Spatial- and Spatiotemporal-Autoregressive Probit Models of Interdependent Binary Outcomes

Robert J. Franzese, Jr., Jude C. Hays and Scott J. Cook

Political Science Research and Methods / Volume 4 / Issue 01 / January 2016, pp 151 - 173
DOI: 10.1017/psrm.2015.14, Published online: 07 September 2015

Political Science Research and Methods Vol 4, No. 1, 151–173 January 2016

doi:10.1017/psrm.2015.14

Spatial- and Spatiotemporal-Autoregressive Probit Models of Interdependent Binary Outcomes*

ROBERT J. FRANZESE, JR., JUDE C. HAYS AND SCOTT J. COOK

Spatial/spatiotemporal interdependence—that is, that outcomes, actions or choices of some unit-times depend on those of other unit-times—is substantively important and empirically ubiquitous in binary outcomes of interest across the social sciences. Estimating and interpreting binary-outcome models that incorporate such spatial/spatiotemporal dynamics directly is difficult and rarely attempted, however. This article explains the inferential challenges posed by spatiotemporal interdependence in binary-outcome models and recent advances in their estimation. Monte Carlo simulations compare the performance of one of these consistent and asymptotically efficient methods (maximum simulated likelihood, using recursive importance sampling) to estimation strategies naïve about (inter-) dependence. Finally, it shows how to calculate, in terms of probabilities of outcomes, the estimated spatial/spatiotemporal effects of (and response paths to) hypotheticals of substantive interest. It illustrates with an application to civil war in Sub-Saharan Africa.

SPATIAL, TEMPORAL AND SPATIOTEMPORAL INTERDEPENDENCE 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 contexts, and widely across the political and social sciences, the outcomes in some units depend substantively and theoretically importantly on those of other units.[1] Scholars argue, for example, that whether and for whom citizens vote depend on whether and how their neighbors or social networks vote; that each legislator's vote depends on how s/he expects or observes others to vote; that governments' policy choices depend on others' policies via competition and learning; that states' entry to and involvement in external wars, international organizations, and treaties are heavily conditioned by whether and which ones other states join; and that regime change at home or, as in our own empirical application below, nations' internal wars are often spurred by example, fomentation or otherwise from abroad. Nevertheless, in most research areas, *spatial* (that is, *cross-unit*) interdependence in discrete outcomes receives very little direct empirical attention.[2] Even when

* Robert J. Franzese, Jr. is a Professor of Political Science, University of Michigan, Ann Arbor, MI 48109 (franzese@umich.edu). Jude C. Hays is an Associate Professor of Political Science, University of Pittsburgh, Pittsburgh, PA 15260 (jch61@pitt.edu). Scott J. Cook is an Assistant Professor of Political Science, Texas A&M University, College Station, TX 77840 (sjcook@tamu.edu). Though many more provided helpful feedback at various stages of this project, we are particularly indebted to Patrick Brandt, Scott McClurg, Eric Neumayer, Thomas Plümper, Lena Schaffer, Curt Signorino, Vera Troeger and the participants of the 2013 Spatial Models of Politics conference at Texas A&M. Thanks are also in order to the reviewers and editor for their useful comments and suggestions. All remaining errors are our own. To view supplementary material for this article, please visit http://dx.doi.org/10.1017/psrm.2015.14.

[1] Appendix I is an extensive topically organized bibliography of interdependence studies across the social sciences.

[2] Perhaps the most notable exceptions are work on (1) diffusion of state policy/institutions and (2) micro-behavioral work on the social determinants of respondent behaviors and opinions (see Franzese, Hays and Cook 2012 and Appendix I).

researchers have attempted to explicitly account interdependence empirically, the inappropriate estimation strategies usually employed have rendered their estimates unreliable. Unlike cross-unit interdependence, the over-time dependence of binary outcomes has received considerable attention—and, consequently, has seen more-effective indirect empirical-modeling strategies developed—yet modeling temporal *auto*-dependence directly raises these same methodological challenges. *Auto*-dependence in binary outcomes, across space and/or over time, requires strategies for modeling simultaneous spatial and/or temporal lags of the *latent* outcomes (that is, the underlying propensities toward the binary outcome), which is methodologically challenging to do. This article discusses a simulation-based strategy for estimating these auto-dependence parameters directly in spatiotemporally autoregressive binary-outcome models *and*, we believe for the first time, for calculating substantively meaningful conditional spatiotemporal effects and response paths in terms of outcome probabilities.

