# The Spatial Probit Model of Interdependent Binary Outcomes: Estimation, Interpretation, and Presentation

Robert J. Franzese, Jr. (franzese@umich.edu)
Associate Professor of Political Science, University of Michigan, Ann Arbor

Jude C. Hays (jchays@uiuc.edu)
Assistant Professor of Political Science, University of Illinois, Urbana-Champaign

20 July 2007

[Prepared for the 24th Annual Summer Meeting of the Society for Political Methodology (@ Penn State)]

**ABSTRACT:** We have argued and shown elsewhere the ubiquity and prominence of spatial interdependence, i.e., interdependence of outcomes among cross-sectional units, across the theories and substance of political and social science, and we have noted that much previous practice neglected this interdependence or treated it solely as nuisance, to the serious detriment of sound inference. These earlier studies considered only linear-regression models of spatial/spatio-temporal interdependence. For those classes of models, we (1) derived analytically in simple cases the biases of non-spatial and spatial least-squares (LS) under interdependence, (2) explored in simulations under richer, more realistic circumstances the properties of the biased non-spatial and spatial least-squares estimators and of the consistent and asymptotically efficient spatial method-of-moments (i.e., IV, 2SLS, GMM) and spatial maximum-likelihood estimators (ML), and (3) showed how to calculate, interpret, and present effectively the estimated spatial/spatio-temporal effects and dynamics of such models, along with appropriate standard errors, confidence intervals, hypothesis tests, etc. This paper begins a like set of tasks for binary-outcome models. We start again by stressing the ubiquity and centrality substantively and theoretically of interdependence in binary outcomes of interest to political and social scientists. We note that, again, this interdependence has typically been ignored in most contexts where it likely arises and that, in the few contexts where it has been acknowledged, or even rather centrally emphasized, those of policy diffusion and of social networks, the endogeneity of the spatial lag used (appropriately) to model the interdependence has only rarely been recognized. Next, we note and explain some of the severe challenges for empirical analysis posed by spatial interdependence in binary-outcome models, and then we follow recent advances in the spatial-econometric literature to suggest Bayesian or recursive-importance-sampling (RIS) approaches for tackling the estimation demands of these models. In brief and in general, the estimation complications arise because among the RHS variables is an endogenous weighted spatial-lag of the unobserved latent outcome, $y^*$, in the other units; Bayesian or RIS techniques facilitate the complicated nested optimization exercise that follows from that fact. We show how to calculate estimated spatial effects (as opposed to parameter estimates) in such models and how to construct confidence regions for those, adopting simulation strategies for these purposes, and then how to present such estimates effectively.

## I. Introduction to Spatial Probit

Many phenomena that social scientists study are inherently or by measurement discrete choices. Canonical political-science examples include citizens' vote-choices and turnout, legislators' votes, governments' policy-enactments, wars among or within nations, and regime type or transition. In all these cases and most others, substantively and theoretically, the choices/outcomes of/in some units depend on those of others. Whether and for whom citizens vote depends on whether and how their neighbors or social networks vote, legislators' votes depend on how they expect or observe others to vote, governments' policies depend on others' policies *via* competition or learning, nations' internal wars arise in some part through contagion from others' conflicts, whether and which other nations join conflicts heavily condition states' involvement in external wars, and regime change at home is often spurred by example, fomentation, or otherwise from abroad.

Indeed, interdependence seems almost inherent to *social*-science discrete-*choices*. Nevertheless, outside of a few topical areas, interdependence in discrete outcomes receives very little theoretical or empirical attention. Perhaps the most-extensive and longest-standing exception in political science surrounds the diffusion of policies or institutions across national or sub-national governments. The study of policy diffusion across U.S. States in particular has deep roots and much contemporary interest (e.g., Crain 1966, Walker 1969, 1973, Gray 1973, Knoke 1982, Caldiera 1985, Lutz 1987, Berry & Berry 1990, 1999, Case et al. 1993, Berry 1994, Rogers 1995, Mintrom 1997ab, Mintrom & Vergari 1998, Mossberger 1999, Godwin & Schroedel 2000, Balla 2001, Mooney 2001, Bailey & Rom 2004, Boehmke & Witmer 2004, Daley & Garand 2004, Grossback et al. 2004, Shipan & Volden 2006, Volden 2006). Similar innovation-learning mechanisms underlie some comparative studies of policy diffusion (Schneider & Ingram 1988, Rose 1993, Meseguer 2004, 2005, Gilardi 2005). Interest in institutional or even regime diffusion, too, is long-standing and recently much invigorated in comparative and international politics. Dahl's (1971) classic *Polyarchy*, for instance, (implicitly) references international diffusion among his list of democracy's eight causes; Starr's "Democratic Dominoes" (1991) and Huntington's *Third Wave* (1991) accord it a central role; and O'Loughlin et al. (1998) and Gleditsch & Ward (2006, 2007) have recently estimated its empirical extent. Eising (2002), Brune et al. (2004), Simmons & Elkins (2004), Brooks (2005), Elkins et al. (2006), Simmons et al. (2006), and others likewise stress international diffusion in recent economic liberalizations. The other major area of extensive interest in interdependence is micro-behavioral, where some of the long-standing and recently surging interest in contextual effects surrounds the effects on a respondent's behaviors or opinions of aggregates of others' behaviors and opinions—e.g., those of her region, community, or social network. Within this large literature on *contextual effects* in political behavior (Huckfeldt & Sprague 1993 review), recent contributions that stress interdependence include Braybeck & Huckfeldt (2002ab), Cho (2003), Huckfeldt et al. (2005), Lin et al (2006), Cho & Gimpel (2007), and Cho & Rudolph (2007).

The substantive range of important spatial-interdependence effects on discrete outcomes extends well beyond inter-governmental/interstate diffusion and social-network effects, however, spanning the subfields and substance of political science. Inside democratic legislatures, representatives' votes depend on others' (expected) votes; in electoral studies, candidate qualities or strategies and citizens votes and election outcomes in some contests depend on (expectations of) those in others; outside legislative and electoral arenas, the probabilities and outcomes of coups, revolutions, and/or riots in one unit depend in substantively crucial ways on (expectations of) those in others. In international relations, the interdependence of states' actions essentially defines the subfield. Whether states enter wars, alliances, treaties, or international organizations, e.g., depends greatly on how many and who else (are expected to) enter. Interdependence is substantively crucial in comparative and international political economy too; for example, globalization, arguably today's most-notable (and indisputably the most-noted) political-economic phenomenon, refers directly to the interdependence of domestic politics, policies, and policymakers. International economic integration is widely considered a root cause of the recent cross-national spread of economic liberalization and the so-called *Washington Consensus*, and many commentators even see international waves of partisan governments and votes as a result of interdependence in mass opinion and vote choices (but *cf.* Kayser 2007).

The ubiquity and substantive/theoretical centrality of interdependence across political-science discrete-choice contexts notwithstanding, studies that accord interdependence and diffusion explicit attention are uncommon. The rare exceptions include the references given above on cross-national and interstate policy-diffusion; Ward, Gleditsch, and colleagues (Shin & Ward 1999, Gleditsch & Ward 2000, Gleditsch 2002, Ward & Gleditsch 2002, Hoff & Ward 2004, Gartzke & Gleditsch 2006, Salehyan & Gleditsch 2006, Gleditsch 2007) and Signorino and coauthors (Signorino 1999, 2002, 2003, Signorino & Yilmaz 2003, Signorino & Tarar 2006) on interdependence in international relations; Li & Thompson (1975), Govea & West (1981), and Brinks & Coppedge (2006) on coup, riot, and revolution contagion, respectively; Schofield et al. (2003) on citizens' votes; Lacombe & Shaughnessy (2005) on legislators' votes; and Mukherjee & Singer (2007) on inflation targeting.

Likewise, despite the common centrality of interdependence in social-science theories of discrete choices, assumptions of ***in***dependence pervade almost all empirical analyses of them, even in those research areas that give interdependence greater substantive and theoretical weight. E.g., empirical models of war in which the dependence of one state's choice on those of others enters explicitly are rare; the Ward, Gleditsch, and colleagues and the Signorino and coauthor citations above are among the few exceptions. Even in the policy- or institutional/regime-diffusion literatures, where inclusion among the explanators of the average or sum (possibly weighted) of other units' outcomes (e.g., the number of states adopting a policy or treaty in policy- or treaty-adoption studies) explicitly accounts interdependence in empirical models, the endogeneity of this *spatial lag* is rarely confronted.

