1} - U_{ik0} \\
&= \gamma'(\omega_{ik1} - \omega_{ik0}) + (\alpha_1 - \alpha_0)' s_{ik} + (\theta_{i1} - \theta_{i0}) + (\varepsilon_{ik1} - \varepsilon_{ik0}) \\
&= x_{ik}' \beta + \theta_i + \varepsilon_{ik} \qquad (2)
\end{aligned}
$$

with *parameter vector*, $\beta = (\gamma', \alpha_1' - \alpha_0')'$, and *attribute vector*, $x_{ik} = (\omega_{ik1}' - \omega_{ik0}', s_{ik}')'$, and with $\theta_i = \theta_{i1} - \theta_{i0}$ and $\varepsilon_i = \varepsilon_{i1} - \varepsilon_{i0}$, then it follows from the utility-maximization hypothesis that

$$
\Pr(Y_{ik} = 1) = \Pr(U_{ik1} > U_{ik0}) = \Pr(z_{ik} > 0) \qquad (3)
$$

At this point it should be emphasized that model [(2), (3)] has many alternative interpretations. Perhaps the most general interpretation is in terms of linear models with limited information: if the elements $x_{ikj}, j = 1, \ldots, q[= (\omega + s)]$, are regarded as general explanatory variables, then model [(2), (3)] can be interpreted as a standard linear model with 'grouped observations' in which only the events '$z_{ik} > 0$' and '$z_{ik} \leq 0$' are observed [as in Albert and Chib, 1993 for example]. However, we shall continue to appeal to the above choice-theoretic interpretation in motivating subsequent details of the model.

Turning next to the unobserved components of the model, it is postulated that all unobserved dependencies between the utility differences for individuals in *separate* regions are captured by dependencies between the regional effects $(\theta_i : i = 1, \ldots, m)$. In particular, the unobserved utility-difference aspects common to individuals in a given region i may be similar to those for individuals in neighboring regions. This is operationalized by assuming that the interaction-effects vector, θ, exhibits the following *spatial autoregressive* structure[1]

$$
\theta_i = \rho \sum_{j=1}^{m} w_{ij} \theta_j + u_i, \quad i = 1, \ldots, m \qquad (4)
$$

where nonnegative elements of the weights, w_{ij} are taken to reflect the degree of 'closeness' between regions i and j. In addition, it is assumed that $w_{ii} \equiv 0$ and row sums, $\sum_{j=1}^{m} w_{ij}$ are normalized to one[2], so that ρ can be taken to

[1] This *simultaneous* autoregressive specification of regional dependencies follows the spatial econometrics tradition [as for example in Anselin (1988) and McMillen (1992)]. An alternative specification is the *conditional* autoregressive scheme employed by Besag (1974,1991).

[2] Note that if a given region i is isolated (with no neighbors) then $w_{ij} = 0$ for all $j = 1, \ldots, m$. In this case, the 'normalized' weights are also zero.

reflect the overall degree of spatial influence (usually nonnegative). Finally, the residuals, u_i are assumed to be *iid* normal variates, with zero means and common variances σ^2. Now, if we let $\theta = (\theta_i : i = 1, \ldots, m)$ denote the *regional effects vector*, and similarly let $u = (u_i : i = 1, \ldots, m)$, then these assumptions can be summarized in vector form as

$$\theta = \rho W \theta + u, \quad u \sim N(0, \sigma^2 I_m) \tag{5}$$

where $W = (w_{ij} : i, j = 1, \ldots, m)$ and where I_m denotes the $m-$square identity matrix for each integer $m > 0$. It is convenient to solve for θ in terms of u which we will rely on in the sequel. Let

$$B_\rho = I_m - \rho W \tag{6}$$

and assume that B_ρ is nonsingular, then from (5):

$$\theta = B_\rho^{-1} u \Rightarrow \theta | (\rho, \sigma^2) \sim N[0, \sigma^2 (B_\rho' B_\rho)^{-1}] \tag{7}$$

Turning next to the individualistic components, ε_{ik}, observe that without further evidence about specific individuals in a given region i, it is reasonable to treat these components as *exchangeable* and hence to model the ε_{ik} as conditionally *iid* normal variates with zero means[3] and common variance v_i, given θ_i. In particular, regional differences in the v_i's allow for possible *heteroscedasticity* effects in the model.[4] Hence, if we now denote the vector of individualistic effects of region i by $\varepsilon_i = (\varepsilon_{ik} : k = 1, \ldots, n_i)'$, then our assumptions imply that $\varepsilon_i | \theta_i \sim N(0, v_i I_{n_i})$.

We can express the full *individualistic effects* vector $\varepsilon = (\varepsilon_i' : i = 1, \ldots, m)'$ as

$$\varepsilon | \theta \sim N(0, V) \tag{8}$$

where the full covariance matrix V is shown in (9).

$$V = \begin{pmatrix} v_1 I_{n_1} & & \\ & \ddots & \\ & & v_m I_{n_m} \end{pmatrix} \tag{9}$$

[3] This zero-mean convention allows one to interpret the beta coefficient corresponding to the regional fixed-effect column, $\delta_i(\cdot)$ as the implicit mean of each ε_{ik}.

[4] It should be noted that the presence of regional dependencies (i.e., nonzero off-diagonal elements in B_ρ also generates heteroscedasticity effects (as discussed for example in McMillen, 1992). Hence the variances, v_i are implicitly taken to reflect regional heteroscedasticity effects other than spatial dependencies.

We emphasize here that as motivated earlier, all components of ε are assumed to be conditionally independent given θ.

Expression (2) can also be written in vector form by setting $z_i = (Z_{ik} : k = 1, \ldots, n_i)'$ and $X_i = (x_{ik} : k = 1, \ldots, n_i)'$, so the utility differences for each region i take the form:

$$z_i = X_i\beta + \theta_i \mathbf{1}_i + \varepsilon_i, \quad i = 1, \ldots, m \tag{10}$$

where $\mathbf{1}_i = (1, \ldots, 1)'$ denotes the n_i-dimensional unit vector. Then by setting $n = \sum_i n_i$ and defining the $n-$vectors $z = (z_i' : i = 1, \ldots, m)'$ and $X = (X_i' : i = 1, \ldots, m)'$,[5] we can reduce (10) to the single vector equation,

$$z = X\beta + \Delta\theta + \varepsilon \tag{11}$$

where

$$\Delta = \begin{pmatrix} \mathbf{1}_1 & & \\ & \ddots & \\ & & \mathbf{1}_m \end{pmatrix} \tag{12}$$

If the vector of *regional variances* is denoted by $v = (v_i : i = 1, \ldots, m)$, then the covariance matrix V in (8) can be written using this notation as

$$V = \text{diag}(\Delta v) \tag{13}$$

Finally, if $\delta(A)$ denotes the indicator function for each event A (in the appropriate underlying probability space), so that $\delta(A) = 1$ for all outcomes in which A occurs and $\delta(A) = 0$ otherwise, then by definition

$$\begin{aligned} \Pr(Y_{ik} = 1 | z_{ik}) &= \delta(z_{ik} > 0) \\ \Pr(Y_{ik} = 0 | z_{ik}) &= \delta(z_{ik} \leq 0) \end{aligned} \tag{14}$$

If the outcome value $Y = (Y_{ik} \in 0, 1)$, then [following Albert and Chib (1993)] these relations may be combined as follows:

$$\Pr(Y_{ik} = y_{ik}) = \delta(y_{ik} = 1)\delta(z_{ik} > 0) + \delta(y_{ik} = 0)\delta(z_{ik} \leq 0) \tag{15}$$

Hence, letting $Y = (Y_{ik} : k = \ldots, n_i, i = 1, \ldots, m)$, it follows that for each possible observed set of choice outcomes, $y \in \{0, 1\}^n$,

[5] Note again that by assumption X always contains m columns corresponding to the indicator functions, $\delta(\cdot), i = 1, \ldots, m$.

$$\Pr(Y = y|z) = \prod_{i=1}^{m} \prod_{k=1}^{n_i} \{\delta(y_{ik} = 1)\delta(z_{ik} > 0) + \delta(y_{ik} = 0)\delta(z_{ik} \leq 0)\} \quad (16)$$

2.2 Hierarchical Bayesian Extension

While this model could in principle be estimated using EM methods similar to McMillen (1992), the following Bayesian approach is more robust with respect to small sample sizes, and allows detailed analysis of parameter distributions obtained by simulating from the posterior distribution of the model. As with all Bayesian models, one begins by postulating suitable prior distributions for all parameters $(\beta, v, \sigma^2, \rho)$, and then derives the corresponding conditional posterior distributions given the observed data. In the analysis to follow it is convenient to represent v equivalently using the covariance matrix V and to write the relevant parameter vector as $(\beta, V, \sigma^2, \rho)$. The prior distributions employed for these parameters are taken to be *diffuse* priors wherever possible, and *conjugate* priors elsewhere. As is well known (see for example Gelman, et al. 1995), this choice of priors yields simple intuitive interpretations of the posterior means as weighted averages of standard maximum-likelihood estimators and prior mean values (developed in more detail below). The following prior distribution hypotheses are standard for linear models such as (12) [see for example Geweke, (1993) and LeSage, (1999)]:

$$\begin{aligned}
\beta &\sim N(c, T) && (17)\\
r/v_i &\sim \mathrm{ID}\chi^2(r) && (18)\\
1/\sigma^2 &\sim \Gamma(\alpha, \nu) && (19)\\
\rho &\sim U[(\lambda_{\min}^{-1}, \lambda_{\max}^{-1}] && (20)
\end{aligned}$$

Here β is assigned a *normal* conjugate prior, which can be made 'almost diffuse' by centering at $c = 0$ and setting $T = tI_q$, for some sufficiently large t. More generally, the mean vector, c and covariance matrix T are used by the investigator to reflect subjective prior information assigned as part of the model specification. The variances, σ^2 together with $(v_i : i = 1, \ldots, m)$, are given (conjugate) *inverse gamma* priors. A diffuse prior for σ^2 would involve setting the parameters $(\alpha = \nu = 0)$.

The prior distribution for each v_i is the *inverse chi-square* distribution, which is a special case of the inverse gamma. This choice has the practical advantage of yielding a simple $t-$distribution for each component of ε (as discussed in Geweke, 1993). Here the choice of values for the hyperparameter r is more critical in that this value plays a key role in the posterior estimates of heteroscedasticity among regions, which we discuss below.

We employ a *uniform prior* on ρ that is diffuse over the relevant range of ρ values for the model in (5). In particular, if $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalues of W, then (under our assumptions on W) it is well known that $\lambda_{\min} < 0, \lambda_{\max} > 0$, and that ρ must lie in the interval $[\lambda_{\min}^{-1}, \lambda_{\max}^{-1}]$ (see for example Lemma 2 in Sun et al., 1999). The densities corresponding to (17), (19), and (20) are given respectively by:

$$\begin{aligned} \pi(\beta) &\propto \exp[-\frac{1}{2}(\beta - c)'T^{-1}(\beta - c)] &(21)\\ \pi(\sigma^2) &\propto (\sigma^2)^{-(\alpha+1)}\exp\left(-\frac{\nu}{\sigma^2}\right) &(22)\\ \pi(\rho) &\propto 1 &(23)\end{aligned}$$

where the inverse gamma density in (22) can be found in standard Bayesian tests such as Gelman et al. (1995, p. 474). Note also that the diffuse density for σ^2 with $\alpha = \nu = 0$ is of the form $\pi(\sigma^2) \propto 1/\sigma^2$. Finally, the prior density of each $v_i, i = 1, \ldots, m$, can be obtained by observing from (8) that the variate, $\lambda = \lambda(v_i) = r/v_i$, has chi-square density

$$f(\lambda) \propto \lambda^{\frac{r}{2}-1}\exp\left(-\frac{\lambda}{2}\right) \tag{24}$$

This together with the Jacobian expression, $|d\lambda/dv_i| = r/(v_i^2)$, then implies that

$$\begin{aligned} \pi(v_i) &= f[\lambda(v_i)] \cdot \left|\frac{d\lambda}{dv_i}\right| \\ &= \left(\frac{r}{v_i}\right)^{\frac{r}{2}-1} \exp\left(-\frac{r/v_i}{2}\right) \cdot \frac{r}{v_i^2} \\ &\propto v_i^{-(\frac{r}{2}+1)}\exp\left(-\frac{r}{2v_i}\right) &(25)\end{aligned}$$

which is seen from (8) to be an inverse gamma distribution with parameters $\alpha = \nu = r/2$, [as in expression (6) of Geweke (1993)].

