Bayesian spatial probit estimation: a primer and an application to HYV rice adoption

Article

Accepted Version

Bayesian spatial probit estimation: a primer and an application to HYV rice adoption

Article

Published Version

BAYESIAN SPATIAL PROBIT ESTIMATION: A PRIMER AND AN APPLICATION TO HYV RICE ADOPTION

Garth Holloway,[a] **Bhavani Shankar**[b] **and Sanzidur Rahman**[c]

Abstract

Increasingly, spatial econometric methods are becoming part of the standard toolkit of applied researchers in agricultural, environmental and development economics. Nonetheless, applications in discrete-choice settings remain few and, despite its appeal, applications of the Bayesian paradigm in these settings are still fewer. We provide a primer to the Bayesian spatial probit with objectives to broaden the scope of Bayesian applications, in general, and to make accessible to non-users a class of iterative estimation methods that have become fairly routine in Bayesian circles, offer an extremely powerful addition to applied researchers toolkits, and are essential in Bayesian implementation of spatial econometric models. We demonstrate the methods and apply them to estimate the 'neighborhood effect' in high-yielding variety adoption among Bangladeshi rice producers. A neighborhood effect exists when a propensity to adopt by one farm causes another to adopt. We estimate the strength of this relationship using a standard, spatial probit model and compare the policy conclusions with and without the neighborhood effect included. (162 words)

Keywords: Bayesian spatial probit, Markov-chain Monte Carlo methods, HYV rice adoption, Bangladesh. (11 words)

Journal of Economic Literature Classifications: 013, 031.

---

[a]Department of Agricultural and Food Economics, School of Agriculture, Policy and Development, PO Box 237, University of Reading, Reading RG6 6AR, UK, phone: (44)-(0)-(118)-931-6775, fax: +975-6467; email: g.holloway@reading.ac.uk. [b]Department of Agricultural and Food Economics, University of Reading, [c]School of Economic Studies, University of Manchester, Manchester MI3 9PL, UK. We are grateful to Jim LeSage for patiently answering all our queries and for pointing us to additional literature. We are also grateful to the editor and two anonymous referees for numerous helpful suggestions. Lucy Lapar donated help with editorial corrections and discussions with John Geweke, Bill McCausland, Dale Poirier and Richard Tiffin proved especially fruitful. Full responsibility for error remains with the authors.

1. Introduction

Discrete choice models abound in areas of agricultural economics, including technology adoption and land-use decision-making. Inevitably, these problems are characterised by some form of spatial dependence. Although accounting for spatial interactions is becoming prevalent in applications involving continuous dependent variables, such aspects are rarely incorporated into discrete choice models.

The importance of accounting for spatial dependence in these situations cannot be over-emphasized. For example, the spatial autoregression parameter (the 'neighborhood effect') in a technology adoption setting contains important policy information for public policy planning (Case, 1992). Knowledge of the location and scale of its distribution can be important in informing extension agents and planners about the likelihood that initial investments will generate further 'secondary' or 'copy' adoption in a locality. And this information, in turn, can aid decision-making so that research portfolio and public investment schedules are optimized.

One reason likely for the paucity of spatial discrete choice modelling is the complexity that it entails. Most of the available methods involve multidimensional integration. Bayesian techniques incorporating Markov chain Monte Carlo (MCMC) methods provide a powerful means to circumvent these problems. The advantages of this approach compared to available alternatives include non-reliance on asymptotic properties to ensure validity and generation of standard errors as a by-product of the estimation algorithm. The Bayesian approach provides a powerful alternative to conventional sampling theory techniques in handling the many tricky issues that confront applied spatial research. Currently in agricultural economics, however, application of the Bayesian paradigm is not widespread.

Our objective in this paper is to provide a step-by-step approach to the Bayesian spatial probit demonstrating its reliance on basic building blocks with which many of us are familiar and, as a by-product, demonstrating the full power of a class of techniques that are now becoming common place in other areas of applied research.

We focus attention on one of two Bayesian models that have been applied recently to implement truncated and discrete-choice data, namely the spatially autoregressive probit (SARP) model, which is the framework of choice for modelling new technology adoption and is the model applied in the empirical section of the paper. Pedagogically, the Bayesian spatial probit is but a step-wise generalization of the MCMC routine that is required to estimate the standard, normal linear model. Because it represents the cornerstone of almost all MCMC work in applied Bayesian science and our main objectives are pedagogic, the normal model provides a natural starting-point from which to incorporate additional computational detail. A focus on normal data provides, thus, ideal motivation for understanding the additional difficulties that spatial dependencies entail.

Section two presents an introduction to MCMC methods using normal data and outlines the Gibbs sampling algorithm that is the backbone of the spatial probit algorithm. This introductory section is intended for readers who are unfamiliar with Markov chain methods and, perhaps, the Bayesian view. Section three extends the basic method to a spatial econometric model with continuous left-hand-side data and incorporates two trivial modifications to derive the spatial probit algorithm. Section four introduces some institutional detail about the empirical application and introduces the data and section five presents the results of the spatial probit algorithm applied to the Bangladeshi data. Conclusions are offered in section six. Throughout the emphasis is on routine application of MCMC to solve complexities arising due to spatial dependence.

2. Demonstrations Using Normal Data

Suppose data $\mathbf{y} \equiv (y_1, y_2, .., y_N)'$ are normal with unknown mean, μ, and unknown variance σ^2 so that the data generating model is

(1) $$y_i = \mu + \varepsilon_i,$$

$i = 1, 2, .., N$; where μ denotes the mean of the distribution for $\mathbf{y}$; ε_i is a random error term that is normally distributed with mean zero and variance σ^2; and, hence, we may write, in a standard notation, $\varepsilon_1, \varepsilon_2, .., \varepsilon_N \sim$ iid $N(\mu, \sigma^2)$. We observe the data, $\mathbf{y}$, but do not observe the errors, $\mathbf{e} \equiv (\varepsilon_1, \varepsilon_2, .., \varepsilon_N)'$, nor the parameters μ and σ; and the objective is to make efficient use of $\mathbf{y}$ in deriving inferences about μ and σ.

Equation (1), together with the distributional assumption on the error term, is sometimes referred to as the normal-means model. Because this data-generating model is so familiar it serves as a natural starting point from which to introduce a generic notation which we retain throughout the demonstration. We continue to use $\mathbf{y} \equiv (y_1, y_2, .., y_N)'$ to denote data but use $f^A(b|c,d,..,z)$ to denote a probability density function (pdf) for the data where the symbols 'A,' 'b,' 'c,' 'd,' and 'z' denote, respectively, the form of the density, its argument, and any parameters that are needed to charaterize its location and scale, skewness, kurtosis, and so on.

It is important to note for later developments that the function $f(\cdot)$ is in the form of a conditional probability density function. Sometimes this conditioning will play a pivotal role in deriving efficient estimation strategies and sometimes it will not, and we make a point of symbolizing when conditioning is important through the notation $f(\cdot|\cdot)$ (the variables preceding the slash depending on the ones that follow). In the context of the data generating model $f^A(b|c, d, .., z)$, our task is typically to use observable data 'b' to make inferences

about variables (parameters) 'c,' 'd,' .., 'z,' which we will refer to collectively as the unknowns, or, compactly, by $\mathbf{Q}$, $\mathbf{Q} \equiv (c, d, .., z)'$.

The point of departure between the present contribution and others within this Special Issue, is the single observation that, because the parameters in $\mathbf{Q}$ are unknown they are, of course, random and have, therefore, associated probability distributions.

It is this step and this single step alone from which all subsequent developments emanate and it is important to bear this mind as we develop our MCMC strategy to estimate the Bayesian spatial probit model. This application requires two inputs, namely a likelihood for the entire data, which we denote $\ell(\mathbf{Q}|\mathbf{y})$, and a prior probability density function characterizing uncertainty about the unknown parameters, which we denote $\pi(\mathbf{Q})$. The assumption that the errors are independent allows us to multiply the individual normal densities comprising the likelihood into the form $\ell(\mathbf{Q}|\mathbf{y}) \equiv \prod_i f^N(y_i|\mu,\sigma)$, which, when viewed as a function of the complete data $\mathbf{y}$, is observed to have the normal form

$$\ell(\mathbf{Q}|\mathbf{y}) \equiv f^N(\mathbf{y}|\mathbf{i}_N\mu,\mathbf{I}_N\sigma), \tag{2}$$

where $\mathbf{i}_N$ denotes an N-dimensional unit vector and $\mathbf{I}_N$ denotes the N×N identity matrix. To draw inferences about $\mathbf{Q}$, we will update between the prior pdf $\pi(\mathbf{Q})$ and the posterior pdf $\pi(\mathbf{Q}|\mathbf{y})$ making use of Bayes' rule,

$$\pi(\mathbf{Q}|\mathbf{y}) \propto \ell(\mathbf{Q}|\mathbf{y})\,\pi(\mathbf{Q}). \tag{3}$$

The right-side of equation (3) omits the scale factor that makes the integral of the left side equal to one and, hence, justifies its interpretation as a true probability density function. The fact that we are able to avoid the computations implied by the integrals

$$f(\mathbf{y}) = \int_{\underline{\theta}_k}^{\bar{\theta}_k} \int_{\underline{\theta}_{k-1}}^{\bar{\theta}_{k-1}} .. \int_{\underline{\theta}_2}^{\bar{\theta}_2} \int_{\underline{\theta}_1}^{\bar{\theta}_1} \ell(\theta_1, \theta_2, .., \theta_k|\,\mathbf{y})\,\pi(\theta_1, \theta_2, .., \theta_k)\,d\theta_1\,d\theta_2\,..\,d\theta_{k-1}\,d\theta_k, \tag{4}$$

is worth stressing because it is precisely these computations that are the major stumbling blocks to the widespread application of the Bayesian paradigm. Although noteworthy exceptions exist, this development seems to have been more retarded in the agricultural economics literature than elsewhere. The advent of MCMC has, of course, changed this situation elsewhere and will do the same in agricultural economics—once the full power of the technique is widely accepted. With this goal in mind, it is useful to note that the term on the right-hand-side of this expression is the marginal likelihood for the data, a quantity that plays a pivotal role in model comparisons. We will not undertake model comparisons in this paper, although it is useful to note that the methods presented can be extended in a simple way to assess model probabilities (Chib, 1995; Chib and Jeliazkov, 2001). But the main point for computational gains is that, because $f(\mathbf{y})$ is not a function of $\mathbf{Q}$ we can ignore it in subsequent developments concerning $\mathbf{Q}$. Those developments typically involve the characterization of marginal pdfs for the model parameters and the difficult task confronting us is the derivation of these marginal quantities from the joint posterior through the integrations

$$(5) \qquad \pi(\theta_j|\mathbf{y}) = \int_{\underline{\theta}_k}^{\bar{\theta}_k} \int_{\underline{\theta}_{k-1}}^{\bar{\theta}_{k-1}} .. \int_{\underline{\theta}_{j+1}}^{\bar{\theta}_{j+1}} \int_{\underline{\theta}_{j-1}}^{\bar{\theta}_{j-1}} .. \int_{\underline{\theta}_2}^{\bar{\theta}_2} \int_{\underline{\theta}_1}^{\bar{\theta}_1} \pi(\theta_1, \theta_2, .., \theta_k|\, \mathbf{y})\, d\theta_1\, d\theta_2\, ..\, d\theta_{j-1}\, d\theta_{j+1}\, ..\, d\theta_{k-1}\, d\theta_k.$$

Derivation of the marginal distribution of an unknown quantity of interest is usually the target of any Bayesian investigation. Problems arise when the marginal density may not exist (not considered here); or, if the marginal density exists but may not have moments that do (considered in the empirical section); or, when the marginal pdf exists but may not have a form for which the integrating constant (the constant that makes the area beneath the density sum to one) is available in closed form. It is precisely this latter situation in which MCMC and, a special case, the Gibbs sampler, have particular advantages in exploiting conditional dependencies that prevail in almost all statistical settings. When the marginal distributions

are not available in closed form, but the fully conditional distributions, $\pi(\theta_j|\theta_1, \theta_2, .., \theta_{j-1}, \theta_{j+1}, .., \theta_k, \mathbf{y})$ satisfy weak regularity conditions (Gelfand and Smith, 1990) and are easy to sample from, the Gibbs sampler provides an extremely powerful and easy-to-implement approach to simulate draws from the marginal pdf. In short, the Gibbs sampler provides a route for sampling from the marginal pdf even though we cannot express it explicitly. And when the fully conditional distributions are not all available in closed form a generalization of the Gibbs sampler known as the Metropolis-Hastings algorithm can then be used to simulate from a target distribution. Although its requires more computation time than the Gibbs sampler it is just as powerful and is more versatile due to the fewer conditions that are required for its use. Two examples will help to demonstrate.