Working under the incorrect assumption of independence, of course, threatens over-confidence or inefficiency in the best of circumstances, and usually bias and inconsistency as well. Inclusion of a spatial lag to reflect interdependence when it is present would seem advisable, but such lags covary with residuals, i.e., are endogenous, and so introduce simultaneity biases.[1] For the linear-regression case, we have argued and shown elsewhere (2004, 2006, 2007abc) that serious omitted-variable biases arise when spatial lags are excluded in the presence of interdependence. We also showed that redressing this by incorporating interdependence explicitly is generally of first-order importance relative to the problems induced by the spatial-lag endogeneity, although these latter simultaneity biases do become appreciable as interdependence strengthens. Accordingly, we covered in these previous works some methodologies for gauging that strength, for redressing the simultaneity issues of spatial lags, and for calculating and presenting estimates of spatially/spatio-temporally dynamic effects and their certainty, but (almost) all in the linear-regression context (2007c briefly introduced and illustrated the spatial-probit model). This paper begins a similar set of efforts for binary-outcome models, where, as elaborated below, the substantive and theoretical importance of interdependence, the empirical problems created by its omission, and the methodological challenges raised by the endogeneity of its appropriately explicit inclusion by spatial lags are all likely to be even greater.

## II. The Econometric Problem

Methods for properly estimating and analyzing models of interdependent categorical or limited dependent-variables have received significant attention in the spatial-econometric literature recently. Most of this research considers the probit model with spatial dependence in the latent-variable, i.e., in the unobserved argument to the probit-modeled probability of a one on the binary outcome.[2]

---

[1] The inclusion of weighted sums of other units' outcomes also introduces measurement error insofar as interdependence truly arises through *expectations* of other units' outcomes. Substantively, alternative interdependence mechanisms may suggest diffusion either of outcomes or expected-outcomes, but only the latter mechanism can be identified logically.

[2] Spatial logit has also been suggested (e.g., Dubin 1997; Lin 2003; Autant-Bernard 2006), but spatial probit dominates the methodological and applied literatures, likely due to the relatively greater feasibility of working with $n$-dimensional

Models of spatial sample-selection (i.e., spatial Tobit or Heckit: McMillen 1995, Smith & LeSage 2004, Flores-Lagunes & Schnier 2006), spatial multinomial-probit (McMillen 1995, Bolduc et al. 1997), and spatial discrete-duration (Phaneuf & Palmquist 2003), all three of which closely resemble the spatial probit, and models of survival with spatial frailty (Banerjee et al. 2004, Darmofal 2007) and of spatial count (Bhati 2005), including a zero-inflated-count model (Rathbun & Fei 2006) have also been suggested. Spatial probit is by far the most common of these spatial qualitative-dependent-variable models in applied research (e.g., Holloway et al. 2002, Beron et al. 2003, Coughlin et al. 2003, Murdoch et al. 2003, Novo 2003, Schofield et al. 2003, Garrett et al. 2005, Lacombe & Shaughnessy 2005,Autant-Bernard 2006, Rathbun & Fei 2006, Mukherjee & Singer 2007).

Several estimation strategies have been suggested for the spatial-probit model. McMillen (1992) first suggested an EM algorithm, which innovation rendered the spatial-probit's non-additively-separable log-likelihood (see below) estimable for the first time, but the strategy also did not provide standard-errors for the crucial spatial-dependence parameter and required arbitrary parameterization of the heteroscedasticity induced by that dependence (see below). McMillen (1995) and Bolduc et. al. (1997) applied simulated-likelihood strategies to estimate their spatial-multinomial-probit models, and Beron et al. (2003) and Beron & Vijverberg (2004) advanced a recursive-importance-sampling (RIS) estimator in that line. LeSage (1999, 2000) introduced a Bayesian strategy of Markov-Chain-Monte-Carlo (MCMC) by Gibbs and Metropolis-Hastings sampling. Fleming (2004) reviews these two families and simpler, if approximate, strategies of estimating linear or nonlinear probability models[3] by nonlinear least-squares (NLS), generalized linear-models (GLM), or generalized linear-mixed-models (GLMM). Pinkse & Slade's (1998) two-step GMM estimator has also seen some use in the applied literature, but the RIS and Bayesian strategies have dominated applications. In the rest

---

normal (as opposed to extreme-value) distributions as necessary to incorporate the interdependence directly.

[3] Even the linear-probability model becomes nonlinear in parameters given the spatial multiplier, $(\mathbf{I}-\rho\mathbf{W})^{-1}$.

of section, we consider the spatial-probit model and RIS and Bayesian strategies for estimating it.

The structural model for the spatial probit takes the form:

$$\mathbf{y}^* = \rho \mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (1),$$

which can be written in reduced form as:

$$\mathbf{y}^* = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad (2),$$

where $\mathbf{u} = (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$, and $\mathbf{y}^*$ is a latent variable that links to the observed binary-outcome, $\mathbf{y}$, through the measurement equation:

$$y_i = \begin{cases} 1 \text{ if } y_i^* > 0 \\ 0 \text{ if } y_i^* \leq 0 \end{cases}. \qquad (3).$$

The marginal probabilities are calculated as follows:

$$p(y_i = 1 \mid \mathbf{X}) = p\left(\left[(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]_i + \left[(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}\right]_i > 0\right)$$

where subscripts *i* indicate the $i^{\text{th}}$ row of the subscripted vector, or, more conveniently:

$$p(y_i = 1 \mid \mathbf{x}_i) = p\left(u_i < \left[(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]_i \big/ \sigma_i\right) \qquad (4).$$

As in the standard probit, the right-hand-side probability that the systematic component of the latent variable, $[(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_i$, exceeds the stochastic component, $u_i$, is read from a normal distribution. However, the interdependence of the $y_i^*$ in spatial probit induces a non-sphericity of the stochastic components, leaving $\mathbf{u}$ distributed *n*-dimensional *multivariate* normal, with mean-vector $\mathbf{0}$ and variance-covariance matrix $[(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})]^{-1}$. The probability that $[(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_i / \sigma_i$ exceeds $u_i$ is read from the $i^{\text{th}}$ marginal distribution of this multivariate normal, which requires integrating the joint distribution over the other $n-1$ dimensions. Also, the denominator, $\sigma_i$, is the square root of the $ii^{\text{th}}$ element of that variance-covariance, and does not equal one as in the standard-probit. I.e., spatial interdependence induces heteroscedasticity. This heteroscedasticity and, more fundamentally, the interdependence (i.e., the ***non***-independence) of the $u_i$, render standard probit inappropriate for the spatial model. One does not maximize the sum of the log of *n* one-dimensional probabilities as because each unit's outcome is dependent and so their joint distribution is not the product of the marginals. Rather, one must calculate the log of one *n*-dimensional normal probability.

The spatial-error version of the probit model is easier to express, taking the form:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad (5),$$

with $\mathbf{u} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon}$, and having the marginal probabilities:

$$p(y_i = 1 \mid \mathbf{x}_i) = p\left(u_i < \mathbf{x}_i\boldsymbol{\beta}/\sigma_i\right) \qquad (6),$$

but this too must be read from the $i^{\text{th}}$ marginal distribution of a multivariate normal with means zero and variance-covariance matrix $[(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})]^{-1}$. Thus, spatial-error probit models entail the same estimation and interpretation complications as do spatial-lag models. Mixed spatial-lag/spatial-error models are also possible, although the literature has not paid them much attention.

Some special circumstances might allow standard-probit estimation of spatial-lag models, but these are unlikely. For instance, Anselin (2006) notes that, in the conditional counterpart of (1):

$$y_i^* = \rho \sum\nolimits_j w_{ij} E(y_j^* \mid \mathbf{X}) + \mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (7),$$

$E(y_j^* \mid \mathbf{X})$ can be estimated by $\sum_j w_{ij} y_j$, the spatially weighted average of actual outcomes in units *j*.[4] This spatial lag could be included as a regressor under some stringent conditions, but these other units' observations *j* cannot be jointly determined with those of *i*. Thus, Anselin states: "While the standard probit model remains valid, coding methods must be employed to ensure that the sample does not contain these neighbors." Phrased differently, this means that, for any units *j* such that diffusion from those *j* to any *i* in the sample is non-negligible (at any order spatial-lag), those *j* must

[4] Note that the row-normalization here means that the resulting seeming replication of the summed weights is irrelevant.

be excluded from the sample but used exclusively in constructing the $\mathbf{Wy}$ spatial lag for the retained observations *i*. Alternatively, all *j*'s neighboring *i* according to $\mathbf{W}$ must be exogenous to all *i* in the sample; i.e., feedback must be directional and orderable from *j*'s to *i*'s only, severing feedback from *i* back to itself. Relatedly, in some substantive/theoretical contexts, researchers might wish to model interdependence as propagating through the actual outcome rather than through the latent variable. Unfortunately, this is in general impossible because, indirectly via feedback, $y_i$ would generate $y_i^*$ but also, directly, $y_i$ is generated by $\Phi(y_i^*)$.[5] These cannot be made consistent unless the indirect generation of $y_i$ by $y_i^*$ is severed by the circumstances just described. In practice, these limitations are typically prohibitive, although contexts where such directional ordering exists and such omissions may be comfortably sustained are imaginable. For instance, Swank (2006, 2007) argues that, in tax competition, U.S. tax policies exclusively lead others' policies, and he excludes all U.S. data in his empirics. If arguments like these are true, then such exclusions will allow standard-probit estimation in spatial binary-outcome contexts.