These prior parameter densities imply corresponding prior conditional densities for θ, ε, and z. To begin with observe from (7) that the prior conditional density of θ given (ρ, σ^2) is of the form

$$\pi(\theta|\rho,\sigma^2) \sim (\sigma^2)^{-m/2}|B_\rho|\text{exp}\left(-\frac{1}{2\sigma^2}\theta' B_\rho' B_\rho \theta\right) \tag{26}$$

and similarly (8) implies that the conditional prior density of ε given (θ, V) is

$$\pi(\varepsilon|V) \sim |V|^{-1/2}\text{exp}\left(-\frac{1}{2}\varepsilon' V^{-1}\varepsilon\right) \tag{27}$$

This in turn implies that the conditional prior density of z given $(\beta, \sigma^2, \theta)$ has the form

$$\begin{aligned} \pi(z|\beta,\theta,V) &\propto |V|^{-1/2}\text{exp}\left\{-\frac{1}{2}(z - X\beta - \Delta\theta)' V^{-1}(z - X\beta - \Delta\theta)\right\} \\ &= \prod_{i=1}^{m}\prod_{k=1}^{n_i}\left\{v_i^{-1/2}\text{exp}\left[-\frac{1}{2v_i}(z_{ik} - x_{ik}'\beta - \theta_i)^2\right]\right\} \end{aligned} \tag{28}$$

3 Estimating the model

Estimation will be achieved via Markov Chain Monte Carlo methods that sample sequentially from the complete set of conditional distributions for the parameters. To implement the MCMC sampling approach we need to derive the complete conditional distributions for all parameters in the model. Given these, we proceed to sample sequential draws from these distributions for the parameter values. Gelfand and Smith (1990) demonstrate that MCMC sampling from the sequence of complete conditional distributions for all parameters in the model produces a set of estimates that converge in the limit to the true (joint) posterior distribution of the parameters.

To derive the conditional posterior distributions, we use the basic Bayesian identity and the the prior densities from section 2,

$$p(\beta,\theta,\rho,\sigma^2,V,z|y)\cdot p(y) = p(y|\beta,\theta,\rho,\sigma^2,V,z)\cdot\pi(\beta,\theta,\rho,\sigma^2,V,z) \tag{29}$$

where $p(\cdot)$ indicates posterior densities (i.e., involving the y observations). This identity together with the assumed prior independence of β, ρ, σ^2, and V implies that the posterior joint density $p(\beta,\theta,\rho,\sigma^2,V,z|y)$ is given up to a constant of proportionality by

$$\begin{aligned} p(\beta,\theta,\rho,\sigma^2,V,z|y) &\propto p(y|z)\cdot\pi(z|\beta,\theta,V)\cdot\pi(\theta|\rho,\sigma^2) \\ &\cdot\ \pi(\beta)\cdot\pi(\rho)\cdot\pi(\sigma^2)\cdot\pi(V) \end{aligned} \tag{30}$$

Using this relation, we establish the appropriate conditional posterior distributions for each parameter in the model in sections 3.1 through 3.6.

3.1 The Conditional Posterior Distribution of β

From (30) it follows that

$$\begin{aligned} p(\beta|\star) &= \frac{p(\beta,\theta,\rho,\sigma^2,V,z|y)}{p(\theta,\rho,\sigma^2,V,z|y)} \propto p(\beta,\theta,\rho,\sigma^2,V,z|y) \\ &\propto \pi(z|\beta,\theta,V)\cdot\pi(\beta) \end{aligned} \tag{31}$$

where we use $\star$ to denote the conditioning arguments: $\theta, \rho, \sigma^2, V, z, y$. This together with (28) and (21) implies that

$$\begin{aligned} p(\beta|\star) &\propto \exp\left\{-\frac{1}{2}(z - X\beta - \Delta\theta)'V^{-1}(z - X\beta - \Delta\theta)\right\} \\ &\quad \cdot\exp\left\{-\frac{1}{2}(\beta - c)'T^{-1}(\beta - c)\right\} \end{aligned} \tag{32}$$

But since

$$\begin{aligned} -&\ \frac{1}{2}(z - X\beta - \Delta\theta)'V^{-1}(z - X\beta - \Delta\theta) - \frac{1}{2}(\beta - c)'T^{-1}(\beta - c) \qquad (33) \\ =&\ -\frac{1}{2}[\beta'X'V^{-1}X\beta - 2(z-\Delta\theta)'V^{-1}X\beta + \beta'T^{-1}\beta - 2c'T^{-1}\beta + \mathrm{C}] \\ =&\ -\frac{1}{2}\left\{\beta'(X'V^{-1}X + T^{-1})\beta - 2[X'V^{-1}(z-\Delta\theta) + T^{-1}c]\beta\right\} + \mathrm{C} \end{aligned}$$

where C includes all quantities not depending on β. It follows that if we now set

$$A = (X'V^{-1}X + T^{-1}) \tag{34}$$

and

$$b = X'V^{-1}(z - \Delta\theta) + T^{-1}c \tag{35}$$

and observe that both A and b are independent of β, then expression (32) can be rewritten as

$$
\begin{aligned}
p(\beta|\star) &\propto \exp[-\frac{1}{2}(\beta' A\beta - 2b'\beta)] \\
&\propto \exp[-\frac{1}{2}(\beta' A\beta - 2b'\beta + b'A^{-1}b)] \\
&\propto \exp[-\frac{1}{2}(\beta - A^{-1}b)'A(\beta - A^{-1}b)] \qquad (36)
\end{aligned}
$$

Therefore, the conditional posterior density of β is proportional to a multinormal density with mean vector $A^{-1}b$, and covariance matrix, A^{-1}, which we express as:

$$\beta|(\theta, \rho, \sigma^2, V, z, y) \sim N(A^{-1}b, A^{-1}) \qquad (37)$$

This can be viewed as an instance of the more general posterior in expression (13) of Geweke (1993) where his G, Ω^{-1}, y and g are here given by $I, V^{-1}, z-\theta$, and c respectively.

As is well known (see for example the discussion in Gelman, et al. 1995, p. 79), this posterior distribution can be viewed as a weighted average of prior and sample data information in the following sense. If one treats $z - \Delta\theta$ in (11) as 'data' and defines the corresponding maximum-likelihood estimator of β for this linear model by

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}(z - \Delta\theta) \qquad (38)$$

then it follows from (34) and (35) that the posterior mean of β takes the form

$$
\begin{aligned}
E(\beta|\theta, \rho, \sigma^2, V, z, y) &= (X'V^{-1}X + T^{-1})^{-1}[X'V^{-1}(z - \Delta\theta) + T^{-1}c] \\
&= (X'V^{-1}X + T^{-1})^{-1}[X'V^{-1}\hat{\beta} + T^{-1}c] \qquad (39)
\end{aligned}
$$

For the case of a single explanatory variable in X where $q = 1$, the right hand side of (39) represents a simple convex combination of c and $\hat{\beta}$. More generally, this posterior mean represents a *matrix-weighted average* of the prior mean c, and the maximum-likelihood estimate, $\hat{\beta}$. Note that as the quality of sample data information increases (i.e., the variances v_i become smaller) or the quantity of sample information increases (i.e., sample sizes n_i become larger) the weight placed on $\hat{\beta}$ increases.

3.2 The Conditional Posterior Distribution of θ

As will become clear below, the conditional posterior for θ is in many ways similar to that for β. Here we let $\star$ represent the conditioning arguments $\beta, \rho, \sigma^2, V, z, y$. First, note that using the same argument as in (30) and (31), together with (26) and (28) we can write

$$\begin{aligned}
p(\theta|\star) &\propto \pi(z|\beta,\theta,V)\cdot\pi(\theta|\rho,\sigma^2) \\
&\propto \exp\left\{-\frac{1}{2}[\Delta\theta-(z-X\beta)]'V^{-1}[\Delta\theta-(z-X\beta)]\right\}\cdot \\
&\quad \exp\left\{-\frac{1}{2\sigma}\theta' B_\rho' B_\rho \theta\right\} \\
&= \exp\left\{-\frac{1}{2}[\theta'\Delta' V^{-1}\Delta\theta - 2(z-X\beta)'V^{-1}\Delta\theta + \theta'(\sigma^{-2}B_\rho' B_\rho)\theta]\right\} \\
&= \exp\left\{-\frac{1}{2}[\theta'(\sigma^{-2}B_\rho' B_\rho + \Delta' V^{-1}\Delta)\theta - 2(z-X\beta)'V^{-1}\Delta\theta]\right\} \quad (40)
\end{aligned}$$

A comparison of (40) with (33) shows that by setting $A_0 = \sigma^{-2}B_\rho' B_\rho + \Delta' V^{-1}\Delta$ and $b_0 = \Delta' V^{-1}(z - X\beta)$, the conditional posterior density for θ must be proportional to a multinormal distribution

$$p(\theta|\beta,\rho,\sigma^2,V,z,y) \sim N(A_0^{-1}b_0, A_0^{-1}) \quad (41)$$

where the mean vector is $A_0^{-1}b_0$ and the covariance matrix is A_0^{-1}. Unlike the case of β however, the mean and covariance matrix of θ involve the inverse of the mxm matrix A_0 which depends on ρ. This implies that this matrix inverse must be computed on each MCMC draw during the estimation procedure. Typically thousands of draws will be needed to produce a posterior estimate of the parameter distribution for θ, suggesting that this approach to sampling from the conditional distribution of θ may be costly in terms of time if m is large. In our illustration in section 5 we rely on a sample of 3,110 US counties and the 48 contiguous states, so that $m = 48$. In this case, computing the inverse was relatively fast allowing us to produce 2,500 draws in 37 seconds using a compiled c-language program on an Anthalon 1200 MHz. processor.

In the Appendix we provide an alternative approach that involves only univariate normal distributions for each element θ_i conditional on all other elements of θ excluding the ith element. This approach is amenable to computation for much larger sizes for m, but suffers from the need to evaluation m univariate conditional distributions to obtain the vector of θ parameter estimates on each pass through the MCMC sampler. This slows down the computations, but it does not suffer from the need to manipulate or invert large matrices.