While efforts to model spatial interdependence directly in binary outcomes have recently expanded in some contexts, the strategies employed often fail to account for the endogeneity induced by including spatial lags. Using our example application from below to illustrate: Senegal is on Guinea-Bissau's right-hand side, the former affecting the probability of internal conflict in the latter, but, simultaneously, Guinea-Bissau is on Senegal's right-hand side, affecting its probability of civil war. Moreover, placing the actual binary outcomes of other units (or weighted sums or averages of actual outcomes) *simultaneously* on the right-hand side is not algebraically consistent (Heckman 1978). Such simultaneity can only logically operate through the *latent* variables or errors, which raises additional econometric and computational challenges. Diffusion researchers often time lag these spatial lags in an attempt to bypass these issues. While this *can* suffice to evade the simultaneity bias (Beck, Gleditsch and Beardsley 2006), its efficacy in doing so rests on the assumptions (we believe frequently untenable) that: (1) the interdependence does not occur instantaneously, that is, within an observational period, (2) the actual periodicity and lag structure match that of the empirical observations and specification and (3) the empirical model of spatiotemporal dynamics is adequate to prevent the past bleeding into the present through mismeasurement and/or misspecification.[3] Should any of these conditions fail, time lagging spatial lags will be technically inappropriate and practically insufficient to redress the spatial-lag endogeneity.

That outcomes will auto-correlate over time requires no parallel introduction or argument. No one would argue or pretend that time-serial observations were temporally independent in almost any context. Yet because temporal auto-dependence affords some simple evasions that spatial interdependence does not, empirical applications that model temporal *auto*-dependence directly are rarer still. Instead, researchers mostly follow Beck, Katz and Tucker's (1998) or Carter and Signorino's (2010) advice to sidestep the estimation challenges by modeling temporal *trends* with flexible functions of time-since-event instead. Such time dependence is not quite *auto*-dependence of a unit's current propensity on its previous propensities, so while these approaches may adequately correct the biases and exaggerated certainty estimates that stem from falsely pretending that time-serial binary observations are independent, they do not offer a direct model of temporal-autoregressive processes.[4] Some researchers have instead shifted

[3] As Beck, Gleditsch and Beardsley (2006), e.g., note, adequacy of the spatiotemporal-dynamic model can and should be tested. However, we have not seen such tests applied in the diffusion literature, and there are few signs that researchers are aware of the issue.

[4] Common (across units) time dependence—such as polynomials in, counters of or dummies for time (as opposed to time-since event)—would miss units' temporal *auto*-dependence more completely. More crucially in our view, none of these alternative approaches yields the theoretically interesting and substantively likely dynamic responses that autoregressive processes do (at least not directly or easily).

strategy and focus to model, not the outcomes directly, but transition matrices of state-switching probabilities (that is, probabilities of switching or staying in states 0 or 1) conditional on the previous observed state (for example, Jackman 2000; Przeworski et al. 2000; Beck et al. 2001).[5] This *is* autoregression in the binary outcome (and therefore a fuller strategy for temporal auto-dependence) but modeling interdependence directly in the latent propensities may be more appealing substantively in some contexts, and affords estimation of models in more-familiar standard binary-outcome format, without shifting strategy and focus.

Working under the incorrect assumption of spatial, temporal or spatiotemporal *in*dependence yields overconfidence and inefficiency at best, and usually bias and inconsistency too. Moreover, while including spatial and/or temporal lags to reflect (inter)dependence directly is advisable, simultaneous spatial lags are endogenous, and thus introduce biases if entered as right-hand-side variables in estimation procedures, like standard logit/probit, that assume their exogeneity (for example, Franzese and Hays 2004, 2007, 2008).[6] Therefore, we suggest a simulation-based approach for estimating models with simultaneous spatial and temporal lags (STP) of the latent outcomes using maximum simulated likelihood (MSL) by recursive-importance sampling (RIS), stated together as STP-MSL-by-RIS. This strategy enables us to recover not only accurate parameter estimates but also, more importantly, estimates of substantive conditional spatiotemporal effects, response paths and long-run-steady-state responses. Ultimately, we are interested in how probabilities of outcomes in this and other units and times respond to hypothetical changes in explanatory factors (**X**'s), propensities ($\mathbf{y}^*$), or outcomes (**y**) in this and other units and times. Yet extant empirical studies, even those few works that attempt to model spatial interdependence directly and, within those, even those rarer ones that apply appropriate estimation strategies to do so, tend to ignore such substantive *effects*, instead simply relying on parameter estimates (and significance) as evidence of spatial-effect *existence*, which, even when properly estimated, offer little sense of *effect* magnitude given the non-linearities of the spatial multiplier and the binary-outcome probability. Likewise, current strategies for modeling time dependence do not (or at least do not easily) enable us to calculate response paths over time or long-run effects in substantive terms. Simulation-based strategies like STP-MSL-by-RIS enable direct estimation of spatiotemporal autoregressive processes and of substantive spatiotemporal effects, response paths and steady states, all in terms of (conditional) probabilities of outcomes. We can calculate, for example, how a hypothetical increase in the propensity toward internal conflict or change in the realized outcome itself in Senegal (and/or elsewhere) affects the probability of civil war in Guinea-Bissau (and/or elsewhere) simultaneously, over time, and in the long run, including all spatiotemporal feedback between each other and around the Sub-Saharan state system in all time periods.