We focus on the (unconditional, simultaneous) spatial-lag model in most of the rest of the paper, although we do present some results pertaining to spatial-error models. We ignore the conditional model as it will usually not be applicable or advisable. We also do not discuss specification tests (spatial-*lag* vs. spatial-*error* vs. non-spatial) here, although they are surely essential, especially given the complexity and computational intensity of appropriate estimation strategies for spatial probit. On these, we refer the interested reader to Pinkse & Slade (1998), Pinkse (1999), Kelejian & Prucha (2001), and, for a recent review, Anselin (2006). In addition to standard probit estimation with the endogenous spatial lag $\mathbf{Wy}$ included as a regressor, which is current standard-practice in empirical work in political science insofar as interdependence of binary outcomes is at all addressed, we

[5] Notice that the same limitation does not quite obtain for temporal dependence since time is unidirectional, so the indirect feedback from $y_t$ to $y_{t-1}$ does not occur (given full and proper modeling of the temporal dynamics).

consider estimation by RIS methods and by Bayesian MCMC methods.

## III. The RIS and Bayesian Estimators for Spatial Probit

LeSage (1999, 2000) suggests using Bayesian Markov Chain Monte Carlo (MCMC) methods to surmount the estimation complications introduced by the *n*-dimensional normal in the spatial-probit likelihood (posterior). The basic idea of Monte Carlo (simulation) methods is straightforward:[6] if one can characterize the joint distribution (likelihood, posterior) of the quantities of interest (parameters), then, given modern computing capabilities, one can simply sample (take random draws) from that distribution and calculate the desired statistics in those samples. With sufficient draws, the sample statistics can approximate the population parameters[7] they aim to estimate arbitrarily closely. In basic Monte-Carlo simulation, the draws are independent and the target distribution is directly specified. In Markov-Chain Monte-Carlo, each draw is dependent in some amount on the previous one in a manner which allows generation of samples with properties mirroring those of the model parameters from just the conditional distributions of the parameters, which is useful where the joint distribution is not expressible directly or, as with spatial probit, is sufficiently complex as to make direct sampling from the joint distribution prohibitively difficult and/or time-consuming.

Gibbs sampling is the simplest and most-common of the MCMC family, and its proceedings are also simple to describe. Given distributions for each parameter conditional on the other parameters, one can cycle through draws from those conditional distributions, eventually reaching a *convergent* state past which all subsequent draws are from the targeted posterior distribution. To elaborate, first: characterize the distribution for each parameter conditional on all the others, then choose (arbitrary) starting values for those parameters and draw a new value for the first parameter conditional on the

[6] Our simple introduction draws heavily from Gill's (2002) wonderful text on Bayesian methods.
[7] Recall that the "population parameters" that can be arbitrarily closely approximated will usually be some *estimates* in an application, like the spatial-probit parameter-estimates, not the "true parameters" of course (which latter concept is somewhat awkward in Bayesian terminology anyway).

others' starting values. Then, conditional on this new draw of the first parameter and starting values for the rest, draw a new value for the second parameter from its conditional distribution. Continue thusly until all parameters have their first set of drawn values, then return to the first parameter and draw its second simulated value conditional on the others' first draws. Cycle thusly for some large number of iterations, and, under rather general conditions, the limiting (asymptotic) distribution of this set parameter draws is the desired joint posterior-distribution. Therefore, after having gathered some very large set of parameter-vector values from this process, discard some large initial set of draws (the *burn-in*) and base inferences on sample statistics from the remaining. A typical burn-in might be 1000 draws and inferences might be based on the next 5000 or 10,000.

The drawbacks of MCMC approaches may be obvious from what we have and have not said. First, no universal tests exist to assure that convergence has occurred, so one's burn-in may appear sufficient in that the next 5000 drawn parameter-vectors seem to follow some circumscribed bounds and behavior of the unknown target distribution (i.e., to have *settled down*) only to have the 5001$^{st}$ vector leap into a new range. Second, adjacent draws, despite their Markov-Chain origins, are asymptotically serially uncorrelated but this *asymptopia* may not arrive within practical limits. The starting values likewise also do not matter asymptotically if the given set of conditional distributions properly could come from a valid joint distribution, but given the first pair of caveats, starting values may matter short of convergence the arrival at which cannot be verified. Third, the conditional distributions must be expressible and sufficiently standard to make these many draws a practicality.

These limitations of MCMC should concern careful researchers, and many diagnostics and tests for (non)convergence, remaining serial correlation, or starting-value sensitivity, and some strategies for ameliorating them, exist (all imperfect but useful still); and these concerns do not overturn the remarkably flexible utility of the Gibbs sampler, neither in general nor specifically for spatial-probit estimation. The conditional distributions for the spatial-probit-model parameters (given below) are all standard except one, and therefore the Gibbs sampler is useful. The crucial spatial-lag-coefficient parameter, $\rho$, has a non-standard conditional-distribution however. The core difference of Gibbs from Metropolis-Hastings sampling is the latter's *seeding* or *jump* distribution from which values are drawn and accepted as new values in the sampling or rejected from doing so, depending on how they compare to a suitably transformed expression of the target distribution.[8] LeSage's Bayesian spatial-probit estimator uses Metropolis-Hastings within Gibbs to simulate the non-standard conditional distribution of $\rho$. Obviously, this step adds greatly to the estimator's computational intensity.

With this brief introduction to Bayesian MCMC estimation by Gibbs and Metropolis-Hastings sampling, we now introduce their application to the spatial-probit model. We follow LeSage (2000) to state the likelihood in terms of the latent (unobserved) outcome, $\mathbf{y}^*$—an additional conditional distribution to be added later will apply (3) to convert unobserved $\mathbf{y}^*$ to observed $\mathbf{y}$[9]—for the spatial-lag model (1) as:

$$L\left(\mathbf{y}^*, \mathbf{W} \mid \rho, \boldsymbol{\beta}, \sigma^2\right) = \frac{1}{2\pi\sigma^{2(n/2)}} \left|\mathbf{I}_n - \rho\mathbf{W}\right| e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})} \quad (8),$$

where $\boldsymbol{\varepsilon} = \left(\mathbf{I}_n - \rho\mathbf{W}\right)\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}$. The spatial-error probit model,

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \;\; ; \;\; \mathbf{u} = \rho\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \quad (9),$$

may be expressed with same likelihood but with $\boldsymbol{\varepsilon} = \left(\mathbf{I}_n - \rho\mathbf{W}\right)\left(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}\right)$. Diffuse priors may now be specified that will produce the following joint posterior-density:

$$p\left(\rho, \boldsymbol{\beta}, \sigma \mid \mathbf{y}^*, \mathbf{W}\right) \propto \left|\mathbf{I}_n - \rho\mathbf{W}\right| \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})} \quad (10).$$

[8] To elaborate: to sample from some non-standard density *f( )*, let $x_0$ be the current draw from *f( )*, beginning with an arbitrary starting value. Consider a candidate next value, $x_1$, for $x$ given by $x_1 = x_0 + cZ$ with $Z$ being drawn from a standard-normal distribution and $c$ a given constant. Then, we assign a probability of accepting this candidate as the next value of $x$ in our MCMC as $p = \min\{1, f(x)/f(x_0)\}$. I.e., we draw from a Uniform(0,1) distribution, and, if $U<p$, the candidate $x_1$ becomes the next $x$; if $U>p$ the next $x$ remains $x_0$. Metropolis-Hastings is thus one type of *rejection sampling*.

[9] This stratagem enables LeSage to express the spatial-Tobit model with this same likelihood, adding later a conditional distribution to generate latent variables $z$ for censored observations instead of one to generate y=(0,1) for the probit.