The Gibbs Sampler

In the normal-means set-up, the marginal distribution for μ, is a t distribution which, in its non-standardized form, is characterized by its mean, $\hat{\mu} \equiv (\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\mathbf{y}$; its degrees-of-freedom, $\nu \equiv$ N-1; and its scale, $\hat{\sigma}^2 \equiv (\mathbf{y}-\mathbf{i}\hat{\mu})'(\mathbf{y}-\mathbf{i}\hat{\mu})/\nu$. The marginal distribution for σ is an inverse-gamma distribution with degrees of freedom, $\nu \equiv$ N-1 and scale $\hat{\sigma}^2$. The two marginal distributions $\pi(\mu|\mathbf{y}) \equiv f^{T}(\mu|\hat{\mu},\nu,\hat{\sigma}^2)$ and $\pi(\sigma|\mathbf{y}) \equiv f^{IG}(\sigma|\nu,\hat{\sigma}^2)$ offer a complete description of the unknown quantities and are the target of the exercise. Although these marginal distributions are easily obtained by direct integration, we are interested in characterizing them through the Gibbs sample, for which we require the full conditional distributions. To derive these distributions we need to first establish the form of the joint posterior,

$$\pi(\mu,\sigma|\mathbf{y}) \propto \sigma^{-(N+1)} \exp\{\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{i}\mu)'(\mathbf{y}-\mathbf{i}\mu)\}, \tag{6}$$

which evolves from combining the non-informative prior $\pi(\mu, \sigma) \propto \sigma^{-1}$ (Jeffreys; Zellner, 1996, pp. 708), with the likelihood in (2) via (3). The full conditionals are $\pi(\mu|\sigma,\mathbf{y}) \equiv$

$f^N(\mu|\hat{\mu},\sigma)$ and $\pi(\sigma|\mu,\mathbf{y}) \equiv f^{IG}(\sigma|\nu,s^2)$, $s^2 \equiv (\mathbf{y}-\mathbf{i}\mu)'(\mathbf{y}-\mathbf{i}\mu)/\nu$, which we obtain from the joint distribution simply by viewing it solely as a function of a single unknown quantity and then identifying that the resulting form is well-known. The Gibbs sampler operates by iterating sequentially between these two conditional distributions drawing, in turn, an inverse-gamma random variable and a normal random variable. For S sufficiently large, and a starting value $\mu = \mu^0$, the Gibbs samples $\{\sigma^{(s)}, s = 1, 2, .., S\}$ and $\{\mu^{(s)}, s = 1, 2, .., S\}$, obtained by the sequence $\sigma^1 \sim f^{IG}(\sigma|\nu,\mu^0)$, $\mu^1 \sim f^N(\mu|\hat{\mu},\sigma^1)$, .., $\sigma^s \sim f^{IG}(\sigma|\nu,\mu^{s-1})$, $\mu^s \sim f^N(\mu|\hat{\mu},\sigma^s)$ provides accurate estimates of posterior moments and, indeed, the marginal distributions themselves. Figures 1-3 present results for the Gibbs sampler applied to the normal data in table 1. The data y are generated from $f^N(y_i|\mathbf{i}, \mathbf{I}_{10})$ and generate a posterior mean $\hat{\mu} = 99$ and sample variance $\hat{s}^2 = 99$. Figure

1 presents plots of the first 50 iterations in the Gibbs sample based on the start value $\mu^0 = (\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\mathbf{y}$ and figures 2 and 3 compare the frequencies of draws obtained from the first 1,000 iterations of the sample with frequencies generated by the target pdfs. The figures illustrate three important points. First, independent of the starting values, the draws mimic the actual draws one expects to obtain in draws from the true marginal pdfs. Second, very few iterations are required before the Gibbs sequence converges to the true pdf's. Third, when it is emphasized that there are only ten observations in the sample, convergence is obtained under very limited information.

(Insert figures 1-3 about here.)

Random-Walk Metropolis-Hastings Sampling

When the full conditional distributions are not available in closed form a more general set of iteration methods must be invoked. One of these—the focus in empirical work—is the Random-Walk Metropolis-Hastings (RW) algorithm. Although its roots are old (Metropolis,

et al., 1953; Hastings, 1970), it is difficult to locate applications of the technique in agricultural economics. The RW algorithm is but one of many variants of a basic accept-reject procedure which are suitable to model spatial dependence. Space limitations prevent reporting results across an array of models that the authors experimented with in the course of this project; but the RW algorithm proved considerably superior in terms of execution time and provided accurate estimates of system parameters with minimal fuss. Like the Gibbs sequence, the RW sequence generates a Markov chain with desirable convergence properties; but, unlike the Gibbs sampler only a subset of the proposed draws in the algorithm are accepted. This key difference makes the search for 'efficient' strategies to improve the basic algorithm desirable and this goal is an ongoing focus in statistical research (see, in particular, the papers by Raftery and Lewis, 1992; Robert, 1995; Gilks, Richardson and Spiegelhalter, 1996, among many others). Continuing with notation developed previously, and the normal-means example, suppose that we wish to simulate a draw from the target density $f(\mu|\cdot)$, which is not of a standard form. We obtain a draw—a 'proposal'—from another distribution that, among other properties, is known (known integrating constant) and is easy to sample from and we accept and reject the proposals based on a probability rule that results in the accepted sequence of draws generating a Markov chain that, eventually, converges to the target distribution. Use $\underline{m}$ to denote the proposal value and use $f^P(\underline{m}|\cdot)$ to denote the proposal density. In general, the proposal density can be conditional on a prior draw for the parameter of interest and, hence, let us use $f^P(m|\mu)$ top denote this fact. The defining feature of the RW algorithm is that the current draw for the parameter, $\underline{m}$, depends on the previous draw, μ, through a random walk. In other words, the proposal and current values are related through the condition

(7) $$\underline{m} = \mu + \varepsilon,$$

where ε is a random perturbation with distribution $f^P(\varepsilon)$ that is independent of μ (and, therefore, $\underline{m}$). It follows from (7) that the proposal distribution has the form $f^P(\underline{m}-\mu)$ and we consider a second experiment. Suppose that the distribution for ε is the standard normal distribution $f^P(\varepsilon) = f^N(\varepsilon|0,1)$. The distribution for $\underline{m}$ is $f^N(\underline{m}|\mu,1)$ and we proceed by successively drawing from the normal distribution with mean the current draw μ and variance one. Then, the following steps simulate draws from the target distributions of interest (Robert and Casella, p. 245): Step 1: Generate a starting value $\mu = \mu'$. Step 2: Generate $m \sim f^P(m|\mu)$. Step 3: Generate $u \sim f^U(u|0,1)$. Step 4: If $u \leq f(\underline{m})/f(\mu)$ set $\mu = m$. Step 5: Return to step 2. Figure 4 presents plots of the first 50 iterations of the Gibbs sequence with the draw for μ simulated by the RW step. The draws are quite similar to the draws from the Gibbs sample and they mimic the values of the actual parameter values used to generate the data. Figures 5 and 6 present plots of the histograms generated by the first 1000 draws. The simulated densities are quite close to the actual target densities. In summary, the RW algorithm provides a versatile alternative to the Gibbs sampler to simulate draws from the two target distributions.

(Insert figures 4-6 about here.)

The methods applied in this section are the building blocks of almost all the variants of the Gibbs sampler and the Metropolis-Hastings algorithm that appear in the literature. Collectively, these steps provide an extremely powerful tool kit from which the investigator can launch more sophisticated analyses. Nowadays, Bayesian research is not constrained by the need to provide numerical approximations to difficult integral calculations by area methods such as Simpson's Rule. Freed from the constraints that these integrations have, in past, bounded investigations, we are now in a position to extend the normal-means framework to the target setting of spatial probit estimation. A manipulation that aids this link

and is important in subsequent developments is to write $\mathbf{x} \equiv \mathbf{i}$ and $\beta \equiv \mu$ in the data generating model in (3) and recall the definitions of the means and variances in the original model, namely $\hat{\mu} \equiv (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} \equiv \hat{\beta}$ and $\hat{\sigma}^2 \equiv (\mathbf{y}-\mathbf{x}\beta)'(\mathbf{y}-\mathbf{x}\beta)/\nu$, with $\mathbf{x}$, an N×K matrix of observations on a set or relevant covariates and, hence, β a K-vector. And so, the normal linear regression model has also been accurately simulated through the Gibbs and Random-Walk Metropolis-Hastings algorithms.

3. Spatial Models

The previous section based on the normal-means (normal-linear regression) model serves to illustrate the powerful way in which the various algorithms serve as alternatives to conventional approaches (direct integration in the case of normal data) to derive inferences with respect to a target marginal distribution. But the normal data environment is, of course, simplistic and the results of the demonstrations, particularly the accuracy of the derived distributions, can be questioned. Such scepticism is, we show, unmerited because the two spatial models of interest—the spatial autoregressive (SAR) model and the spatial-autoregressive probit (SARP) model—are but simple extensions of the normal-linear model.

A principal source of information for learning about these models in the Bayesian environment are the papers (LeSage, 1997, 2000, 2002) and much of what we present in this section is a 'resampling' of some of this work, together with several personal communications about various sampling issues (LeSage, 2002b, 2002c). We now continue a step-by-step development of the extensions from the normal regression model that are needed in spatial inference. Unlike that model, spatial problems generate distributions for which the marginal pdfs are unavailable in closed form and require application of MCMC methods. In this case, measures of accuracy are now no longer available for all of the parameters in

question. Nevertheless, it is possible with what we have established so far to give a heuristic indication of what a correctly implemented algorithm should produce.

Before presenting results from simulated probit data, we consider the standard spatial framework, where the dependent variable is a known continuous measure of observed data. We are concerned with the model

(8) $\mathbf{y} = \rho\,\mathbf{w}\mathbf{y} + \mathbf{x}\,\mathbf{b} + \mathbf{e}$,

where $\mathbf{y}_{(N\times1)} \equiv (y_1, y_2, .., y_N)'$ denotes observations on a dependent variable of interest across spatially delimited units, i = 1, 2, ..N; $\rho_{(1\times1)}$ denotes correlation between units; $\mathbf{w}_{(N\times N)}$ denotes a spatial-weight matrix, defined in more detail, subsequently; $\mathbf{x}_{(N\times K)} \equiv (\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_K)$, $\mathbf{x}_{1(N\times1)} \equiv (x_{11}, x_{21}, .., x_{N1})'$, $\mathbf{x}_{2(N\times1)} \equiv (x_{12}, x_{22}, .., x_{N2})'$, .., $\mathbf{x}_{K(N\times1)} \equiv (x_{1K}, x_{2K}, .., x_{NK})'$ denotes observations on the covariates; $\mathbf{b}_{(K\times1)} \equiv (\beta_1, \beta_2, .., \beta_K)'$ denotes the ceteris paribus relationship between $\mathbf{x}$ and $\mathbf{y}$; and $\mathbf{e}_{(N\times1)} \equiv (\varepsilon_1, \varepsilon_2, .., \varepsilon_N)'$ denotes random error, which, we assume, is distributed normally with mean zero and covariance $\sigma^2\mathbf{I}_N$; in other words, $\mathbf{e} \sim f^N(\mathbf{e}|\mathbf{0}_N,\sigma^2\mathbf{I}_N)$, where $\mathbf{0}_N$ is the length-N null vector and $\mathbf{I}_N$ is the dimension-N identity matrix. The econometrician observes $\mathbf{x}$, $\mathbf{w}$ and $\mathbf{y}$ and makes inferences about ρ, $\mathbf{b}$ and σ.

