Spatial, Temporal, and Spatiotemporal Autoregressive Probit Models of Binary Outcomes: Estimation, Interpretation, and Presentation

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

Jude C. Hays (jch61@pitt.edu)
Associate Professor of Political Science, University of Pittsburgh

Lena M. Schaeffer (lena.schaffer@ir.gess.ethz.ch)
Ph.D. Candidate, *Internationale Beziehungen*, ETH—*Eldgenössische Technische Hochschule Zürich*

30 August 2010

[Prepared for the Annual Meeting of the American Political Science Association, Sep 2010]

ABSTRACT: Spatial/Spatiotemporal interdependence—i.e., that the outcomes, actions, or choices of some unit-times depend on those of others—is substantively and theoretically ubiquitous and central in binary outcomes of interest across the social sciences. However, most empirical applications omit spatial interdependence and, at best, treat temporal dependence as nuisance to be "kludged"; indeed, even theoretical and substantive discussion usually ignores (inter)dependence. Moreover, in the few contexts where spatial interdependence has been acknowledged or emphasized, such as in the social-network and policy-diffusion literatures, empirical models either do not fully reflect the simultaneity of the outcomes across units, or they do not recognize the endogeneity of the spatial lags which are used (appropriately) to model the interdependence. This paper notes and explains some of the severe challenges posed by spatiotemporal interdependence in binary-outcome models and then follows recent spatial-econometric advances to suggest two simulation-based approaches for surmounting the computational intensiveness of these models: classical recursive-importance-sampling (*RIS*) or Bayesian Markov-chain Monte-Carlo (*MCMC*). Serial autocorrelation in binary outcomes raises essentially the same challenges, so these strategies offer effective approach temporal dependence as well. We provide Monte-Carlo comparisons of the performance of these alternative estimators for spatial probit, including comparisons to estimation-strategies blind to or naïve about (inter)dependence—i.e., omitting spatial lags or including them but treating them as exogenous regressors in standard probit estimation—and then we show how to apply related simulation methods to calculate estimated spatial effects of hypothetical shocks in terms of outcomes or probabilities of outcomes (with associated confidence/credibility regions) rather than only in parameter-estimate or latent-variable terms as in all prior spatial-probit applications. We illustrate with applications to U.S. states' adoptions of legislative term-limits and to great-power decisions to enter World War I.

Spatial, Temporal, and Spatiotemporal Autoregressive Probit Models of Binary Outcomes: Estimation, Interpretation, and Presentation

ABSTRACT: Spatial/Spatiotemporal interdependence—i.e., that the outcomes, actions, or choices of some unit-times depend on those of others—is substantively and theoretically ubiquitous and central in binary outcomes of interest across the social sciences. However, most empirical applications omit spatial interdependence and, at best, treat temporal dependence as nuisance to be "kludged"; indeed, even theoretical and substantive discussion usually ignores (inter)dependence. Moreover, in the few contexts where spatial interdependence has been acknowledged or emphasized, such as in the social-network and policy-diffusion literatures, empirical models either do not fully reflect the simultaneity of the outcomes across units, or they do not recognize the endogeneity of the spatial lags which are used (appropriately) to model the interdependence. This paper notes and explains some of the severe challenges posed by spatiotemporal interdependence in binary-outcome models and then follows recent spatial-econometric advances to suggest two simulation-based approaches for surmounting the computational intensiveness of these models: classical recursive-importance-sampling (*RIS*) or Bayesian Markov-chain Monte-Carlo (*MCMC*). Serial autocorrelation in binary outcomes raises essentially the same challenges, so these strategies offer effective approach temporal dependence as well. We provide Monte-Carlo comparisons of the performance of these alternative estimators for spatial probit, including comparisons to estimation-strategies blind to or naïve about (inter)dependence—i.e., omitting spatial lags or including them but treating them as exogenous regressors in standard probit estimation—and then we show how to apply related simulation methods to calculate estimated spatial effects of hypothetical shocks in terms of outcomes or probabilities of outcomes (with associated confidence/credibility regions) rather than only in parameter-estimate or latent-variable terms as in all prior spatial-probit applications. We illustrate with applications to U.S. states' adoptions of legislative term-limits and to great-power decisions to enter World War I.

I. Spatial Interdependence and Temporal Dependence in Binary-Outcome Models

Many phenomena that social scientists study are inherently, or by measurement, discrete choices. Canonical political-science examples include citizens' vote and turnout choices, legislators' votes, governments' policy-enactments, wars among or within nations, and regime type or transition. In all these political contexts, and widely across the social sciences,[1] substantively and theoretically, the choices/outcomes of/in some units depend on those of other units. 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' policy choices depend on others' policies *via* competition or learning; nations' internal wars may arise in some part through contagion from others' conflicts; states' entry to and involvement in external wars, international organizations, and treaties are heavily conditioned by whether and which other states join; 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,

[1] For an extensive, topically organized bibliography of interdependence studies across the social sciences, see appendix to Hays et al. (2010): http://www-personal.umich.edu/~franzese/StatMeth.mSTAR.WebAppendix.pdf.

beyond 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.[2] 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 much invigorated recently in comparative and international politics. Dahl's (1971) classic *Polyarchy*, e.g., (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. Graham et al. (2008) offer an excellent recent review of these diffusion literatures.

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 effects on respondent behaviors or opinions of aggregates of others'—e.g., those of her region, community, or social network. Within the large *contextual-effects* literature in political behavior (Huckfeldt & Sprague 1993 review), recent work stressing 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

[2] 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. See also the extended bibliography linked in note 1.

the subfields and substance of political science. Inside democratic legislatures, representatives' votes depend on others' (expected) votes. In electoral studies, candidate qualities or strategies, 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; globalization, for instance, 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 resulting from some interdependence of mass opinion and vote choices (but *cf.* Kayser 2007).

The substantive/theoretical ubiquity and centrality of interdependence across political-science discrete-choice contexts notwithstanding, studies that accord interdependence explicit attention are uncommon. The rare exceptions include the policy-diffusion and contextual-effect literatures cited above; Ward, Gleditsch, and colleagues[3] and Signorino and coauthors[4] in international relations; Li & Thompson (1975), Govea & West (1981), and Brinks & Coppedge (2006) on coups, riots, and revolutions, respectively; Schofield et al. (2003) on citizens' votes and Lacombe & Shaughnessy (2005) on legislators' votes; and Mukherjee & Singer (2007) on inflation targeting.

Likewise, despite the manifest interdependence in social-science discrete-choices, assumptions

[3] Shin & Ward 1999, Gleditsch & Ward 2000, Gleditsch 2002, Ward & Gleditsch 2002, Hoff & Ward 2004, Gartzke & Gleditsch 2006, Salehyan & Gleditsch 2006, Gleditsch 2007.
[4] Signorino 1999, 2002, 2003, Signorino & Yilmaz 2003, Signorino & Tarar 2006.

of *in*dependence pervade almost all empirical analyses of them, even in those research areas that give interdependence greater substantive and theoretical weight. Empirical models of war in which the dependence of one state's choices on those of others enters explicitly are rare.[5] Empirical models of policy, institution, or regime diffusion often do account interdependence explicitly by including as explanators (weighted) averages or sums of other units' outcomes (e.g., the number of other states that have adopted a policy or treaty), but the endogeneity of this *spatial lag* is rarely confronted. Typically, diffusion researchers time-lag these spatial lags, as in the sophisticated event-history analyses of modern applications for example, and this *can* suffice to evade the simultaneity bias (see, e.g., Beck et al. 2006), but only *if* and insofar as (*i*) actual interdependence transpires only with a lag, (*ii*) with actual lag periodicity and lag structure matching that of the empirical observations and specification, and (*iii*) that the empirical model of spatiotemporal dynamics is adequate to prevent the past bleeding into present through mismeasurement/misspecification.[6] However, placing the actual binary outcomes of other units (or their weighted sums or averages) *simultaneously* on the right-hand side is not algebraically possible (as discussed further below: see Heckman 1978); such simultaneity can only logically operate through the latent variables or errors. This paper presents and shows how to estimate and interpret such simultaneous spatial-lag models.

Similarly, empirical network analyses, including the most recent and exciting contributions in random-graph (e.g., Robins et al. 2007; Hunter et al. 2008), longitudinal-network (Snijders 2005), and/or network-coevolution (e.g., Snijders et al. 2007) models,[7] fail to address fully and/or directly the interdependence of their binary outcomes. These network-formation models proceed, instead, by modeling network ties conditional on some summary-statistic(s) of the network reflecting particular

[5] Ward, Gleditsch, and colleagues and the Signorino and coauthors are among the few exceptions (see note 3).
[6] As Beck et al. (2006), e.g., note, adequacy of the spatiotemporal dynamic model can and should be tested. We have not seen these tests conducted in the diffusion literature though, nor, usually, sign that researchers are aware of the issue.
[7] The Bayesian latent-space approach of Hoff et al (2002), Hoff & Ward (2004), and Hoff & Westveld (2007) is related, with related appeals and limitations.

behavioral tendencies, some of which might implicate dependencies: e.g., a tendency toward transitive-triplets (A-C and A-B ties increase the likelihood of B-C ties). Longitudinal-network and network-coevolution models typically apply these conditional-independence assumptions and also often apply temporal-sequencing strategies similar in essence to those of the diffusion literatures. By any of these approaches, the network-tie decisions and/or behavioral choices of units are assumed independent of each other and of other units' decisions and choices, conditional upon the existing network and set of units' characteristics as given in the summary statistic of the pre-existing set of network and behavioral choices intended to reflect the modeled dependency. That is, none of these approaches allows a direct, simultaneous interdependence of binary outcomes (where *simultaneous* means, as elaborated in points *i-iii* above, *within observational period, as effectively modeled*).[8]

That outcomes will autocorrelate over time requires no parallel introduction or argument. No one would argue or pretend that time-serial observations were temporally independent in virtually any context. Yet empirical applications that model temporal *auto*-dependence directly are perhaps even rarer than those addressing spatial interdependence. This is because binary-outcome models that fully properly reflect autocorrelation of the latent propensities in one unit-time with the preceding (same-unit) periods raise the same computational challenges as spatial-lag simultaneity raises, and some simple evasions are on offer for the temporal-dependence case. To sidestep the estimation challenges, researchers today mostly follow Beck et al.'s (1998) advice to model temporal *trends* with cubic splines in time-since-event instead.[9] Such time dependence—the binary equivalent of the time-dependent hazards of, e.g., Weibull duration models—is not quite *auto*-dependence of a unit's current propensity on its previous propensities, so while such kludges may mostly redress the exaggeration of the information in the data from falsely pretending time-serial binary observations

---

[8] We would also note in passing here, although this is an argument for fuller development in another venue, that ubiquitous interdependence typically violates the crucial SUTVA assumption of so-called *causal inference*, invalidating, or at least seriously complicating, such less structural/parametric approaches to empirical inference.

[9] Signorino & Carter (2010) have more-recently suggested even simpler polynomials in time function as well or better.

are independent, they do not offer direct model of the temporal *auto*-dependent process. (A common time-dependence likely misses some temporal autocorrelation also.) We suspect most importantly: the kludges do not yield the theoretically interesting and substantively likely dynamic responses that autoregressive processes do. Alternatively, some researchers have shifted strategy and focus to model instead transition matrices of state-switching probabilities (i.e., probabilities of switching or staying in states 0 or 1) conditional on the previous observed state (e.g., Przeworski et al. 2000, Beck et al. 2001). This is autoregression in the binary outcome (which, unlike in the *simultaneous* spatial-interdependence case, is algebraically possible), and therefore a fuller strategy for temporal auto-dependence, but modeling dependence directly in the latent propensities may be more appealing substantively in some contexts and more-easily affords estimation of models in more-familiar standard binary-regression format, without shifting strategy and focus.[10]

Working under the incorrect assumption of spatial, temporal, or spatiotemporal independence, of course, threatens over-confidence or inefficiency in the best of circumstances, and usually bias and inconsistency as well. Inclusion of spatial and/or temporal lags to reflect (inter)dependence would seem advisable, but models with temporal or time-lagged spatial lags raise formidable estimation challenges. Furthermore, simultaneous spatial-lags are endogenous and so introduce biases if entered in estimation procedures that assume independence.[11] For the linear-regression case, we have argued and shown elsewhere[12] that serious omitted-variable biases arise when spatial lags are excluded in the presence of interdependence and that redressing this issue by explicit inclusion of spatial lags to

[10] Beck et al. (2001) describe the lagged-latent binary-outcome model and its estimation by Bayesian MCMC. They correctly identify the final distribution in their sampler as a *T*-dimensional truncated normal, and they offered the same Gibbs-within-Gibbs solution to that challenge offered by LeSage & Pace (2009). This paper shows a classical simulated-likelihood strategy that is at least as computationally feasible and performs at least as well in mean-squared-error terms.
[11] Including other units' *outcomes* also introduces measurement error insofar as interdependence truly arises through the *propensities* or *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 simultaneously identified for binary outcomes. See further explanation below (and original exposition in Heckman 1978).
[12] Franzese & Hays (2003ab, 2004ab, 2005ab, 2006abc, 2007abc, 2008abcd, 2009abcd), Franzese et al. (2009, 2010), Hays (2009ab), Hays & Colaresi (2009), Hays & Kachi (2009), Hays et al. (2010).

reflect interdependence is generally of first-order benefit relative to the problems induced by spatial-lag endogeneity. However, these simultaneity biases do become appreciable as interdependence strengthens, so we also covered in these previous works methods for gauging that strength, for redressing the simultaneity issues of spatial lags, and for calculating and presenting estimates of spatially/spatiotemporally dynamic effects and their certainty, but almost exclusively in the linear-regression context. This paper begins a similar exploration of spatial, temporal, and spatiotemporal autoregressive models of binary outcomes, where the substantive and theoretical importance of (inter)dependence, the empirical problems created by its omission, and the methodological challenges raised by the endogeneity of its appropriately explicit inclusion are all at least as great.