One can now derive the conditional posterior densities for $\rho, \boldsymbol{\beta},$ and $\sigma$ for the sampler:

$$p\left(\sigma \mid \rho, \boldsymbol{\beta}\right) \propto \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})} \qquad (11).$$

Notice that conditioning on $\rho$ allows $\left|\mathbf{I}_n - \rho\mathbf{W}\right|$ to be subsumed into the constant of proportionality and that (11) implies $\sigma^2 \quad \chi_n^2$, a standard distribution facilitating the Gibbs sampler. Next,

$$p\left(\boldsymbol{\beta} \mid \rho, \sigma\right) \quad N\left[\boldsymbol{\beta}, \sigma_\varepsilon^2 (\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{X})^{-1}\right] \qquad (12),$$

where, in spatial-lag models, $\mathbf{C} = \mathbf{I}_n$ and $\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_n - \rho\mathbf{W}\right)\mathbf{y}^*$, whereas $\mathbf{C} = \left(\mathbf{I}_n - \rho\mathbf{W}\right)$ and $\boldsymbol{\beta} = (\mathbf{X}'(\mathbf{I}_n - \rho\mathbf{W})'(\mathbf{I}_n - \rho\mathbf{W})\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_n - \rho\mathbf{W})'(\mathbf{I}_n - \rho\mathbf{W})\mathbf{y}^*$ in spatial-error models. The conditional multivariate-normality of $\boldsymbol{\beta}$ will again allow the Gibbs sampler. The conditional distribution of $\rho$, however, is non-standard, obligating Metropolis-Hastings sampling as already noted:

$$p\left(\rho \mid \boldsymbol{\beta}, \sigma\right) \propto \left|\mathbf{I}_n - \rho\mathbf{W}\right| \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})} \qquad (13),$$

with $\boldsymbol{\varepsilon}$ defined differently as above for the spatial-error and the spatial-lag models.[10] Finally, we add the conditional distribution, namely a truncated normal, that translates $\mathbf{y}^*$ to $\mathbf{y}$:

$$f\left(z_i \mid \rho, \boldsymbol{\beta}, \sigma\right) \quad N(y_i, \sigma_i^2), \text{ left- or right-truncated at 0 as } y_i = 1 \text{ or } 0 \qquad (14),$$

where $y_i$ is the predicted value of $y_i^*$ (the $i^{\text{th}}$ row of $\left(\mathbf{I}_n - \rho\mathbf{W}\right)^{-1}\mathbf{X}\boldsymbol{\beta}$ for spatial-lag and of $\mathbf{X}\boldsymbol{\beta}$ for spatial-error models), and the variance of $y_i$ is $\sum_i \omega_{ij}^2$ with $\omega_{ij}$ the $i^{\text{th}}$ row of $\left(\mathbf{I}_n - \rho\mathbf{W}\right)^{-1}\boldsymbol{\varepsilon}$.[11]

With the complete set of conditional distributions in hand, we can now implement the MCMC to

---

[10] Anselin (1988) shows that the minimum and maximum eigenvalues of a standardized spatial-weight matrix, $\mathbf{W}$, bound $\rho$ to $1/\lambda_{\min} < \rho < 1/\lambda_{\max}$. One could add this constraint to the rejection sampling, but our preliminary simulations seem to suggest that the model-estimates have better properties if one instead applies the wider bounds of (-1,1).

[11] Spatial Tobit replaces (14) with: $f\left(z_i \mid \rho, \boldsymbol{\beta}, \sigma\right) \quad \begin{cases} [1 - \Phi(y_i^*/\sigma_i)]^{-1} \exp[-(z_i - y_i^*)^2 / 2\sigma_i], & \text{if } z_i \le 0 \\ 0, & \text{if } z_i \le 0 \end{cases}$. The Tobit allows $\sigma_i^2 = \sigma_{\varepsilon_i}^2 \sum_i \omega_{ij}^2$, but the probit must scale $\sigma_{\varepsilon_i}^2$ to 1, it and $\boldsymbol{\beta}$ not being separately identified for binary outcome models.

estimate the model,[12] proceeding thus:[13]

1. Use expression (11) to draw $\sigma_1$ using starting values $\rho_0, \boldsymbol{\beta}_0, \sigma_0$.
2. Use $\sigma_1$, $\rho_0$, and expression (12) to draw $\boldsymbol{\beta}_1$.
3. Use $\sigma_1$, $\boldsymbol{\beta}_1$, and expression (13) to draw $\rho_1$ by Metropolis-Hastings sampling.
4. Sample the outcome, $z_i$, using the censoring distribution (14) and $\sigma_1$, $\boldsymbol{\beta}_1$, and $\rho_1$.
5. Return to step 1 incrementing the subscript counters by one.

After a sufficient burn-in—our simulation and application experiences so far suggest at least 1000 is advisable—the distributions of $\sigma_1$, $\boldsymbol{\beta}_1$, and $\rho_1$ will have reached convergence and subsequent draws on the parameters may be used to give their estimates (as means or medians of some large number of draws) and estimates of their certainty (as standard deviations or percentile ranges).

A frequentist approach has also been suggested, Recursive Importance-Sampling (RIS), which also uses simulation to approximate probabilities difficult to calculate analytically, and can also be used to estimate spatial-lag or spatial-error probit. To introduce RIS, following Vijverberg's (1997) notation, to approximate a probability from an *n*-dimensional multivariate-normal distribution,

$$p = \int_{-\infty}^{\mathbf{x}_0} f_n(\mathbf{x})\, d\mathbf{x}\,, \tag{15}$$

where $f_n(\mathbf{x})$ and $[-\infty, \mathbf{x}_0]$ are the density and interval over which we want to integrate respectively, we first choose a *n*-dimensional sampling distribution with well-known properties. The *importance*

[12] LeSage (2000) first assigns diffuse priors to the parameters. He also relaxes the assumption of homoskedasticity in the latent-variable stochastic-term, allowing it to vary arbitrarily by observation. This will allow exploration of variation in model fit and identification of and robustness to potential outliers. Relaxing homoskedasticity so fully as to allow each observation its own relative variance creates as many parameters to estimate as observations, an issue LeSage evades by specifying an informative prior for those relative-variance parameters, specifically one suggested by Geweke (1993) that, *inter alia*, has the useful property of yielding a distribution of $\varepsilon$ consistent with a probit choice-model as the Gewekian-distribution parameter, *q*, goes to infinity, and that at $q \approx 7.5$ yields a choice-model approximating logit. The posterior-estimates of *q*, may be evaluated to test logit versus probit (versus un-named possibilities generated by $q<7$).

[13] Allowing arbitrary relative-variance as in note 12 requires the additional (informative) Gewekian prior described there and a (diffuse) hyper-prior on its parameter, *q*; produces more complicated expressions for the conditional distributions of $\sigma$, $\rho$, $\boldsymbol{\beta}$; and adds a conditional distribution (fortunately standard: $\chi^2_{q+1}$) for the relative variances, $v_i$. The steps below would now also include conditioning on starting values for, and then the previous draws of, $\mathbf{v}$, and a step inserted between 2 and 3 would draw the next $\mathbf{v}$ from $\chi^2_{q+1}$ conditional on the current $\sigma$, $\rho$, $\boldsymbol{\beta}$. Notice that a hyper-prior on *q* set determinately to a large number (or 7.5), spatial probit (or logit) without heteroscedasticity/outlier-robustness results.

distribution is a truncated version of this sampling distribution with support over $[-\infty, \mathbf{x}_0]$. Define $g_n^c(\mathbf{x})$ as the density for the *n*-dimensional importance distribution, and then multiply and divide the right-hand-side of the integral by this density:

$$p = \int_{-\infty}^{\mathbf{x}_0} \frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})} \, g_n^c(\mathbf{x}) \, d\mathbf{x} \tag{16}$$

The solution to this integral is by definition a mean, namely of the ratio of densities, which we can estimate with a sample of *R* draws from the importance distribution. Note that each of the *R* draws gives a random vector that contains *n* elements. Formally, the estimate is

$$p = E\left[\frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})}\right] \approx \frac{1}{R}\sum_{r=1}^{R} \frac{f_n(\mathbf{x}_r)}{g_n^c(\mathbf{x}_r)} \equiv \hat{p} \tag{17}$$

Again, the standard probit-model assumes the errors independent, which makes the probabilities in the likelihood relatively easy to compute; in spatial probit, however, errors are interdependent, so we have to calculate a single probability from an *n*-dimensional normal distribution:

$$p(\mathbf{u} < \mathbf{v}) \tag{18}$$

where $\mathbf{u}$ is an $n \times 1$ vector of errors distributed $MVN(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = (\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})^{-1}$ and $\mathbf{v}$ also an $n \times 1$ vector defined as $\mathbf{v} = \mathbf{Q}(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X\beta}$, $\mathbf{Q}$ being a diagonal matrix with $q_i = 2y_i - 1$ along the diagonal. To implement the RIS estimator, we draw from the *n*-dimensional importance-distribution, which, for illustration, we assume a truncated multivariate (*n*-dimensional) normal.[14]

Since the variance-covariance matrix for $\mathbf{u}$ is positive definite, a Cholesky decomposition exists such that $\mathbf{\Sigma}^{-1} = \mathbf{A}'\mathbf{A}$, where $\mathbf{A}$ is an upper triangular matrix. The transformation $\boldsymbol{\eta} = \mathbf{Au}$ produces *n* independent standard normal variables. Substituting $\mathbf{u} = \mathbf{A}^{-1}\boldsymbol{\eta} \equiv \mathbf{B}\boldsymbol{\eta}$ into (18) gives:

[14] In this case, RIS is equivalent to the better-known GHK (Geweke-Hajivassiliou-Keane) simulation estimator.

$$p\left[\mathbf{B\eta} < \mathbf{v}\right] = p\left[\mathbf{\eta} < \mathbf{B}^{-1}\mathbf{v}\right],$$

where (19).

$$\mathbf{B}^{-1}\mathbf{v} = \begin{bmatrix} b_{1,1}^{-1} & b_{1,2}^{-1} & b_{1,3}^{-1} & & b_{1,n}^{-1} \\ 0 & & & & \\ 0 & & & & b_{n-1,n}^{-1} \\ & & & & b_{n-1,n}^{-1} \\ 0 & & 0 & 0 & b_{n,n}^{-1} \end{bmatrix} \times \begin{bmatrix} v_1 \\ \\ \\ \\ v_n \end{bmatrix}$$