The article proceeds as follows: first, we briefly explain (with technical details and elaboration relegated to web appendices): (1) the severe inferential challenges posed by spatial, temporal or spatiotemporal interdependence in binary outcomes and (2) recent spatial-econometric

[5] Jackman (2000) and Beck et al. (2001) describe the lagged-latent binary-outcome model and a Bayesian Markov-Chain-Monte-Carlo (MCMC) estimator for it. They correctly identify the final distribution in their sampler as a T-dimensional truncated normal, and they offer the same Gibbs-within-Gibbs solution to that challenge as LeSage and Pace (2009). We use a classical simulated-likelihood strategy that is also computationally feasible and does at least as well in mean-squared error.

[6] Including other units' *outcomes* also creates measurement error insofar as the interdependence truly works through *propensities* or *expectations* of other units' outcomes. Alternative mechanisms may suggest interdependence of outcomes or of expected or latent outcomes, but only the latter are algebraically consistent in a simultaneous autoregressive framework for binary outcomes (as explained further below and, originally, in Heckman 1978).

advances using simulation-based methods to surmount the analytical challenges that arise in such problems. Next, Monte Carlo simulations compare the performance of one such consistent and asymptotically efficient method (STP-MSL-by-RIS) to estimation strategies naïve about (inter) dependence—that is, including lags but treating them as exogenous regressors in standard probit estimation. Penultimately, we show—we believe for the first time—how to calculate spatiotemporal response paths in outcome probabilities (with associated uncertainty estimates) for some units from hypothetical changes to *outcomes* in other units. Finally, we illustrate with an application to civil war in Sub-Saharan Africa and conclude.

THE ECONOMETRIC PROBLEM

Methods for properly estimating and analyzing models of interdependent binary outcomes have received significant attention in the spatial-econometric literature recently. Most of this research considers the spatial-probit model with interdependence in the latent variable—that is, in the unobserved argument to the probit-modeled probability of a binary outcome.[7] Several estimation strategies have been suggested. We focus on simulated-likelihood strategies and, in particular, the RIS estimator advanced by Beron, Murdoch and Vijverberg (2003) and Beron and Vijverberg (2004).[8] In brief, *inter*dependence in binary outcomes necessitates simulation-based estimation strategies, STP-MSL or MCMC, because the dependence across observations invalidates the typical process of deriving the overall likelihood (posterior) to be maximized by multiplying the probabilities (adding the log-likelihoods) of each observation; *in*dependence is required for that step. Interdependence instead leaves one n-dimensional distribution in the likelihood (posterior)—in the spatial-probit case, a truncated cumulative multivariate normal distribution—to maximize, which is difficult and computationally intense. We now review the spatial-probit model closely and then show how probit with temporal or spatiotemporal dependence in the latent variable is very similar in form, indicating that the same empirical strategies can apply there.

The structural model for the latent variable of the spatial probit takes the form:

$$\mathbf{y}* = \rho\mathbf{W}\mathbf{y}* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{Wy}*$ is the spatial lag, and $\mathbf{W}$ the all-important spatial-connectivity matrix, with element w_{ij} giving the relative connectivity from unit j to i. The ρ, which is to be estimated, is the strength of interdependence by that pattern given in $\mathbf{W}$. The model (1) can be rewritten in reduced form as:

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

Latent-variable $\mathbf{y}*$ links to the observed binary outcome, $\mathbf{y}$, through the measurement equation:

$$y_i = \{1 \text{ if } y_i* > 0; 0 \text{ if } y_i* \leq 0\}. \tag{3}$$

[7] Appendices II and III elaborate this introduction of the issues, methods and models, including a broader literature review, closer tracing of the technical details of the RIS and MCMC estimators, and a brief description of several alternative, less comprehensive, strategies: McMillen's original (1992) EM, Pinkse and Slade's (1998) and Klier and McMillen's (2005) GMM, and Fleming's (2004) NLS and GLM estimation strategies.