II. The Econometric Problem

Methods for properly estimating and analyzing models of interdependent qualitative or limited dependent variables (henceforth: *QualDep* models) have received significant attention in the spatial-econometric literature recently. Most of this research considers the spatial-probit model with inter-dependence in the latent-variable, i.e., in the unobserved argument to the probit-modeled probability of a binary outcome.[13] Models of spatial sample-selection (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 of which closely resemble the spatial probit, have also been suggested, as have models of interdependent survival (Hays & Kachi 2009) or of survival with spatial "frailty" (i.e., error components: Banerjee et al. 2004, Darmofal 2007) and of spatial counts (e.g., Bhati 2005, Franzese & Hays 2009a), including a zero-inflated-count model (e.g., Rathbun & Fei 2006). Spatial probit is far the most-

[13] See, e.g., McMillen 1992, 1995, 2005; Bolduc et. al. 1997; Pinkse & Slade 1998; LeSage 1999, 2000, LeSage&Pace 2004, 2009; Beron et al. 2003; Beron & Vijverberg 2004. Spatial logit has also been suggested (e.g., Dubin 1997; Lin 2003; Autant-Bernard 2006), but spatial probit dominates the methodological and applied literatures, perhaps because the *n*-dimensional normal is relatively easier to manage than the *n*-dimensional extreme-value distribution.

common *S-QualDep* model in applied research, however.[14]

Several estimation strategies have been suggested for the spatial-probit model. McMillen (1992) suggested an EM algorithm, which first rendered the spatial-probit's non-additively-separable log-likelihood (see below) estimable, but the strategy also did not provide standard-errors for the crucial spatial-dependence parameter and required arbitrary parameterization of the heteroscedasticity that dependence induces (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 Metropolis-Hastings-within-Gibbs sampling. (LeSage & Pace 2009 corrects a crucial error in the earlier formulations of the estimator.) Fleming (2004) reviews these two families and simpler, if approximate, strategies allowing spatial interdependence in linear or nonlinear probability models[15] estimable by nonlinear least-squares, generalized linear-models, or generalized linear-mixed-models. Pinkse & Slade's (1998) two-step GMM estimator for spatial-error probit has seen some use in the literature, as has McMillen's (2005) GMM for linearized spatial-lag logit or probit and Pinkse et al.'s (2006) one-step (continuously updating) GMM for spatial-probit, but the first is inconsistent for the spatial-lag model and all three, being instrumental-variable estimations of linear approximations around zero interdependence, work well only in large samples with weak interdependence. The RIS and Bayesian strategies do not have these limitations[16] and (so) have dominated recent applications.

The remainder of this section considers the spatial-probit model and RIS and Bayesian strategies for estimating it, and then shows that a probit model with temporal dependence in latent variables is

---

[14] E.g., Holloway et al 2002, Beron et al 2003, Coughlin et al 2003, Murdoch et al 2003, Novo2003, Schofield et al 2003, Garrett et al. 2005, Lacombe & Shaughnessy 2005, Autant-Bernard 2006, Rathbun&Fei 2006, Mukherjee&Singer 2007.

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

[16] The instrumented-approximation approaches, on the other hand, are massively more efficient computationally, with estimation times orders of magnitude quicker, which becomes a dominant consideration in samples of thousands, plus.

very similar in form, indicating applicability of the same estimation strategies there. The structural model for the latent variable of 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:

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

Latent-variable **y*** links to the observed binary-outcome, **y**, through the measurement equation:

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

The probability that the i^{th} observation is one is calculated as follows:

$$\begin{aligned} p(y_i = 1 \mid \mathbf{X}) &= p\left(\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]_i + \left[(\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}\right]_i > 0\right) \\ &= p\left(u_i < \left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]_i \big/ \sigma_i\right) = \boldsymbol{\Phi}_i\left\{\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]_i \big/ \sigma_i\right\} \end{aligned} \qquad (4).[17]$$

Thus, as in the standard probit, a cumulative-normal distribution, $\boldsymbol{\Phi}\{\bullet\}$, gives the probability that the systematic component, $[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_i / \sigma_i$, exceeds the stochastic component, u_i. However, in spatial probit, the interdependence of the y_i^* induces a non-sphericity of the stochastic components; specifically, $\mathbf{u}$ is distributed *n*-dimensional *multivariate* normal with variance-covariance matrix $[(\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})]^{-1}$ (and mean **0**). Intuitively, $\boldsymbol{\varepsilon}$ is *multivariate* normal with mean **0** and spherical variance-covariance $\sigma^2\mathbf{I}$, with σ^2 normalized to 1 as usual for a probit model; therefore:

$$\begin{aligned} V[\mathbf{u}] \equiv \boldsymbol{\Sigma} = [(\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}] &= [(\mathbf{I}-\rho\mathbf{W})^{-1}]' V(\boldsymbol{\varepsilon}) [(\mathbf{I}-\rho\mathbf{W})^{-1}] \\ &= [(\mathbf{I}-\rho\mathbf{W})^{-1}]'\mathbf{I}[(\mathbf{I}-\rho\mathbf{W})^{-1}] = [(\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})]^{-1} \end{aligned} \qquad (5).$$

The probability that $[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_i / \sigma_i$ exceeds u_i is read from the i^{th} marginal distribution of this multivariate cumulative-normal, denoted $\boldsymbol{\Phi}_i\{\bullet\}$, which requires integrating that joint distribution over the other $n-1$ dimensions. Also, in (4), σ_i^2 is the ii^{th} element of variance-covariance (5), which

[17] In the middle step, note that the symmetry about zero of $\boldsymbol{\varepsilon}$, and so of $\mathbf{u}$, implies that $p(-u_i<x) = p(u_i<x)$ for any x.

is not a constant, 1, as in 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. Because the outcomes are interdependent, their joint distribution is not the product of the n marginal distributions, so one does not maximize sums of logs of n additively separable one-dimensional probabilities. They are interdependent, so one maximizes the log of *one* non-separable n-dimensional distribution. Finally, notice also that the i^{th} observation probability depends on the entire matrix $\mathbf{X}$ and vector $\boldsymbol{\varepsilon}$. This follows from the nonlinearity of the sigmoidal probit function, which implies that effects depend on where along the S-shape they occur, and where on that S-curve one lies depends on all of $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ given the dependence of y_i^* on $\mathbf{W}\mathbf{y}^*$.

The spatial-error version of the probit model is slightly simpler, taking the form:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad (6),$$

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

$$p(y_i = 1 \mid \mathbf{x}_i) = p\left(u_i < \mathbf{x}_i\boldsymbol{\beta}/\sigma_i\right) = \boldsymbol{\Phi}_i\left\{\mathbf{x}_i\boldsymbol{\beta}/\sigma_i\right\} \qquad (7),$$

where $\mathbf{x}_i$ the i^{th} row of $\mathbf{X}$. Again, these u_i are heteroskedastic and the probability derives from the i^{th} marginal distribution of a multivariate cumulative-normal with means $\mathbf{0}$ and variance-covariance $[(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})]^{-1}$, so spatial-error probit models entail the same estimation and interpretation complications as spatial-lag models. (Mixed spatial-lag/spatial-error models are also possible, but they have received little attention.) In the spatial-error model, because the interdependence operates only through $\boldsymbol{\varepsilon}$ and not all of $\mathbf{y}^*$, the position of the i^{th} observation on the sigmoidal probit-function depends on the entire vector $\boldsymbol{\varepsilon}$ but only on that observation's independent-variable values, $\mathbf{x}_i$.

Special circumstances might allow standard-probit estimation of spatial-lag models, but we view these as highly atypical. For instance, Anselin (2006) notes that, in the conditional counterpart of (1),

$$y_i^* = \rho \sum_j w_{ij} E(y_j^* \mid \mathbf{X}) + \mathbf{x}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (8),$$

$E(y_j^* \mid \mathbf{X})$ could be estimated by $\sum_j w_{ij} y_j$, the spatially weighted average of actual outcomes in units *j*. However, this spatial lag could be included as a regressor without introducing endogeneity problems only under stringent conditions that ensure other units' observations *j* are not jointly determined with those of *i*, and that "coding methods ensure that the sample does not contain these neighbors" (Anselin 2006). This means that any units *j* from which diffusion to any *i* in the sample is non-negligible (at any order spatial-lag) must be excluded from the sample but used in constructing the $\mathbf{Wy}$ spatial lag for the retained observations *i*. Alternatively, all *i*'s neighboring *j* according to $\mathbf{W}$ must be exogenous to *i* for 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. Moreover, while some substantive-theoretical contexts might suggest that interdependence propagates through the actual outcome rather than the latent variable, a simultaneous such model is not generally possible because, indirectly via feedback, y_i would generate y_i^* but also, directly, y_i is generated by $\Phi(y_i^*)$.[18] Conditions like those described above allow direct inclusion of $\mathbf{Wy}$ because they sever such indirect generation of y_i^* by y_i. These limitations are usually prohibitive practically, though contexts where such directional ordering and such omissions of certain *j* may be defensible are imaginable. Swank (2006, 2007), e.g., argues that U.S. tax policies exclusively lead others' tax policies, and he excludes all U.S. data in his tax-competition empirics, reserving those U.S. data solely for the role of spatial lag. If valid, these

[18] The requirement applies to any simultaneous feedback among endogenous qualitative variables, as perhaps first noted by Heckman (1978) in the context of a system of 2 endogenous equations, at least one of them being qualitative and modeled by a latent variable crossing a threshold. He states: "A necessary and sufficient condition for [sensibility of such a system of endogenous latent-variable equations is] that the probability of the event d_i=1 is not a determinant of the event… …[This] principal assumption essentially requires that the latent variable *y** and not the measured variable *y* appears [on the right-hand side of the] structural equation" (pp. 936-7). The same limitation does not quite obtain for temporal dependence, however. Since time is unidirectional, one may be able to rely on pre-determinedness of y_{t-1}, i.e., the indirect feedback from y_t to y_{t-1} does not occur (given sufficiently full and accurate specification of the temporal dynamics). Still, conditions for proper identification of just a temporally dynamic model with lagged binary-dependent-variables remain less than straightforward (see, e.g., Chamberlain 1993, Honore & Kyriazidou 2000).

arguments and sample-exclusions would allow standard-probit estimation.

Consider now how similar a probit model with temporal autocorrelation, say an AR(1) process, in the latent variable is to the spatial probit. Start with the structural model in matrix notation:

$$\mathbf{y}^* = \phi\mathbf{A}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (9).$$

With $\mathbf{y^*}$ a T×1 vector of latent variables, and $\mathbf{A}$ a matrix of zeros except for all ones on the lower first-minor diagonal (the diagonal just below the prime diagonal), and dropping the first observation, this gives a standard first-order temporal autoregressive model. The reduced form is again:

$$\mathbf{y}^* = (\mathbf{I} - \phi\mathbf{A})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{u} \ \text{, with } \mathbf{u} = (\mathbf{I} - \phi\mathbf{A})^{-1}\boldsymbol{\varepsilon} \qquad (10),$$

and, again, this implies a nonspherical variance-covariance of the form:

$$V[\mathbf{u}] \equiv \boldsymbol{\Omega} = [(\mathbf{I} - \phi\mathbf{A})^{-1}\boldsymbol{\varepsilon}] = [(\mathbf{I} - \phi\mathbf{A})^{-1}]' V(\boldsymbol{\varepsilon})[(\mathbf{I} - \phi\mathbf{A})^{-1}] = [(\mathbf{I} - \phi\mathbf{A})'(\mathbf{I} - \phi\mathbf{A})]^{-1} \qquad (11),$$

and so the rest of the discussion regarding the estimation and interpretation complications that come with this inseparable T-dimensional cumulative nonspherical normal apply *mutatis mutandis*. (The AR(1) temporally autocorrelated *error* model is likewise analogous to the spatial-error model.)

We will mostly focus on the unconditional, simultaneous spatial-lag model next because it raises the estimation and interpretation issues fully, and the temporal and spatiotemporal autoregressive analogue and extension follow entirely straightforwardly. We ignore the conditional spatial model, as it is usually inapplicable and anyway raises fewer interpretation and no estimation complications. We will not discuss the time-lagged spatial-lag model further because the conditions discussed above for the practical adequacy of the strategy seem restrictive for many social-science applications and because, even if otherwise adequate, the strategy evades little of the estimation complications, which arise even for merely time-lagged binary-dependent-variables (as discussed). We also do not discuss tests of the adequacy of time-lagged spatial-lag models or specification tests of spatial-*lag* vs. spatial-*error* vs. non-spatial models here, though these tests are important to consider, especially given the complexity and computational intensity of valid estimation strategies for full, simultaneous spatial probit.[19] For starts on these discussions, we refer the reader to Pinkse & Slade (1998), Pinkse (1999), Kelejian & Prucha (2001), and, for a recent review, Anselin (2006). Our considerations focus on spatial-probit estimation by RIS and by Bayesian MCMC methods, and their comparison to standard probit estimation with the endogenous spatial-lag, $\mathbf{Wy}$, included as a regressor, which is current standard-practice in empirical work where interdependence of binary outcomes is addressed.

III. The RIS and Bayesian Estimators for Simultaneous Spatial Probit

LeSage (1999, 2000) suggests using Bayesian Markov-Chain-Monte-Carlo (MCMC) methods to surmount the estimation complications introduced by the n-dimensional cumulative-normal in the spatial-probit likelihood (posterior). The basic idea of Monte Carlo (simulation) methods is simple:[20] if one can characterize the joint distribution (likelihood or posterior) of the quantities of interest (parameters), then 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 they aim to estimate arbitrarily closely.[21] In basic Monte-Carlo simulation, the draws are independent and the target distribution is specified directly. In MCMC, each draw is dependent on the previous one in a manner that generates samples with properties mirroring those of the joint population, using just the conditional distribution of each parameter. This is useful where the joint distribution is not expressible directly or, as with spatial probit, where its complexity makes direct sampling from the joint distribution prohibitively difficult and/or time-consuming.