Since the elements of the $n \times 1$ vector $\mathbf{\eta}$ are independent, the probability in (19) can be calculated by taking the product of independent draws from truncated-normal distributions with upper bounds that are determined recursively beginning with the last observation. More specifically, the sampling proceeds as follows: First, calculate the upper bound for the truncated-normal distribution of the $n^{th}$ observation in the dataset. Take a draw from this distribution and use it to calculate the upper bound for the truncated-normal distribution for the $n$-1 observation. The first two draws are used to calculate the upper bound for the n-2 observation and so on until all $n$ upper bounds are determined. Formally, the recursive process for calculating the upper bounds is

$$\begin{aligned} \eta_n &< b_{n,n}^{-1} v_n \equiv \upsilon_n \\ \eta_{n-1} &< b_{n-1,n-1}^{-1}\left[v_{n-1} - b_{n-1,n}\eta_n\right] \equiv \upsilon_{n-1} \\ \eta_{n-2} &< b_{n-2,n-2}^{-1}\left[v_{n-2} - b_{n-2,n-1}\eta_{n-1} - b_{n-2,n}\eta_n\right] \equiv \upsilon_{n-2} \end{aligned} \quad (20),$$

which can be written more compactly as:

$$\eta_j < b_{j,j}^{-1}\left[v_j - \sum_{i=j+1}^{n} b_{j,i}\eta_i\right] \equiv \upsilon_j \quad (21).$$

Denote the bounds calculated with draws from the importance distribution as $\bar{\boldsymbol{\upsilon}}$. The probability of observing a given sample of ones and zeros is determined by evaluating the cumulative distribution function at each of these bounds and then multiplying these probabilities. This process is repeated $R$

times, giving the RIS estimate for the joint probability (the likelihood) as the mean joint probability:

$$\hat{l} = (1/R)\sum_{r=1}^{R}\left[\prod_{j=1}^{n}\Phi\left(\bar{\upsilon}_{j,r}\right)\right] \qquad (22).$$

Standard optimization-routines can be used to maximize this function, and standard methods for calculating the variance-covariance matrix can be employed.

## IV. Monte Carlo Analyses of Standard-Probit vs. Bayesian-Spatial-Probit Estimation

We explore the small sample properties of the ML and Gibbs estimators for the spatial-lag probit model using a data-generation process that closely follows Beron and Vijverberg (2004):

$$\mathbf{y}^* = \left(\mathbf{I}_n - \rho\mathbf{W}\right)^{-1}\left(\mathbf{x}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \qquad (23),$$

where

$$\mathbf{x} = \left(\mathbf{I}_n - \theta\mathbf{W}\right)^{-1}\mathbf{z} \qquad (24),$$

which generates $\mathbf{X}$ with spatial interdependence of the same $\mathbf{W}$ as $\mathbf{y}^*$, but different coefficient, and

$$\mathbf{z}, \boldsymbol{\varepsilon} \sim N\left(0,1\right) \qquad (25).$$

To generate the observed $y$'s:

$$y_i = \begin{cases} 1 \text{ if } y_i^* \geq 0 \\ 0 \text{ if } y_i^* < 0 \end{cases} \qquad (26).$$

We employed a row-standardized version of the binary contiguity spatial weights matrix for the 48 contiguous American states, set the parameters $\rho$ and $\beta$ to 0.5 and 1.0 respectively, and used two sample sizes, 48 and 144, and two values of $\theta$, 0.0 and 0.5, giving four experiments in total.[15] In Table 1, we report the results for 1,000 trials using the ML estimator with two spatial lags, $\mathbf{Wy}$ and

[15] To create the weights matrix for the larger sample size we took the Kronecker product of the original 48x48 weights matrix with a 3x3 identity matrix.

$\mathbf{Wy}^*$, and the Bayesian Gibbs-sampler estimator. The ML with a $\mathbf{Wy}^*$ regressor cannot be used in practice, $\mathbf{y}^*$ being unobserved, but these results provide a nice comparison since these estimates are not affected by measurement error and attenuation bias. We do not evaluate the RIS estimator since its properties are examined thoroughly in Beron and Vijverberg (2004) (and because RIS requires significantly greater computer-time). For the Gibbs sampler (which takes considerable computer-time itself), we discard the first 1000 draws and retain the next 1000.

**Table 1: Simulation Results**

| | **ML with Wy** | | **ML with Wy*** | | **Bayesian Gibbs** | |
|---|---|---|---|---|---|---|
| | β | ρ | β | ρ | β | ρ |
| **<u>Experiment #1</u>:** n=48, θ=0.0 | | | | | | |
| Coefficient Estimate | 1.02 | 0.32 | 1.13 | 0.74 | 1.23 | 0.30 |
| SD Sampling Distribution | 0.33 | 0.69 | 0.41 | 0.36 | 0.28 | 0.16 |
| SE Estimate | 0.30 | 0.41 | 0.35 | 0.30 | 0.42 | 0.21 |
| **<u>Experiment #2</u>:** n=48, θ=0.5 | | | | | | |
| Coefficient Estimate | 1.22 | 0.35 | 1.13 | 0.69 | 1.21 | 0.28 |
| SD Sampling Distribution | 0.56 | 0.76 | 0.61 | 0.33 | 0.24 | 0.14 |
| SE Estimate | 0.36 | 0.46 | 0.42 | 0.29 | 0.39 | 0.20 |
| **<u>Experiment #3</u>:** n=144, θ=0.0 | | | | | | |
| Coefficient Estimate | 0.94 | 0.42 | 1.01 | 0.68 | 1.14 | 0.34 |
| SD Sampling Distribution | 0.17 | 0.27 | 0.19 | 0.16 | 0.15 | 0.10 |
| SE Estimate | 0.16 | 0.22 | 0.18 | 0.15 | 0.22 | 0.12 |
| **<u>Experiment #4</u>:** n=144, θ=0.5 | | | | | | |
| Coefficient Estimate | 1.08 | 0.48 | 0.97 | 0.64 | 1.13 | 0.32 |
| SD Sampling Distribution | 0.19 | 0.29 | 0.21 | 0.16 | 0.14 | 0.09 |
| SE Estimate | 0.18 | 0.23 | 0.20 | 0.15 | 0.21 | 0.12 |

Perhaps the most surprising result from our experiments is the relatively poor performance of the Bayesian estimator with Gibbs sampling in terms of bias, although its performance is notably better when measured in mean-squared-error terms. With respect to bias, the standard ML estimator using **Wy** as the spatial lag (the proxy spatial lag) actually weakly dominates the Bayesian estimator.[16] The standard ML estimator with the proxy spatial-lag clearly suffers two biases, but those fortuitously

[16] We recognize that our experiments evaluate the Bayesian estimator using frequentist standards and that some may find this objectionable on philosophical grounds.

somewhat offset in this example. The first is a simultaneity bias that also plagues the ML estimator with the true spatial-lag; this bias inflates the $\hat{\rho}$ estimate. The second is an attenuating bias in the $\hat{\rho}$ estimates due to measurement error in the proxy spatial-lag, $\mathbf{Wy}$ instead of $\mathbf{Wy}^*$. The simultaneity bias increases with $\rho$, but the attenuation bias is not a function of sample size in this example (given this particular **W**). When $\rho$ is small, the measurement-error attenuation-bias dominates, and the net bias is negative. When $\rho$ is large, the simultaneity inflation-bias dominates, and net bias positive, with the biases exactly canceling at some intermediate $\rho$ (near our fourth experiment, apparently).

In our smallest sample, the standard errors from the ML with proxy spatial-lag are overconfident with respect to the estimator's precision. The standard-error estimates for $\hat{\rho}$ are under-estimated in both cases by 40%. The downward bias in the standard-error estimate for $\hat{\beta}$ (56% of the true value) is only noticeable where the *x*'s are spatial interdependent. Finally, we note that, under the conditions like those of our second experiment, Beron and Vijverberg (2004, Tables 8.3 and 8.4) report that RIS overestimates $\beta$ by 10% and underestimates $\rho$ by 18%, which compares favorably to the same numbers for the standard ML estimator with proxy spatial-lag (22% and 30%).