Equation (8) is the data-generating equation for the spatial autoregressive (SAR) model. The SAR model is one of two conventional specifications discussed in the literature. The other model allows for correlations among errors across the spatial units. Our main interest in the empirical part of this paper concerns correlation across the dependent variables (not the errors) and, so, for this reason and the interests of space, we focus on the SAR formulation. Developments for both models are present neatly in LeSage (2000, equations (1) through (7)). The empirical application to follow motivates the spatial correlation parameter, ρ, and the important spatial-weight matrix, $\mathbf{w}$. Suppose that our interest lies in estimating the correlation across crop yields in contiguous settings (a situation only slightly

dissimilar from our empirical application) and, suppose that, perhaps for policy purposes, we are interested in estimating the extent to which yields in associated areas are correlated; areas with high yields, presumably, associated with contiguous high-yielding areas, and so on. In this setting **w** represents an assignment of ones and zeros corresponding to contiguous areas, such that, $w_{ij} = 1$ if observations i and j are in the same 'location' and $w_{ij} = 0$, otherwise, i, j = 1, 2, .., N, $i \neq j$. We assume that the correlation between an observation unit and itself is zero and, hence, that $w_{ii} = 0$, i = 1, 2, .., N. In most practical applications of this model it is customary to normalize each row of **w** such that each of the N_i (< N) contiguous units affecting observation i has one N_i^{th} contribution to the total impact on i; but this normalization is not necessary. The condition $\rho \neq 0$ implies correlation within locations and forces endogenity (qua simultaneity) between the crop yields; and the condition $\rho = 0$ implies that there is no correlation and forces the model to collapse to the standard multivariate model which can be handled without the need to Gibbs sample (Zellner, 1996, pp. 224-236). The anology in the non-spatial context is simply the difference between a simultaneous-equations set-up and its reduced-form and the presence of non-zero ρ is the defining distinction, indeed the raison d'être for the MCMC approach. But more precisely, it is the combination of the combined impacts of **w** and ρ that lead to endogeneity, and whereas ρ denotes an unobserved parameter with an associated probability distribution that must be estimated, **w** denotes given data. It is useful to note in passing that it would be desirable to estimate the elements of **w** but that most useful parameterizations lead to identification problems. We suspect that the development of more robust specifications of **w** is, perhaps, the most fruitful avenue for advances in spatial methods, at least where adoption issues are the focus. Finally, this statement assumes deeper meaning when it is recognized that (at least, in experimental

settings) the choices of $\mathbf{w}$ and ρ are not independent. In fact, as Anselin (1988) shows any draw for ρ in a Metropolis scheme must satisfy

$$(9) \qquad \underline{\rho} \leq \rho \leq \bar{\rho},$$

where $\underline{\rho}$ is the inverse of the minimum eigenvalue of $\mathbf{w}$ and $\bar{\rho}$ is the inverse of the maximum eigenvalue.

With reference to (8), in the case $\rho = 0$, we have the normal-linear model and all of the previous results go through, with the reinterpretation that $K = 1$ and $\mu_{(1\times1)} = \mathbf{b}_{(K\times1)}$ and $\mathbf{x}_{(N\times K)} = \mathbf{i}_{(N\times1)}$. Consequently, the extension to consider spatial effects rests importantly on the distribution for ρ. We will follow the practice outlined above and use a non-informative prior pdf for the unknowns $\mathbf{Q} \equiv (\rho, \mathbf{b}, \sigma)'$, form the likelihood for the unknowns conditional on the data $\mathbf{y}$, and study the form of the resulting posterior as a starting point to formal analyses. The posterior is

$$(10) \qquad \pi(\mathbf{Q}|\mathbf{y}) \propto |\mathbf{A}|\, \sigma^{-(N+1)} \exp\{ \frac{1}{2\sigma^2} (\mathbf{Ay}-\mathbf{xb})' (\mathbf{Ay}-\mathbf{xb})' \},$$

where $\mathbf{A}_{(N\times N)} \equiv \mathbf{I}_N - \rho\mathbf{w}$ arises from the Jacobian of the transformation between $\mathbf{y}$ and $\mathbf{e}$. It is important for later developments to recognize that the matrix $\mathbf{A}$ contains the unknown parameter ρ and that, when $\rho = 0$, $\mathbf{A} = \mathbf{I}_N$ and the model reduces to the normal-linear regression model. However, even for non-zero correlation, the joint posterior for $\mathbf{Q}$ is very similar to the posterior for the normal-means model and, so, many of those same concepts prevail. Because there are now three components of interest, we must fix two of these constant when developing the Gibbs strategy for estimation. First, in deriving the conditional distribution for σ, we treat $\mathbf{b}$ and ρ as known constants. Given $\mathbf{b}$ and ρ, inspection of (10) and comparison with Zellner (1996, p. 371, equation (a.37b)) reveals that the posterior for σ has the form

(11) $\pi(\sigma|\mathbf{b}, \rho, \mathbf{y}) \propto f^{IG}(\sigma|\nu, s^2)$,

an inverse-gamma distribution with $\nu \equiv N$ and $s^2 \equiv (\mathbf{Ay}-\mathbf{xb})'(\mathbf{Ay}-\mathbf{xb})/\nu$. Hence, the dependence of σ on $\mathbf{b}$, ρ and $\mathbf{y}$ is through the expression s^2. Second, holding σ and ρ constant, the posterior is in an identical form to the multivariate regression model with the dependent variable redefined to be $\mathbf{z}_{(N\times 1)} = \mathbf{Ay}$. Consequently, all of Zellner's results (pp. 65-66) go through with this reinterpretation and we find that the conditional distribution for $\mathbf{b}$, has the form

(12) $\pi(\mathbf{b}|\rho, \sigma, \mathbf{y}) \propto f^{N}(\mathbf{b}|\hat{\beta}, \mathbf{V}_{\hat{\mathbf{b}}})$,

$\hat{\beta} \equiv (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{z}$ and $\mathbf{V}_{\hat{\mathbf{b}}} \equiv \sigma^2(\mathbf{x}'\mathbf{x})^{-1}$. Thus, the full conditional distribution for $\mathbf{b}$ is multivariate normal with mean $\hat{\beta}$ and covariance matrix $\mathbf{V}_{\hat{\mathbf{b}}}$. Finally, with σ and $\mathbf{b}$ assumed fixed, we observe that the form of the posterior for ρ is precisely the form of the joint posterior (10). That is, $\pi(\rho|\sigma,\mathbf{b},\mathbf{y}) \propto \pi(\mathbf{Q}|\mathbf{y})$ and no further simplification is possible. Due to the appearance of the determinant resulting from the Jacobian transformation, this density does not have a well-known form and, hence, we are in the situation motivated previously of requiring the application of MCMC methods.

As we outlined above, in the case of the RW algorithm we will need a proposal density from which to generate draws. The normal distribution is a natural choice in each of the three cases due to the fact that, net of the Jacobian term, $|\mathbf{A}|$, the fully conditional density for ρ is normal with known mean and variance. More precisely, by completing the square in ρ (see the text above equation (7)) it is possible to write for ρ

(13) $\pi(\rho|\sigma,\mathbf{b},\mathbf{y}) \propto |\mathbf{A}| \exp\{ \frac{1}{2\sigma^2} (\rho-\hat{\rho})'(\mathbf{wy})'(\mathbf{wy})(\rho-\hat{\rho})\}$,

where $\hat{\rho} \equiv ((\mathbf{wy})'(\mathbf{wy}))^{-1} (\mathbf{wy})' (\mathbf{y}-\mathbf{xb})$. Hence, $\pi(\rho|\sigma,\mathbf{b},\mathbf{y}) \propto |\mathbf{A}| f^{N}(\sigma|\hat{\rho}, \mathbf{V}_{\hat{\rho}})$, where $\mathbf{V}_{\hat{\rho}} \equiv \sigma^2 ((\mathbf{wy})'(\mathbf{wy}))^{-1}$. The fact that the full conditional distribution for ρ contains a normal component makes it sensible to choose as the proposal distribution a normal density. And this is what we do in the demonstration that follows.

In the experiment we use the data in table 1 (column 2), which are simulated from the model (8) with $\mathbf{Q} = (\rho, \beta, \sigma)' = (-0.75, -1.0, 1.0)'$ and (column 3) $\mathbf{x} \sim f^{N}(\mathbf{x}|\mathbf{0}_N, 10\mathbf{I}_N)$. In each of the three cases we sample sequentially with the insertion of an additional step to simulate the draw for ρ, which is not available as a known pdf. We use r to denote the candidate draws from the proposal density and use ρ to denote an accepted draw. Some experiments suggested that an acceptance rate of around fifty percent produced stable estimates in a timely manner and we therefore endogenized the step size of the RW algorithm by allowing the standard deviation in the random walk error, ξ, to increase (respectively, decrease) by a scale factor 1.1 whenever the acceptance rate exceeded the upper bound from below (exceeded the lower bound from above) in a band set at acceptance rate limits of forty and sixty percent.

Figure 7 plots the first 100 draws in the Gibbs sequence with the RW step inserted. The sequence was quick to converge and produced draws for each of the three parameters that are close to the given values used to generate the data, namely $(\rho, \beta, \sigma)' = (-.75, -1, 1)$. The sample was obtained in less than a minute of real time. Figures 7 presents plots of the histograms for ρ generated from the experiment. The distribution is centered close to the given parameter value (-0.75) and is approximately normal. Experiments with different start values generated almost identical distributions and we conclude that the RW-Gibbs sequence produces robust estimtes of the SAR model parameters.

(Insert figures 7 and 8 about here.)

Random-Walk Metropolis Sampling The Bayesian Spatial Probit

Our final demonstration is the framework we apply to the empirical model introduced in the next section. Having provided a heuristic justification for the methods in the context of continuous sample data **y**, we consider the application of the RW algorithm to the spatial probit model. With reference to (8), we are now concerned with the model

(14) $$\mathbf{z} = \rho\, \mathbf{w}\mathbf{z} + \mathbf{x}\mathbf{b} + \mathbf{e},$$

and we observe $y_i = 1$ if $z_i > 0$ and $y_i = 0$, otherwise. Hence, the components of $\mathbf{z} \equiv (z_1, z_2, .., z_N)'$ as opposed to $\mathbf{y} \equiv (y_1, y_2, .., y_N)'$ are latent and, in terms of our desire for step-by-step development, we have introduced one additional unknown into the model. Consequently, the Gibbs sampling algorithm will require one additional step for its implementation and that step—as might now be expected—is to obtain a draw for the latent endogenous variable, **z**, from its fully conditional distribution. This distribution is obtained by interpreting the posterior solely as a function of the unknown vector **z** and by completing the square in the exponential part of the normal kernel. The product is the truncated, multivariate-normal distribution

(15) $$\pi(\mathbf{z}|\rho,\sigma,\mathbf{b},\mathbf{y}) \propto f^{TN}(\mathbf{z}|\hat{\mathbf{z}}, \mathbf{V}_{\hat{z}}),$$

where $\hat{\mathbf{z}} \equiv (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{x}\mathbf{b}$ and $\mathbf{V}_{\hat{z}} \equiv \sigma^2(\mathbf{A}'\mathbf{A})^{-1}$ and the truncation satisfies the conditions stated in the data generating model (14). Although this single step is but a slight complication over the SAR algorithm, drawing from this conditional distribution can pose problems. This is because a simple acceptance scheme whereby the latent **z** is accepted if each component satisfies the inequality constraints has a very small chance of acceptance. Except for very small problems, this method is computationally impractical. The approach suggested by Geweke (1992) and adopted previously by LeSage (2002) is to use the acceptance scheme on the fully conditional distributions for each of the components of the vector **z**. However, this

approach can still result in an unreasonably large number of rejected draws. The alternative is to retain the conditional approach but use efficient one-for-one draws by applying the probability integral transform (e.g., Mood, Graybill and Boes, 1974, p. 202). We found that both methods generated accurate estimates, but that the probability integral transform method was far superior in terms of execution time.