We can describe Gibbs sampling, the simplest and most-common of the MCMC family, thusly: 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 point

[19] Monte Carlo simulation exploring the sensitivity of the time-lagging spatial-dependence strategy to validity of the lagged-interdependence-only assumption, to the periodicity-matching assumption, and to the empirical adequacy of the spatiotemporal dynamic model and tests thereof are also important analyses that remain for the future.

[20] Our simple introduction draws heavily from Gill's (2002) wonderful text on Bayesian methods.

[21] The *population parameters* thusly *arbitrarily closely approximated* are usually some *estimates* in an application, like the spatial-probit parameter-estimates, not the *true parameters* (a foreign concept in Bayesian terminology anyway).

all subsequent draws will be from the targeted posterior joint-distribution. To elaborate: first express 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 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 of parameter draws is the desired joint posterior-distribution. Thus, after having gathered some very large set of parameter-vector values by this process, discard some large initial set of draws (the *burn-in*) and base inferences on sample statistics from the remaining set of parameter vectors. A typical burn-in might be 1000 draws, and inferences might be based on the next 5000 or 10,000. Also, since each draw is conditional on the previous one's drawn values, autocorrelation typically remains, so "thinning" the post-burn-in sample by using every, say, third or fifth draw may boost efficiency.

The drawbacks of MCMC may be obvious from what we have said and declined to say. First, no universal tests exist to verify that *convergence* has occurred, so a burn-in may appear sufficient in that the next 5000 drawn parameter-vectors seem to follow some circumscribed bounds and behavior of some unknown target distribution (i.e., the sampler may seem to have *settled down*) only to have the 5001st leap into a new range and proceed toward convergence elsewhere. Second, despite their Markov-Chain origins, adjacent draws are asymptotically serially uncorrelated, but this *asymptopia* may not arrive within practical limits, and thinning may be insufficient help or too computationally costly. Third, the starting values are likewise asymptotically irrelevant, assuming the supplied set of conditional distributions properly could come from a valid joint distribution, but, as the previous two caveats imply, starting values may matter short of convergence, arrival at which is not verifiable.[22] These issues concern careful researchers, and many diagnostics and tests for non-convergence, serial correlation, or starting-value sensitivity, and numerous strategies for ameliorating them, exist (all imperfect, but useful still). However, the concerns do not outweigh the remarkably flexible utility of the Gibbs sampler, either in general or specifically in its application to spatial-probit estimation.

All but one of the conditional distributions for the spatial-probit-model parameters (given below) are standard, so the Gibbs sampler is useful for them. The crucial spatial-lag-coefficient, ρ, has the lone non-standard conditional-distribution; for it, Metropolis-Hastings sampling is used. Metropolis-Hastings differs from Gibbs sampling in the former's *seeding* or *jump* distribution from which values are drawn and then accepted or rejected as the next sampled parameters, depending on how they compare to a suitably transformed expression of the target distribution.[23] The Bayesian spatial-probit estimator (LeSage1999, 2000) uses Metropolis-Hastings for ρ within the Gibbs sampler procedure for the other parameters. Of course, this step adds some 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 outcome, $\mathbf{y}^*$—an additional conditional distribution will later apply (3) to convert unobserved $\mathbf{y}^*$ to observed $\mathbf{y}$[24]—for the spatial-lag model (1) as:

$$L\left(\mathbf{y}^*,\mathbf{W}\mid\rho,\boldsymbol{\beta},\sigma^2\right)=\frac{1}{2\pi\sigma^{2(n/2)}}\left|\mathbf{I}_n-\rho\mathbf{W}\right|e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})} \qquad (12),$$

where $\boldsymbol{\varepsilon}=\left(\mathbf{I}_n-\rho\mathbf{W}\right)\mathbf{y}^*-\mathbf{X}\boldsymbol{\beta}$. (The likelihood for spatial-error probit model (6) is the same but with

[22] The conditional distributions must also be expressible and sufficiently tractable to make so many draws a practicality.
[23] 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 $p=\min\{1, f(x_1)/f(x_0)\}$. 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*.
[24] This stratagem also enables LeSage to express the spatial-Tobit model by this same likelihood, adding a conditional distribution later to generate latent variables z for censored observations instead of one to generate y=(0,1) for the probit.

$\boldsymbol{\varepsilon} = \left(\mathbf{I}_n - \rho\mathbf{W}\right)\left(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}\right)$, where ρ serves here for (6)'s λ.) Diffuse priors yield joint posterior-density:

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

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

$$p\left(\sigma \mid \rho, \boldsymbol{\beta}\right) \propto \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})} \qquad (14).$$

Notice that conditioning on ρ allows $\left|\mathbf{I}_n - \rho\mathbf{W}\right|$ to be subsumed into the constant of proportionality and that (14) implies $\sigma^2 \sim \chi_n^2$, a standard distribution facilitating the Gibbs sampler. Next,

$$p\left(\boldsymbol{\beta} \mid \rho, \sigma\right) \sim N\left[\tilde{\boldsymbol{\beta}}, \sigma_\varepsilon^2 (\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{X})^{-1}\right] \qquad (15),$$

where, in spatial lag, $\mathbf{C} = \mathbf{I}_n$ and $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{I}_n - \rho\mathbf{W}\right)\mathbf{y}^*$, and, in spatial error, $\mathbf{C} = \left(\mathbf{I}_n - \rho\mathbf{W}\right)$ and $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{y}^*$. The conditional multivariate-normality of $\boldsymbol{\beta}$ allows the Gibbs sampler for it also, but ρ has non-standard conditional distribution, requiring Metropolis-Hastings sampling:

$$p\left(\rho \mid \boldsymbol{\beta}, \sigma\right) \propto \left|\mathbf{I}_n - \rho\mathbf{W}\right| \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})} \qquad (16),$$

with $\boldsymbol{\varepsilon}$ defined as given above for the spatial-error and the spatial-lag models.[25]

Finally, LeSage (1999, 2000) erroneously added the conditional distribution, namely a truncated normal, that translates $\mathbf{y}^*$ to $\mathbf{y}$, as a *univariate* truncated normal:

$$f\left(z_i \mid \rho, \boldsymbol{\beta}, \sigma\right) \sim N(\tilde{y}_i, \sigma_i^2), \text{ left- or right-truncated at 0 as } y_i = 1 \text{ or } 0 \qquad (17),$$

where $\tilde{y}_i$ is the predicted value of y_i^* (the i^{th} element of $\left(\mathbf{I}_n - \rho\mathbf{W}\right)^{-1}\mathbf{X}\boldsymbol{\beta}$ for spatial-lag or of $\mathbf{X}\boldsymbol{\beta}$ for spatial-error models) and the variance of $\tilde{y}_i$ is $\sum_i \omega_{ij}^2$ with ω_{ij} the i^{th} element of $\left(\mathbf{I}_n - \rho\mathbf{W}\right)^{-1}\boldsymbol{\varepsilon}$. In addition to producing biased estimates, this mistake, which earlier versions of this paper followed, gave the false impression that the Bayesian MCMC estimation-strategy was simpler and much faster

[25] Anselin (1988) shows that the minimum and maximum eigenvalues of a standardized spatial-weight matrix, **W**, bound ρ to $1/\lambda_{\min} < \rho < 1/\lambda_{\max}$. Adding this constraint to the rejection sampling should be beneficial.

than the classical simulated-likelihood (RIS) strategy. Lesage & Pace (2009) corrects the mistake, replacing *univariate* (17) with the properly multivariate truncated normal distribution:

$$f\left(\mathbf{z} \mid \rho, \boldsymbol{\beta}, \sigma\right) \sim \mathbf{MVN}(\tilde{\mathbf{y}}, \boldsymbol{\Sigma}), \text{ with each } i \text{ left- or right-truncated at 0 as } y_i = 1 \text{ or } 0 \qquad (18).$$ [26]

That is, the Bayesian MCMC estimator must also confront the multidimensional-normal integration that is the major complication raised by (inter)dependence in probit models.

Since the choice rule is $p(y_i = 1 \mid \mathbf{X}) = p(\mathbf{u} < [(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}])_i$, the cutpoint that gives $p(y_i=1)$, μ_i, depends on all the y_j*. The stochastic-component draws must therefore come from the nonspherical truncated multivariate normal (TMVN) with variance-covariance $\boldsymbol{\Sigma}$ and bounds $(-\infty, \mu_i)$ for $y_i=1$ and $(\mu_i, +\infty)$ for $y_i=0$. Following Geweke (1991) on drawing from a TMVN, the correct Bayesian MCMC estimator for the spatial-probit model adds another *m* step Gibbs sampler within the overall sampler, drawing each cutpoint, z_i, conditional on all the $z_{\sim i}$, from the conditional distributions for this *n*-variate TMVN. This parallels closely the computation intensity of the classical RIS strategy, which must also simulate the integration of this same multidimensional, cumulative, nonspherical truncated normal (and uses the Geweke-Hajivassiliou-Keane (GHK) simulator to do so).[27]

With all the conditional distributions, we can implement MCMC to estimate the model thus:[28]

---

[26] If we have correctly generated the multivariate analogue to the erroneously univariate expression in Lesage (2000) and Smith & LeSage (2004), spatial Tobit would replace (18) with: $f\left(\mathbf{z} \mid \rho, \boldsymbol{\beta}, \sigma\right) \sim \begin{cases} [1 - \boldsymbol{\Phi}(\boldsymbol{\Sigma}^{-.5}\mathbf{y}^*)]^{-1} \exp[-2\boldsymbol{\Sigma}^{-.5}(\mathbf{z} - \mathbf{y}^*)^2], & \text{for } z_i > 0 \\ 0, & \text{for } z_i \le 0 \end{cases}$.

[27] While one doesn't need near as many *m* on this Gibbs-within-Gibbs sampler as the thousands recommended for outer Gibbs sampler, but even *m*=10 for, say, a sample of the 3000 US counties yields 30,000 draws within each of the outer thousands of draws. For instance, LeSage & Pace (2009) report that, for just *m*=1 and merely 1000 outer draws for the 3000 US counties, their "relatively slow laptop" required 45 minutes for one spatial-probit estimation.

[28] In assigning diffuse priors to the parameters, LeSage (2000) also relaxes the assumption of homoskedasticity in $\boldsymbol{\varepsilon}$, allowing $V(\varepsilon)$ to vary arbitrarily by observation *i*. This allows exploration of variation in model fit and identification of and robustness to potential outliers, but creates as many parameters to estimate as observations. LeSage circumvents that issue 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 $\boldsymbol{\varepsilon}$ 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 therefore be used to test logit versus probit (versus un-named possibilities $q \neq 7.5$ and $q \neq \infty$).

Allowing arbitrary relative-variance requires the additional (informative) Gewekian prior and a (diffuse) hyper-prior on its parameter, q; produces more complicated expressions for the conditional distributions of σ, ρ, $\boldsymbol{\beta}$; and adds a conditional distribution (fortunately standard: χ^2_{q+1}) for the relative variances, υ_i. The steps below would now also include conditioning on starting values for, and then the previous draws of, $\boldsymbol{\upsilon}$, and a step inserted between 2 and 3 would

1. Use expression (14) to draw σ_1 using starting values $\rho_0, \boldsymbol{\beta}_0, \sigma_0$.
2. Use σ_1, ρ_0, and expression (15) to draw $\boldsymbol{\beta}_1$.
3. Use σ_1, $\boldsymbol{\beta}_1$, and expression (16) to draw ρ_1 by Metropolis-Hastings sampling.
4. SUBLOOP: Use an *m*-step Gibbs sampler to sample the outcomes, **z**, using the conditional distributions from the multivariate censoring distribution (18) and σ_1, $\boldsymbol{\beta}_1$, and ρ_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 σ, $\boldsymbol{\beta}$, and ρ 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).[29]

A classical approach, Recursive Importance-Sampling (RIS), has also been suggested to estimate simultaneous spatial-lag or spatial-error probit; RIS also uses simulation to approximate probabilities difficult to calculate analytically. We introduce RIS following Vijverberg's (1997) notation. To approximate an *n*-dimensional cumulative multivariate-normal distribution, e.g.,

$$p = \int_{-\infty}^{\mathbf{x}_0} f_n(\mathbf{x})\, d\mathbf{x}\,, \qquad (19),$$

where $f_n(\mathbf{x})$ is the density and $[-\infty, \mathbf{x}_0]$ the interval over which we want to integrate, we first choose a *n*-dimensional sampling-distribution with well-known properties and label a truncation of this sampling distribution with support over $[-\infty, \mathbf{x}_0]$ the *importance distribution*. Defining $g_n^c(\mathbf{x})$ as the density for this *n*-dimensional importance distribution, we then multiply and divide the right-hand-side of the integral we wish to calculate, (19), by this density, which simply restates (19) as:

$$p = \int_{-\infty}^{\mathbf{x}_0} \frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})}\, g_n^c(\mathbf{x})\, d\mathbf{x} \qquad (20).$$

By definition, the solution to this integral is a mean because $g_n^c(\mathbf{x})$ is a valid *pdf* over the integral's

---

draw the next $\boldsymbol{\upsilon}$ from χ^2_{q+1} conditional on the current $\sigma, \rho, \boldsymbol{\beta}$. Notice that setting the hyper-prior for q determinately to a large number (or 7.5) yields spatial probit (or logit) without heteroscedasticity/outlier-robustness.

[29] Thinning may also be advisable, although we have not yet explored that or found relevant discussion in the literature.

range, so (20) gives the probability sought, p, as the mean of $f_n(\mathbf{x})/g_n^c(\mathbf{x})$, which we can estimate using a sample of R draws of the $n\times 1$ vector $\mathbf{x}$ from the importance distribution. Formally:

$$p = E\left[\frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})}\right] \approx \frac{1}{R}\sum_{r=1}^{R}\frac{f_n(\tilde{\mathbf{x}}_r)}{g_n^c(\tilde{\mathbf{x}}_r)} \equiv \hat{p} \qquad (21).$$

To implement the RIS estimator, we draw $\mathbf{x}$ from the importance-distribution, for which we will use a truncated multivariate (independent) normal,[30] and calculate $f_n(\mathbf{x})/g_n^c(\mathbf{x})$.