## V. Calculating and Presenting Estimated Spatial Effects with Certainty Estimates

Properly estimating *parameters* such as *coefficients* and their certainties is obviously essential to valid inference, but our ultimate aims usually are to estimate, draw inferences regarding, interpret and present (ideally: causal) *effects* or *predictions*, i.e., changes in outcomes associated with (ideally: caused by) changes in explanatory factors or expectations of outcomes given levels of explanators. That is, typically, our direct aim is not to estimate *coefficients* like $\rho$ and $\boldsymbol{\beta}$ *per se*, but in service of estimating *effects* like $\frac{\partial y_i^*}{\partial x_i}$ or $\frac{\Delta y_i^*}{\Delta x_i}$—the marginal or discrete-change effects of some explanatory factor, $x$, in unit $i$ on the latent-variable outcome in $i$—or, even better, $\frac{\partial p(y_i=1)}{\partial x_i}$ or $\frac{\Delta p(y_i=1)}{\Delta x_i}$—the effects of $x_i$ on the probability of choice or outcome one in $i$.[17] Indeed, in spatial-analysis contexts, our interests likely extend centrally to the cross-unit effects, such as $\frac{\partial y_j^*}{\partial x_i}$, $\frac{\Delta y_j^*}{\Delta x_i}$, $\frac{\partial p(y_j=1)}{\partial x_i}$, $\frac{\Delta p(y_j=1)}{\Delta x_i}$, that are the content of diffusion, interdependence, or spatial interaction (the terms are roughly synonymous). As must always be stressed, only in purely linear and additively separable models, like the canonical linear-regression model, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \varepsilon$, are *coefficients* and *effects* (of unitary or marginal changes in $x$ on $y$) identical. Even in implicitly nonlinear-additive models, like spatial-autoregressive linear-regression models, *effects* involve spatial-feedback multipliers and so involve (nonlinear) combinations of coefficients. Thus, even if we were content to confine our interpretation and presentation to the latent-component argument to the probability of outcomes/choices, we could not read effects directly from the usual table of coefficients. Instead, calculation, interpretation, and presentation of estimated effects and their certainties follow the spatial linear-regression case that we have discussed extensively elsewhere (Franzese & Hays 2004, 2006, 2007abc). Namely:

$$\mathbf{y}^* = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$
$$= \begin{bmatrix} 1 & -\rho w_{1,2} & & & -\rho w_{1,n} \\ -\rho w_{2,1} & 1 & & & \\ & & & & \\ & & & 1 & -\rho w_{(n-1),n} \\ -\rho w_{n,1} & & & -\rho w_{n,(n-1)} & 1 \end{bmatrix}^{-1} \left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \equiv \mathbf{S}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \qquad (27),$$

so, denoting the $i^{\text{th}}$ column of $\mathbf{S}$ as $\mathbf{s}_i$ and their estimates as $\hat{\mathbf{S}}$ and $\hat{\mathbf{s}}_i$, the estimated effect of

[17] We may also wish to offer empirical estimates of models relating explanators to outcomes for use in interpretation and/or prediction of in- and out-of-sample realizations of such phenomena. That is, our aims might include empirical estimates of relatively compact theoretical models usefully portable for interpreting or predicting complex reality. Thus, we may also want to describe expected outcomes as some function of the levels of explanators, i.e., $E(y_i \mid \mathbf{x}_i, \mathbf{x}_j)$ as some "usefully portable" $f(\mathbf{x}_i, \mathbf{x}_j)$. The issues discussed in this section would apply for such interests also.

explanatory variable $k$ in unit $i$, $\Delta\left(x_k\right)_i$, on the outcomes in all units, $i$ and all $j$, is $\frac{\Delta\hat{\mathbf{S}}\mathbf{X}\hat{\boldsymbol{\beta}}}{\Delta x_{i,k}}$ which is simply, $\hat{\mathbf{s}}_i\hat{\beta}_k$. The standard-error calculation, using the delta method approximation, is

$$\hat{V}\left(\hat{\mathbf{s}}_i\hat{\beta}_k\right)\approx\left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]\hat{V}\left(\hat{\boldsymbol{\theta}}\right)\left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]', \text{ where } \hat{\boldsymbol{\theta}}=\begin{bmatrix}\hat{\rho}\\ \hat{\beta}_k\end{bmatrix} \text{ and } \left[\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]=\begin{bmatrix}\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\rho}} & \hat{\mathbf{s}}_i\end{bmatrix} \quad (28).$$

The vector $\frac{\partial\hat{\mathbf{s}}_i\hat{\beta}_k}{\partial\hat{\rho}}$ is the $i^{\text{th}}$ column of $\beta_k\frac{\partial\hat{\mathbf{S}}}{\partial\hat{\rho}}$. Since $\mathbf{S}$ is an inverse matrix, the derivative in equation (28) is $\frac{\partial\hat{\mathbf{S}}}{\partial\hat{\rho}}=-\hat{\mathbf{S}}\frac{\partial\hat{\mathbf{S}}^{-1}}{\partial\hat{\rho}}\hat{\mathbf{S}}=-\hat{\mathbf{S}}\frac{\partial\left(\mathbf{I}-\hat{\rho}\mathbf{W}\right)}{\partial\hat{\rho}}\hat{\mathbf{S}}=-\hat{\mathbf{S}}(-\mathbf{W})\hat{\mathbf{S}}=\hat{\mathbf{S}}\mathbf{W}\hat{\mathbf{S}}$. In Franzese & Hays (2004, 2006, 2007abc), we showed how to use these and related expressions to generate grids, tables, or maps of responses across units to various counterfactuals, along with appropriate indicators of the estimated certainties of these estimated spatial effects.[18] We also showed in the spatio-temporal context how to estimate and graph spatio-temporal response-paths and estimate and tabulate or array in grids long-run-steady-state spatio-temporal responses to counterfactuals, along with certainty estimates thereof.

All of these techniques could be used in the spatial-probit context if we confine our attention to the latent variable, but for most purposes such confinement would be quite unsatisfactory. Also, even if remaining at that level, several issues regarding the application of delta-method asymptotic linear-approximation may trouble us, the intrinsic appeal of analytic solutions notwithstanding. The adjectives appended to *delta-method* just now give the first set of concerns: *approximate*, *asymptotic*, and *linear*. Being based upon a linearization, the certainty estimates only approximate validly in some proximity of the estimated *nonlinear* expression, and we do not know in general how small a range. Being asymptotic, they only approximate validly for large samples, and we do not know in

[18] We did not show how to indicate certainties of estimates reflected in spatial-feedback maps, and in general presenting certainty information in maps is difficult, but one could, for example, superscript asterisks indicating significance levels to map unit-labels.

general how large. And they are in any event an approximation. Moreover, using the approximately estimated standard errors to generate confidence intervals and hypothesis tests in the usual manners generally assumes (multivariate) normality of the parameter estimates. In maximum-likelihood contexts, this is not especially problematic since all ML estimates are at least asymptotically normal, although sample-size concerns may arise, perhaps regarding estimates involving $\hat{\rho}$, which is where spatial complications tend to arise. Given all this, we suspect that, in general, the asymptotic linear-approximations we have been recommending may have been larger than need be even in the linear-regression context. For those spatial linear-regression contexts, simple simulation strategies may be more effective. The parameter estimates' asymptotic normality, and their likely near-normality even in smaller samples, suggests that sampling coefficient estimates from a multivariate normal with the estimated means and variance-covariances, calculating the quantities of interest from these draws, and generating the desired indicators of certainty from the resulting samples may be more effective.

Even greater causes for concern arise in the spatial-probit context because the nonlinearity of the estimates of interest is more severe and even asymptotic normality is not assured. In fact, the (kernel of the) posterior joint-distribution of the parameters is not normal (as seen in (10), due to the $|\mathbf{I}_n - \rho\mathbf{W}|$ term), and, of the poster conditional-distributions, only that of $\boldsymbol{\beta}$ is exactly normal. Thus, we suggest using the same MCMC processes that yielded the parameter estimates and their certainty estimates to estimate by simulation the quantities on interest and their certainties. To elaborate, recall that, after sufficient burn-in, LeSage's Metropolis/Hastings-within-Gibbs sampler generates draws from the poster joint-distribution of the parameters. We gave the parameter estimates as the sample-means of these draws, and derived certainty estimates for those parameter-estimates from variances or percentile-ranges of those draws. Since one property of the Gibbs sampler is that it converges to the correct joint-posterior of the parameters, we could simply calculate any quantity of interest for each element of the (post-burn-in) sample *vector* of parameters, supplying whatever counterfactual values of interest for whatever variables are involved in the expression of interest.