3.3 Conditional Posterior Distribution for ρ

To determine the conditional posterior for ρ, observe that using (30) we have:

$$\begin{aligned} p(\rho|\beta,\theta,\sigma^2,V,z,y) &\propto \frac{p(\rho,\beta,\theta,\sigma^2,V,z|y)}{p(\beta,\theta,\sigma^2,V,z|y)} \propto p(\rho,\beta,\theta,\sigma^2,V,z|y) \\ &\propto \pi(\theta|\rho,\sigma^2)\cdot\pi(\rho) \end{aligned} \tag{42}$$

which together with (26) and (23) implies that

$$p(\rho|\beta,\theta,\sigma^2,V,z,y) \propto |B_\rho|\text{exp}\left(-\frac{1}{2\sigma^2}\theta' B_\rho' B_\rho \theta\right) \tag{43}$$

where $\rho \in [\lambda_{\min}^{-1}, \lambda_{\max}^{-1}]$. As noted in LeSage (2000) this is not reducible to a standard distribution, so we might adopt a Metropolis-Hastings step during the MCMC sampling procedures. LeSage (1999) suggests a normal or $t-$ distribution be used as a transition kernel in the Metropolis-Hastings step. Additionally, the restriction of ρ to the interval $[\lambda_{\min}^{-1}, \lambda_{\max}^{-1}]$ can be implemented using a rejection-sampling step during the MCMC sampling.

Another approach that is feasible for this model is to rely on univariate numerical integration to obtain the the conditional posterior density of ρ. The size of B_ρ will be based on the number of regions, which is typically much smaller than the number of observations, making it computationally simple to carry out univariate numerical integration on each pass through the MCMC sampler. Specifically, we can use the properties of the inverted gamma distribution to integrate out the nuisance parameter σ obtaining:

$$p(\rho|\beta,\theta,V,z,y) \propto |B_\rho| \left[(m)^{-1}\theta' B_\rho' B_\rho \theta\right]^{-m/2} \pi(\rho) \tag{44}$$

The conditional posterior distribution over a grid of ρ values can be obtained numerically using univariate integration, to find the normalizing constant, where the limits of integration extend over $\rho \in [\lambda_{\min}^{-1}, \lambda_{\max}^{-1}]$. Having achieved a grid approximation to the conditional posterior for ρ, we then draw from this using inversion. An advantage of this approach over the Metropolis-Hastings method is that each pass through the sampler produces a draw for ρ, whereas acceptance rates in the Metropolis-Hastings method are usually around 50 percent requiring twice as many passes through the sampler to produce the same number of draws for ρ.

Note that we can construct a vector of log determinant values for $\log|B_\rho|$ based on a grid of values for $\rho \in [\lambda_{\min}^{-1}, \lambda_{\max}^{-1}]$ prior to beginning our MCMC sampler. On each pass through the sampler we then need only compute the quantity $(m)^{-1}\theta' B_\rho' B_\rho \theta$, over this grid of ρ values, which can be done rapidly for a small spatial matrix W reflecting connectivity relations between the regions in our model.

For our applications presented in section 4, the number of observations was based on $n = 3,110$ US counties and the regions $m = 48$, based on US states. An MCMC sampler implemented in the c-language produced 2,500 draws in 30 seconds on a 1200 MHz. Anthalon desktop computer. Using the interpreted MATLAB language, the MCMC sampler produced the same 2,500 draws in 107 seconds. In a typical application 2,500 draws would suffice for convergence of the sampler to produce adequate posterior inferences.

3.4 The Conditional Posterior Distribution of σ^2

To determine the conditional posterior of σ^2, the same argument as in (42) along with (26) and (22) implies that

$$\begin{aligned} p(\sigma^2|\beta,\theta,\rho,V,z,y) &\propto \pi(\theta|\rho,\sigma^2)\cdot\pi(\sigma^2) \\ &\propto (\sigma^2)^{-m/2}\exp\left(-\frac{1}{2\sigma^2}\theta' B_\rho' B_\rho\theta\right)\cdot \\ &\quad (\sigma^2)^{-(\alpha+1)}\exp\left(-\frac{\nu}{\sigma^2}\right) \end{aligned} \tag{45}$$

Hence, we have

$$p(\sigma^2|\beta,\theta,\rho,V,z,y) \propto (\sigma^2)^{-(\frac{m}{2}+\alpha+1)}\exp\left[-\theta' B_\rho' B_\rho\theta + \frac{2\nu}{2\sigma^2}\right] \tag{46}$$

which is seen from (22) to be proportional to an inverse gamma distribution with parameters $(m/2)+\alpha$ and $\theta' B_\rho' B_\rho\theta + 2\nu$. Following Geweke (1993), we may also express this posterior in terms of the chi-square distribution as follows. Let $\lambda = [\theta' B_\rho' B_\rho\theta + 2\nu]/\sigma^2$, so that $\sigma^2 = [\theta' B_\rho' B_\rho\theta + 2\nu]/\lambda$ implies $|d\sigma^2/d\lambda| = [\theta' B_\rho' B_\rho\theta + 2\nu]/(\lambda^2)$, then it follows that

$$f(\lambda) = \pi[\sigma^2(\lambda)]\cdot\left|\frac{d\sigma^2}{d\lambda}\right|$$

$$\begin{aligned}
&= \left(\frac{\theta' B'_\rho B_\rho \theta + 2\nu}{\lambda}\right)^{-(\frac{m}{2}+\alpha+1)} \exp\left(-\frac{\lambda}{2}\right) \cdot \frac{\theta' B'_\rho B_\rho \theta + 2\nu}{\lambda^2} \\
&\propto \lambda^{\frac{m}{2}+\alpha+1-2} \exp\left(-\frac{\lambda}{2}\right) \\
&= \lambda^{\left(\frac{m+2\alpha}{2}\right)-1} \exp\left(-\frac{\lambda}{2}\right) \qquad (47)
\end{aligned}$$

Hence the density of λ is proportional to a chi-square density with $m + 2\alpha$ degrees of freedom, and we may also express the conditional posterior of σ^2 as

$$\frac{\theta' B'_\rho B_\rho \theta + 2\nu}{\sigma^2} |(\beta, \theta, \rho, V, z, y) \sim \chi^2(m + 2\alpha) \qquad (48)$$

3.5 The Conditional Posterior Distribution of v

To determine the conditional posterior distribution of $v = (v_i : i = 1 \ldots, m)$, we observe that from the same argument as in (30) and (31), together with (28), (18), and (25) that if we let $v_{-i} = (v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_m)$ for each i, let $e = z - X\beta - \Delta\theta$ and also let $\star$ represent the conditioning arguments (in this case: $\beta, \theta, \rho, \sigma^2, v_{-i}, z, y$), then

$$\begin{aligned}
p(v_i|\star) &\propto \pi(z|\beta, \theta, V) \cdot \prod_{i=1}^{m} \pi(v_i) \\
&\propto |V|^{-1/2} \exp\left(-\frac{1}{2} e' V^{-1} e\right) \cdot \pi(v_i) \\
&\propto |V|^{-1/2} \exp\left(-\frac{1}{2} e' V^{-1} e\right) \cdot v_i^{-\left(\frac{r}{2}+1\right)} \exp\left(-\frac{r}{2v_i}\right) \qquad (49)
\end{aligned}$$

But since $V = \text{diag}(\Delta v) \rightarrow |V|^{-1/2} = \prod_{i=1}^{m} (v_i^{-n_i/2})$ and $e' V^{-1} e = \sum_{i=1}^{m} \sum_{k=1}^{n_i} e_{ik}^2 / v_i = \sum_{i=1}^{m} e'_i e_i / v_i$, where $e_i = (e_{ik} : k = 1, \ldots, n_i)$ we have

$$\begin{aligned}
p(v_i|\star) &\propto \prod_{j=1}^{m} (v_j^{-n_j/2}) \cdot \prod_{j=1}^{m} \exp\left(-\frac{e'_j e_j}{2v_j}\right) \cdot v_j^{-\left(\frac{r}{2}+1\right)} \exp\left(-\frac{r}{2v_j}\right) \\
&\propto v_i^{-n_i/2} \exp\left(-\frac{e'_i e_i}{2v_i}\right) \cdot v_i^{-\left(\frac{r}{2}+1\right)} \exp\left(-\frac{r}{2v_i}\right) \\
&= v_i^{-\left(\frac{r+n_i}{2}+1\right)} \exp\left(-\frac{e'_i e_i + r}{2v_i}\right) \qquad (50)
\end{aligned}$$

and may conclude from (22) that the conditional posterior distribution of each v_i is proportional to an inverse gamma distribution with parameters $(r+n_i)/2$ and $(e_i'e_i+r)/2$. As with σ^2, this may also be expressed in terms of the chi-square distribution as follows. If we let $\lambda = (e_i'e_i+r)/v_i$, so that $v_i(\lambda) = (e_i'e_i+r)/\lambda_i$ implies $|dv_i/d\lambda_i| = (e_i'e_i+r)/\lambda_i^2$, then it follows that

$$\begin{aligned} f(\lambda_i) &= \pi[v_i(\lambda_i)] \cdot \left|\frac{dv_i}{d\lambda_i}\right| \\ &= \left[\frac{e_i'e_i+r}{\lambda_i}\right]^{-\left(\frac{r+n_i}{2}+1\right)} \exp\left(-\frac{\lambda_i}{2}\right) \cdot \frac{e_i'e_i+r}{\lambda_i^2} \\ &= \lambda_i^{\left(\frac{r+n_i}{2}\right)-1} \exp\left(-\frac{\lambda_i}{2}\right) \end{aligned} \tag{51}$$

which is proportional to a chi-square density with $r+n_i$ degrees of freedom. Hence in a manner similar to (48) we may express the conditional posterior of each v_i as

$$\frac{e_i'e_i+r}{v_i}|(\beta,\theta,\rho,\sigma^2,v_{-i},z,y) \sim \chi^2(r+n_i) \tag{52}$$

In this form, it is instructive to notice that the posterior mean of v_i has a 'weighted average' interpretation similar to that of β discussed above. To see this, note first that from (18) v_i/r has an inverse chi-squared prior distribution with r degrees of freedom, and the mean of the inverse chi-square with ν degrees of freedom is given by $1/(\nu-2)$, it follows that the prior mean of v_i is $\mu_i = E(v_i) = rE(v_i/r) = r/(r-2)$ for $r > 2$. Next observe from (52) that the random variable $v_i/(e_i'e_i+r)$ is also conditionally distributed as inverse chi-square with $r+n_i$ degrees of freedom, so that

$$\begin{aligned} \frac{1}{e_i'e_i+r}E(v_i|\beta,\theta,\rho,\sigma^2,v_{-i},z,y) &= E(\frac{v_i}{e_i'e_i+r}|\beta,\theta,\rho,\sigma^2,v_{-i},z,y) \\ &= \frac{1}{(n_i+r)-2} \end{aligned} \tag{53}$$

But if the maximum-likelihood estimator for v_i given the 'residual' vector e_i is denoted by $\hat{v}_i = (1/n_i)e_i'e_i$, then it follows from (53) that

$$\begin{aligned} E(v_i|\beta,\theta,\rho,\sigma^2,v_{-i},z,y) &= \frac{e_i'e_i+r}{(n_i+r)-2} \\ &= \frac{n_i\hat{v}_i+(r-2)\mu_i}{n_i+(r-2)} \end{aligned} \tag{54}$$

From this we see that the posterior mean of v_i is a weighted average of the maximum-likelihood estimator, $\hat{v}_i$, and the prior mean, μ_i of v_i. Moreover, we have the case where more weight is given to the sample information embodied in $\hat{v}_i$ as the sample size, n_i increases. Even for relatively small sample sizes in each region, these posterior means may be expected to capture possible heteroscedasticity effects between regions. Note also that the value of the hyperparameter, r is critical here. In particular large values of r would result in the heteroscedasticity effects being overwhelmed. LeSage (1999) suggests that the range of values $2 < r \leq 7$ is appropriate for most purposes and recommends a value $r = 4$ as a rule-of-thumb.