Again, in the standard probit-model with independent errors, the numerator would simply sum n univariate cumulative standard-normal distributions, which is manageable. In spatial probit, with its interdependent errors, however, the numerator is a single n-dimensional cumulative-normal:

$$p(\mathbf{u} < \mathbf{v}) \qquad (22),$$

with $\mathbf{u}$ the $n\times 1$ vector of errors distributed $MVN(\mathbf{0},\boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} = (\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})^{-1}$, and with $\mathbf{v}$ an $n\times 1$ vector $\mathbf{v} = \mathbf{Q}(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}$, where $\mathbf{Q}$ is a diagonal matrix with diagonals $q_i = 2y_i - 1$.[31] The RIS estimator for spatial probit exploits that, as a variance-covariance matrix, $\boldsymbol{\Sigma}$ is positive definite, so a Cholesky decomposition exists such that $\boldsymbol{\Sigma}^{-1} = \mathbf{A}'\mathbf{A}$, with $\mathbf{A}$ being an upper-triangular matrix and $\boldsymbol{\eta} = \mathbf{A}\mathbf{u}$ giving n independent standard-normal variables, $\boldsymbol{\eta}$. (This exploitation is familiar as the same one applied in GLS.) Let $\mathbf{B} \equiv \mathbf{A}^{-1}$; substituting $\mathbf{u} = \mathbf{A}^{-1}\boldsymbol{\eta} \equiv \mathbf{B}\boldsymbol{\eta}$ into (22) then gives:

$$\Pr(\mathbf{B}\boldsymbol{\eta} < \mathbf{v}) = \Pr\left(\begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & \cdots & b_{1,n} \\ 0 & b_{2,2} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_{n\text{-}1,n\text{-}1} & b_{n\text{-}1,n} \\ 0 & \cdots & 0 & 0 & b_{n,n} \end{bmatrix}\begin{bmatrix} \eta_1 \\ \vdots \\ \vdots \\ \eta_{n\text{-}1} \\ \eta_n \end{bmatrix} < \begin{bmatrix} v_1 \\ \vdots \\ \vdots \\ v_{n\text{-}1} \\ v_n \end{bmatrix}\right) \qquad (23).$$

---

[30] Other importance distributions, such as a t or a uniform may be used. With a normal importance-distribution, RIS is equivalent to the better-known GHK (Geweke-Hajivassiliou-Keane) simulation estimator.

[31] Note that q=2y_i-1 is 1 for y_i=1 and -1 for y_i=0; thus, multiplying by $\mathbf{Q}$ serves to select the right sign on the systematic component up to which to integrate the distribution of the stochastic component $\mathbf{u}$. See, e.g., Greene (2008:778).

The elements of the $n \times 1$ vector $\boldsymbol{\eta}$ are independent, so the probability in (23) can be calculated by first evaluating the cumulative-normal distribution function at the implied upper bounds, which are determined recursively starting with the last observation, and then multiplying these probabilities. To determine these upper bounds, start by solving the inequalities in (23) for the vector $\boldsymbol{\eta}$:

$$\Pr\left(\begin{bmatrix} \sum_{i=1}^{n} b_{1,i}\eta_i \\ \vdots \\ b_{n\text{-}1,n\text{-}1}\eta_{n\text{-}1} + b_{n\text{-}1,n}\eta_n \\ b_{n,n}\eta_n \end{bmatrix} < \begin{bmatrix} v_1 \\ \vdots \\ v_{n\text{-}1} \\ v_n \end{bmatrix}\right) = \Pr\left(\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_{n\text{-}1} \\ \eta_n \end{bmatrix} < \begin{bmatrix} b_{1,1}^{-1}\left(v_1 - \sum_{i=2}^{n} b_{1,i}\eta_i\right) \\ \vdots \\ \vdots \\ b_{n\text{-}1,n\text{-}1}^{-1}\left(v_{n\text{-}1} - b_{n\text{-}1,n}\eta_n\right) \\ b_{n,n}^{-1}v_n \end{bmatrix}\right) \quad (24)$$

First, calculate the upper bound for the truncated-normal distribution of the n^{th} observation, which is $b_{n,n}^{-1}v_n$. Call the cumulative standard-normal evaluated at this upper bound p_n. Then take a draw from the standard-normal distribution truncated at $b_{n,n}^{-1}v_n$; call that draw $\tilde{\eta}_n$ and use it to calculate the upper bound for the truncated-normal distribution for the $(n\text{-}1)^{th}$ observation conditional on the n^{th} as $b_{n\text{-}1,n\text{-}1}^{-1}\left[v_{n\text{-}1} - b_{n\text{-}1,n}\tilde{\eta}_n\right]$. Evaluate the cumulative standard-normal at this upper bound and call it $p_{n\text{-}1}$. Then use the first two draws to calculate the $(n\text{-}2)^{th}$ upper bound and calculate $p_{n\text{-}2}$ analogously, and so on through all n observations. Formally, this recursive process for calculating the upper bounds is:

$$\left.\begin{array}{c} \eta_n < b_{n,n}^{-1}v_n \equiv \upsilon_n \\ \eta_{n\text{-}1} < b_{n\text{-}1,n\text{-}1}^{-1}\left[v_{n\text{-}1} - b_{n\text{-}1,n}\tilde{\eta}_n\right] \equiv \upsilon_{n\text{-}1} \\ \eta_{n\text{-}2} < b_{n\text{-}2,n\text{-}1}^{-1}\left[v_{n\text{-}2} - b_{n\text{-}2,n\text{-}1}\tilde{\eta}_{n\text{-}1} - b_{n\text{-}2,n}\tilde{\eta}_n\right] \equiv \upsilon_{n\text{-}2} \\ \vdots \end{array}\right\} \Rightarrow \eta_j < b_{j,j}^{-1}\left[v_j - \sum_{i=j+1}^{n} b_{j,i}\tilde{\eta}_i\right] \equiv \upsilon_j \quad (25).$$

The probability of observing a given sample of ones and zeros can now be found by evaluating the *univariate* cumulative-normal distribution function at each of these bounds, p_i, and then multiplying those probabilities: $\prod_{j=1}^{n} p_j = \prod_{j=1}^{n} \Phi\left(\upsilon_j\right)$. Repeating the entire process R times and averaging gives the

RIS estimate for the joint probability, i.e., the simulated likelihood, as this mean:

$$\hat{l} = (1/R)\sum_{r=1}^{R}\left[\prod_{j=1}^{n}\Phi\left(\upsilon_{j,r}\right)\right] \tag{26}$$

One can then maximize this simulated likelihood by any standard optimization routine to estimate parameters and apply the standard ML estimator for the variance-covariance (i.e., $-[\mathrm{H}(\hat{l})]^{-1}$).

IV. Monte Carlo Analyses of Standard-Probit vs. Bayesian MCMC & Classical RIS Spatial-Probit Estimation

We explore the small-sample properties of standard ML-probit and Bayesian MCMC and classical RIS estimators for the spatial-lag probit model using a data-generation process (DGP) that closely follows Beron & Vijverberg's (2004) Monte Carlo exploration of the RIS estimator:

$$\mathbf{y}^* = \left(\mathbf{I}_n - \rho\mathbf{W}\right)^{-1}\left(\mathbf{x}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right), \text{ where } \mathbf{x} = \left(\mathbf{I}_n - \theta\mathbf{W}\right)^{-1}\mathbf{z} \text{ and } \mathbf{z}, \boldsymbol{\varepsilon} \sim N(0,1) \tag{27}$$

We apply (3) to generate **y** from these **y***. Note that **x** and **y*** exhibit the same pattern of spatial interdependence, **W**, but with strengths θ *vs.* ρ. For **W**, we use a row-standardized binary-contiguity matrix for the 48 contiguous U.S. states. We set ρ to 0.5 and β to 1.0, and consider sample sizes n={48,144}[32] and θ values 0.0 and 0.5, giving four experiments total. Table 1 reports preliminary results for 100 trials using the standard-probit ML estimator with spatial lags **Wy** or **Wy***, the Bayesian MCMC spatial-probit estimator, and the classical RIS spatial-probit estimator. For the Bayesian estimator, we use LeSage's (2009) MatLab code, with q set to the default probit value determinately (see note 28), and a burn-in of 1000 trials, retaining the next 1000 for our simulation sample. For the classical RIS estimator we set R to 100 (see equation (21)). Standard ML-probit with a **Wy*** regressor is not practicable because **y*** is unobserved, but those results provide us valuable comparison. Only the simultaneity of the spatial lag biases those estimates, whereas the incorrect

[32] To create the weights matrix for the larger sample size we took the Kronecker product of the original 48×48 weights matrix with a 3×3 identity matrix. This could reflect, e.g., three observations of outcomes in each of the 48 states.

Wy spatial-lag used in current standard-practice incurs both simultaneity, with its likely inflation-bias, and measurement error (of **Wy** vs. **Wy***), with its attenuation bias.

Table 1: Simulation Results (100 Trials)

	ML with Wy		**ML with Wy***		**Bayesian MCMC**		**MSL with RIS**	
	β	ρ	β	ρ	β	ρ	β	ρ
Experiment #1: n=48, θ=0.0								
Mean Coefficient Estimate	0.90	0.35	1.08	0.81	1.05	0.41	0.94	0.36
Root Mean-Squared Error	0.30	0.48	0.52	0.48	0.36	0.20	0.30	0.20
Actual Std Dev of Estimates	0.29	0.46	0.51	0.37	0.36	0.18	0.30	0.16
Mean of Reported Std Err	0.27	0.37	0.33	0.32	0.33	0.21	0.28	0.20
Experiment #2: n=48, θ=0.5								
Mean Coefficient Estimate	1.07	0.42	1.07	0.76	1.18	0.41	1.06	0.37
Root Mean-Squared Error	0.31	0.49	0.38	0.50	0.41	0.25	0.30	0.23
Actual Std Dev of Estimates	0.30	0.49	0.37	0.42	0.37	0.23	0.30	0.19
Mean of Reported Std Err	0.30	0.40	0.34	0.31	0.36	0.20	0.30	0.18
Experiment #3: n=144, θ=0.0								
Mean Coefficient Estimate	0.88	0.40	1.02	0.73	0.99	0.43	0.93	0.34
Root Mean-Squared Error	0.18	0.26	0.16	0.30	0.17	0.14	0.17	0.19
Actual Std Dev of Estimates	0.14	0.22	0.16	0.16	0.17	0.12	0.15	0.10
Mean of Reported Std Err	0.15	0.20	0.17	0.17	0.17	0.13	0.16	0.11
Experiment #4: n=144, θ=0.5								
Mean Coefficient Estimate	1.00	0.47	0.98	0.69	1.04	0.45	0.99	0.37
Root Mean-Squared Error	0.16	0.28	0.19	0.26	0.18	0.14	0.16	0.17
Actual Std Dev of Estimates	0.16	0.28	0.19	0.18	0.18	0.13	0.16	0.11
Mean of Reported Std Err	0.16	0.21	0.18	0.15	0.18	0.12	0.16	0.10

Comparing the performances of the common-practice ML-probit with spatial-lag **Wy** strategy and the more sophisticated Bayesian MCMC and RIS estimators for the strength of interdependence ($\hat{\rho}$) reveals the relative inefficiency of the former approach. The simple estimator never outperforms the RIS or MCMC estimators by the root mean-squared error criterion.[33] Moreover, the standard error estimates are overconfident in all four experiments—by 20%, 18%, 9%, and 25% respectively. The relatively good bias properties of the common-practice estimator are attributable to the fact that it suffers two biases that happen fortuitously to offset somewhat in these experimental conditions. First is the simultaneity bias, which also plagues the ML estimator with the true spatial-lag, **Wy***. In that latter, this is the only source of bias, and, indeed, those columns show strong inflation of $\hat{\rho}$. The

[33] We recognize that this evaluates a Bayesian estimator by frequentist standards, but we think those standards worth considering nonetheless.

second bias of the common-practice standard ML-probit is an attenuation $\hat{\rho}$ due to measurement error in proxy spatial-lag, **Wy**, compared to true spatial-lag, $\mathbf{Wy}^*$. The simultaneity inflation-bias increases with ρ, but the impact of the attenuation bias instead decreases with sample size (for this particular **W** at least). Therefore, when ρ and n are small, measurement-error attenuation dominates, leaving a net-negative bias. When ρ and n are large, the simultaneity inflation-bias dominates, and net bias is positive. This implies that at some ρ and n between (somewhere near the conditions of our fourth experiment, apparently) the biases cancel. Further preliminary experiments varying ρ and n, using fewer trials for speed, so far confirm these intuitions.

Given the small numbers of trials in our experiments and draws for the RIS simulator ($R = 100$) (again, for speed in this current draft) we are hesitant to draw strong conclusions about the performance of the classical estimator, although we do note that our results largely mirror those reported by Beron and Vijverberg (2004, Tables 8.3 & 8.4). Under conditions like those of our second experiment, they report that RIS overestimates β by 10% and underestimates ρ by 18%. We find similar biases of 6% and 26% respectively. On the positive side, the RIS estimator seems to be relatively efficient, performing well by RMSE criteria, and, importantly, the standard error estimates are reasonably accurate. Perhaps the most troubling aspect of our results concerns this small sample bias of both the Bayesian MCMC and classical RIS estimators.