Most usually, our interests will surround levels or changes in $\hat{p}_i$ and $\hat{p}_j$'s (or most generally, $\hat{\mathbf{p}}$), the (vector of) probabilities of outcomes 1 in *i* and *j* units induced by hypothetical levels or changes in some $\left(x_k\right)_i$ and $\left(x_k\right)_j$ (or most generally, $\mathbf{X}$). For instance, using (4), we could calculate the effects on the estimated probability of an outcome of 1 in unit *i* of some change in $\mathbf{x}_i$ as:

$$\frac{\Delta\left[p(y_i=1)\right]}{\Delta\mathbf{X}} = p\left(u_i < \frac{\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}_1\boldsymbol{\beta}\right]_i}{\sigma_i}\right) - p\left(u_i < \frac{\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}_0\boldsymbol{\beta}\right]_i}{\sigma_i}\right) \qquad (29).$$

where $\Delta\mathbf{X}=\mathbf{X}_1-\mathbf{X}_0$ is the hypothetical change being considered in some *x* or *x*'s in some unit(s). Notice that to calculate the effect even of a change in one *x* in unit *i* on the probability of outcome 1 in *i*, the researcher must specify not only the from/to levels of that change and the levels of all the $\mathbf{x}_i$, as in the standard probit, but also all the levels of all the *x* in all the other units. Intuitively, this is because not only do all the $\mathbf{x}_i$ affect where we are on the probit sigmoid curve, but all the $y_j^*$ also affect that positioning via spatial feedbacks, and those in turn depend on all $\mathbf{x}_j$ (and all the other $y_{\sim j}^*$, including $y_i^*$, and so on). These expressions and procedures hold for any *i* and $\Delta\mathbf{X}$, so the effect on some *j* of changes in *i* is calculated by substituting the $j^{th}$ and $jj^{th}$ elements for all *i* and $ii^{th}$ and considering such $\Delta\mathbf{x}_i$. All of this seems doable and worth doing, although the need to specify all of $\Delta\mathbf{X}$ for any counterfactual may be a bit daunting, however, as previously noted, the larger challenge is that the probability that $u_i$ is less than $[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_i/\sigma_i$ (with $\sigma_i=\left\{[(\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})]_{ii}^{-1}\right\}^{.5}$) comes from the $i^{th}$ marginal distribution of a multivariate normal with means zero and variance-covariance $[(\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})]^{-1}$. Obtaining that marginal distribution would involve integrating this $n$-variate normal distribution over the $n-1$ other dimensions. Just to obtain the estimated effects on probabilities, therefore, is computationally burdensome. Then, too, the variance-covariance of that distribution involve estimates, $\hat{\rho}$, and so this procedure would need to be repeated enough times to include that variability also in any estimate of the certainty of these estimated effects.

In principle, then, we can calculate the $\Delta p_i$ responses in all units for any hypothetical change, $\Delta \mathbf{X}$, by this formula:

$$\frac{\Delta \mathbf{p}}{\Delta \mathbf{X}} = \mathbf{\Phi}_n\left(\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}_1\boldsymbol{\beta}\right] \quad \left[\left\{\sigma_i^{-1}\right\}\right]\right) - \mathbf{\Phi}_n\left(\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}_0\boldsymbol{\beta}\right] \quad \left[\left\{\sigma_i^{-1}\right\}\right]\right) \qquad (30).$$

where $\Delta\mathbf{p}$ is the $n\times 1$ vector of $\Delta\left[p(y_i=1)\right]$ across all $i$, $\mathbf{\Phi}_n(\ )$ refers the cumulative distributions, evaluated element-by-element at the values of its $n\times 1$ vector argument, from the $n$-variate normal distribution with means zero and variance-covariance $[(\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})]^{-1}$, $\left[\left\{\sigma_i^{-1}\right\}\right]$ is the $n\times 1$ vector of the previously defined scalars $\sigma_i$, and indicates element-by-element multiplication. In principle, too, for given $\hat{\rho}$, these integrals could be calculated numerically. Certainty estimates for these effect estimates could then be obtained by repeating the process for many draws from the Gibbs-sampled joint-posterior for the parameters and calculating variances or percentile ranges.

We think a simpler expedient to evade the integration of the n-dimensional multivariate-normal may be to draw coefficients from the multivariate posterior-distribution $\sigma$, $\rho$, and $\boldsymbol{\beta}$, and draw $\boldsymbol{\varepsilon}$ from its independent-normal distributions, and then calculate $\hat{\mathbf{y}}_1^*$ and $\hat{\mathbf{y}}_0^*$ using $\left(\mathbf{I}_n-\rho\mathbf{W}\right)^{-1}\left(\mathbf{X}\hat{\boldsymbol{\beta}}+\boldsymbol{\varepsilon}\right)$ for some fixed $\mathbf{X}_1$ and $\mathbf{X}_0$; then simply apply (3) to convert to vectors of ones and zeros, $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_0$. Do this for large number of draws, and the share of ones should be $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_0$. Also, the average of $\hat{\mathbf{y}}_1-\hat{\mathbf{y}}_0$ should then be the desired vector of estimated effects, and the variance-covariance of those differences should be the variance-covariance of those estimated effects.[19]

## VI. Reanalysis: Diffusion of U.S. State Child Health Insurance Program (CHIP) Premiums

As illustration, Table 2 reports estimates of probit models with spatial-lag regressors (or spatial error-dependence) by standard maximum-likelihood (ML) methods that erroneously assume spatial lags exogenous, by MCMC or RIS methods but maintaining the same erroneous assumption, and by true spatial-lag probit (and also spatial-error probit) using the MCMC and RIS methods described in Section III. Using a strong recent example from the productive literature on policy diffusion among U.S. states, we will replicate Volden's (2006) study of the *Children's Health Insurance Program*, CHIP. The dependent variable indicates whether a state's CHIP requires recipients to pay monthly premia. The explanatory variables (assumed exogenous) are five-year averages of state (unit-level) demographic, economic, and political factors expected to affect social-insurance generosity—namely the state poverty-rate, the average monthly wage in retail, government ideology (0 to 100 right to left), degree of interparty competition (measured by two-party vote-share in state-office election, 0.5 to 1.0 competitive to non-competitive), tax effort (revenue as a percentage of state GDP), and an (assumed exogenous) external condition—the federal-government-paid portion of AFDC benefits. Most crucially to us, the models also properly include a spatial lag (or a spatial-error component) to reflect possible diffusion. We use a standardized binary contiguity-weights matrix, $\mathbf{W}$, which codes $w_{ij} = (1,0)$ for whether states $i$ and $j$ border and then row-standardizes (as commonly done in spatial-econometrics) the resulting matrix by dividing each element by its row's sum. This gives $(\mathbf{Wy})_i$ as the unweighted average of the dependent variable in $i$'s bordering states.

The first two columns report models estimated assuming that the spatial lags are exogenous. The

[19] We say *in principle* throughout this paragraph because we have not yet implemented any of this and know that it will be extremely computationally demanding.

probit model in the first column is estimated by standard ML techniques. The parentheses contain estimated standard errors and the hypothesis tests assume that the asymptotic *t*-statistics are normally distributed. The models in columns two and three are estimated using MCMC methods with diffuse zero-mean priors. The reported coefficient-estimates are the mean of the posterior distribution based on 10,000 cycles of the sampler after a 1000-cycle burn-in. The parentheses report sample standard-deviations of the posterior distributions. The *p*-values are also calculated directly from this posterior, without calculating or assuming anything about *t*-statistics. The results in columns two and three are very similar, as they should be given the latter's diffuse priors and asymptotic normality. However, because the probit-MCMC estimator used in column two, as with probit-ML, incorrectly treats the spatial lag as exogenous (i.e., as any other right-hand-side variable), the likelihood is misspecified so the sampler draws from the wrong posterior distribution for the spatial coefficient $\hat{\rho}$. As we have seen, these specification errors seriously compromise inferences from either of these models about the strength and importance of spatial interdependence.

**Table 2: CHIP Premiums in U.S. States, Estimation Results**

| | Probit-ML | Probit-MCMC | Spatial-Lag Probit | Spatial-Error Probit |
|---|---|---|---|---|
| *Constant* | -4.978<br>(6.260) | -5.163<br>(6.292) | -5.606<br>(10.159) | -5.531<br>(7.337) |
| *Poverty Rate* | -.244<br>(.153) | -.265**<br>(.156) | -.374**<br>(.231) | -.243*<br>(.157) |
| *Retail Wage* | .004<br>(.003) | .004*<br>(.003) | .006*<br>(.004) | .004*<br>(.003) |
| *Government Ideology* | .011<br>(.013) | .011<br>(.013) | .014<br>(.020) | .014<br>(.014) |
| *Inter-party Competition* | 2.174<br>(3.388) | 2.108<br>(3.478) | 1.473<br>(6.134) | 2.636<br>(3.794) |
| *Tax Effort* | -.014<br>(.019) | -.014<br>(.019) | -.020<br>(.034) | -.017<br>(.021) |
| *Federal Share* | .045<br>(.063) | .048<br>(.064) | .065<br>(.095) | .043<br>(.066) |
| *Spatial lag or error-lag* | .079<br>(.798) | .102<br>(.815) | .200***<br>(.148) | .297***<br>(.196) |
| **Pseudo-$R^2$** | .222 | .220 | .607 | .574 |
| **Observations** | 48 | 48 | 48 | 48 |

*Notes:* The first two columns' models are estimated assuming the spatial lags exogenous. The first column estimates are from the standard ML estimator. Its parentheses contain estimated standard errors; its hypothesis tests assume asymptotic normality of calculated *t*-statistics. The models in columns two through four apply MCMC methods with diffuse priors. The reported coefficient estimates are the posterior-density means based on 10,000 samples after 1000-sample burn-ins. The parentheses contain sample standard-deviations of these posteriors. The *p*-values are calculated directly from the posterior density without calculating *t*-statistics or assuming normality. The last two columns report estimates from true spatial estimators described in the text. In column three, 30 of the 10,000 sampled spatial-lag coefficients were negative; in column four, none of the 10,000 were negative. ***p-value <.01, **p-value<.05, *p-value <.10.