One additional modification to the previous algorithm is required prior to implementation. This adjustment is to fix one of the unknown parameters in order to identify the other unknowns in the model. This is the familiar scaling problem that exists in conventional probit estimation and arises because the probit model is valid only up to a scalar transformation. The usual practice, which we adopt here, is to fix the variance at one. Hence, (13) is implemented by imposing the restriction $\sigma = 1$.

An experiment is implemented using the second and third columns of table 1. Note in column 2 that four of the endogenous values are positive and six are negative. If we simulate draws assuming that these binary outcomes represent the observed data we expect to obtain draws for the latent data that are centered about the table values. This, essentially, is the case and, although space prohibits reporting the probability distribution so obtained, the histogram for ρ is only slightly different from the distribution obtained from the spatial model with observable, continuous response data. Figure 9 compares the draws for the spatial econometric model (SAR) and the spatial probit model (SARP). The simulated frequencies from the SARP model are the unfilled bars and the frequencies simulated by the SAR model are the filled bars. Both sets of frequencies are generated from a Gibbs sample of 1,000 accepted draws. The start values for ρ in the two models are the same and we follow a suggestion in Albert and Chib using $\mathbf{z}^0 = \mathbf{y}$ as the start value in the spatial probit algorithm. Experiments with other start values suggest that the results are independent of this choice.

Except for some skewness in the draws from the probit model, the frequencies generated by the two models are remarkably similar.

(Insert figure 9 about here.)

In summary, the Gibbs sampler provides an extremely powerful technique for simulating from a marginal distribution that is not available in closed form. The addition of a Metropolis step within the algorithm provides considerable versatility that facilitates estimation of spatial econometric and spatial probit models. The technique is attractive and, when viewed as a set of logical extensions to the normal linear model, is mostly rather simple. Its practical implementation met with few obstacles in a hardware/software environment that is widely available to other researchers. The technique provides accurate estimates of spatial model parameters and appears to be extremely robust, working well in a limited-information environment (ten observations). The success of the technique in the experimental settings raises considerable scope for its application in empirical work.

4. Empirical Application

Application of the SAR probit model, equation (14) to neighborhood effects in HYV rice adoption evolves from previous work by Case (1992). Readers are referred to that source for further details and motivation. Briefly, a farmer's expected profit from adopting an HYV plant, in place of a local variety (LV) depends upon a set of a set of price variables (input and output prices), a set of fixed factors (say, farm assets, land holding), a set of socio-economic characteristics (for example, education, wealth), and neighborhood influences (expected profits to neighbours from adoption). The first three sets of characteristics are, of course, standard fare in adoption models. They are accounted for by the matrix $\mathbf{x}_{(N\times K)}$ in equation (14). The fourth effect is, of course, modelled through the combination of the spatial weight matrix $\mathbf{w}_{(N\times N)}$ and the spatial correlation parameter ρ. In Case's application of to sickle-

harvester adoption in Java, the term 'neighbors' refers to all other farmers in the same district. All neighbors are weighted equally, and the neighborhood effects for each farmer are normalised to 1. In other words, the row restriction $\sum_j \mathbf{W}_{ij} = 1$ is imposed column-wise on the rows. We follow these procedures in the this empirical application.

In the Bangladeshi context, rice is the staple for the vast majority of the population, and the predominantly agrarian economy revolves around the production of rice year-round, in three seasons. Food security continues to be a predominant concern, with the population expanding by 2.2 million a year. The Bangladesh Rice Research Institute has released dozens of varieties of HYVs over the years, and these modern varieties are known to enable substantially better yields than local varieties. In spite of such varietal development activity and progress in irrigation provision, Bangladesh has one of the lowest HYV adoption rates in Asia (Azam, 1998). The adoption issue is, thus, a critical one for Bangladesh.

A literature does exist on HYV adoption in Bangladesh, mostly employing OLS or probit regressions of adoption on variables such as farm size and farmer education (see, for example, Hossain (1989), Ahmed and Hossain (1990) and Allaudin and Tisdell (1991)). However, as in the broader technology adoption literature, this literature has not considered the role played by the adoption attitudes of 'neighbors' in influencing the adoption decisions of individual farmers. As Case (1992) has argued, ignoring neighborhood influences not only biases the estimated parameters in standard adoption models, but also sacrifices important policy-relevant information. For example, a key principle in rural extension activity in many developing countries is to disseminate information to a critical mass of farmers such that positive externalities in the form of 'secondary' or 'copy' adoption in the locality carry forward the momentum generated by the initial investment. The size of this externality constitutes important data for policymakers operating under limited budgets and wishing to maximise returns to extension investment.

There is evidence that a village-level synergy exists in technology adoption in Bangladesh. One example of this is the experience of INTERFISH, a large agriculture/aquaculture extension project designed to promote the spread of rice-fish cropping systems in rural Bangladesh. An external team reviewing the achievements of the project found that 'copy farmers' (secondary adopters) abounded in the areas where the project was based (Best et al. 1998). These secondary adopters were seen to commit increasing amounts of land to the new technology following positive adoption decisions by neighboring farmers. The review team estimated that the spatial reach via secondary adoption could be a radius of two to three kilometres and, on this basis, suggested that future project activities should allow sufficient space between sites in order to maximize returns to the project's investment.

The empirical application applies the spatial lag model to cross-sectional, primary data for the Aman (monsoon rice) crop of 1996 in Bangladesh. The data were collected in an intensive farm-survey coordinated by one of the authors. Multistage random sampling techniques were used in selecting study locations as well as the sample respondents. In our application, neighborhoods are defined by villages; in other words, it is assumed that the attitude towards HYV adoption for farmer i depends not only on its own internal characteristics, but also on the influence of other farmers in the village. The effects of farmers in surrounding villages is assumed to be negligible. The survey, conducted in 1996, had strict constraints on budgets and personnel, and was not designed with spatial estimation in mind. Thus surveys were carried out in clusters of villages in each of the 3 districts. The three clusters had 8, 7 and 6 villages, respectively, making 21 villages in all in our sample. The districts (clusters) themselves are hundreds of kilometres apart from each other, and therefore can safely be considered not to be in each other's neighborhoods. Within each cluster, individual villages are between 2.5 and 8 kilometres apart. While these are not apparently great distances, our experience of Bangladesh enables us to be reasonably

comfortable with the assumption that none of the villages is in the other's 'neighborhood' (interactions terms between villages in the **w** matrix are zero). Agrarian activities in Bangladesh are known to be intricately linked with the socio-economic dynamics of individual villages (Herbon, 1994). At the same time, villages are typically located around small waterbodies, and the agriculture is strongly adapted to the local micro-relief. It is therefore not surprising to find very different mixes of crops and modes of production in areas just a few miles away from each other. In light of these facts, and given the geographical makeup of our data, we have adopted the 'village' definition of 'neighbors'.

The survey collected information on varietal choice; input and output prices; levels of fixed factors; and socio-economic characteristics of the farm families. A total of 406 observations on local varietal use (76 observations) and modern varieties (330 observations) constitute the sample. The variables included in the model are the following: DISTRICT: Dummy variable representing the district in which the farm is located. NGO: Dummy variable indicating whether the household received assistance from NGOs. ASSETS: Value of farm assets in thousands of Taka. EDUCATION: Number of years of schooling for household head. EXPERIENCE: Years of farming experience for household head. HOUSEHOLD SIZE: Number of people in household. FARM SIZE: Total size of holding in hectares. RENTED HECTARES: Hectares of land rented in hectares. MARKET: Distance from nearest market ('growth center'), in kilometres. RICE MILL: Distance from nearest rice mill. EXTENSION: Distance from nearest Department of Agriculture Extension office.[3] Table 2 presents summary statistics by adoption status.

(Insert table 2 about here.)

5. Results

Estimation results both with and without neighbourhood influences are presented in table 3. Confidence intervals (highest posterior density intervals) at the 95% percentile are reported in parentheses. Because the qualitative effects (signs of coefficients) of most of the covariates remain the same between the two models, it is worth contemplating the interpretation of the qualitative effects before discussing differences in coefficient magnitudes between models.

Of the human capital variables (education, experience), only education is significant. While the negative and significant estimate for education may appear counter-intuitive, it is consistent with the findings of a previous study based on a simple probit estimate from earlier data. Rosenzweig (1981) postulates that education can affect new technology adoption in different ways. On the one hand, it can encourage adoption by lowering learning costs. On the other, it may discourage adoption since education provides more profitable off-farm employment opportunities, and new technologies may reduce the ability of farm operators to substitute their time inputs away from cultivation. Although education has been found to positively affect HYV adoption in other rice economies such as Indonesia (Pitt and Sumodiningrat, 1991), we are able to strengthen the evidence for the opposite trend in the case of Bangladesh.

NGO contact and asset values of the farm have insignificant coefficients. Family size is insignificant too; however it becomes significant at the 10% level (not reported here). With the exception of planting and harvesting periods in which all family members contribute to operations, routine laboring is undertaken by adult males. But in the planting and harvesting periods, there is an acute shortage of labor (Metzel and Ateng, 1993), and every spare hand, including children, is pressed into farm work. This phenomeon is even more acute in HYV cultivation, where crop management is generally more labor intensive. In this

regard, a larger family size is more conducive to HYV adoption, and this observation is reflected in the table 3 estimates. A consumption-based explanation for this phenomenon has also been advanced for the case of Bangladesh. Hossain (1989) finds a similar result and interprets it as a confirmation of the Chayanovian hypothesis that higher subsistence pressure leads to greater adoption of new technology.

Farm size and rented hectares both have significant coefficients. Smaller farms appear to have a greater propensity for HYV adoption. Once again, a 'subsistence pressure' argument fits well with this finding. Land rental imposes an additional payment burden in cash or crop-share, which may provide an incentive for the adoption of higher surplus yielding varieties.

Turning to the variables measuring infrastructural underdevelopment (distances to markets, rice mills and extension offices) we find that their coefficients are all insignificant. While the sign on the coefficient for distance from extension is counterintuitive in addition to being insignificant, coefficients for distances from markets have the expected sign. With distances from markets and rice mills increasing, one would expect incentives for HYV adoption to be depressed. The insignificance of coefficients for these variables possibly indicates that the underlying variables are not well measured. In capturing the effects of extension activity on adoption, for instance, one would ideally like to have actual measures of extension contact, such as numbers of visits from extension agents. Unfortunately, these data are unavailable.

Two measures of primary importance in the study are the signs and magnitudes of the neighborhood correlation coefficient, ρ. The posterior means estimate of this parameter is 0.54, and the confidence interval presented in table 1 suggests that the estimate is significantly different from zero. Figure 10 presents the complete distribution of draws from the Gibbs sample. This distribution, recall, is derived from a random-walk Metropolis step.

The distribution is uni-modal and appears to be almost symmetric. Its key fature is its location. Very little of the density resides in the negative part of the real line. Hence, with few caveats, there exists a strong, positive neighborhood effect among the Bangladeshi respondents.