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 *effects* (ideally: causal ones), i.e., changes in the expectations of *outcomes* associated with (ideally: caused by) changes in explanatory factors or other counterfactual shocks. Ultimately, we estimate *coefficients* like ρ and $\boldsymbol{\beta}$ for the purpose of estimating *effects* like $\frac{\partial y_i^*}{\partial x_i}$ or $\frac{\Delta y_i^*}{\Delta x_i}$, i.e., the effects of a marginal or discrete change in some explanatory factor in unit i, x_i, on the latent-variable outcome in i, y_i^*, or, better, the effects of x_i on the probability of i's choice or outcome, $\frac{\partial p(y_i=1)}{\partial x_i}$ or $\frac{\Delta p(y_i=1)}{\Delta x_i}$. Given interdependence, even these sorts of *within-unit* counterfactuals—of x_i on y_i—involve feedback from i through other units j back to i. In fact, in diffusion, interdependence, or spatial- or network-interaction contexts (roughly synonyms), our interests usually extend centrally to cross-unit feedback effects, such as $\frac{\partial y_j^*}{\partial x_i}$, $\frac{\Delta y_j^*}{\Delta x_i}$, $\frac{\partial p(y_j=1)}{\partial x_i}$, or $\frac{\Delta p(y_j=1)}{\Delta x_i}$. Either within or across units, we could also wish to consider some generic shocks to y_i^*, the linear *propensity* toward outcome y_i=1, rather than shock to some x_i. For these purposes, we expand the latent model to include some unspecified unit-specific factor, i.e., *unit effects* (fixed not random), $\boldsymbol{\eta}$, to $\mathbf{y}^* = \rho\mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}$. Finally, in interdependent binary-outcome contexts, we are also likely to want estimates of the effects on the probability of some unit(s) i's choices/outcomes of counterfactual shocks to choices/outcomes of other unit(s) j. For instance, the effect on the probability Michigan enacts some policy of Illinois and/or Ohio enacting it. We denote this sort of counterfactual effect as $\frac{\partial p(\mathbf{y}_i)}{\partial \mathbf{y}_j}$ or $\frac{\Delta p(\mathbf{y}_i)}{\Delta \mathbf{y}_j}$. In contexts of spatially interdependent binary outcomes, none of these substantive *effects* is simple to estimate; indeed, the difficulty of calculation tends to increase with the centrality of their substantive interest.

To begin, we remind and emphasize that only in purely linear and additively separable models, like the canonical regression, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \varepsilon$, are *coefficients* and *effects* (of changes in x on y) identical. Even in models only implicitly non linear-additive, like spatial-autoregressive linear-regression, *effects* involve (often nonlinear) combinations of coefficients and variables, *via* the spatial-feedback multipliers in that case. Thus, even if we were content to confine our interpretation and presentation to the latent-variable arguments, y^*, to the probabilities of actual interest, $\hat{\mathbf{p}}$, we could not read *effects* directly from the usual table of *coefficients*. Instead, calculation, interpretation, and presentation of estimated effects on latent variables and their certainties would ensue as in the spatial linear-regression that we have discussed extensively elsewhere (see note 12). To review:

$$\mathbf{y}^* = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

$$= \begin{bmatrix} 1 & -\rho w_{1,2} & \cdots & \cdots & -\rho w_{1,n} \\ -\rho w_{2,1} & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 1 & -\rho w_{n\text{-}1,n} \\ -\rho w_{n,1} & \cdots & \cdots & -\rho w_{n,n\text{-}1} & 1 \end{bmatrix}^{-1} \left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \equiv \mathbf{S}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \quad (28).$$

Thus, denoting the i^{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 explanatory variable *k* in unit *i*, $\Delta x_{i,k}$, 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

$$\hat{V}\left(\hat{\mathbf{s}}_i\hat{\beta}_k\right) \approx \left[\frac{\partial \hat{\mathbf{s}}_i\hat{\beta}_k}{\partial \hat{\boldsymbol{\theta}}}\right] \hat{V}\left(\hat{\boldsymbol{\theta}}\right) \left[\frac{\partial \hat{\mathbf{s}}_i\hat{\beta}_k}{\partial \hat{\boldsymbol{\theta}}}\right]', \text{ where } \hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\rho} \\ \hat{\beta}_k \end{bmatrix} \text{ and } \left[\frac{\partial \hat{\mathbf{s}}_i\hat{\beta}_k}{\partial \hat{\boldsymbol{\theta}}}\right] = \begin{bmatrix} \frac{\partial \hat{\mathbf{s}}_i\hat{\beta}_k}{\partial \hat{\rho}} & \hat{\mathbf{s}}_i \end{bmatrix} \quad (29).$$

The vector $\frac{\partial \hat{\mathbf{s}}_i\hat{\beta}_k}{\partial \hat{\rho}}$ is the i^{th} column of $\beta_k \frac{\partial \hat{\mathbf{S}}}{\partial \hat{\rho}}$. Since $\mathbf{S}$ is an inverse matrix, the derivative in equation (29) is $\frac{\partial \hat{\mathbf{S}}}{\partial \hat{\rho}} = -\hat{\mathbf{S}}\frac{\partial \hat{\mathbf{S}}^{-1}}{\partial \hat{\rho}}\hat{\mathbf{S}} = -\hat{\mathbf{S}}\frac{\partial\left(\mathbf{I} - \hat{\rho}\mathbf{W}\right)}{\partial \hat{\rho}}\hat{\mathbf{S}} = -\hat{\mathbf{S}}(-\mathbf{W})\hat{\mathbf{S}} = \hat{\mathbf{S}}\mathbf{W}\hat{\mathbf{S}}$. Elsewhere, 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. We also showed in the spatiotemporal 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.

If we confine our attention to the latent variable, **y***, all of these techniques could apply in the spatial-probit context exactly as previously described, but, for most purposes, interpretation in terms of latent **y*** is unsatisfactory. Furthermore, several issues regarding the application of delta-method asymptotic linear-approximation increasingly trouble us, the intrinsic appeal of analytic solutions notwithstanding. First, deriving from 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 general how large, and they are in any event an approximation. Finally, using the approximately estimated standard errors to generate confidence intervals and hypothesis tests in the usual manners 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, though sample-size concerns may arise, perhaps especially regarding estimates involving $\hat{\rho}$, which is exactly where the spatial complications tend to arise. Given all this, we increasingly suspect that 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, including interpretation in terms of latent **y*** in the present binary-outcome case, simple simulation strategies—i.e., sampling coefficient estimates from multivariate normal with the estimated means and variance-covariance, calculating the quantities of interest from those draws, and then generating the desired indicators of certainty from the resulting sample—may be more effective.[34]

Even greater concerns arise in the spatial-probit context because the nonlinearity of the estimates of interest is more severe and asymptotic normality may be more distant. In fact, the (kernel of the) posterior joint-distribution of the parameters is not normal (as seen in (13), due to the $|\mathbf{I}_n - \rho\mathbf{W}|$

[34] Note: the average of the simulated quantities of interest and their standard deviation will not generally coincide exactly with the quantity of interest calculated at the ML parameter estimates and their (Delta Method approximated) standard errors. The former are $E(f(\hat{\boldsymbol{\theta}}))$ and $V(f(\hat{\boldsymbol{\theta}}))$, with $\hat{\boldsymbol{\theta}} \sim^{A} N(\boldsymbol{\theta}, -\boldsymbol{\mathcal{H}}^{-1})$ whereas the latter are $f(\hat{\boldsymbol{\theta}}_{ML})$ and $V(f(\hat{\boldsymbol{\theta}}_{ML}))$. By definition of *maximum likelihood* and its invariance property, the latter should correspond to the modal estimate of the quantity of interest and the asymptotic variance of the linear-approximation to that modal estimate, whereas the former is average and variance of the quantity of interest calculated at draws from normal sampling/poster distribution.

term), and, of the posterior conditional-distributions, only that of $\boldsymbol{\beta}$ is exactly normal. Thus, we suggest using the same MCMC (or RIS) processes that yielded the parameter estimates and their certainty estimates to estimate by simulation the quantities of interest and their certainties. To elaborate, recall that, after sufficient burn-in, LeSage's Gibbs-within-Metropolis-Hastings-and-Gibbs sampler generates draws from the poster joint-distribution of the parameters. The parameter estimates are the sample-means of these draws, and certainty estimates for those parameter-estimates are 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 using the (post-burn-in) sample of parameter vectors and supplying whatever counterfactual values of interest for whatever variables enter that quantity of interest. RIS-simulated likelihoods would support the same procedure, but a serious complication would yet remain in either case.

To understand the complications, consider our interests in levels or changes of $\hat{p}_i$ and $\hat{p}_j$'s or, most generally, $\hat{\mathbf{p}}$, the vector of probabilities of 1's in units *i* and *j* induced by hypothetical levels or changes in some $x_{i,k}$ or $x_{j,k}$, or, most generally, **X**. For instance, using (4), we could calculate the effects of some change in **X** on the estimated probability of an outcome of 1 in unit *i* as:

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

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 one unit *i* on the outcome in just that *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 standard probit, but also all the levels of all the $\mathbf{x}_j$ 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, as usual, 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 cross-unit effects on some *j* of changes in some *i* are calculated by the same formula, applying the desired $\Delta\mathbf{X}$ and changing the subscripts to refer to *j*th elements. This seems feasible, although the need to specify all of $\Delta\mathbf{X}$ for any counterfactual may be a bit daunting, but a far larger challenge is still looming.

Just as in the estimation problem, the $p(u_i < \left[(\mathbf{I}-\hat{\rho}\mathbf{W})^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}\right]_i / \hat{\sigma}_i)$ of interest here emerge from a multivariate cumulative-normal with means **0** and variance-covariance $[(\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})]^{-1}$. In the case of estimation, we sought to maximize a likelihood conditional on the data, i.e., **y** and **X**, which implied that we needed to evaluate one *n*-dimensional cumulative normal rather than multiply *n* unidimensional cumulative normals. To understand exactly how the same issue arises in estimating our counterfactual effects, consider the following spatial-probit model, simplified to a bivariate case:

$$\begin{aligned} y_1^* &= \rho w_{12} y_2^* + \beta_1 x_1 + \eta_1 + \varepsilon_1 \\ y_2^* &= \rho w_{21} y_1^* + \beta_2 x_2 + \eta_2 + \varepsilon_2 \end{aligned} \quad (31),$$

with η_i a fixed effect specific to y_i^* and $\varepsilon_i \sim N(0,1)$. The reduced form of the model is:

$$\begin{aligned} y_1^* &= \frac{\beta_1}{1\text{-}\rho^2 w_{12} w_{21}} x_1 + \frac{\rho w_{12}\beta_2}{1\text{-}\rho^2 w_{12} w_{21}} x_2 + \frac{1}{1\text{-}\rho^2 w_{12} w_{21}} \eta_1 + \frac{\rho w_{12}}{1\text{-}\rho^2 w_{12} w_{21}} \eta_2 \frac{1}{1\text{-}\rho^2 w_{12} w_{21}} \varepsilon_1 + \frac{\rho w_{12}}{1\text{-}\rho^2 w_{12} w_{21}} \varepsilon_2 \\ y_2^* &= \frac{\rho w_{21}\beta_1}{1\text{-}\rho^2 w_{12} w_{21}} x_1 + \frac{\beta_2}{1\text{-}\rho^2 w_{12} w_{21}} x_2 + \frac{\rho w_{21}}{1\text{-}\rho^2 w_{12} w_{21}} \eta_1 + \frac{1}{1\text{-}\rho^2 w_{12} w_{21}} \eta_2 + \frac{\rho w_{21}}{1\text{-}\rho^2 w_{12} w_{21}} \varepsilon_1 + \frac{1}{1\text{-}\rho^2 w_{12} w_{21}} \varepsilon_2 \end{aligned} \quad (32).$$

Latent y_i^* still links to the observed binary variable y_i through measurement equation (3), implying:

$$y_i = \begin{cases} 1 \text{ if } \varepsilon_i + \rho w_{ij}\varepsilon_j < \beta_i x_i + \rho w_{ij}\beta_j x_j + \eta_i + \rho w_{ij}\eta_j \\ 0 \text{ if } \varepsilon_i + \rho w_{ij}\varepsilon_j \geq \beta_i x_i + \rho w_{ij}\beta_j x_j + \eta_i + \rho w_{ij}\eta_j \end{cases} \quad (33)$$

The joint probability of any y_1 and y_2 is the product of a marginal and conditional probability; e.g.:

$$\Pr(y_1 = 1 \cap y_2 = 1) = \Pr(y_2 = 1 \mid y_1 = 1) \times \Pr(y_1 = 1) \quad (34)$$

For estimation purposes, given sample observations on y_1 and y_2, we apply the appropriate version of (34)'s right-hand side to calculate the joint likelihood for the pair of observations. One sees this product of one marginal and *n*-1 conditional distributions directly in the RIS estimator's equation (23), for example. If we wanted to calculate the marginal probability for y_1, $\Pr(y_1 = 1)$, i.e., the probability $y_1 = 1$ unconditional on ε_2, i.e., unconditional on the other unit, i.e., unconditional on y_2 or the $\Pr(y_2 = 1)$, we would integrate over ε_2. Then, because $\rho w_{21} \int_{-\infty}^{\infty} \varepsilon_2 f(\varepsilon_2)\, d\varepsilon_2 = 0$, the $\rho w_{21}\varepsilon_2$ term of (33) drops from the calculation, which means the simple univariate cumulative normal could be evaluated at the right-hand-side value to obtain $\Pr(y_1 = 1)$. That is, the marginal probability for y_1 depends on x_2 and η_2 (and x_1 and η_1, of course), but not on ε_2, the disturbance term from y_2^*.

However, the essence of interdependence would suggest that we are not particularly interested in these marginal probabilities, substantively. We want to consider counterfactual shocks to **X** or **η**, including the feedback represented in $\mathbf{Wy}^*$, which means conditional on ε_2. Calculating conditional probabilities like $\Pr(y_1 = 1 \mid y_2 = 1)$ is more complicated because this probability depends on the disturbance term from y_2^*. Since we are conditioning on $y_2 = 1$ (in this example), the possible error term from y_2^*, call it $\tilde{\varepsilon}_2$, is a random variable that comes from a truncated normal distribution with support over the range $\left[-\infty, \beta_1 x_1 + \rho w_{12}\beta_2 x_2 + \eta_1 + \rho w_{12}\eta_2\right]$. Since these distribution are truncated at the cutpoints for the conditional effects, the $\rho w_{21}\varepsilon_2$ term of (33) does *not* drop from the calculation, and we must compute the *n*-dimensional cumulative normal, just as in the original estimation stage.