[8] The other main strategy currently in use is Bayesian MCMC by Metropolis-Hastings-within-Gibbs sampling, introduced by LeSage (1999, 2000; with LeSage and Pace 2009 correcting a crucial error in these earlier formulations) and separately, correctly, by Jackman (2000) and Beck et al. (2001).

The probability that the ith observation equals 1 is calculated as follows:

$$p(y_i = 1\,|\,\mathbf{X}) = p\Big(\Big[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\Big]_i + \Big[(\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}\Big]_i > 0\Big)$$
$$= p\Big(u_i < \Big[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\Big]_i\Big) = \Phi\Big\{\Big[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\Big]_i \Big/ \sigma_{u_i}\Big\}. \qquad (4)^{9}$$

Thus, as in the standard probit, a cumulative-normal distribution, $\Phi\{\}$, gives the probability that the systematic component, $[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{XP}]_i$, exceeds the stochastic component, u_i. (The division by $\sigma^2_{u_i}$ occurs here because $\mathbf{u}$, unlike $\boldsymbol{\varepsilon}$, is heteroskedastic.) The interdependence of the y_i* in spatial probit induces a nonsphericity of the reduced-form stochastic components, $\mathbf{u}$. Specifically, $\mathbf{u}$ is distributed n-dimensional *multivariate* normal with mean 0 and variance-covariance $[(\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})]^{-1}$. Intuitively, start with the standard (nonspatial) probit: $\boldsymbol{\varepsilon}$ is multivariate normal with mean 0 and spherical variance-covariance $\sigma^2\mathbf{I}$, and σ^2 is normalized to 1. Then:

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

Thus the $\sigma^2_{u_i}$ in (4) is the iith element of variance-covariance (5), which is not a constant (such as 1 in standard probit). That is, interdependence also induces heteroskedasticity. This heteroskedasticity and, more crucially, the interdependence (that is, the <u>*non*</u>-independence) of the u_i render standard probit inappropriate and create the computational intensity. With the outcomes interdependent, their joint likelihood to be maximized is not the product (sum) of n independent univariate marginal (log) distributions; instead, one must maximize the log of *one* non-separable n-dimensional distribution. Finally, notice also that the ith observation probability depends on the entire matrix $\mathbf{X}$ and vector $\boldsymbol{\varepsilon}$. The probability of Civil War in Senegal depends on the propensities toward internal conflict, and thus the full set of systematic and stochastic features *in the entire Sub-Saharan state system*. 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 all of $\boldsymbol{\varepsilon}$, given the dependence of y_i^* on $\mathbf{Wy}^*$.[10]

Consider now the similarity of a probit model with temporal auto-dependence, say an AR(1) process, in the latent variable. Start with the structural model in matrix notation:

$$\mathbf{y}^* = \phi\mathbf{L}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \qquad (6)$$

With $\mathbf{y}^*$ a $T\times 1$ vector of latent variables, $\mathbf{L}$ a matrix of 0's except for all 1's on the lower first-minor diagonal (the diagonal just below the prime diagonal) *and dropping the*

[9] In the middle step, note that the symmetry about 0 of $\boldsymbol{\varepsilon}$, and so of $\mathbf{u}$, implies that $p(-u_i < x) = p(u_i < x)\ \forall x$. In the last line, σ_{u_i} is the standard deviation of heteroskedastic $\mathbf{u}$, of u_i, not of homoskedastic ε, for which $\sigma_{u_i} = \sigma_u = \sigma_\varepsilon = 1$.