(Insert figure 10 about here).

With significant local synergies in adoption confirmed, the question arises about the extent to which ignoring these influences biases policy conclusions. The marginal probabilities reported in table 3 do not seem to vary greatly between the spatial and non-spatial models for most variables. For example, the presence of one additional family member increases the probability of adoption by 8% according to the traditional model. In contrast, inclusion of neighborhood effects results in a marginal probability of 11%. However, the difference in predictions is seen to be substantial for a few variables. For instance, both models predict increased distances from markets to depress adoption probabilities. But the non-spatial model overstates this effect very significantly, predicting that ceteris paribus every additional kilometre from the market reduces adoption probability by 12%, while the spatial model estimates the same effect to be only 6%.

Finally, we consider the implications of ignoring the neighborhood effects in one situation that has significant relevance in the Bangledeshi context. This is the notion of some form of optimal size in the structure of farm units. From the estimates for the probit and spatial probit models we note that there is a significantly negative response to farm size. The larger the farm the less inclined an operator is to adopt HYV inputs. But, in policy discussions concerning this potentially important effect it is natural to ask the size of farm at which adoption status changes. In other words, we seek the size of the farm operation at which a non-adopter decides to adopt and the size at which an adopter decides to reject the HYV input. This quantity will vary among respondents and may be useful for planning

purposes and land-use strategies and it is desirable to have an estimate of this quantity in policy discussions.

Our sample is quite 'unbalanced.' About 80% of the sample consists of adopters. This imbalance makes policy prediction more difficult. Nevertheless, as we change farm size, we can imagine a spectrum of 'reservation values' (specified in terms of hectares of land holding) at which each respondent changes adoption status. These reservation values are estimable for each farmer through the insertion of one additional step in the Gibbs algorithm. This step is to find the level of the covariate in question for which the dependent variable in the regression model is exactly zero or, in other words, the level $\hat{x}_i$ such that

$$\hat{x}_i = \frac{-\mathbf{x}_{-j}\mathbf{b}_{-j}}{\beta_j}, \tag{16}$$

Here $\mathbf{x}_{-j}$ denotes the covariate matrix with the column corresponding to the land variable (column j) excluded, $\mathbf{b}_{-j}$ denotes the corresponding coefficient vector, and β_j denotes the coefficient of the land variable in the original regression. Due to the appearance of the latter in the denominator the left hand side of (16) does not have a form that enables direct simulation. But, once again, using the Gibbs sequence we are able to generate a sample of draws for $\hat{x}_i$ and, in so doing, characterize its location and scale. Although the estimates themselves may be extremely important for policy purposes, in the spirit of the methodological contributions of the paper we are mostly interested in how these estimates are affected by the exclusion of neighborhood effects.

Figure 11 presents two sets of estimates of 95% highest-posterior density zones for the quantities in (16) from the standard probit model (dotted lines) and the spatial probit model (solid lines). The two distributions are dissimilar with the estimates obtained from the SARP model considerably more precise. The importance of allowing for neighbourhood impacts appears, thus, to play an important role in the Bangladeshi data and, to reiterate the

cautionary remarks offered by Case, their exclusion significantly biases empirical results and, thus, the policy conclusions that evolve from them. A strong, positive neighbourhood effect is present in the Bangladeshi data and we must take care to account for it in devising policy prescriptions.

(Insert figure 11 about here.)

5. Conclusions

Markov chain Monte Carlo methods have completely revolutionized Bayesian inference. Problems that were not manageable just a decade ago have become routine and, with them, Bayesian philosophy is making inroads in many fields of empirical research. In this paper we provide a stepping-stone primer to Bayesian spatial probit estimation and demonstrate its importance in policy formation. Policy conclusions are affected by the propensity of adoption decisions by neighbors to affect others and we find a strong, positive neighbourhood effect in the Bangladeshi data. This conclusion is obtained robustly through a simple extension of a basic algorithm used to estimate the normal linear model. The algorithm is implemented with hardware and software that is widely available to other researchers and generates precise estimates of policy parameters, efficiently, robustly and with few computational demands. If adopted by our neighbors, the class of techniques presented here are likely to stimulate additional advances in the growing field of applied spatial inference.

Footnotes

[1] The convention followed in sampling theory of presenting t-statistics associated with regression parameters is not followed here for two reasons. First, because the simulated distributions are, themselves approximations to t-distributions the result that normalization by standard errors brings the estimate into a t-distribution no longer holds. Second, the highest posterior density regions are the conventional statistics in Bayesian applications.

[2] As a referee points out, it is possible in some settings that adoption encourages land rental instead of vice-versa, because the surplus generated by HYV adoption may prompt farm expansion by rental. In the Bangladesh context, however, it has been our observation that renting is common for marginal, entrant farmers as well as larger, more commercial farms. Indeed, land is so coveted and in such short supply that renting is often the only way in which landless labourers can become cultivators themselves.

[3] Unfortunately, we do not have data on one key variable, soil/land quality, which could potentially cause spatial correlations. However, a recent study by Barr (2000) in Bangladesh has found that soil properties do not seem to affect cropping decisions by farmers.

[4] Additional references on Bayesian estimation, especially in a spatial context are available at Jim LeSage's web site (www.spatial-econometrics.com).

References

Ahmed, R., Hossain, M. 1990. Developmental Impact of Rural Infrastructure in Bangladesh, International Food Policy Research Institute, Washington,D.C.

Albert J, Chib S. 1993. Bayesian Analysis of Binary and Polychotomous Response Data. Journal of the American Statistical Association 88: 669-79.

Alauddin, M., Tisdell, C. 1991. The Green Revolution and Economic Development. MacMillan Press Ltd, London.

Azam, J-P.,1996. The impact of Floods on the Adoption Rate of High Yielding Rice Varieties in Bangladesh. Agricultural Economics 13,179-189.

Barr,J., 2000. Investigation into Livelihood Strategies and resource use Patterns in Floodplain Production Systems in Bangladesh. Project report to the Natural Resource Systems Programme, Department for International Development, UK.

Best, J., McKemey, K., Underwood, M., 1998. CARE INTERFISH – Output to Purpose Review. Agricultural Extension and Rural Development Department, University of Reading, UK.

Case, A. 1992. Neighborhood Influence and Technological Change. Regional Science and Urban Economics 22, 491-508.

Chib, S. Marginal Likelihood From the Gibbs Output. Journal of the American Statistical Association, (1995), 90, 1313-1321.

Chib, S. and I. Jeliazkov. Marginal Likelihood from the Metropolis-Hastings Output. Journal of the American Statistical Association, (2001), 96, 270-281

Dubin, R. 1992. Estimating Logit Models With Spatial Dependence. In New Directions in Spatial Econometrics, Anselin, A. and R. J. G. M. Florax. Berlin: Springer-Verlag.

Eunus, M. 2001. Crop Production in Bangladesh: Constraints and Opportunities. Advances in Agronomic Research in Bangladesh. 5, 1-16.

Gelfand A E, Smith A F M. 1990. Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association. 85: 398- 409.

Geweke, J. Personal communication, June 5, 2002.

Gilks, W. R., S. Richardson and D. J. Spiegelhalter. Markov Chain Monte Carlo in Practice. London: Chapman & Hall, 1996.

Gilley, O. W. and R. K. Pace. "On the Harrison and Rubinfield Data." Journal of Environmental Economics and Management 31 (1996)): 403-05.

Gilks, W. R., Richardson, S. T. and Spiegelhalter, D. J. (eds.) Markov chain Monte Carlo in Practice. London: Chapman&Hall, 1996.

Hossain, M.,1989. Green Revolution in Bangladesh : Impact on Growth and Distribution of Income, Dhaka, The University Press Ltd.

Jeffreys, H. 1961. Theory of probability. Ocford University Press, Oxford, 3rd ed.

Kashem, M.A. 1987.. Small Farmers' Constraints to the Adoption of Modern Rice Technology. Bangladesh Development Studies 15, 119-130.

LeSage, J. P. 1993. Spatial Modelling of Agricultural Markets. American Journal of Agricultural Economics 75,1211-16.

LeSage, J. P. "Bayesian Estimation of Spatial Autoregressive models." International Regional Science Review 20: 1 and 2 (1997): 113-29.

LeSage, J. P. "Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models." Geographical Analysis 32: 1 (January 2000): 19-35.

Lesage, J. P. "Application of Bayesian Methods to Spatial Econometrics." Mimeograph, Department of Economics, Toledo, Ohio, March 2002.

LeSage, J. P. Personal communication, Februrary, 2002.

LeSage, J. P. Personal communication, June 28, 2002.

Metzel, J. and Ateng, B. 1993. Constraints to Diversification in Bangladesh: A Survey of Farmer's Views. Bangladesh Development Studies 21, 39-72.

Mitchell, D., Islam, T. 1998. Growth Potential in Bangladesh Agriculture. Unpublished Report, World Bank, Washington, D.C.

Mood, A. M., F. A. Graybill and D. C. Boes. 1974 Introduction to the Theory of Statistics. New York: McGraw-Hill.

Pitt, M. and Sumodiningrat, G. (1991). Risk, Schooling and the Choice of Seed Technology in Developing Countries: A Meta-Profit Function Approach. International Economic Review, 32: 457:473.

Poirier, D. Personal communication. June 5, 2002.

Raftery, A and S. M. Lewis. "How Many Iterations In The Gibbs Sampler?" In Bayesian Statistics, Volume 4, Berger, J. O., A. P. Dawid and A. F. M. Smith (eds.) Oxford: Oxford University Press, 1992.

Robert, C. P. 1995. Convergence Control Techniques for Markov chain Monte Carlo algorithms." Statistical Science

Robert, C and G. Casella. 1999. Monte Carlo Statistical Methods. New York, Springer.

Rosenzweig, M. R. (1982), Agricultural Development, Education and Innovation, in Greschwitz, M., Diaz-Alejandro, Ranism G. and Rosenzweig, M, R. (eds.), The Theory and Experience of Economic Development: Essays in Honor of Sir Arthur Lewis, George Allen Unwin.

Tanner M H, Wong W H. 1987. The Calculation of Posterior Distributions By Data Augmentation. Journal of the American Statistical Association 82: 528-50.

Zellner A. 1996. An Introduction to Bayesian Inference in Econometrics. Wiley Classics Library. New York: John Wiley and Sons.

Tables

Table 1. Experimental data

Normal Means	Regression	
y_i	y_i	x_i
-0.26	-5.35	4.82
-0.36	1.93	-5.17
0.61	10.69	-11.05
-0.70	19.56	-14.32
-2.38	1.11	-1.44
0.44	10.45	-10.89
-3.60	-8.33	8.37
-1.79	-5.46	-2.77
-1.13	6.40	-3.08
-0.56	-4.24	1.58
-0.97	2.67	-3.39
$\hat{y} = -0.97$	$\hat{y} = 8.98$	$\hat{x} = -3.39$
$\hat{\sigma} = 1.30$	$\hat{\sigma} = 2.67$	$\hat{\sigma} = 7.24$

Table 2: Summary statistics by adoption status.

Variable	Non-dopters Mean	Non-adopters Std dev.	Adopters Mean	Adopters: Std. Dev
NGO	0.26	0.44	0.22	0.42
ASSET	4620	10058	4762	14285
EDUCATION	5.29	4.58	3.26	4.26
EXPERIENCE	23.13	13.84	25.87	15.23
FAMILY SIZE	6.28	2.54	5.84	2.43
FARM SIZE	0.67	0.69	0.65	0.55
RENTED LAND	0.09	0.18	0.19	0.29
MARKET	3.77	2.56	2.78	1.49
RICE MILL	6.42	4.97	8.74	5.29
EXTENSION	10.08	5.47	12.94	5.04

Table 3. Equation estimates and marginal effects.