More specifically, the marginal probabilities are

$$\begin{aligned}\Pr(y_1 = 1 \mid \mathbf{x}, \boldsymbol{\eta}) &= \Pr(\varepsilon_1 < \beta_1 x_1 + \rho w_{12}\beta_2 x_2 + \eta_1 + \rho w_{12}\eta_2) \\ &= \Phi\left[\beta_1 x_1 + \rho w_{12}\beta_2 x_2 + \eta_1 + \rho w_{12}\eta_2\right]\end{aligned} \quad (35),$$

and the conditional probabilities are

$$\begin{aligned}\Pr(y_2 = 1 \mid y_1 = 1; \mathbf{x}, \boldsymbol{\eta}) &= \Pr(\rho w_{21}\tilde{\varepsilon}_1 + \varepsilon_2 < \beta_2 x_2 + \rho w_{21}\beta_1 x_1 + \eta_2 + \rho w_{21}\eta_1) \\ &= \Phi\left[\beta_2 x_2 + \rho w_{21}\beta_1 x_1 + \eta_2 + \rho w_{21}\eta_1 - \rho w_{21}\tilde{\varepsilon}_1\right]\end{aligned} \quad (36).$$

Both of these cumulative distribution-functions are of the unidimensional standard-normal but, in the case of conditional (36), because the left-hand-side term in line one involving $\tilde{\varepsilon}_1$ has been transformed by partial differencing (and multiplication by the denominator in (32), which would be retained on the right-hand side also). To get an unbiased estimate of the conditional probability (36), we can take a draw from the truncated normal distribution for values of $\tilde{\varepsilon}_1$. Taking R draws, and averaging the probabilities, enhances the efficiency of this maximum simulated-likelihood estimator.

To reiterate, to calculate counterfactual effects of shocks to some units on probabilities of outcomes in some units, we could consider marginal or conditional probabilities. The marginal probabilities for $\mathbf{y}_i$ do not depend on $\mathbf{y}_j$, though they do depend on the full matrix $\mathbf{X}$ and vector $\boldsymbol{\eta}$. The conditional probabilities for $\mathbf{y}_i$, which have greater substantive meaning, do depend on $\mathbf{y}_j$, as well as the full matrix $\mathbf{X}$ and vector $\boldsymbol{\eta}$. These are more difficult, though we will show still possible, to calculate. In short, estimating effects in terms of probabilities of outcomes, i.e., in terms of the substantive quantity of interest, is as computationally burdensome as obtaining the estimates, for exactly the same reason. Then, to estimate the variance-covariance of these effect estimates, the entire effect-estimate procedure must be repeated many times.

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

$$\frac{\Delta\mathbf{p}}{\Delta\mathbf{X}} = \boldsymbol{\Phi}_n\left(\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}_1\boldsymbol{\beta}\right]\odot\left[\left\{\sigma_i^{-1}\right\}\right]\right)-\boldsymbol{\Phi}_n\left(\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}_0\boldsymbol{\beta}\right]\odot\left[\left\{\sigma_i^{-1}\right\}\right]\right) \qquad (37),$$

with $\Delta\mathbf{p}$ the $n\times1$ vector of $\Delta\left[p(y_i=1)\right]$ across all i; $\boldsymbol{\Phi}_n(\bullet)$ the cumulative-normal distributions, evaluated element-by-element at the values of its $n\times1$ vector argument, from the n-variate normal distribution with means zero and variance-covariance $[(\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})]^{-1}$; $\left[\left\{\sigma_i^{-1}\right\}\right]$ the $n\times1$ vector of the previously defined scalars σ_i^{-1}; and $\odot$ indicating element-by-element multiplication (i.e., Hadamard product). In principle, for given $\hat{\rho}$, these integrals could be calculated numerically using RIS or Gibbs sampling techniques, and certainty estimates for these effect estimates could then be obtained by repeating the process for many draws. However, calculating effects this way would take *c* times as long as the estimation procedure, with *c* the number of effect-estimates from which the estimated variance of the effect-estimate derives; computational intensity would be prohibitive.[35]

The derivative calculation for the marginal effect of X_i on the probability that y_i equals one avoids the multivariate integral. According to Beron and Vijverberg (2004) this effect is

$$\frac{d\,\Pr[y_i = 1 \mid X, \mathbf{W}]}{dX_i} = \phi\left(\Omega_{\alpha,ii}^{-1/2}[\Gamma_\alpha X\beta]_i\right)\Omega_{\alpha,ii}^{-1/2}\Gamma_\alpha\beta \qquad (38)$$

where ϕ is the univariate density function for the standard normal distribution, $\Gamma_\alpha = \frac{dy^*}{d(X_i\beta)}$, and $[\Gamma_\alpha X\beta]_i$ is the i^{th} element of the vector $\Gamma_\alpha X\beta$. Note that changes in X_j also effect y_i^*, so the quantity $\frac{d\,\Pr[y_i = 1 \mid X, \mathbf{W}]}{dX_j}$ is of interest too. Unfortunately, these formulas do not allow us to condition the effects of changes in *X* on the probability y_i equals one on the other y_j. In fact, Beron and Vijverberg argue that it is inappropriate to do so because the y_j are responding endogenously to the changes in *X*. This conclusion seems unnecessarily restrictive given that, after we estimate the model, we can easily sample from the distribution of disturbances using the reduced-form, generate *y*'s according to the measurement equation, and calculate exactly these conditional frequencies. In

[35] A simpler expedient may exist to evade integration of the *n*-dimensional multivariate-normal by drawing coefficients from the multivariate posterior or simulated-likelihood of σ, ρ, and $\boldsymbol{\beta}$. (If one wishes to include estimated inherent-uncertainty as well as estimation-uncertainty in these counterfactuals, then one should also draw $\boldsymbol{\varepsilon}$ from its independent-normal distributions, adding it to the $\mathbf{X}\hat{\boldsymbol{\beta}}$ in the next term.) Calculate $\hat{\mathbf{y}}_1^*$ and $\hat{\mathbf{y}}_0^*$ using $(\mathbf{I}_n - \rho\mathbf{W})^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}$ for some fixed $\mathbf{X}_1$ and $\mathbf{X}_0$, then simply apply (3) to convert those to vectors of ones and zeros, $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_0$. For a large number of draws, the averages of $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_0$ will be $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_0$, and $\hat{\mathbf{p}}_1$-$\hat{\mathbf{p}}_0$ will be the desired vector of estimated effects, and the variance-covariance of those differences will be the variance-covariance of those estimated effects. We are still working through some difficulties with implementing this conjecture, however.

other words, we can use the model to generate counterfactual values of the dependent variable for a given set of **X** and **W**, and then estimate the probabilities $\Pr[y_i = 1 | X, \mathbf{W}, y_j = 1]$ and $\Pr[y_i = 1 | X, \mathbf{W}, y_j = 0]$ using relatively frequencies. Once we have the parameter estimates and a specific counterfactual, the computation costs of proceeding in this way are relatively low. We give an example of this approach to counterfactual analysis in the illustration that follows.

In interpretation, as in estimation, the challenges raised by temporal auto-dependence in binary-choice models are analogous to those of spatial effects.

VI. Illustrations: Diffusion of Legislative Term-Limits among the U.S. States and the Great Powers' World War I Entry-Decisions

As a first illustration, we draw on the policy-diffusion and states-as-laboratories literatures in the study of U.S. politics (e.g., Volden 2006, Morehouse & Jewell 2004). Specifically, we consider a spatial-lag model of term-limit adoption in the states. The spatial lag allows us to consider whether states learn from or are otherwise influenced by their neighbors. Many studies examine the effects of term limits on the composition and functioning of legislatures or on individual legislators' behavior (e.g., Carey et al. 1998, Cain & Levin 1999), but why states might adopt limits in the first place has received much less scholarly attention, adding to the interest of the example. The dependent variable in our analysis indicates (0,1) whether a state has adopted term limits. From 1990 to 2000, 21 states adopted term limits.[36] The principal determinant of term-limit adoption is whether a state allows ballot initiatives or popular referenda (I&R) to consider state-level statutes and/or constitutional amendments (i.e., direct democracy). Simple reasoning likely underlies this strong relationship.

State legislators, particularly career politicians, are less likely than the public to want term limits, and direct democracy allows the electorate to bypass the legislature (Cain & Levin 1999). Indeed,

[36] In six of these states, term limits have either been overturned by state supreme courts or repealed by state legislatures. We code the dependent variable as 1 in these six cases.

only one state that does not allow such direct democratic processes, Louisiana, has term limits. Table 2 lists the 27 states (of the 48 contiguous) that either have term-limits or allow some form of direct democracy. The relationship is extremely strong: 21 of the 22 states without I&R do not have term limits, and 20 of the 26 states with I&R do have them.[37] The six states with I&R but without term limits are Illinois, Kentucky, Maryland, Mississippi, New Mexico, and North Dakota.

Table 2: Term Limits and Direct Democracy

State	Term-Limits (Year)	Repealed (Year)	Ballot Initiatives	Popular Referenda
Arizona	Yes (1992)	No	Yes	Yes
Arkansas	Yes (1992)	No	Yes	Yes
California	Yes (1990)	No	Yes	Yes
Colorado	Yes (1990)	No	Yes	Yes
Florida	Yes (1992)	No	Yes	No
Idaho	Yes (1994)	Yes (2002)*	Yes	Yes
Illinois	No	—	Yes	Yes
Kentucky	No	—	No	Yes
Louisiana	Yes (1995)	No	No	No
Maine	Yes (1993)	No	Yes	Yes
Maryland	No	—	No	Yes
Massachusetts	Yes (1994)	Yes (1997)	Yes	Yes
Michigan	Yes (1992)	No	Yes	Yes
Mississippi	No	—	Yes	No
Missouri	Yes (1992)	No	Yes	Yes
Montana	Yes (1992)	No	Yes	Yes
Nebraska	Yes (2000)	No	Yes	Yes
Nevada	Yes (1996)	No	Yes	Yes
New Mexico	No	—	No	Yes
North Dakota	No	—	Yes	Yes
Ohio	Yes (1992)	No	Yes	Yes
Oklahoma	Yes (1990)	No	Yes	Yes
Oregon	Yes (1992)	Yes (2002)	Yes	Yes
South Dakota	Yes (1992)	No	Yes	Yes
Utah	Yes (1994)	Yes (2003)*	Yes	Yes
Washington	Yes (1992)	Yes (1998)	Yes	Yes
Wyoming	Yes (1992)	Yes (2004)	Yes	Yes

Notes: In Idaho and Utah, term limits were repealed by their state legislatures (*). Term-limits were overturned by state supreme courts in MA, OR, WA, and WY. Our sample only includes the contiguous 48 states. Alaska allows both ballot initiatives and popular referenda, but has never adopted term limits. Hawaii allows neither ballot initiatives nor popular referenda and has never adopted term limits.

Our empirical models include two other explanatory variables. The first indicates (0,1) whether the state voted for Clinton in the 1992 presidential election. Some argue that Democrats, because of

[37] We can easily reject the null hypothesis that these two variables are independent $(\chi^2_{(1)} = 25.367,\ \text{p-value=.000})$, and Kendall's τ_b, a (-1…+1) correlation-like measure of association, is .727 with asymptotic standard error of .091.

their relatively positive view of government and related support for state intervention, are more accepting of political careerism and therefore more likely to oppose term limits.[38] On the other hand, the populist tendencies of some Democrats may lead them to support term limits as a way to promote citizen participation in government. The second is the average state-level tax effort (state revenue as percent of state GDP) during the 1980s. High state taxes may indicate public support for political centralization and a strong professionalized legislature, or, alternatively, high taxes may provide impetus for an overburdened electorate to "throw the bums out" using term limits.

Table 3: Adoption of Term Limits for State Legislators, Estimation Results

	Probit-ML	Probit-MCMC	Spatial-Lag Probit	Spatial-Error Probit	Probit-RIS
Constant	-.539 (2.579)	-.909 (1.814)	-.598 (3.080)	-.411 (2.533)	.313 (2.223)
I&R	2.320*** (.581)	1.806*** (.481)	3.257*** (.917)	2.650*** (.627)	2.336*** (.620)
Clinton	.273 (.542)	.131 (.476)	.147 (.769)	.056 (.606)	.304 (.518)
Tax Effort	-.178 (.273)	-.068 (.183)	-.176 (.321)	-.146 (.262)	-.214 (.247)
Spatial lag or error-lag	.926 (.801)	.634 (.687)	.144 (.207)	.018 (.279)	.416** (.194)
Pseudo-R^2	.480	.458	.833	.803	—
Log-Likelihood	-17.093	—	—	—	-16.261
Observations	48	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 uninformative 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. ***p-value <.01, **p-value<.05, *p-value <.10. The grayed columns use the incorrect univariate TMVN in the final sampler step, which obviously we intend to fix for next drafts...

Table 3 reports estimates of probit models with spatial-lag regressors (or, in one column, spatial error-dependence) by standard maximum-likelihood (ML) methods that erroneously assume spatial lags exogenous, by Bayesian MCMC methods but maintaining the same erroneous assumption, and by true spatial-lag probit (or spatial-error probit) using the Bayesian MCMC and the frequentist RIS methods described in Section III. We use a standardized binary contiguity-weights matrix, **W**, which

[38] There is individual-level evidence for a relationship between Republican partisanship and support for term limits in several states (see Cain and Levin 1999 for a discussion).

codes $w_{ij} = (1,0)$ for whether states i and j border and then row-standardizes[39] the resulting matrix by dividing each element by its row's sum. This gives $\mathbf{Wy}_i$ as the unweighted average of the outcome in i's bordering states—i.e., the share of bordering states that have term limits—or $\mathbf{Wy}_i$* as the unweighted average propensity of i's neighbors to adopt term limits.