3.6 The Conditional Posterior Distribution of z

Finally, we construct a key posterior distribution for this model, namely that of the utility-difference vector z. By the same argument as in (30) and (31) now taken together with (16) and (28), we see that

$$\begin{aligned} p(z|\beta,\theta,\rho,\sigma^2,V,y) &\propto p(y|z)\cdot\pi(z|\beta,\theta,V) \\ &\propto \prod_{i=1}^{m}\prod_{k=1}^{n_i}\{\delta(y_{ik}=1)\delta(z_{ik}>0)+\delta(y_{ik}=0)\delta(z_{ik}\leq 0)\} \\ &\quad \prod_{i=1}^{m}\prod_{k=1}^{n_i}\left\{v_{ik}^{-1/2}\exp\left[-\frac{1}{2v_i}(z_{ik}-x'_{ik}\beta-\theta_i)^2\right]\right\} \end{aligned} \tag{55}$$

Hence by letting $z_{-ik} = (z_{11},\ldots,z_{i,k-1},z_{i,k+1},\ldots,z_{mn_m})$ for each individual k in region i, it follows at once from (55) that

$$\begin{aligned} p(z_{ik}|\star) &\propto v_i^{-1/2}\exp\left[-\frac{1}{2v_i}(z_{ik}-x'_{ik}\beta-\theta_i)^2\right]\cdot \\ &\qquad \{\delta(y_{ik}=1)\delta(z_{ik}>0)+\delta(y_{ik}=0)\delta(z_{ik}\leq 0)\} \end{aligned} \tag{56}$$

Thus we see that for each ik, the conditional posterior of z_{ik} is a *truncated* normal distribution, which can be expressed as follows:

$$z_{ik}|\star \sim \left\{ \begin{array}{lll} N(x'_i\beta+\theta_i,v_i) & \text{left-truncated at 0,} & \text{if} \quad y_i = 1 \\ N(x'_i\beta+\theta_i,v_i) & \text{right-truncated at 0,} & \text{if} \quad y_i = 0 \end{array} \right. \tag{57}$$

where $\star$ denotes the conditioning arguments, $(\beta,\theta,\rho,\sigma^2,V,z_{-ik},y)$.

3.7 The MCMC sampler

By way of summary, the MCMC estimation scheme involves starting with arbitrary initial values for the parameters which we denote $\beta^0, \theta^0, \rho^0, \sigma^0, V^0$ and the latent variable z^0. We then sample sequentially from the following set of conditional distributions for the parameters in our model.

1. $p(\beta|\theta^0, \rho^0, \sigma^0, V^0, y^0, z)$, which is a multinormal distribution with mean and variance defined in (37). This updated value for the parameter vector β we label β^1.

2. $p(\theta|\beta^1, \rho^0, \sigma^0, V^0, y^0, z)$, which we sample from a multinormal distribution in (41) (or the set of n univariate normal distributions with means and variances presented in (75) of the Appendix.) These updated parameters we label θ^1. Note that we employ the updated value β^1 when evaluating this conditional distribution.

3. $p(\sigma^2|\beta^1, \theta^1, \rho^0, V^0, y^0, z)$, which is chi-squared distributed $n + 2\alpha$ degrees of freedom as shown in (48). Label this updated parameter σ^1 and note that we will continue to employ updated values of previously sampled parameters when evaluating these conditional densities.

4. $p(\rho|\beta^1, \theta^1, \sigma^1, V^0, y^0, z)$, which can be obtained using a Metropolis-Hastings approach described in LeSage (2000) based on a normal candidate density along with rejection sampling to constrain ρ to the desired interval. One can also rely on univariate numerical integration to find the conditional posterior on each pass through the sampler. This was the approach we took to produce the estimates reported in section 5.

5. $p(v_i|\beta^1, \theta^1, \rho^1, \sigma^1, v_{-i}, y^0, z)$ which can be obtained from the chi-squared distribution shown in (52).

6. $p(y|\beta^1, \theta^1, \rho^1, \sigma^1, V^1, z)$, which requires draws from left- or right-truncated normal distributions based on (57).

We now return to step 1) employing the updated parameter values in place of the initial values $\beta^0, \theta^0, \rho^0, \sigma^0, V^0$ and the updated latent variable y^1 in place of the initial y^0. On each pass through the sequence we collect the parameter draws which are used to construct a posterior distribution for the parameters in our model.

In the case of θ and V, the parameters take the form of an mn_m-vector, which is also true of the draws for the latent variable vector y. Storing these values over a sampling run involving thousands of draws when m is large would require large amounts of computer memory. One option is to simply compute a mean vector which doesn't require storage of the draws for these vectors. The posterior mean may often provide an adequate basis for posterior inferences regarding parameters like v_i and the latent variable y. Another option is to write these values to a disk file during the sampling process, which might tend to slow down the algorithm slightly.

4 Some special cases

In this section we set forth distributional results for two cases which might be of special interest. First, we consider the case in which all individuals are interchangeable, i.e., in which *homoscedasticity* is postulated to hold among regions. We then consider a case where spatial dependencies are presumed to occur among individuals themselves, so that each individual is treated as a region.

4.1 The homoscedastic case

This represents a situation where individual variances are assumed equal across all regions, so the regional variance vector, v reduces to a scalar producing the simple form of covariance matrix shown in (58).

$$V = vI_n \tag{58}$$

With this version of the model, the conditional posterior densities for β, θ, ρ, and σ^2 remain the same. The only change worthy of mention occurs in the conditional posterior density for v. Here it can be readily verified by using the same definitions, $e = z - X\beta - \Delta\theta$ and $n = \sum_i n_i$, that the conditional posterior density for v given $(\beta, \theta, \rho, \sigma^2, z, y)$ is identical to (52) with all subscripts i removed, i.e.,

$$\frac{e'e + r}{v} |(\beta, \theta, \rho, \sigma^2, z, y) \sim \chi^2(r + n) \tag{59}$$

In addition, the conditional posterior density for each z_{ik} given $(\beta, \theta, \rho, \sigma^2, v, z_{-ik}, y)$ is identical to (57) with v_i replaced by v. Of course, for large n relative to r this approaches the usual $\chi^2(n)$ distribution for σ^2 in the homoscedastic Bayesian linear model.

4.2 The individual spatial-dependency case

Another special case is where individuals are treated as 'regions' denoted by the index i.. In this case we are essentially setting $m = n$ and $n_i = 1$ for all $i = 1, \ldots, m$. Note that although one could in principle consider heteroscedastic effects among individuals, the existence of a single observation per individual renders estimation of such variances problematic at best. In this case, one might adopt a homoscedasticity hypothesis described in section 4.2 and use v to denote the common individual variance.[6] Here it can be verified that by simply replacing all occurences of $(ik, X_i, \theta_i \mathbf{1}_i, \Delta\theta, \Delta v)$ with $(i, x_i', \theta_i, \theta, v)$ respectively, and again using the definition of V in (58), the basic model in (2) through (16), together with the conditional posterior densities for β, ρ, and σ^2 continue to hold. In this homoscedastic context, the appropriate conditional posterior density for each $\theta_i, i = 1, \ldots, n(= m)$, again has the form (75), where the definitions of a_i and b_i are now modified by setting $v_i = v, n_i = 1$, and $\phi_i = (z_i - x_i'\beta)/v$.

5 Applications of the model

We first illustrate the spatial probit model with interaction effects using a generated data set. The advantage of this approach is that we know the true parameter magnitudes as well as the generated spatial interaction effects. This allows us to examine the ability of the model to accurately estimate the parameters and interaction effects. We provide an applied illustration in section 5.2 using the 1996 presidential election results that involves 3,110 US counties and the 48 contiguous states.

5.1 A generated data example

This experiment used the latitude-longitude centroids of $n = 3,110$ US counties to generate a set of data. The $m = 48$ contiguous states were used as regions. A continuous dependent variable was generated using the following procedure. First, the spatial interaction effects were generated using:

$$\begin{aligned} \theta &= (I_m - \rho W)^{-1}\varepsilon \\ \varepsilon &\sim N(0, \sigma^2) \end{aligned} \tag{60}$$

[6] An alternative (not examined here) would be to consider regional groupings of individuals with possible heteroscedasticity effects between regions, while allowing spatial dependencies to occur at the individual rather than regional level.

where ρ was set equal to 0.7 in one experiment and 0.6 in another. In (60), W represents the 48x48 standardized spatial weight matrix based on the centroids of the states.

Six explanatory variables which we label X were created using county-level census information on: the percentage of population in each county that held high school, college, or graduate degrees, the percentage of non-white population, the median household income (divided by 10,000) and the percent of population living in urban areas. These are the same explanatory variables we use in our application to the 1996 presidential election presented in section 5.2, which should provide some insight into how the model operates in a generated data setting.

The data matrix X formed using these six explanatory variables was centered using the studentize transformation that subtracts means and divides by the standard deviations. Use of a centered data matrix X along with negative and positive values for β ensures that the generated y values have a mean close to zero. Since we will convert the generated continuous y magnitudes to 0,1 z values using the rule in (62), this should produce a fairly equal sample of 0,1 values.

$$\begin{aligned} z &= 0 \text{ if } y <= 0 \qquad (61) \\ z &= 1 \text{ if } y > 0 \end{aligned}$$

The vector of $\Delta\theta$ along with the matrix X and parameters for $\beta = (3,\ -1.5,\ -3,\ 2,\ -1,\ 1)'$ were used to generate a continuous y vector using:

$$\begin{aligned} y &= X\beta + \Delta\theta + u \\ u &\sim N(0, V) \\ V &= I_m \qquad (62) \end{aligned}$$

Generating a continuous y allows us to compare the posterior mean of the draws from the truncated normal distribution which would serve as the basis for inference about the unobserved y values in an applied setting. Another focus of inference for this type of model would be the estimated values for the m-vector of parameters θ.

A set of 100 data samples were generated and used to produce estimates whose mean and standard deviations are shown in Table 1, alongside the true parameter values used to generate the sample data. In addition to the spatial probit model estimates, we also estimated a least-squares model, a non-spatial probit model and the spatial individual effects model based on the continuous y values. This was done using the generated continuous y vector and a minor change in the MCMC sampling scheme that ignored the update step for the latent variable y in step 6) of the sampler. Instead, we rely on the actual values of y, allowing us to see how inferences regarding the parameters are affected by the presence of binary versus continuous y values. Ideally, we would produce similar estimates from these two models, indicating that the latent draws for y work effectively to replace the unknown values.

Given the studentized matrix X the variance of each column, $\sigma_x^2 = 1$, and we set $\sigma^2 = 2$ to create a situation where the relative importance or signal strength of X in generation of y was one-half that of the strength of the individual effects θ. A second experiment with another 100 samples was based on $\sigma_x^2 = 1$ and $\sigma^2 = 0.5$ which creates a situation where the relative importance or signal strength of θ in generation of y is one-half that of X. This experiment used a value of $\rho = 0.6$ rather than the 0.7 value to examine sensitivity to this parameter. Both experiments relied on a homoscedastic prior setting $r = 60$ to reflect the homoscedasticity in the generating process. In applied practice, it is important to note that one can encounter situations where an inference is drawn suggesting the individual effects are not spatially distributed, that is $\rho = 0$. This might be due to a large amount of noise, represented by σ^2 in the generating process used here. It might also be due to a very strong signal in X relative to the signal in the individual effects, resulting in the influence of individual effects begin masked in the estimated outcomes. This situation would be represented by a large σ_x^2 relative to σ^2 in our generating process. Of course, these two influences also depend on the inherent observation noise reflected in $u \sim N(0, V)$, which we controlled by setting $V = I_m$ in our experiments.