Variable	Spatial probit coefficient	Non-spatial probit coefficient	Spatial probit marginal effects	Non-spatial Probit marginal effects
ρ	0.54			
	(0.41, 0.67)			
District 1	0.01	0.22	0.00	0.06
	(-0.70, 0.71)	(-0.50, 0.99)	(-0.25, 0.24)	(-0.14, 0.27)
District 2	0.29	0.36	0.10	0.11
	(-0.84, 1.46)	(-1.12, 1.90)	(-0.29, 0.51)	(-0.35, 0.58)
District 3	-0.35	0.94	-0.12	-0.28
	(-1.85, 1.17)	(-2.88, 0.91)	(-0.64, 0.42)	(-0.93, 0.30)
NGO	-0.20	-0.24	-0.07	-0.07
	(-0.47, 0.04)	(-0.52, 0.02)	(-0.16, 0.02)	(-0.16, 0.01)
ASSET	0.03	0.03	0.01	0.01
	(-0.01, 0.08)	(-0.02, 0.08)	(-0.00, 0.03)	(-0.01, 0.02)
EDUCATION	-0.15	-0.16	-0.05	-0.05
	(-0.28, 0.02)	(-0.29, 0.03)	(-0.10, -0.01)	(-0.09, -0.01)
EXPERIENCE	-0.12	-0.12	-0.04	-0.04
	(-0.36, 0.10)	(-0.35, 0.10)	(-0.12, 0.04)	(-0.10, 0.02)
FAMILY SIZE	0.32	0.25	0.11	0.08
	(-0.02, 0.66)	(-0.08, 0.58)	(-0.01, 0.23)	(-0.02, 0.18)
FARM SIZE	-0.21	-0.18	-0.07	-0.06
	(-0.38, -0.03)	(-0.35, -0.01)	(-0.13, -0.01)	(-0.11, -0.00)

RENTED LAND	0.17	0.17	0.06	0.05
	(0.08, 0.26)	(0.08, 0.26)	(0.03, 0.09)	(0.03, 0.08)
MARKET	-0.17	-0.43	-0.06	-0.12
	(-0.52, 0.19)	(-0.86, -0.02)	(-0.18, 0.06)	(-0.27, 0.00)
RICE MILL	-0.19	-0.27	-0.07	-0.08
	(-0.55, 0.20)	(-0.78, 0.29)	(-0.19, 0.07)	(-0.24, 0.07)
EXTENSION	0.50	1.21	0.17	0.36
	(-0.36, 1.40)	(0.13, 2.36)	(-0.13, 0.47)	(0.03, 0.71)
adoption percent (non-adopters)	70%	77%		
adoption percent (non-adopters)	88%	75%		

Note: 95% highest posterior density regions in parentheses.

Figures

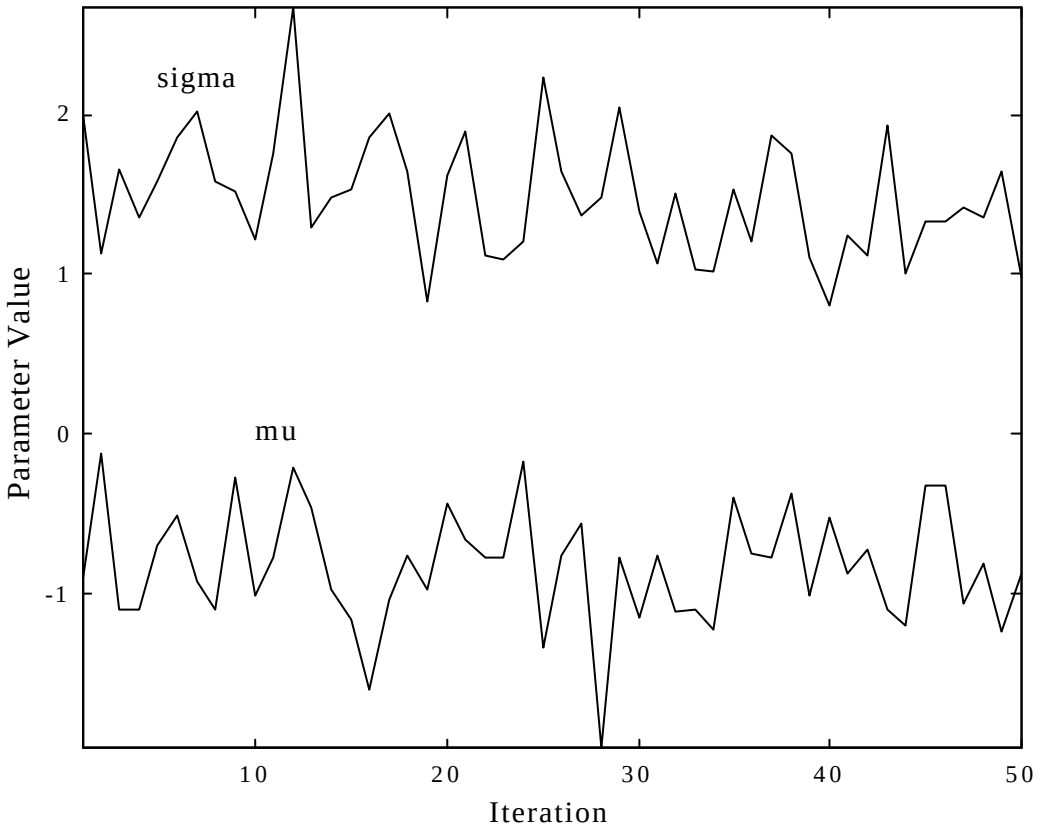

Figure 1. Plots of the first fifty draws in the Gibbs sample based on ten observations drawn from N(-1,1) and start value $\mu^0 = (\mathbf{i}'\mathbf{i})^{-1}\,\mathbf{i}'\mathbf{y}'$.

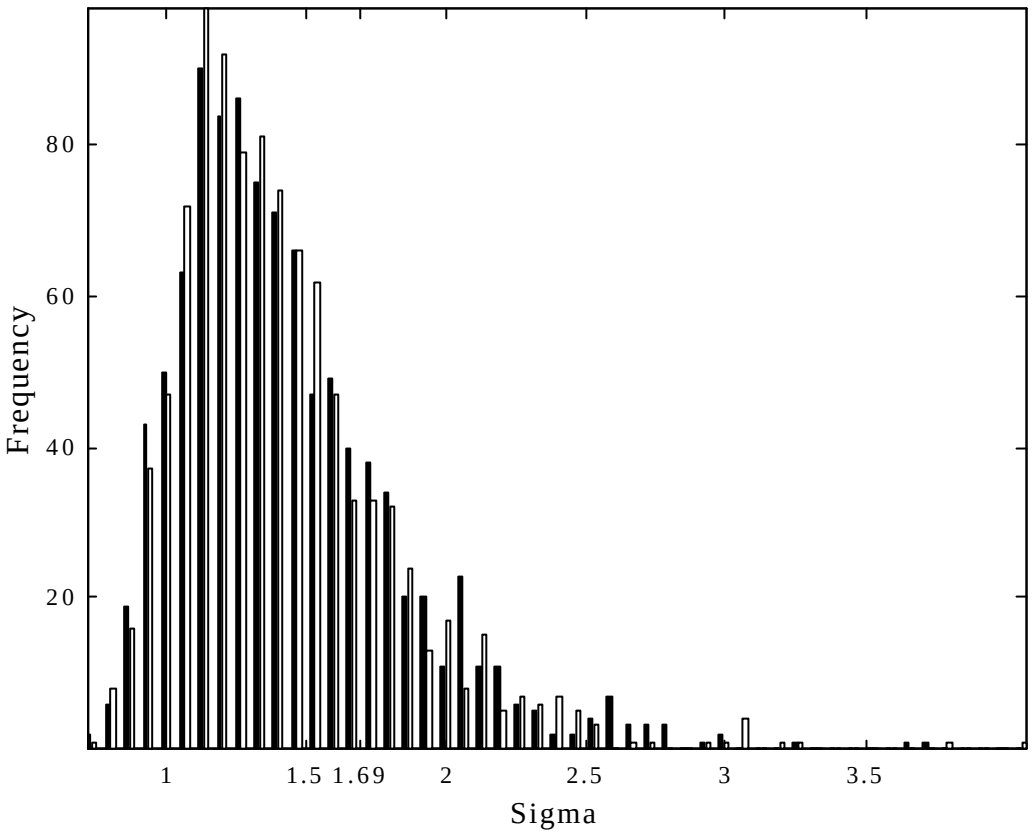

Figure 2. Plots of simulated frequencies (filled bars) for sigma from the Gibbs sample versus true frequencies (unfilled bars) from the inverse-gamma distribution.

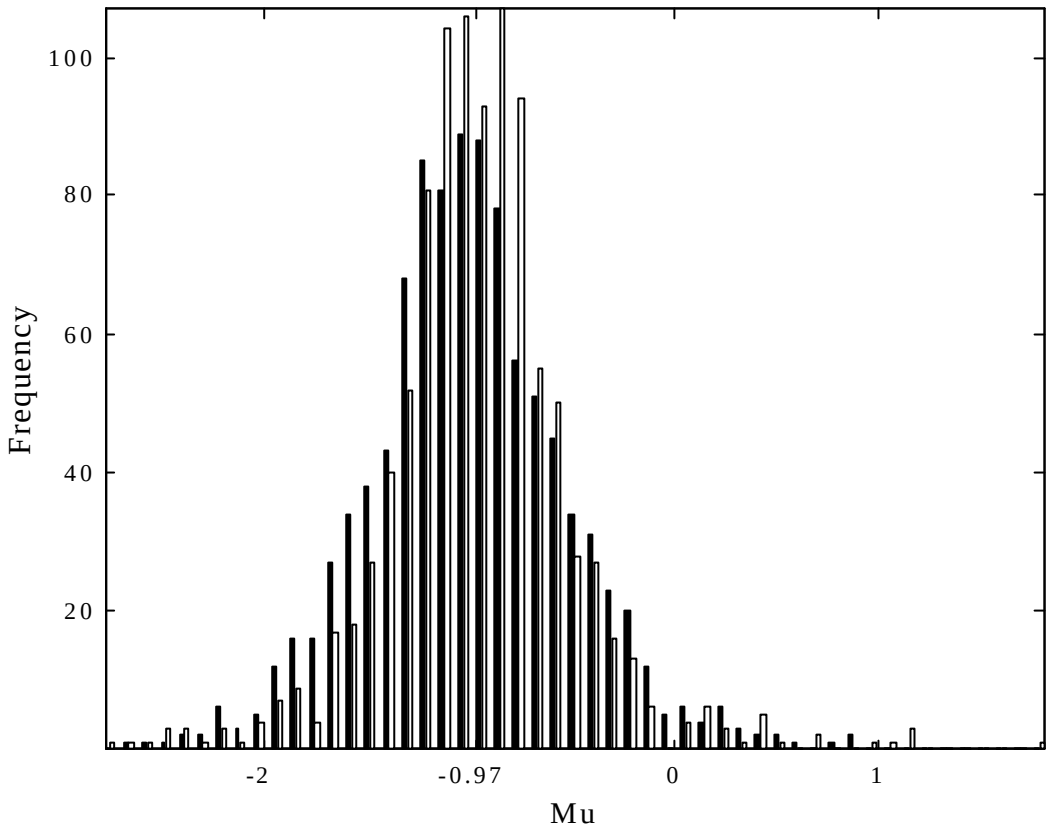

Figure 3. Plots of simulated frequencies (filled bars) for mu from the Gibbs sample versus true frequencies (unfilled bars) from the $f^{T}(\mu|\hat{\mu}, \nu, \hat{\sigma})$ distribution.