The first two columns report models estimated (wrongly) assuming the spatial lags exogenous. The first-column model applies standard-probit ML techniques. The parentheses contain the standard estimated standard errors, with the hypothesis tests assuming the test-statistics asymptotic-normally distributed. The next two columns' models are estimated using MCMC methods with diffuse zero-mean priors, including an uninformative uniform(-1,1) prior on ρ. The reported coefficient-estimates are means of posterior distributions using 10,000 cycles of the sampler after a 1000-cycle burn-in. The parentheses report sample standard-deviations of the posterior distributions, and p-values also emerge directly from the posterior (without calculating or assuming anything about test statistics).

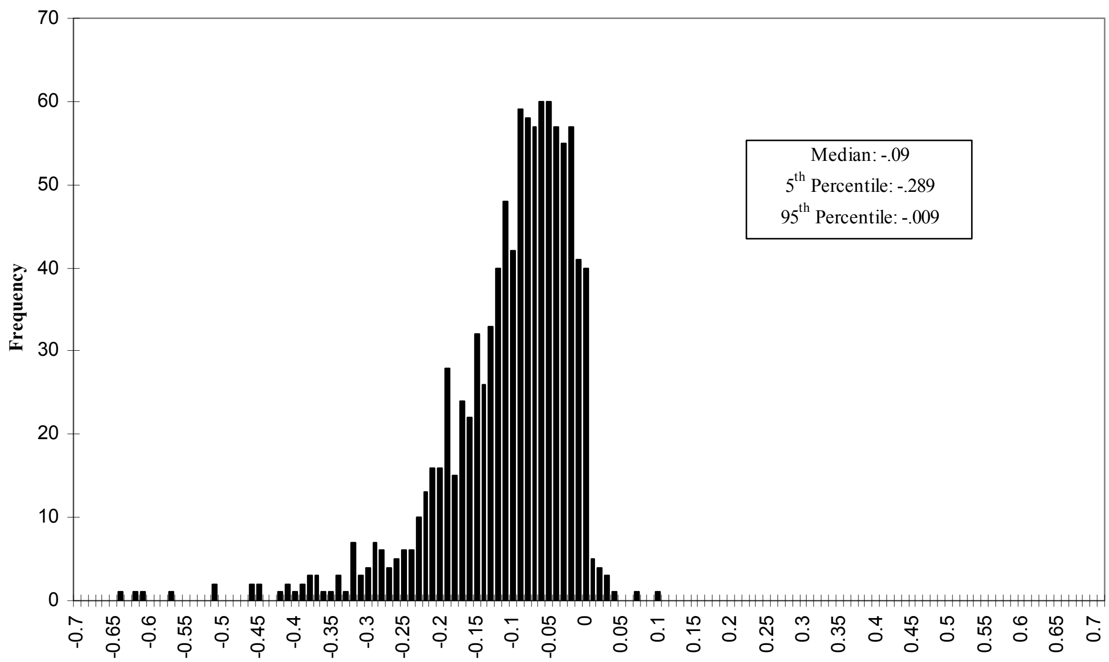
The results in columns one and two are similar, which is not surprising given our use of diffuse priors in column two, which uses the probit-MCMC estimator for a model that, as with probit-ML, incorrectly treats the spatial lag as exogenous (i.e., just as any other right-hand-side variable). This likelihood 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. The result here seems an overestimation from 50% to over 100% of interdependence-strength and a more than three- to over-four-fold overestimation of the uncertainty, judging the first and second columns relative to the fifth, RIS, column. Columns three and four reports the Bayesian-Gibbs spatial-lag and spatial-error probit estimates using the code containing the incorrectly univariate final-step draws, so we will not discuss them here other than to note that the mistaken treatment of the problem as univariate

[39] Row-standardization is standard in spatial econometrics, but it is not necessarily substantively neutral (see, e.g., Pluemper & Neumayer 2008).

apparently (again, judged compared to the correctly specified fifth-column RIS estimates) induced serious deflation bias in the estimated strength of contagion.

Finally, the correctly specified RIS estimates finds $\hat{\rho}$ to be clearly statistically significantly present, and much stronger than the erroneously simpler Bayesian-MCMC estimates, although not nearly so strong as the naïve estimators that erroneously treat the spatial-lag as endogenous and so suffer inflationary simultaneity biases (that here dominate the attenuation biases from misspecifying the spatial lag in binary-outcome rather than latent-propensity terms).

Figure 1: *Parametric Bootstrap:* Counterfactual Effect on Probability Washington State Adopts Term Limits from 2-Unit Decreases (+1 to -1) in Oregon & Idaho's Propensities to Adopt Limits



In Figure 1, we report the results from a certain kind of hypothetical effect-estimate using the RIS parameter-estimates and a parametric bootstrap method similar to that discussed in Section V. We focus on Washington, Oregon, and Idaho, three states that adopted term limits in the early 1990's. Specifically, we ask what happens to the probability Washington adopts term limits when we manipulate the underlying propensity of its neighbors—Oregon and Idaho—to adopt term limits. To calculate this, we start by taking 1,000 draws from a multivariate normal distribution with a mean equal to the vector of parameter estimates and a variance-covariance matrix equal to the estimated information matrix. For each draw, we calculate the probability that each state will adopt term limits when the y_i^*'s for Oregon and Idaho are set 1-unit above their estimated values, using the parameter draws and the observed values for the independent variables, and the y_i^*'s for the other forty-six states are held fixed at their estimated values (i.e., not manipulated). We then calculate these same probabilities when the y_i^*'s for Oregon and Idaho are set 1-unit below their estimated values and take the difference between the first and second vector of probabilities for each set of parameter draws. We do not draw stochastic components in these calculations. Unfortunately, this sort of counterfactual gives the substantively less-fully defensible marginal-distribution effects discussed in the previous section, and leaves half the discussion unfortunately in the realm of the latent variables.

Nonetheless, Figure 1, giving the empirical distribution of the 1000 changes in Washington's marginal probability of adopting term limits given our counterfactual changes to Oregon and Idaho offers some substantive grasp on the estimation-results' meaning. Figure 1 clearly demonstrates that, for the overwhelming majority of trials, the marginal effects of decreasing the propensities that Oregon and Idaho will adopt term limits is to decrease the probability that Washington will adopt term limits. The median effect from our trials is -.09, and a 90% confidence interval runs from -.289 to -.009. Of course, it is possible to conduct similar experiments for any cluster of neighboring states (i.e., change the subjects in our experiment) and to counterfactually manipulate the presence or absence of direct democracy (i.e., change the treatment). Our results suggest that states' decisions to adopt term limits are appreciably influenced by the experiences of their neighbors.

As a second illustration, we model the decisions of states to enter WWI using spatial-lag probit. Treating these participation decisions as independently driven purely by domestic and international structural factors such as regime type, trade exposure, and relative military capabilities seems highly inappropriate. Ultimately, each state's decision surely heavily influenced others' entry decisions, and any empirical analysis should consider this interdependence. Our models incorporate two bases for such interdependence: contiguity and rivalry. I.e., we posit the entry decisions of neighbors and of political rivals influence states' own participation choices. Positive interdependence indicating the contagion of conflict along these bases may seem likely, but negative interdependence, suggesting free-riding behavior, is plausible as well. Consider rivals: a participating state's rivals may stay out of the conflict hoping for a favorable outcome—namely, that the participating rival will lose the war and suffer a decrease in power, at no cost to the free-riding rival standing aside.

Table 4: Great Power WWI-Entry Decisions, Estimation Results

	(1)	(2)	(3)	(4)	(5)
Constant	-.525	-1.33*	-1.65	-1.73**	-1.85
	(.533)	(.681)	(.778)	(.749)	(.888)
Contiguity Spatial Lag	.429**	.207	.110		
	(.189)	(.206)	(.223)		
Rivalry Spatial Lag				.723***	.659
				(.274)	(.402)
National Capabilities	18.57***	20.62**	30.02	18.39**	40.19
	(7.83)	(8.80)	(11.24)	(8.29)	(20.36)
Democracy	.039	-.024	-.023	-.068	-.011
	(.036)	(.047)	(.050)	(.056)	(.062)
Trade	-.072	-.083	-.082	-.101	-.077
	(.116)	(.121)	(.125)	(.138)	(.129)
Europe		1.61**	1.90	2.07***	2.00
		(.639)	(.690)	(.743)	(.868)
Estimator	RIS	RIS	Bayesian	RIS	Bayesian
Observations	44	44	44	44	44
Log-Likelihood	-19.03	-15.32		-14.26	

Notes: The dependent variable reflects participation in WWI (0=No, 1=Yes). Of the 44 sample countries, 15 enter the War. Spatial-weights matrices are row-standardized. *National Capabilities* are the COW CINC index scores. *Democracy* is Polity measures of regime type. *Trade* is the value of total trade in current US dollars (Source: Barbieri 2002). Europe is an indicator equal to 1 for countries located on the continent, but including the U.K. In the RIS columns, parentheses contain standard-error estimates. In the Bayesian columns, parentheses contain the standard deviations of the posterior distributions. ***significant at 1%; **significant at 5%; *significant at 10%.

Table 4 gives the estimation results. Controls include national capabilities, democracy, trade (described in Table 3 notes), and, with the exception of the column 1 model, a European indicator.

Consistent with theories of war as a contagion, we find statistically significant positive interdependence in the first contiguity probit model. Not surprisingly, when we add the Europe dummy variable (column 2), the contiguity-based spatial lag becomes statistically insignificant. Geography was an important factor in the war-joining decision of states, and the dummy variable provides a simple way to incorporate this influence.[40] Therefore, we include this dummy variable as a proxy for geography in the rivalry regression as well.[41] These estimates, provided in column 4, suggest that countries were also influence by their politico-military rivals.[42] For comparison purposes, we provide the Bayesian MCMC estimates in columns 3 and 5, although we prefer the better RMSE performance of the RIS estimator evidenced in Table 1 for sample conditions like these.

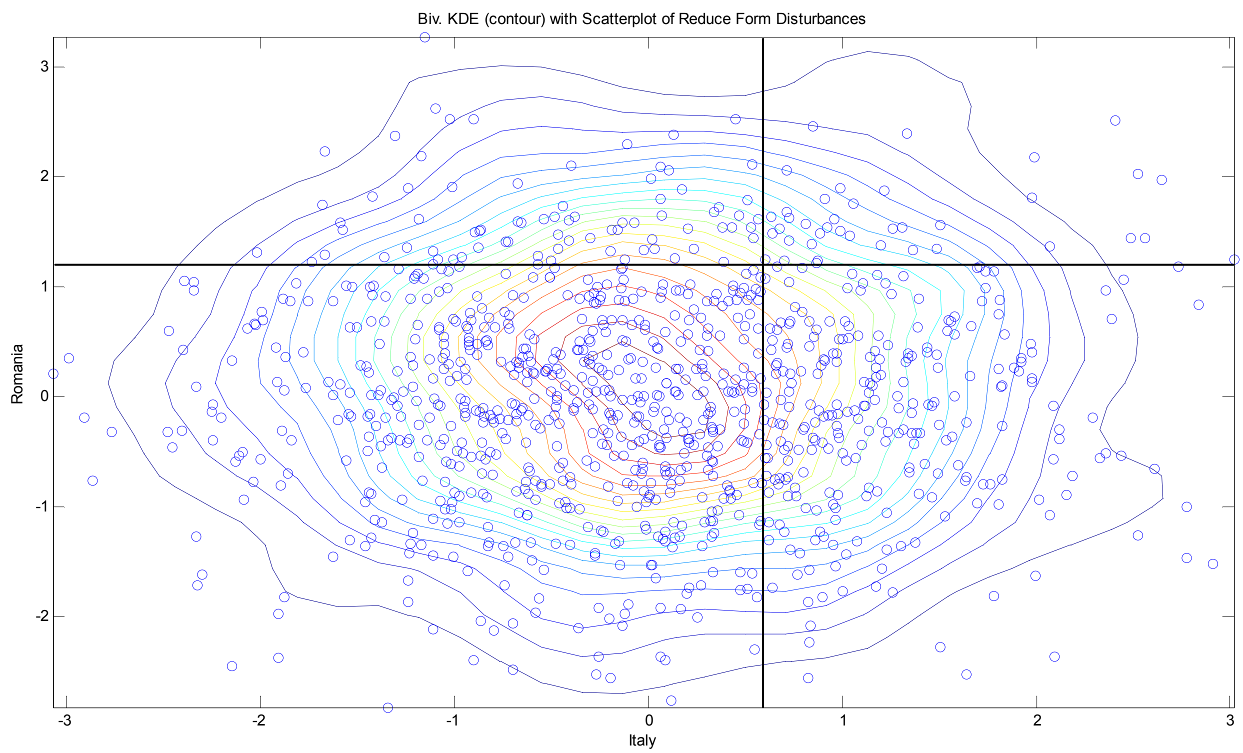
There are many interesting counterfactual questions related to the interdependence of states' WWI participation decisions. How did America's decision to participate affect the participation of others? To what extent did Italy's decision to enter the war in mid-1915 affect the probability that Bulgaria and Romania would be drawn into the conflict before the fighting stopped? These are examples of the sort of *conditional* effects of hypotheticals that we discussed in Section V. To illustrate how we can in fact answer such substantively more-meaningful conditional probabilities, we answer the latter—the effect of Italy's entry on Romania's probability of entering—using the spatial probit estimates from column 1 of Table 4.

In terms of the model, the question becomes, given that Italy's reduced-form disturbance is above or below the negative of its reduced-form cutpoint, what is the probability that Romania's reduced-form disturbance will be above or below the negative of its reduced-form cutpoint? To

---

[40] Of course, this dummy variable specification is unsatisfactory because it tells us nothing about the way in which geography matters. There is simply too little variation in European vs. Non-European entry or the patterns therein to distinguish contiguity contagion from "something (additive) about Europe" as explanations for World War I's location.
[41] This is much easier than the alternative of estimating a probit with multiple spatial lags, although the latter is doable; see Hays et al. (2010) for a linear-regression example and Lacombe (2010) for a multiple-**W** spatial-probit application.
[42] At the onset of WWI, there were fourteen rivalries (see Diehl and Goertz 2001): UK-Germany, UK-Russia, France-Germany, France-Turkey, Germany-Norway, Austria-Hungary-Italy, Austria-Hungary-Serbia, Serbia-Bulgaria, Russia-Bulgaria, Russia-Turkey, Russia-Japan, Turkey-Italy, Turkey-Greece, and Turkey-Bulgaria.

answer this question, we sample from the reduced form disturbances. More specifically, we draw 1,000 times from a $N(0,1)$ for each of the 44 states in the sample. This gives a 44×1000 matrix of i.i.d. standard-normal disturbances. Then we pre-multiply this disturbance matrix by the 44×44 spatial multiplier, which gives $\mathbf{U}=(\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$. Since the counterfactual question involves the participation of Italy and Romania specifically, we take just a bivariate slice of the resulting 44-dimensional multivariate distribution, although the procedure being described here produces the entire vector of 44 states' responses to the hypothetical.