Estimation results for these two illustrations are shown in Table 1 based on 1,500 draws with the first 500 omitted for ‘burn-in’ of the sampler.

Turning attention to the experimental results for the case where $\sigma^2 = 2$ shown in the table, we see that least-squares and probit estimates for β are inaccurate as we would expect. The spatial probit estimates were on average very close to those from the spatial regression model based on non-binary y-values, suggesting that sampling for the latent y works well. It should come as no surprise that estimates for β based on the non-binary dependent variable y are more precise than those from the spatial probit model based on binary z values. From the standard deviations of the estimates over the 100 samples, we see that use of binary dependent variables results in a larger standard deviation in the outcomes, reflecting less precision in the β and σ estimates. This seems intuitively correct, since these estimates are constructed during MCMC sampling based on draws for the latent y values. A graphical depiction of these draws from a single estimation run are shown in Figure 1, where we see that they are centered on the true y, but exhibit dispersion. The $R-$squared between the posterior mean of the latent y draws and the actual y (which we know here) was around 0.9 for this single estimation run. In summary, additional uncertainty arising from the presence of binary dependent variables z that must be sampled to produce latent y during estimation result in increased dispersion or uncertainty regarding the β estimates for the spatial probit model, relative to the non-binary spatial regression model.

The experimental results for the case where $\sigma^2 = 0.5$ show roughly the same pattern in outcomes. This suggests that the estimation procedure will work well for cases where the relative signal strength of X versus θ varies within a reasonable range.

An important use for this type of model would be inferences regarding the character of the spatial interaction effects. Since these were generated here, we can compare the mean of the posterior distribution for these values with the actual magnitudes. Figure 2 shows this comparison, where the average θ estimates from both the spatial probit and spatial regression model are plotted against the average of the true θ values generated during the experiment. In the figure, the individual effect estimates were sorted by magnitude of the average actual θ values for presentation purposes. We see from the figure that the estimates were on average close to the true values and one standard deviation of these estimates were also close to one standard deviation of the actual θ values. The spatial regression estimates were slightly more accurate as we would expect, exhibiting a correlation of 0.97 with the actual θ, whereas the spatial probit estimates had a correlation of 0.91

Figure 3 shows the actual θ values plotted versus the posterior mean of these estimates from a single estimation. Estimates from both the spatial probit model and the spatial regression model are presented and we see that accurate inferences regarding the individual effects could be drawn. The correlation between the actual individual effects used to generate the data and the predictions is over 0.9 for both models.

By way of summary, the spatial probit model performed well in this generated data experiment to detect the pattern of spatial interaction effects and to produce accurate parameter estimates.

5.2 An application to the 1996 presidential election

To illustrate the model in an applied setting we used data on the 1996 presidential voting decisions in each of 3,110 US counties in the 48 contiguous states. The dependent variable was set to 1 for counties where Clinton won the majority of votes and 0 for those where Dole won the majority.[7] To illustrate individual versus regional spatial interaction effects we treat the counties as individuals and the states as regions where the spatial interaction effects occur.

As explanatory variables we used: the proportion of county population with high school degrees, college degrees, and graduate or professional degrees, the percent of the county population that was non-white, the median county income (divided by 10,000) and the percentage of the population living in urban areas. These were the same variables used in the generated data experiments, and we applied the same studentize transformation here as well. Of course, our application is illustrative rather than substantive.

We compare estimates from a least-squares and traditional non-spatial probit model to those from the spatial probit model with a homogeneity assumption and a heteroscedastic assumption regarding the disturbances. The spatial probit model estimates are based on 6,000 draws with the first 1,000 omitted to allow the sampler to achieve a steady-state.[8] Diffuse or conjugate priors were employed for all of the parameters β, σ^2 and ρ in the Bayesian spatial probit models. A hyperparameter value of $r = 4$ was used for the heteroscedastic spatial probit model, and a value of $r = 40$ was employed for the homoscedastic prior. The heteroscedastic value of $r = 4$ implies a prior mean for r equal to $r/(r-2) = 2$ [see discussion surrounding (53)] and a prior standard deviation equal to $\sqrt(2/r) = 0.707$. A two standard deviation interval around this prior mean would range from 0.58 to 3.41, suggesting that posterior estimates for individual states larger than 3.4 would indicate evidence in the sample data against homoscedasticity. The posterior mean for the v_i estimates was greater than this upper level in 13 of the 48 states (shown in Table 2), with a mean over all states equal to 2.86 and a standard deviation equal to 2.36. The frequency distribution of the 48 v_i estimates is shown in Figure 4, where we see that the mean is not representative for this skewed distribution. We conclude there is evidence in favor of mild heteroscedasticity.

[7] The third party candidacy of Perot was ignored and only votes for Clinton and Dole were used to make this classification of 0,1 values.

[8] Estimates based on 1,500 draws with the first 500 omitted were nearly identical suggesting that one need not carry out an excessive number of draws in practice.

Comparative estimates are presented in Table 3, where we see that different inferences would be drawn from the homoscedastic versus heteroscedastic estimates. With the exception of high school graduates, the magnitudes of the coefficients on all other variables are quite different. The heteroscedastic estimates are larger (in absolute value terms) than the homoscedastic results with the exception of population in urban areas which is not significant. In the case of college graduates, the homoscedastic and heteroscedastic results differ regarding the magnitude and significance of a negative impact on Clinton winning. Heteroscedastic results suggest a larger negative influence significant at conventional levels while the homoscedastic results indicate a smaller insignificant influence.

The results also indicate very different inferences would be drawn from the non-spatial probit model versus the spatial probit models. For example, the non-spatial model produced larger coefficient estimates for all three education variables. It is often the case that ignoring spatial dependence leads to larger parameter estimates, since the spatial effects are attributed to the explanatory variables in these non-spatial models. Another difference is that the coefficient on median income is small and insignificant in the non-spatial model whereas it is larger (in absolute value terms) and significant in both spatial models.

The parameter estimates for the spatial interaction effects should exhibit spatial dependence given the estimates for ρ. Figure 5 shows a graph of these estimates along with a ± 2 standard deviation confidence interval. In the figure, the states were sorted by 0,1 values reflecting the 18 states where Dole won the majority of votes versus the 30 states where Clinton won. From the figure we see that in the 30 states where Clinton won there is evidence of predominately positive spatial interaction effects, whereas in the states where Dole won there are negative individual effects.

A comparison of the individual effect estimates from the homoscedastic and heteroscedastic models is shown in Figure 6, where we see that these two sets of estimates would lead to the same inferences.

Figure 7 shows a map of the significant positive and negative estimated individual effects as well as the insignificant effects, (based on the heteroscedastic model). This map exhibits spatial clustering of positive and negative effects, consistent with the positive spatial dependence parameter estimate for ρ.

Finally, to assess predictive accuracy of the model we examined the predicted probabilities of Clinton winning. In counties where Dole won, the model should produce a probability prediction less than 0.5 of a Clinton win. On the other hand accurate predictions in counties where Clinton won would be reflected in probability predictions greater than 0.5. We counted these cases and found the heteroscedastic model produced the correct predictions in 71.82 percent of the counties where Dole won and in 71.16 percent of the counties where Clinton won. The homoscedastic model produced correct predictions for Dole in 73.29 percent of the counties and for Clinton in 69.06 percent of the counties.

6 Conclusion

A hierarchical Bayesian spatial probit model that allows for spatial interaction effects as well as heterogeneous individual effects was introduced here. The model extends the traditional Bayesian spatial probit model by allowing decision-makers to exhibit spatial similarities. In addition to spatial interaction effects, the model also accommodates heterogeneity over individuals (presumed to be located at distinct points in space) by allowing for non-constant variance across observations.

Estimation of the model is via MCMC sampling which allows for the introduction of prior information regarding homogeneity versus heterogeneity as well as prior information for the regression and noise variance parameters.

The model is not limited to the case of limited dependent variables and could be applied to traditional regression models where a spatial interaction effect seems plausible. This modification involves eliminating the truncated normal draws used to obtain latent y values in the case of limited dependent variables. MATLAB functions that implement the probit and regression variants of the model presented here are available at: http://www.spatial-econometrics.com. They rely on a c-language interface available in MATLAB to call an externally compiled c-program from within the MATLAB programming environment. This enhances the speed of the estimation program by a factor of six times over a program written in the interpreted MATLAB matrix programming language.

References

Albert, James H. and Siddhartha Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data", *Journal of the American Statistical Association*, Volume 88, number 422, pp. 669-679.

Amemiya, T. (1985) *Advanced Econometrics*, Cambridge MA, Harvard University Press.

Besag, J. J.C. York, and A. Mollie (1991) "Bayesian Image Restoration, with Two Principle Applications in Spatial Statistics', *Annals of the Institute of Statistical Mathematics*, Volume 43, pp. 1-59.

Gelfand, Allan E. and Adrian F.M Smith (1990), "Sampling-based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, Volume 85, pp. 398-409.

Gelman, A., J.B. Carlin, H.A. Stern and D.R. Rubin (1995) *Bayesian Data Analysis*, Chapman & Hall, London.

Geweke, John. (1993) "Bayesian treatment of the independent Student-*t* linear model", *Journal of Applied Econometrics*, Volume 8, S19-S40.

LeSage, James P. (1997) "Bayesian Estimation of Spatial Autoregressive Models", *International Regional Science Review*, 1997 Volume 20, number 1&2, pp. 113-129.

LeSage, James P. (1999) *The Theory and Practice of Spatial Econometrics*, unpublished manuscript available at:
http://www.spatial-econometrics.com.

LeSage, James P. (2000) "Bayesian Estimation of Limited Dependent variable Spatial Autoregressive Models", *Geographical Analysis*, 2000 Volume 32, number 1, pp. 19-35.

Sun, D., R.K. Tsutakawa, P.L. Speckman (1999) "Posterior distribution of hierarchical models using car(1) distributions", *Biometrika*, Volume 86, pp. 341-350.

Appendix

This appendix derives a sequence of univariate conditional posterior distributions for each element of θ that allows the MCMC sampling scheme proposed here to be applied in larger models. For models with less than $m = 100$ regions it is probably faster to simply compute the inverse of the $m\text{x}m$ matrix A_0 and use the multinormal distribution presented in (41). For larger models this can be computationally burdensome as it requires large amounts of memory.

The univariate conditional distributions are based on the observation that the joint density in (40) involves no inversion of A_0, and hence is easily computable. Since the univariate conditional posteriors of each component, θ_i of θ must be proportional to this density, it follows that each is univariate normal with a mean and variance that are readily computable.