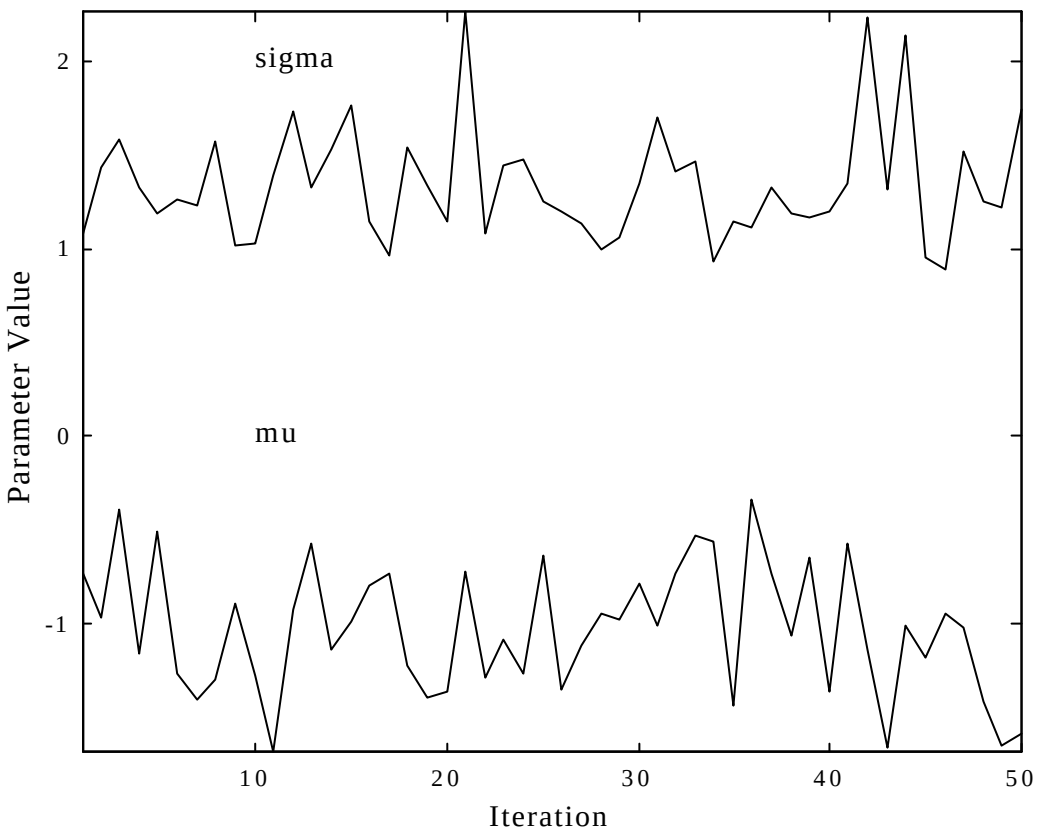

Figure 4. Plots of the first fifty draws in the Gibbs sample based on ten observations drawn from N(-1,1) and start value $\mu^0 = (\mathbf{i}'\mathbf{i})^{-1}\,\mathbf{i}'\mathbf{y}'$.

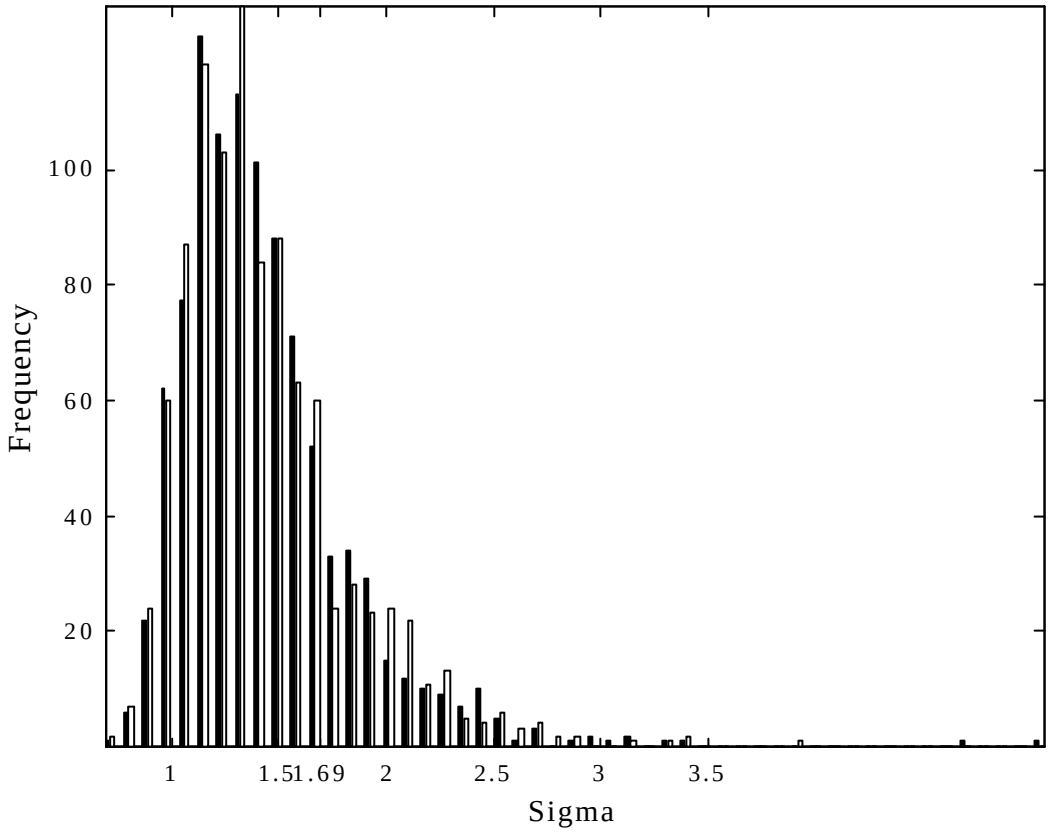

Figure 5. Plots of simulated frequencies (filled bars) for sigma from the Gibbs sample using a random-walk Metropolis step versus true frequencies (unfilled bars) from the inverse-gamma distribution.

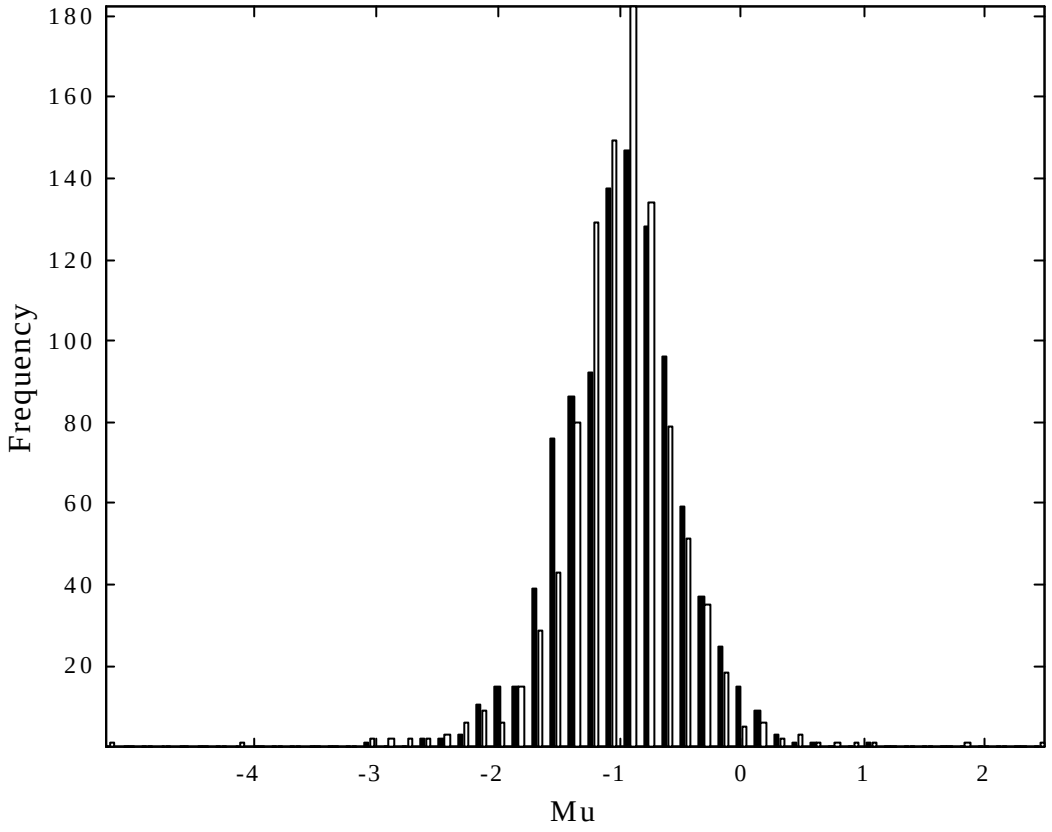

Figure 6. Plots of simulated frequencies (filled bars) for mu from the Gibbs sample using a random-walk Metropolis step versus true frequencies (unfilled bars) from the $f^{T}(\mu|\hat{\mu}, \nu, \hat{\sigma})$ distribution.

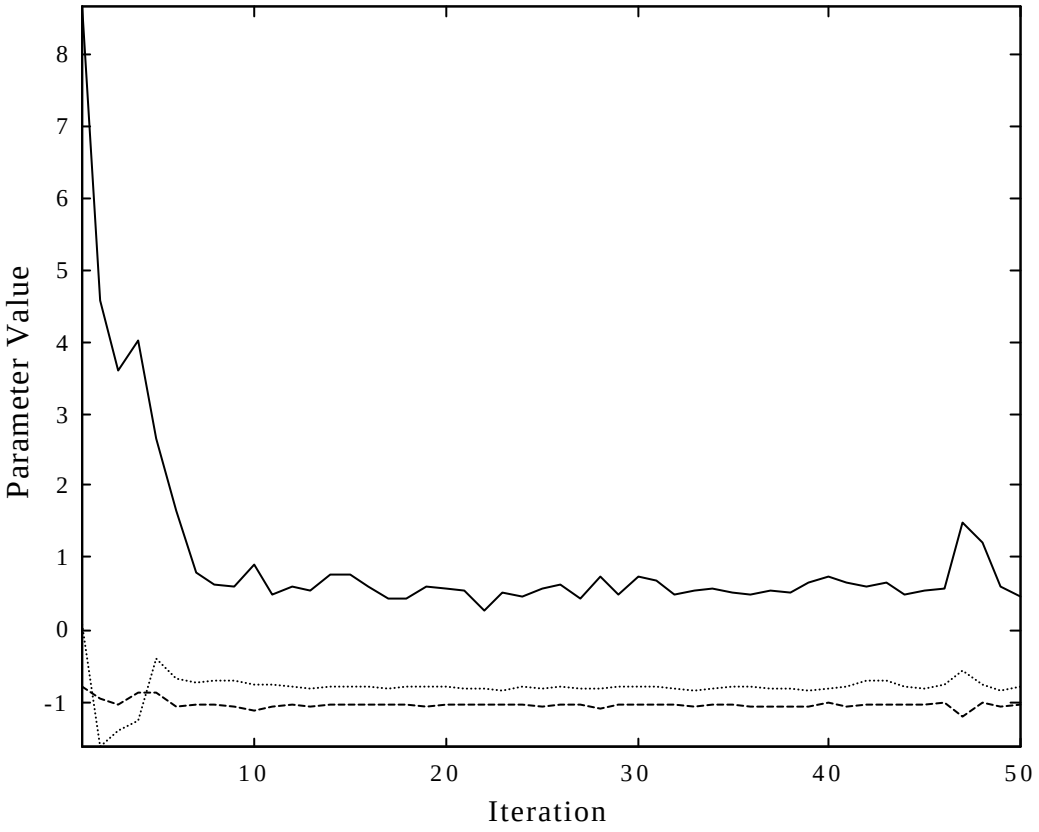

Figure 7. Plots of the first fifty draws in the Gibbs sample for the SAR model with a random-walk metropolis step or σ (solid line), ρ (dotted line) and β (dashed line) based on the ten observations table 1 (columns 2 and 3) and start values $\mu^0 = (\mathbf{i}'\mathbf{i})^{-1}\,\mathbf{i}'\mathbf{y}'$, $\rho = 0$.