The vector of reduced-form cutpoints is calculated as $(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}$. A country participates if its reduced-form disturbance is greater than the negative value of its reduced-form cutpoint. The bivariate pair of these simulated reduced-form disturbances corresponding to Italy and Romania are plotted in Figure 1. Given their covariates, the reduced form cutpoints for Italy and Romania are -.634 and -1.126 respectively, so these countries join when their reduced form disturbances are greater than .634 and 1.126, indicated by the lines in the figure. In the simulations, Romania joins the war 12.3% of the time when Italy stays out—i.e., 12.3% of the points to the left of Italy's cutpoint lie above Romania's cutpoint—and 15.6% of the time when Italy enters—15.5% or the points right of Italy's line lie above Romania's. Thus, the model estimates suggest that Italy's entry to the War is associated with a 3.3% increase in Romania's probability of entering. Since Italy and Romania did not share a border, these are second-order effects.

(Repeating this simulation exercise a large number of times for draws of parameter estimates and of reduced-form residuals would provide certainty estimates for these effects, but we have not yet done this for this draft.)

(Likewise, in proof of concept: Cheibub & Hays (2009) offer preliminary estimates by the methods proposed here of a bivariate system of probit equations, with the jointly endogenous binary dependent-variables being multiparty elections and intrastate violence / civil conflict, and with each dependent variable having temporal auto-dependence. But we have not yet incorporated these demonstrations into this draft.)

VII. Conclusion

Spatial/spatiotemporal (inter)dependence is substantively and theoretically ubiquitous and important across social-science binary-outcomes. Standard ML-estimation of binary-outcome models in the presence of spatial interdependence and/or temporal auto-dependence are badly misspecified if that (inter)dependence is ignored, but they are also misspecified (we suspect less badly, but we have not explored that systematically as yet), if that interdependence is reflected by inclusion of endogenous spatial lags and/or temporally lagged outcomes, as opposed to lags of latent variables, as explanator(s). Spatial-, temporal-, or spatiotemporal-lag probit models are difficult and highly computationally demanding, but not impossible, to estimate with appropriate estimators. We expect future work to demonstrate more fully and clearly the conditions under which expending such effort in estimation merits the gains. It is also possible, and by simulation feasible, to calculate and present properly the estimated spatial, temporal, and/or spatiotemporal *effects* (as opposed to merely probit *coefficients*) on binary outcomes, along with their associated estimates of certainty. Doing so will be of great substantive advantage under any conditions, regardless of whether the sophisticated spatial-probit estimators offer much gain (in terms of bias, efficiency, or standard-error accuracy) from standard-probit estimators with spatial lags pretended to be exogenous.

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.

Beck, N., Katz, J., Tucker, R. 1998. "Taking time seriously: Time-series--cross-section analysis with a binary dependent variables,"*American Journal of Political Science* 42(4):1260-88.

Beck, N., Epstein, D., Jackman, S., O'Halloran, S. 2001. "Alternative Models of Dynamics in Binary Time-Series–Cross-Section Models: The Example of State Failure." Summer Meeting of the Society for Political Methodology (Emory).

Beck, N., Gleditsch, K., Beardsley, K. 2006. "Space is More than Geography: Using Spatial Econometrics in the Study of Political Economy," *International Studies Quarterly* 50(1):27-44.

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.

Cain, B.E., Levin, M.A. 1999. "Term Limits." *Annual Review of Political Science* 2:163-88.

Carey, J.M, Niemi, R.G., Powell, L.W. 1998. "The Effects of Term Limits on State Legislatures." *Legislative Studies Quarterly* 23(2): 271-300.

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

Chamberlain, G. 1993. "Feedback in Panel Data Models," *Harvard Institute of Economic Research Working Paper* #1656.

Cheibub, J., Hays, J. 2009. "Can Elections Be Bad for Democracy?" Presented at the American Political Science Association Annual Meeting.

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. 2003a. "Diagnosing, Modeling, Interpreting, and Leveraging Spatial Relationships in Time -Series-Cross-Section Data," NSF Proposal #0318045. (http://www.umich.edu/~franzese/SpatialCorrelation.NSF.Proposal.pdf).

Franzese, R., Hays, J. 2003b. "Modeling Spatial Relationships in International and Comparative Political Economy: An Application to Globalization and Capital Taxation in Developed Democracies," presented at the Annual Meeting of Midwest Political Science Association. (http://www.umich.edu/~franzese/SpatialCorrelation.CapitalTaxationPaper.032803.pdf)

Franzese, R., Hays, J. 2004a. "Empirical Modeling Strategies for Spatial Interdependence: Omitted-Variable vs. Simultaneity Biases," presented at the 21st Summer Meeting of the Society for Political Methodology. (http://www.umich.edu/~franzese/FranzeseHays.PolMeth.2004.pdf)

Franzese, R., Hays, J. 2004b. "Modeling International Diffusion: Inferential Benefits and Methodological Challenges, with an Application to International Tax Competition," Wissenschaftszentrum-Berlin SP II 2004 – 12. (http://www.umich.edu/~franzese/FranzeseHays.WZB.SPII-12.2004.pdf)

Franzese, R., Hays, J. 2005a. "Modeling Spatial Interdependence in Comparative and International Political Economy with an Application to Capital Taxation," presented at the Annual Meeting of the Midwest Political Science Association. (http://www.umich.edu/~franzese/FranzeseHays_MWPSA05.pdf)

Franzese, R., Hays, J. 2005b. "Spatial Econometric Modeling, with Application to Employment Spillovers and Active-Labor-Market Policies in the European Union," presented at Groningen University, workshop on 'Partisan Politics, Political Autonomy, and Policy Harmonization across Europe'. (http://www.umich.edu/~franzese/FranzeseHays.SpatialEcon.EmploymentSpillovers.ALP.pdf)

Franzese, R., Hays, J. 2006a. "Calculating Spatio-Temporal Dynamics and Effects," presented at the General Conference of the European Consortium of Political Research, Pisa. (http://www.umich.edu/~franzese/FranzeseHays_April06.pdf)

Franzese, R., Hays, J. 2006b. "Spatiotemporal Models for Political-Science Panel and Time-Series-Cross-Section Data," presented at the 23rd Annual Summer Meetings of the Society for Political Methodology. (http://www.umich.edu/~franzese/FranzeseHays.S.ST.EconometricsForPS.PolMeth06.pdf)

Franzese, R., Hays, J. 2006c. "Strategic Interaction among EU Governments in Active-Labor-Market Policymaking: Subsidiarity and Policy Coordination under the European Employment Strategy," *European Union Politics* 7(2):167-89.

Franzese, R., Hays, J. 2007a. "Correlation in European Union Labor-Market Policies: Interdependence or Common Stimuli?" Presented at the Annual Meetings of the American Political Science Association. (http://www.umich.edu/~franzese/Franzese_Hays_APSA07.pdf)

Franzese, R., Hays, J. 2007b. "Empirical Models of International Capital-Tax Competition," in G. Gregoriou, C. Read, eds., *International Taxation Handbook*, Amsterdam: Elsevier, pp. 43-72.

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

Franzese, R., Hays, J. 2008a. "Contagion, Common Exposure, and Selection: Empirical Modeling of the Theories and Substance of Interdependence in Political Science," *Concepts & Methods: Newsletter of the International Political Science Association* 4(2):3-9. (http://www.umich.edu/~franzese/FranzeseHays.Interdependence.IPSA.pdf)

Franzese, R., Hays, J. 2008b. "Empirical Models of Spatial Interdependence," in J. Box-Steffensmeier, H. Brady, D. Collier, eds., *Oxford Handbook of Political Methodology*, Oxford: Oxford UP, pp. 570-604.

Franzese, R., Hays, J. 2008c. "Interdependence in Comparative Politics: Substance, Theory, Empirics, Substance," *Comparative Political Studies* 41(4/5):742-80.

Franzese, R., Hays, J. 2008d. "Spatial Interdependence in Comparative & International Political Economy," presented at Paris 13 (Université Paris), Axe 5: PSE & University of Essex. (http://www.umich.edu/~franzese/FranzeseHays.SpatialInterdepCIPE.pdf)

Franzese, R., Hays, J. 2009a. "A Comparison of the Small-Sample Properties of Several Estimators for Spatial-Lag Count-Models," presented at the Summer Meeting of the Political Methodology Society. (http://www.umich.edu/~franzese/HaysFranzese.ComparisonEstimatorsSpatialCountModels.PolMeth2009.pdf)

Franzese, R., Hays, J. 2009b. "Empirical Modeling of Spatial Interdependence in Time-Series Cross-Sections," in S. Pickel, G. Pickel, H-J. Lauth, D. Jahn, eds., *Methoden der vergleichenden Politik- und Sozialwissenschaft: Neue Entwicklungen und Anwendungen*. (*Methods of Comparative Political and Social Science: New Developments & Applications*.) *Wiesbaden: Westdeutscher Verlag*, pp. 233-62.

Franzese, R., Hays, J. 2009c. *Spatial Econometric Models of Interdependence*. Book manuscript. (http://www.umich.edu/~franzese/FranzeseHays.SpatialEconometricModelsInterdependence.ProspectusOutlineToC.Aug2008.pdf)

Franzese, R., Hays, J. 2009d. "The Spatial Probit Model of Interdependent Binary Outcomes: Estimation, Interpretation, and Presentation," presented at the Annual Meetings of the Public Choice Society. (http://www.umich.edu/~franzese/SpatialProbit.PubChoice09.pdf)

Franzese, R., Hays, J., Kachi, A. 2009. "An m-STAR Model of Dynamic, Endogenous Interdependence – a.k.a. Network-Behavior Coevolution – in the Social Sciences," presented at the 3rd World Conference of the Spatial Econometrics Association. (http://www.umich.edu/~franzese/HaysKachiFranzese.mSTARasNetworkCoevolution.SEA2009.pdf)

Franzese, R., Hays, J., Kachi, A. 2010. "Modeling History Dependence in Network-Behavior Coevolution," presented the 27th Summer Meeting of the Society for Political Methodology. (http://polmeth.wustl.edu/mediaDetail.php?docId=1230)

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.

Graham, E., Shipan, C.R., Volden, C. 2008. "The Diffusion of Policy Diffusion Research," unpublished, The Ohio State University and The University of Michigan.

Hays, J.C. 2009. "Bucking the System: Using Simulation Methods to Estimate and Analyze Systems of Equations with Qualitative and Limited Dependent Variables," SLAMM (St. Louis Area Methods Meetings), Washington University in St. Louis. (http://www.umich.edu/~franzese/ICPSR/Hays.SpatialSystemsQualDeps.SLAMM09.pdf)

Hays, J., Colaresi, M. 2009. "Spatial and Temporal Interdependence," Prepared for the ISA Compendium

Project. Scientific Study of International Processes Section.

Hays, J., Ehrlich, S., Peinhardt, C. 2005. "Government Spending and Public Support for Trade in the OECD: An Empirical Test of the Embedded Liberalism Thesis," *International Organization* 59(2):473-94.

Hays, J., Kachi, A. 2009. "Interdependent Duration Models in Political Science," American Political Science Association Annual Meetings. (http://www.umich.edu/~franzese/ICPSR/HaysKachi.InterdependentDurationModels.APSA2009.pdf)

Hays, J., Kachi, A., Franzese, R. 2010. "A Spatial Model Incorporating Dynamic, Endogenous Network Interdependence: A Political Science Application," *Statistical Methodology* 7(3):406-28.

Heckman, J.J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica* 46:931-59.

Hoff, P., Raftery, A., Handcock, M. 2002. "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association* 97(460):1090-9.

Hoff, P., Ward, M. 2004. "Modeling Dependencies in International Relations Networks." *Political Analysis* 12(2):160-75.

Hoff, P. Westveld, A. 2007. "A Bayesian Mixed Effects Model for Longitudinal Social Network Data." Unpublished, University of Washington.

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.

Honore, B.E., Kyriazidou, E. 2000. "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica* 68(4):839-74.

Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M., Morris, M. 2008. "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software* (URL http://www.jstatsoft.org/v24/i03/).

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.

Morehouse, S.M., Jewell, M.E. 2004. "States As Laboratories: A Reprise." *Annual Review of Political Science* 7: 177-203.

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:

Pluemper, T., Neumayer, E. 2008. "Model Specification in the Analysis of Spatial Dependence," unpublished.

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

Robins, G, Pattison, P., Kalish, Y., Lusher, D. 2007. "An introduction to exponential random graph (p*) models for social networks," *Social Networks* 29(2):173-91.

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.

Snijders, T.A.B. 2005. "Models for Longitudinal Network Data." In P. Carrington, J. Scott and S. Wasserman, eds., *Models and Methods in Social Network Analysis*. New York: Cambridge.

Snijders, T.A.B., Steglich, C., Schweinberger, M. 2007. "Modeling the Co-Evolution of Networks and Behavior," in K. van Montfort, H. Oud, A. Satorra, eds., *Longitudinal Models in the Behavioral and Related Sciences* (Mahwah NJ: Lawrence Erlbaum), pp.41-71.

Volden, Craig. 2006. "States as Policy Laboratories: Emulating Success in the Children's Health Insurance Program." *American Journal of Political Science* 50(2): 294–312.