To formalize these observations, observe first that if for each realized value of θ and each $i = 1, \ldots, m$ we let $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_m)$, then by the same argument as in (31) we see that

$$\begin{aligned} p(\theta_i|\star) &= \frac{p(\theta, \beta, \rho, \sigma^2, V, z, y)}{p(\theta_{-i}, \beta, \rho, \sigma^2, V, z|y)} \propto p(\theta, \beta, \rho, \sigma^2, V, z|y) \\ &\propto \pi(z|\beta, \theta, V) \cdot \pi(\theta|\rho, \sigma^2) \\ &\propto \exp\{-\frac{1}{2}[\theta'(\sigma^{-2}B'_\rho B_\rho + \Delta' V^{-1}\Delta)\theta \\ &- 2(z - X\beta)'V^{-1}\Delta\theta]\} \qquad (63) \end{aligned}$$

This expression can be reduced to terms involving only θ_i as follows. If we let $\phi = (\phi_i : i = 1, \ldots, m)' = [(z - X\beta)'V^{-1}\Delta]'$, then the bracketed expression in (63) can be written as,

$$\begin{aligned} &\theta'(\sigma^{-2}B'_\rho B_\rho + \Delta' V^{-1})\Delta\theta - 2(z - X\beta)'V^{-1}\Delta\theta \\ = &\frac{1}{\sigma^2}\theta'(I - \rho W')(I - \rho W)\theta + \theta'\Delta' V^{-1}\Delta\theta - 2\phi'\theta \\ = &\frac{1}{\sigma^2}[\theta'\theta - 2\rho\theta' W\theta + \rho^2\theta' W' W\theta] + \theta'\Delta' V^{-1}\Delta\theta - 2\phi'\theta \qquad (64) \end{aligned}$$

But by permuting indices so that $\theta' = (\theta_i, \theta'_{-1})$, it follows that

$$\theta' W\theta = \theta' \begin{pmatrix} w_{.i} & W_{-i} \end{pmatrix} \begin{pmatrix} \theta_i \\ \theta_{-i} \end{pmatrix}$$

$$\begin{aligned} &= \theta'(\theta_i w_{.i} + W_{-i}\theta_{-i}) \\ &= \theta_i(\theta' w_{.i}) + \theta' W_{-i}\theta_{-i} \end{aligned} \tag{65}$$

where $w_{.i}$ is the ith column of W and W_{-i} is the mx$(m-1)$ matrix of all other columns of W. But since $w_{ii} = 0$ by construction, it then follows that

$$\begin{aligned} \theta' W \theta &= \theta' \left(\sum_{j\neq i} \theta_j w_{ji} \right) + \begin{pmatrix} \theta_i & \theta'_{-i} \end{pmatrix} \begin{pmatrix} \sum_{j\neq i} \theta_j w_{ij} \\ \mathrm{C} \end{pmatrix} \\ &= \theta_i \sum_{j\neq i} \theta_j (w_{ji} + w_{ij}) + \mathrm{C} \end{aligned} \tag{66}$$

where C denotes a constant not involving parameters of interest. Similarly, we see from (65) that

$$\begin{aligned} \theta' W' W \theta &= (\theta_i w_{.i} + W_{-i}\theta_{-i})'(\theta_i w_{.i} + W_{-i}\theta_{-i}) \\ &= \theta_i^2 w'_{.i} w_{.i} + 2\theta_i (w'_{.i} W_{-i}\theta_{-i}) + \mathrm{C} \end{aligned} \tag{67}$$

Hence, by observing that

$$\theta'\theta = \theta_i^2 + \mathrm{C} \tag{68}$$

$$\theta'\Delta' V^{-1}\Delta\theta = n_i\theta_i^2/v_i + \mathrm{C} \tag{69}$$

$$-2\phi' V^{-1}\theta = -2\phi_i\theta_i + \mathrm{C} \tag{70}$$

where the definition of $\phi = (\phi_i : i = 1, \ldots, m)'$ implies [using the notation in (10)] that each ϕ_i has the form

$$\phi_i = \frac{\mathbf{1}'_i(z_i - X_i\beta)}{v_i}, \quad i = 1, \ldots, m \tag{71}$$

Finally, by substituting these results into (64), we may rewrite the conditional posterior density of θ_i as

$$\begin{aligned} p(\theta_i|\star) \;&\propto\; \exp\{-\frac{1}{2}[(-2\rho\theta_i \sum_{j\neq i} \theta_j (w_{ji} + w_{ij})\theta_i \\ &+\; \rho^2\theta_i^2 w'_{.i} w_{.i} + 2\rho^2\theta_i (w'_{.i} W_{-i}\theta_{-i}))\frac{1}{\sigma^2} + n_i\theta_i^2/v_i - 2\phi_i\theta_i]\} \\ &=\; \exp\{-\frac{1}{2}(a_i\theta_i^2 - 2b_i\theta_i)\} \end{aligned}$$

$$\begin{aligned} &\propto \quad \exp\{-\frac{1}{2}(a_i\theta_i^2 - 2b_i\theta_i + b_i^2/a_i)\} \\ &= \quad \exp\{-\frac{1}{2(1/a_i)}\left(\theta_i - \frac{b_i}{a_i}\right)^2\} \end{aligned} \tag{72}$$

and a_i and b_i are given respectively by

$$a_i \quad = \quad \frac{1}{\sigma^2} + \frac{\rho^2}{\sigma^2}w'_{.i}w_{.i} + \frac{n_i}{v_i} \tag{73}$$

$$b_i \quad = \quad \phi_i + \frac{\rho}{\sigma^2}\sum_{j\neq i}\theta_j(w_{ji} + w_{ij})\theta_j - \frac{\rho^2}{\sigma^2}w'_{.i}W_{-i}\theta_{-i} \tag{74}$$

Thus the density in (72) is seen to be proportional to a univariate normal density with mean, b_i/a_i, and variance, $1/a_i$, so that for each $i = 1, \ldots .m$ the conditional posterior distribution of θ_i given θ_{-i} must be of the form

$$\theta_i|(\theta_{-i}, \beta, \rho, \sigma^2, V, z, y) \sim N(\frac{b_i}{a_i}, \frac{1}{a_i}) \tag{75}$$