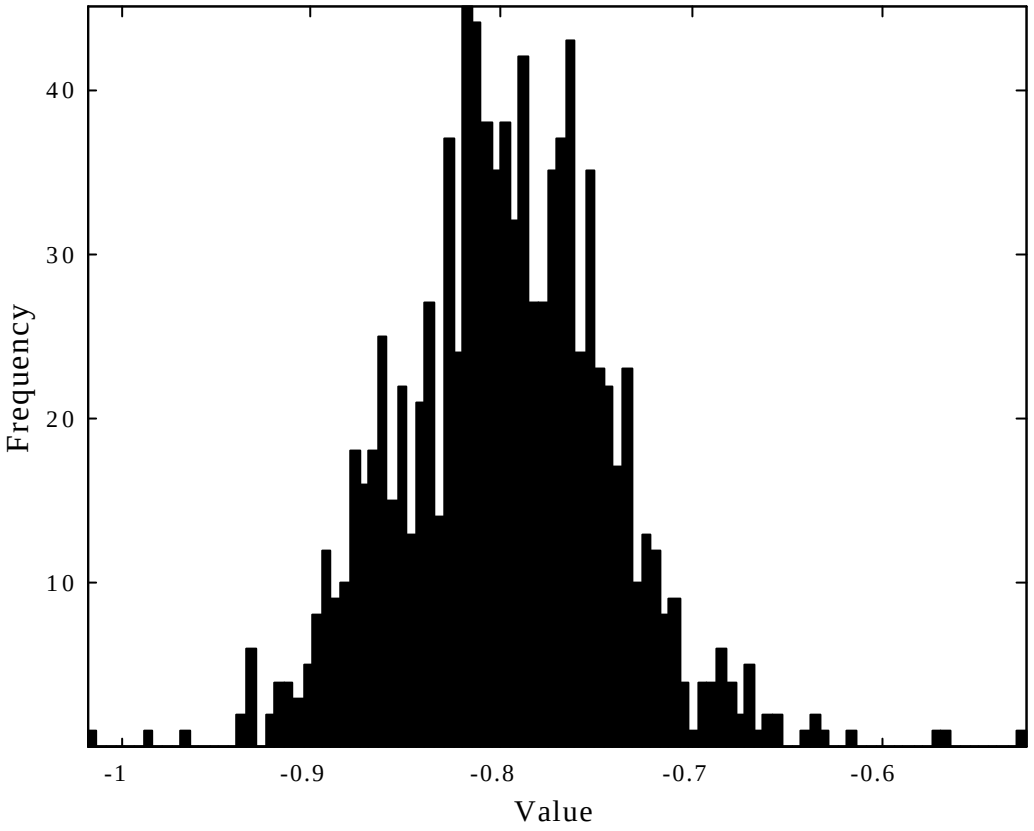

Figure 8. Plots of simulated frequencies for ρ in the SAR simulation using a random-walk Metropolis step.

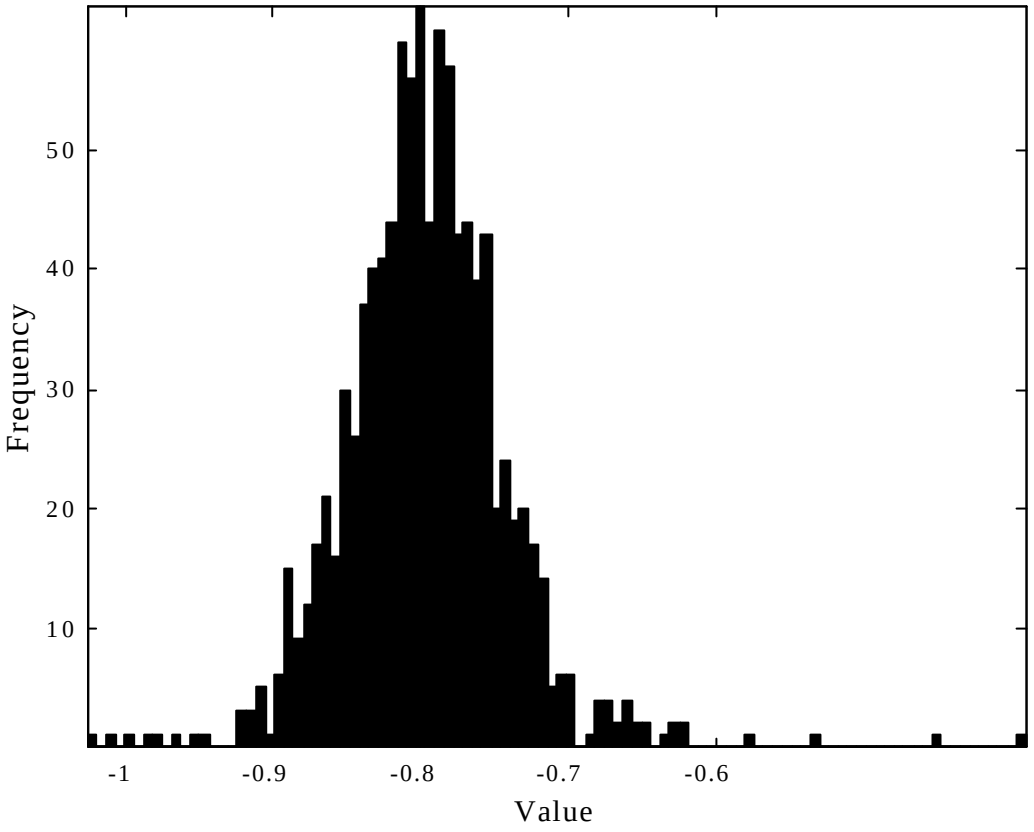

Figure 9. Plots of simulated frequencies for ρ in the SARP simulation using a random-walk Metropolis step.

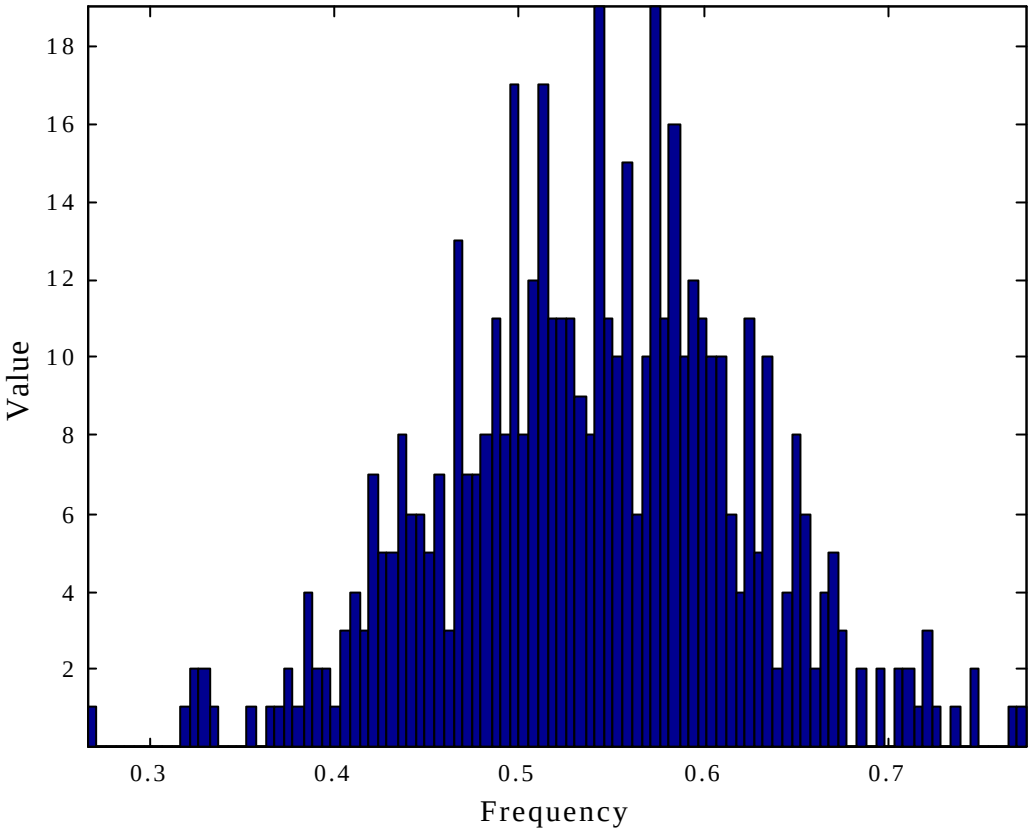

Figure 10. Empirical distribution for ρ (the 'neighbrhood' effect) in Bangladeshi data.

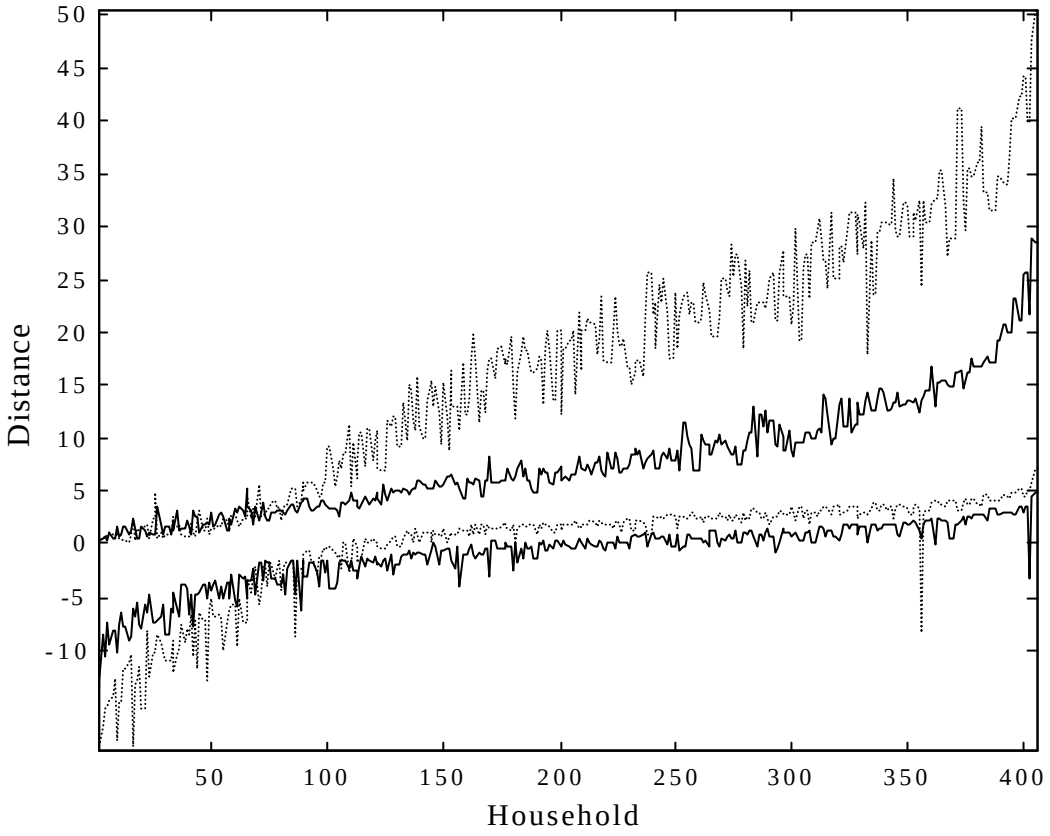

Figure 11. Ninety-five highest-posterior density regions for estimates of the farm size at which adoption status changes using resultd from the probit model (dashed lines) and the spatial probit model (solid lines).