The model in column three is estimated with the true spatial estimator described above. The draws for $\rho$ are taken from the correct (non-standard) posterior distribution using Metropolis-Hastings. In this case only 30 of the 10,000 spatial AR coefficients sampled from the posterior distribution were negative. Thus, there is strong evidence of positive spatial interdependence in states' decisions to include a monthly premium in their CHIP. In addition, these probit results suggest that a state's poverty rate and average monthly retail wage are also important determinants.

## VII. Conclusion

Spatial interdependence is prevalent and substantively and theoretically important in social-science binary-outcomes. Standard ML-estimation of binary-outcome models in the presence of spatial interdependent are badly misspecified if that interdependence is ignored, but they are also misspecified (we suspect less badly, but we have not explored that yet), if that interdependence is reflected by inclusion of an endogenous spatial lag as an explanator. Spatial-lag probit models are difficult and highly computationally demanding, but not impossible, to estimate with appropriate estimators. Interpretation is also complicated by the same considerations, although we have shown how, in principle, they may be calculated directly, and have suggested a far more expedient method that may work sufficiently well. The next important task is to implement and evaluate these ways of calculating and presenting spatial effects along with certainty estimates.

## REFERENCES


Autant-Bernard, C. 2006. "Where Do Firms Choose to Locate Their R&D? A Spatial Conditional Logit Analysis on French Data," *European Planning Studies* 14(9):1187-1208.

Banerjee, S., Carlin, B.P., Gelfand, A.E. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman & Hall.

Beron, K.J., Murdoch, J.C., Vijverberg, W.P.M. 2003. "Why Cooperate? Public Goods, Economic Power, and the Montreal Protocol," *Review of Economics and Statistics* 85(2):286-97.

Beron, K.J., Vijverberg, W.P.M. 2004. "Probit in a Spatial Context: A Monte Carlo Analysis," in L. Anselin, R.J.G.M. Florax, & S.J. Rey, eds., Advances in Spatial Econometrics: Methodology, Tools and Applications. Berlin: Springer-Verlag.

Bhati, A.S. 2005. "Modeling Count Outcomes with Spatial Structures: An Information-Theoretic Approach," unpublished: Justice Policy Center, The Urban Institute. http://www.american.edu/cas/econ/faculty/golan/Papers/Papers05/BhatiPaper.pdf, or http://www.uni-kiel.de/ifw/konfer/spatial/bhati.pdf.

Bolduc, D., Fortin, B., Gordon, S. 1997. "Multinomial Probit Estimation of Spatially Interdependent Choices: An Empirical Comparison of Two New Techniques," *International Regional Science Review* 20:77-101.

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

Cho, W.T., Rudolph, T. 2007. "Emanating Political Participation: Untangling the Spatial Structure behind Participation," *British Journal of Political Science* 37(1):

Coughlin, C.C., Garrett, T.A., Hernández-Murillo, R. 2003. *Spatial Probit and the Geographic Patterns of State Lotteries*. St. Louis Federal Reserve Bank Working Paper 2003-042B. http://research.stlouisfed.org/wp/2003/2003-042.pdf.

Darmofal, D. 2007. "Bayesian Spatial Survival Models for Political Event Processes," Unpublished: http://people.cas.sc.edu/darmofal/DarmofalBayesianSpatialSurvival.pdf.

Dubin, R.A. 1997. "A Note on the Estimation of Spatial Logit Models," *Geographical Systems* 4(2):181-93.

Fleming, M.M. 2004. "Techniques for Estimating Spatially Dependent Discrete-Choice Models," in L. Anselin, R.J.G.M. Florax, & S.J. Rey, eds., *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Berlin: Springer-Verlag.

Franzese, R., Hays, J. 2004. "Empirical Modeling Strategies for Spatial Interdependence: Omitted-Variable vs. Simultaneity Biases," unpublished. Summer meetings of the Political Methodology Society: http://www.umich.edu/~franzese/FranzeseHays.PolMeth.2004.pdf.

Franzese, R., Hays, J. 2006. "Spatio-Temporal Models for Political-Science Panel and Time-Series-Cross-Section Data," unpublished. Summer meetings of the Political Methodology Society: http://www.umich.edu/~franzese/FranzeseHays.S.ST.EconometricsForPS.PolMeth06.pdf.

Franzese, R., Hays, J. 2007a. "Empirical Models of Spatial Interdependence," in Box-Steffensmeier, J., Brady, H., & Collier, D., eds., *Oxford Handbook of Political Methodology* (forthcoming).

Franzese, R., Hays, J. 2007b. "Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data," *Political Analysis* 15(2):140-64.

Franzese, R., Hays, J. 2007c. "Spatial Interdependence in Comparative Politics: Theoretical- and Empirical-Model Specification, Estimation, Interpretation, and Presentation," forthcoming in 40th Anniversary Issue of *Comparative Political Studies*.

Garrett, T.A., Wagner, G.A., Wheelock, D.C. 2005. "A Spatial Analysis of State Banking Regulation," *Papers in Regional Science* 84(4):575-95.

Holloway, G., Shankar, B., Rahmanb, S. 2002. "Bayesian Spatial Probit Estimation: A Primer and an Application to HYV Rice Adoption," *Agricultural Economics* 27(3):383-402.

Kayser, M.A. 2007. "Partisan Waves: International Sources of Electoral Choice," unpublished. University of Rochester. http://mail.rochester.edu/~mksr/papers/PWaves_ECM_070108.pdf.

Kelejian, H.H., Prucha, I.R. 2001. "On the Asymptotic Distribution of the Moran I Test Statistic with Applications," *Journal of Econometrics* 104:219-57.

Lacombe, D.J., Shaughnessy, T.M. 2005. "An Examination of a Congressional Vote Using Bayesian Spatial Probit Techniques." Paper presented at the 2005 Meetings of the Public Choice Society.

LeSage, J.P. 1999. *Spatial Econometrics*.
http://www.rri.wvu.edu/WebBook/LeSage/spatial/wbook.pdf

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

Lin, G. 2003. "A Spatial Logit Association Model for Cluster Detection," *Geographical Analysis* 35(4):329-40.

McMillen, D.P. 1992. "Probit with Spatial Autocorrelation," *Journal of Regional Science* 32:335-48.

McMillen, D.P. 1995. "Selection Bias in Spatial Econometric Models," *Journal of Regional Science* 35(3):417-36.

Murdoch J.C., Sandler, T., Vijverberg, W.P.M. 2003. "The Participation Decision versus the Level of Participation in an Environmental Treaty: A Spatial Probit Analysis," *Journal of Public Economics* 87(2):337-62.

Novo, A. 2003. *Contagious Currency Crises: A Spatial Probit Approach*. Banco de Portugal Working Paper: http://www.bportugal.pt/publish/wp/2003-5.pdf.

Phaneuf, D.J., Palmquist, R.B. 2003. "Estimating Spatially and Temporally Explicit Land Conversion Models Using Discrete Duration," Unpublished:
http://www.aere.org/meetings/0306workshop_Phaneuf.pdf.

Pinkse, J., Slade, M.E. 1998. "Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models," *Journal of Econometrics* 85: 125-54.

Pinkse, J. 1999. *Asymptotic Properties of Moran and Related Tests and Testing for Spatial Correlation in Probit Models*. University of British Columbia, Department of Economics Working Paper:

Rathbun, S.L., Fei, S. 2006. "A Spatial Zero-Inflated Poisson Regression Model for Oak Regeneration," *Environmental and Ecological Statistics* 13(4):409-26.

Schofield, N., Miller, G., Martin, A. 2003. “Critical Elections and Political Realignments in the USA: 1860-2000,” *Political Studies* 51(2):217-40.

Signorino, C. 1999. “Strategic Interaction and the Statistical Analysis of International Conflict,” *American Political Science Review* 93(2):279-98.

Signorino, C. 2002. “Strategy and Selection in International Relations,” *International Interactions* 28:93-115.

Signorino, C., Yilmaz, K. 2003. “Strategic Misspecification in Regression Models,” *American Journal of Political Science* 47(3):551-66.

Signorino, C. 2003. “Structure and Uncertainty in Discrete Choice Models,” *Political Analysis* 11(4): 316-44.

Signorino, C., Tarar, A. 2006. “A Unified Theory and Test of Extended Immediate Deterrence,” *American Journal of Political Science* 50(3):586-605.

Simmons, B., Elkins, Z. 2004. “The Globalization of Liberalization: Policy Diffusion in the International Political Economy,” *American Political Science Review* 98(1):171-89.

Smith, T.E., LeSage, J.P. 2004. “A Bayesan Probit Model with Spatial Dependencies,” in J.P. LeSage & R.K. Pace, eds., *Spatial and Spatio-Temporal Econometrics*. Amsterdam: Elsevier.