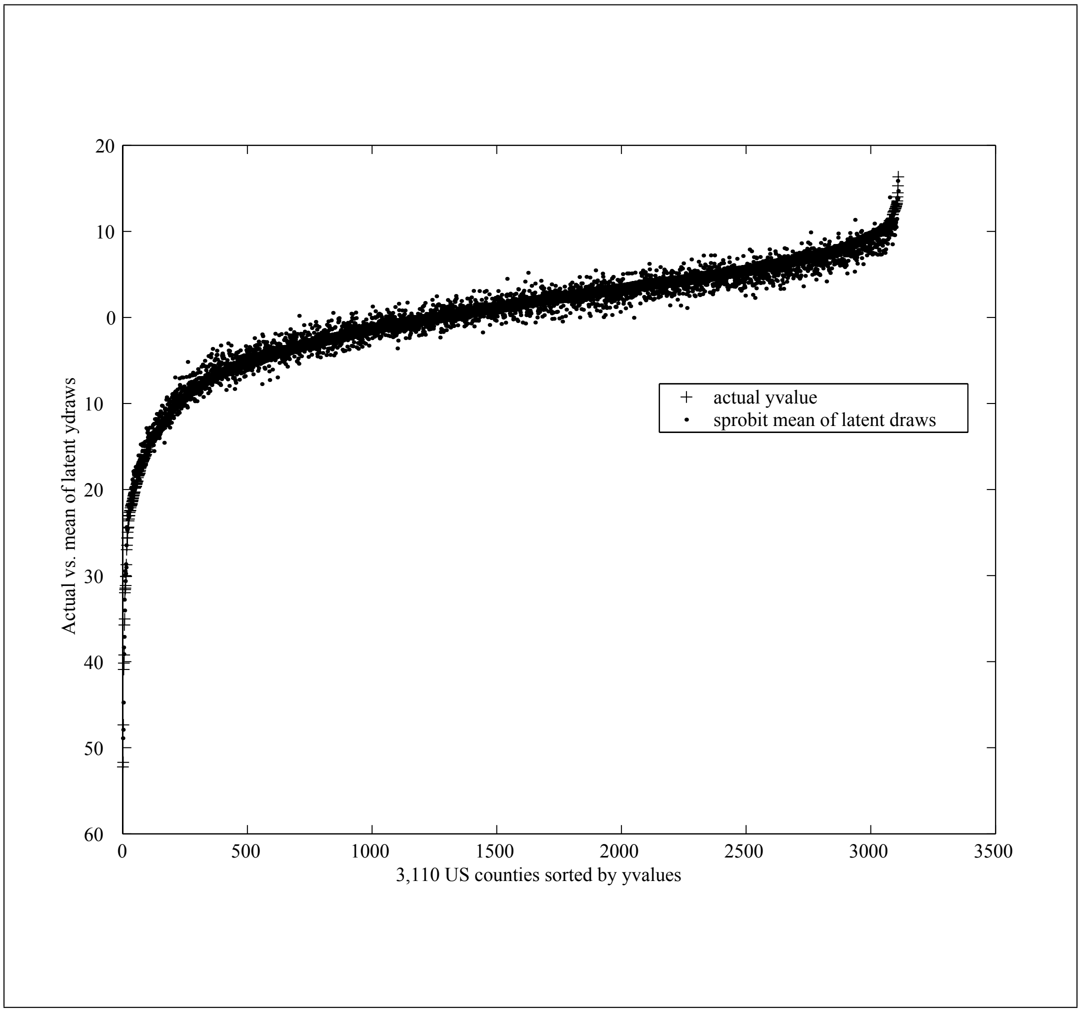

Figure 1: Actual y vs. mean of the latent y-draws

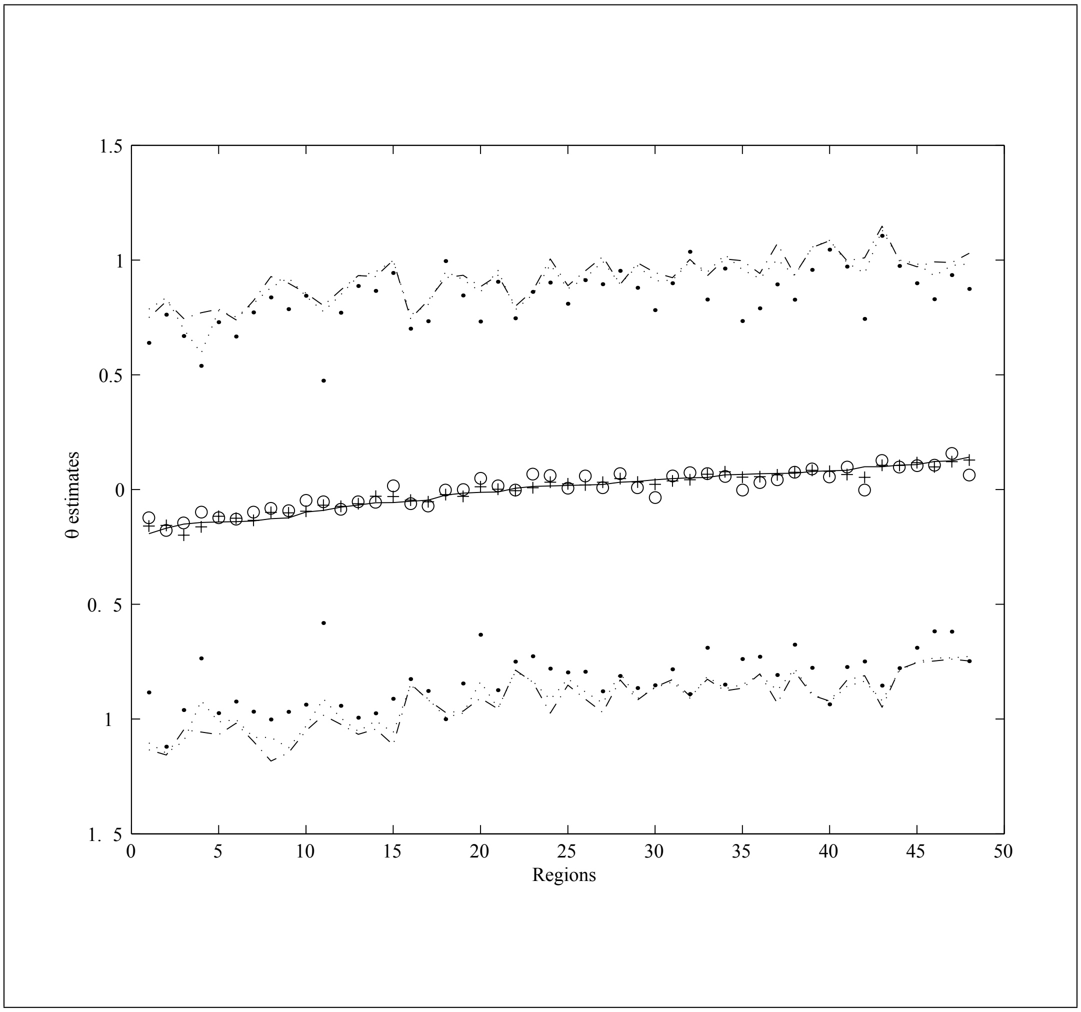

Figure 2: Mean of Actual vs. Predicted θ estimates over 100 samples

Table 1: Generated data results

Experiments using $\sigma^2 = 2$				
Estimates	ols	probit	sprobit	sregress
$\beta_1 = 3$	0.2153	1.5370	2.9766	2.9952
$\beta_2 =$ -1.5	-0.1291	-0.8172	-1.5028	-1.5052
$\beta_3 =$ -3	-0.0501	-1.5476	-2.9924	-2.9976
$\beta_4 = 2$	0.1466	1.0321	2.0019	2.0013
$\beta_5 =$ -1	-0.0611	-0.5233	-0.9842	-1.0013
$\beta_6 = 1$	0.0329	0.5231	0.9890	1.0006
$\rho = 0.7$			0.6585	0.6622
$\sigma^2 = 2$			2.1074	2.0990
Standard deviations	ols	probit	sprobit	sregress
σ_{β_1}	0.0286	0.2745	0.1619	0.0313
σ_{β_2}	0.0434	0.3425	0.1463	0.0393
σ_{β_3}	0.0346	0.4550	0.2153	0.0390
σ_{β_4}	0.0256	0.2250	0.1359	0.0252
σ_{β_5}	0.0176	0.1630	0.1001	0.0293
σ_{β_6}	0.0109	0.1349	0.0819	0.0244
σ_ρ			0.1299	0.1278
σ_σ			0.5224	0.3971
Experiments using $\sigma^2 = 0.5$				
Estimates	ols	probit	sprobit	sregress
$\beta_1 = 3$	0.2312	2.4285	3.0290	2.9983
$\beta_2 =$ -1.5	-0.1312	-1.1601	-1.5017	-1.4966
$\beta_3 =$ -3	-0.0517	-2.4646	-3.0277	-3.0042
$\beta_4 = 2$	0.1513	1.5975	2.0137	1.9984
$\beta_5 =$ -1	-0.0645	-0.8140	-1.0121	-1.0002
$\beta_6 = 1$	0.0348	0.8046	1.0043	0.9994
$\rho = 0.6$			0.5963	0.5886
$\sigma^2 = 0.5$			0.4960	0.5071
Standard deviations	ols	probit	sprobit	sregress
σ_{β_1}	0.0172	0.2058	0.1684	0.0292
σ_{β_2}	0.0219	0.2516	0.1420	0.0392
σ_{β_3}	0.0168	0.3873	0.2215	0.0382
σ_{β_4}	0.0125	0.1652	0.1301	0.0274
σ_{β_5}	0.0103	0.1595	0.1102	0.0329
σ_{β_6}	0.0074	0.1280	0.0899	0.0237
σ_ρ			0.1257	0.1181
σ_σ			0.1584	0.1101

Table 2: Posterior means for V_i parameters indicating heteroscedasticity

State	V_i estimate
Arizona	4.7210
Colorado	8.6087
Florida	3.9645
Georgia	11.1678
Kentucky	8.8893
Missouri	6.5453
Mississippi	4.3855
North Carolina	3.8744
New Mexico	4.7840
Oregon	4.3010
Pennsylvania	3.4929
Virginia	7.9718
Washington	4.7888

Table 3: 1996 Presidential Election results

<table>
<tr><td colspan="4">Homoscedastic Spatial Probit model with individual spatial effects</td></tr>
<tr><td>Variable</td><td>Coefficient</td><td>Std. deviation</td><td>P-level†</td></tr>
<tr><td>high school</td><td>0.0976</td><td>0.0419</td><td>0.0094</td></tr>
<tr><td>college</td><td>-0.0393</td><td>0.0609</td><td>0.2604</td></tr>
<tr><td>graduate/professional</td><td>0.1023</td><td>0.0551</td><td>0.0292</td></tr>
<tr><td>non-white</td><td>0.2659</td><td>0.0375</td><td>0.0000</td></tr>
<tr><td>median income</td><td>-0.0832</td><td>0.0420</td><td>0.0242</td></tr>
<tr><td>urban population</td><td>-0.0261</td><td>0.0326</td><td>0.2142</td></tr>
<tr><td>ρ</td><td>0.5820</td><td>0.0670</td><td>0.0000</td></tr>
<tr><td>σ^2</td><td>0.6396</td><td>0.1765</td><td></td></tr>
<tr><td colspan="4">Heteroscedastic Spatial Probit model with individual spatial effects</td></tr>
<tr><td>Variable</td><td>Coefficient</td><td>Std. deviation</td><td>P-level†</td></tr>
<tr><td>high school</td><td>0.0898</td><td>0.0446</td><td>0.0208</td></tr>
<tr><td>college</td><td>-0.1354</td><td>0.0738</td><td>0.0330</td></tr>
<tr><td>graduate/professional</td><td>0.1787</td><td>0.0669</td><td>0.0010</td></tr>
<tr><td>non-white</td><td>0.3366</td><td>0.0511</td><td>0.0000</td></tr>
<tr><td>median income</td><td>-0.1684</td><td>0.0513</td><td>0.0002</td></tr>
<tr><td>urban population</td><td>-0.0101</td><td>0.0362</td><td>0.3974</td></tr>
<tr><td>ρ</td><td>0.6176</td><td>0.0804</td><td>0.0000</td></tr>
<tr><td>σ^2</td><td>0.9742</td><td>0.3121</td><td></td></tr>
<tr><td colspan="4">Non-spatial Probit model</td></tr>
<tr><td>Variable</td><td>Coefficient</td><td>t-statistic</td><td>t-probability</td></tr>
<tr><td>high school</td><td>0.1961</td><td>6.494</td><td>0.0000</td></tr>
<tr><td>college</td><td>-0.1446</td><td>-3.329</td><td>0.0008</td></tr>
<tr><td>graduate/professional</td><td>0.2276</td><td>5.568</td><td>0.0000</td></tr>
<tr><td>non-white</td><td>0.2284</td><td>8.203</td><td>0.0000</td></tr>
<tr><td>median income</td><td>-0.0003</td><td>-0.011</td><td>0.9909</td></tr>
<tr><td>urban population</td><td>-0.0145</td><td>-0.521</td><td>0.6017</td></tr>
<tr><td colspan="4">† see Gelman, Carlin, Stern and Rubin (1995) for a description of p-levels</td></tr>
</table>

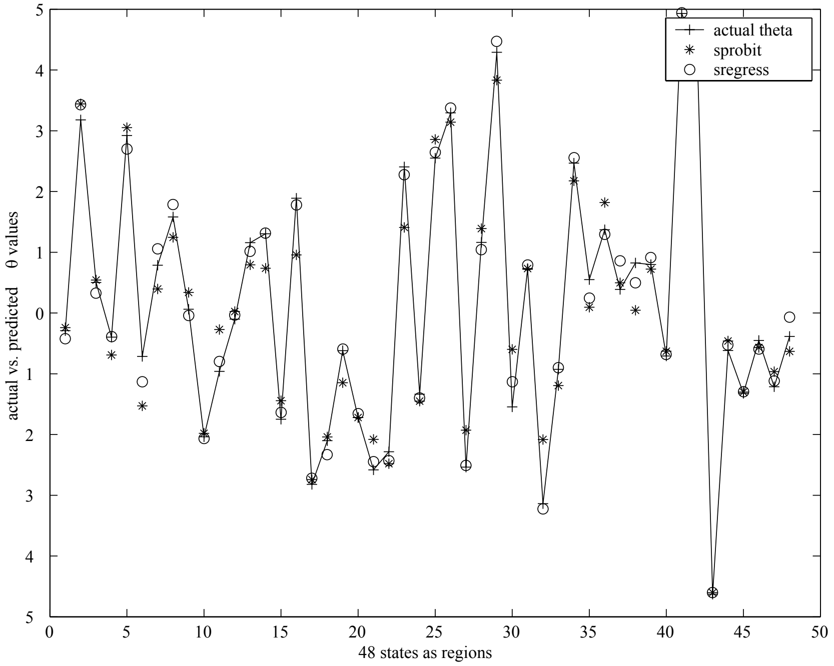

Figure 3: Actual vs. Predicted θ from a single estimation run

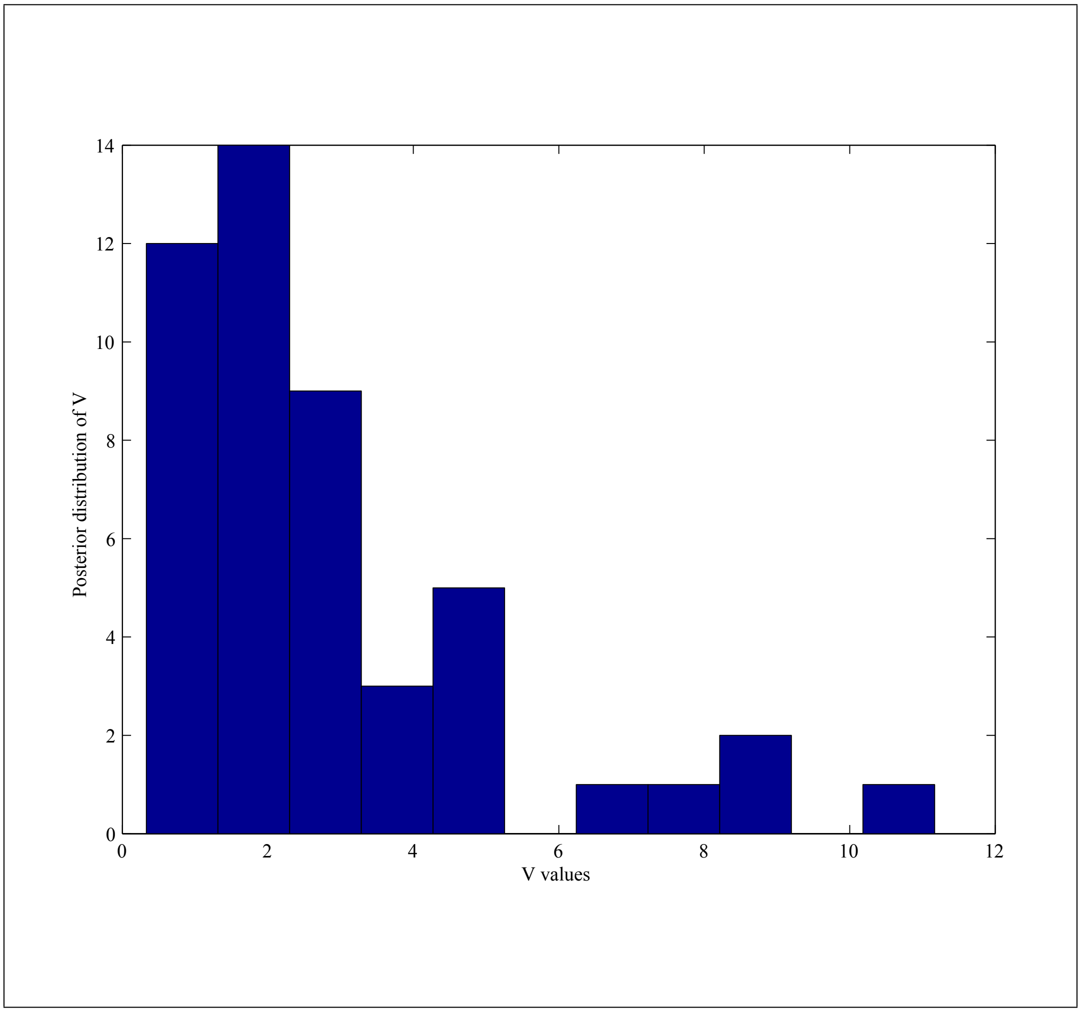

Figure 4: Frequency Distribution of V_i estimates

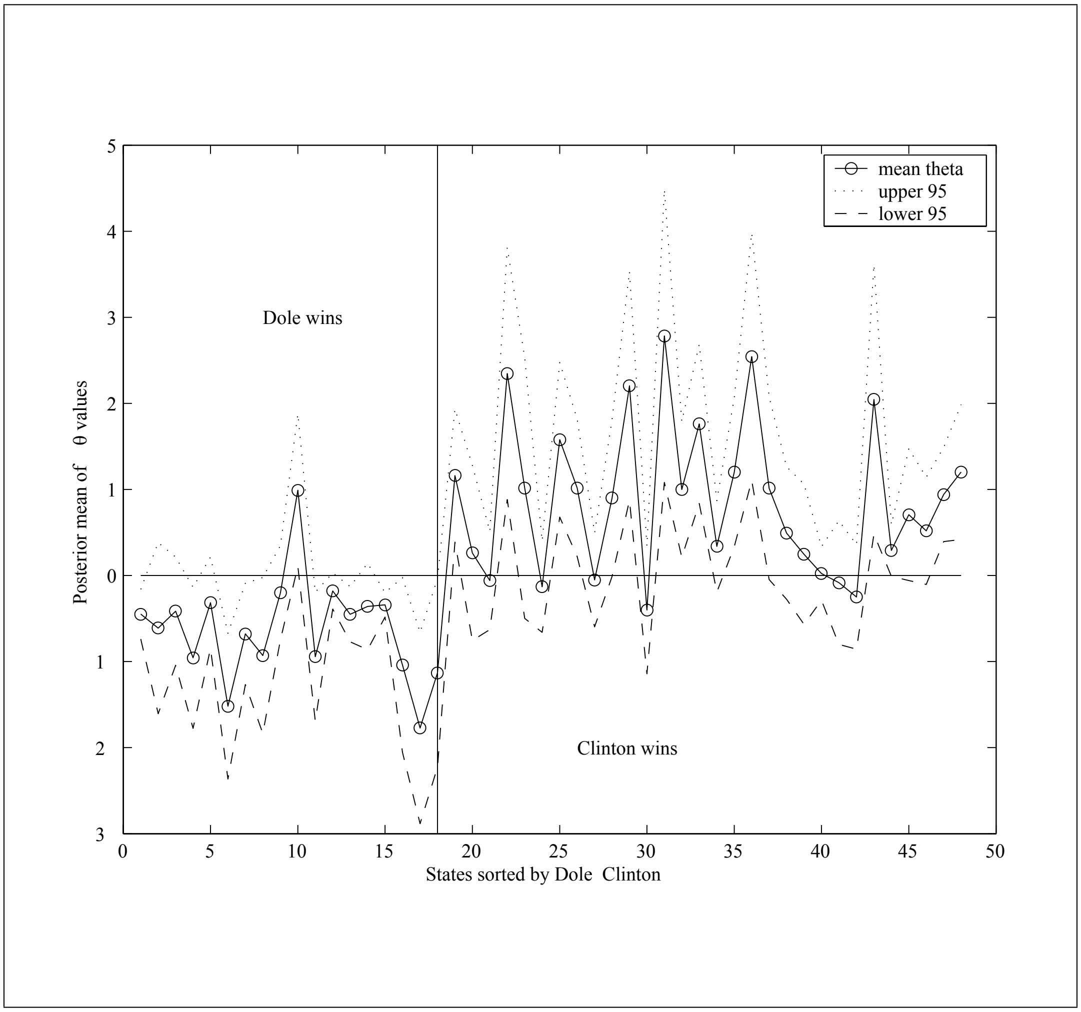

Figure 5: Individual effects estimates for the 1996 presidential election

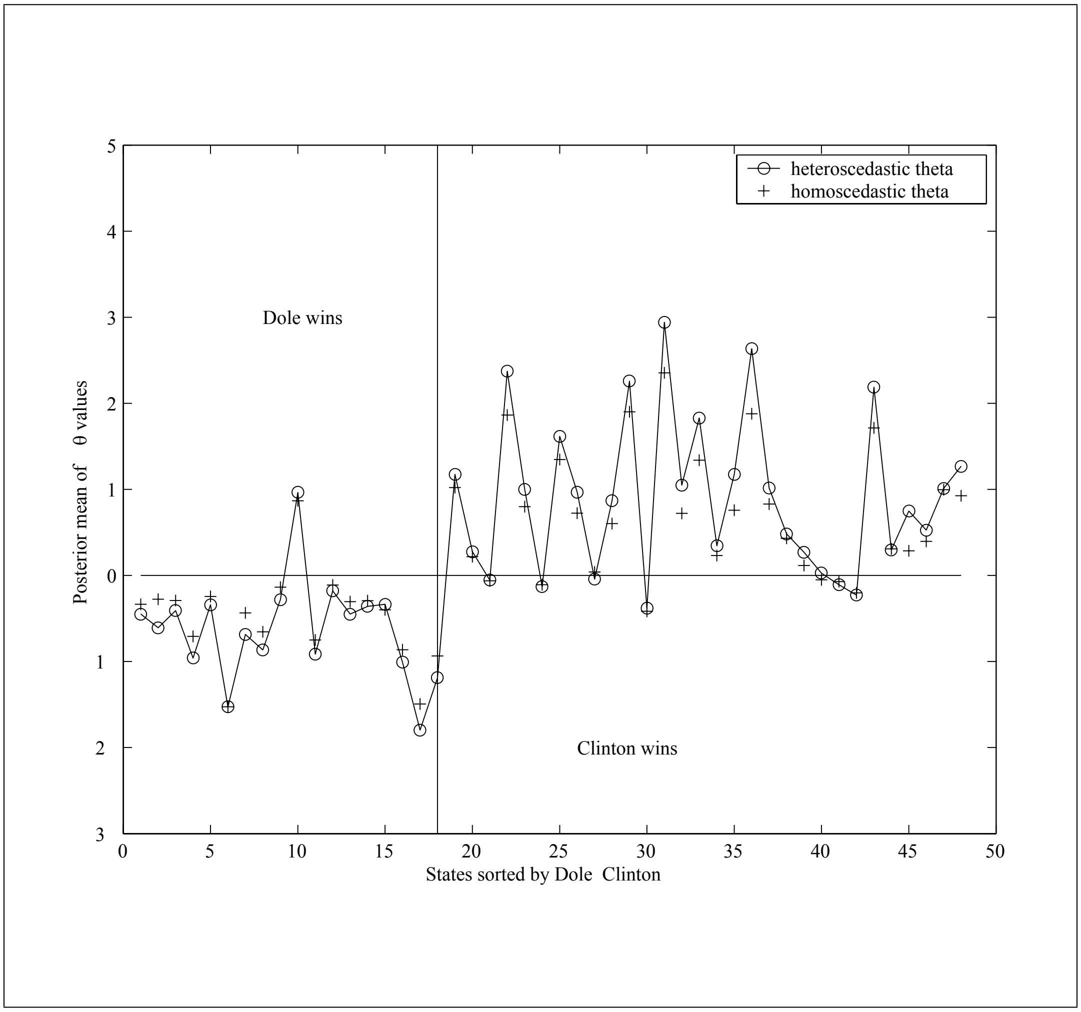

Figure 6: Individual effects estimates from homoscedastic vs. heteroscedastic spatial probit models

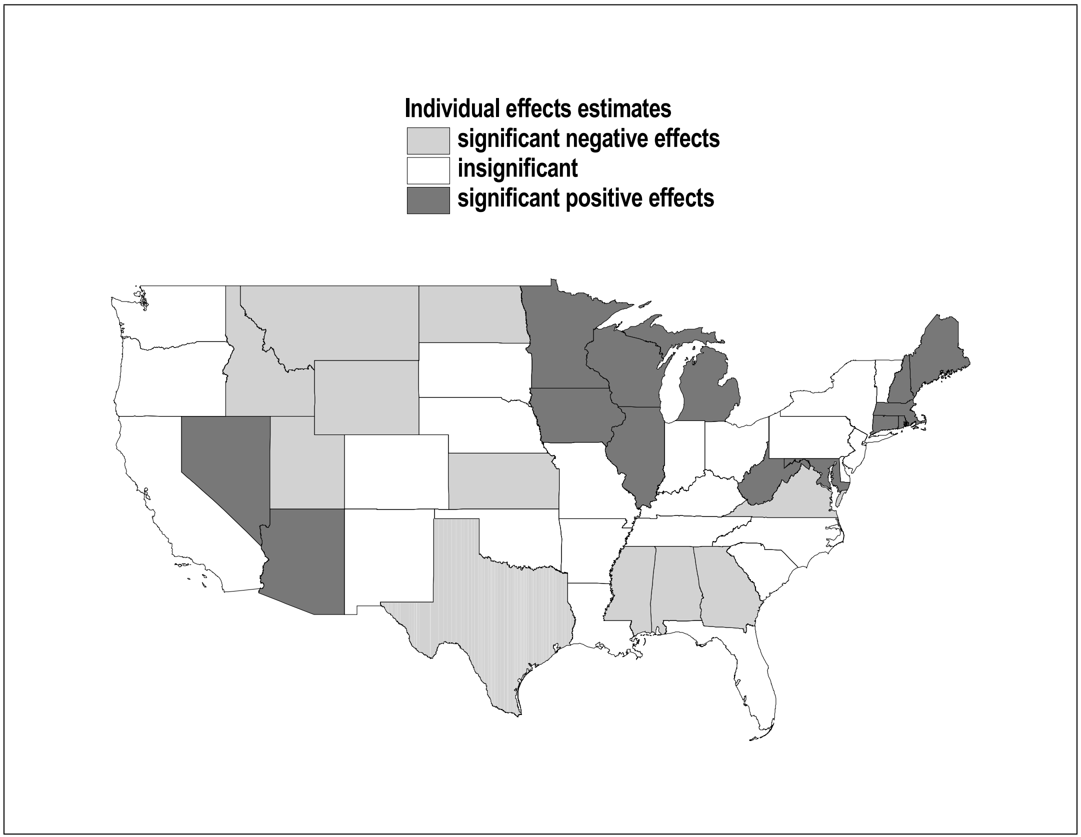
Individual effects estimates
significant negative effects
insignificant
significant positive effects

Figure 7: A map of individual effects estimates from the heteroscedastic spatial probit model