[10] The spatial-error-probit model is only slightly simpler (see Appendix II.) (Mixed spatial-lag/spatial-error models are also possible, but have received little attention.) The same cumulative n-dimensional heteroskedastic normal issues arise, although the position of the ith observation on the sigmoidal probit function depends only on $\mathbf{x}_i$ and not $\mathbf{X}$, because the interdependence operates only through $\boldsymbol{\varepsilon}$ and not all of $\mathbf{y}^*$. Special circumstances might allow standard-probit estimation of spatial-lag-$\mathbf{y}$ models, but these are highly implausible in most cases. Essentially, the interdependence pattern must not involve indirect feedback from outcome y_i to latent-variable y_i* (see Heckman 1978; Anselin 2006, and Appendix II). The unidirectionality of time prevents these indirect paths from y_t to latent-variable y_t* via the time-lagged outcome, which is what allows state-switching models to be properly identified, for instance, yet identification of even just temporally dynamic binary-outcome models remains less than straightforward (see, e.g., Chamberlain 1993; Honore and Kyriazidou 2000).

first observation,[11] this gives a standard first-order temporal autoregressive model, with reduced form:

$$\mathbf{y}* = (\mathbf{I}-\phi\mathbf{L})^{-1}\mathbf{X}\boldsymbol{\beta}+\mathbf{u}, \text{with } \mathbf{u} = (\mathbf{I}-\phi\mathbf{L})^{-1}\boldsymbol{\varepsilon}. \tag{7}$$

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

$$V[\mathbf{u}] \equiv \Sigma = V[(\mathbf{I}-\phi\mathbf{L})^{-1}\boldsymbol{\varepsilon}] = [(\mathbf{I}-\phi\mathbf{L})^{-1}]'V(\boldsymbol{\varepsilon})[(\mathbf{I}-\phi\mathbf{L})^{-1}] = [(\mathbf{I}-\phi\mathbf{L})'(\mathbf{I}-\phi\mathbf{L})]^{-1}, \tag{8}$$

and the rest of the discussion regarding the estimation and interpretation complications that come with this nonseparable T-dimensional cumulative nonspherical normal distributions apply *mutatis mutandis*. For the panel or time-series-cross-section case with spatial and temporal lag, we could write:

$$\mathbf{y}* = \rho\mathbf{W}\mathbf{y}* + \phi\mathbf{L}\mathbf{y}* + \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}. \tag{9}$$

Ordering the data by period, units 1 to N in time 1 through to units 1 to N in time T, and dropping period 0, $\mathbf{W}$ is now an $NT\times NT$ matrix with $N\times N$ submatrices $\mathbf{W}_t$ with elements w_{ij} giving relative connectivity from unit j to unit i, period by period down the block diagonal, and $\mathbf{L}$ is an $NT\times NT$ matrix of 0's except with 1's on the diagonal of the lower block first-minor (that is, with $N\times N$ identity matrices along the blocks just below the prime block-diagonal), making $\mathbf{Ly}*$ the within-unit one-period time lag of $\mathbf{y}*$. The reduced form is now:

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

which again implies a nonspherical variance-covariance for $\mathbf{u}$, similar in form:

$$V[\mathbf{u}] \equiv \Sigma = V[(\mathbf{I}-\rho\mathbf{W}-\phi\mathbf{L})^{-1}\boldsymbol{\varepsilon}] = [(\mathbf{I}-\rho\mathbf{W}-\phi\mathbf{L})'(\mathbf{I}-\rho\mathbf{W}-\phi\mathbf{L})]^{-1}. \tag{11}$$

We will focus on this simultaneous spatiotemporal-lag model (dropping the first time periods) next because it raises the estimation and interpretation issues fully, and the reduction to the simple cross-section or time-series is a straightforward.

THE MAXIMUM SIMULATED-LIKELIHOOD (STP-MSL) BY RECURSIVE IMPORTANCE SAMPLER (RIS) ESTIMATOR FOR SIMULTANEOUS SPATIAL, TEMPORAL OR SPATIOTEMPORAL PROBIT

We discuss STP-MSL strategies, and in particular the RIS estimator for spatial probit advanced by Beron, Murdoch and Vijverberg (2003) and Beron and Vijverberg (2004).[12] The same equations and procedures apply for the spatial-, temporal- and spatiotemporal-autoregressive cases, simply using $(\mathbf{I}-\rho\mathbf{W})^{-1}$, $(\mathbf{I}-\phi\mathbf{L})^{-1}$ or $(\mathbf{I}-\phi\mathbf{L}-\rho\mathbf{W})^{-1}$ as appropriate for the multipliers, as seen in Equations 2–11 above.

RIS uses simulation to approximate probabilities that are difficult to calculate analytically, such as those from the cumulative multivariate normal distributions unavoidably introduced

[11] In effect, this takes the process to have begun with the first observation. In this case, matrix pre-multiplication by $\mathbf{L}$ is identical to the backshift operator, $L(\cdot)$. Taking the first observation as given or fixed, however, leaves estimates of ϕ susceptible to Hurwicz bias, a small-sample attenuation bias of order $1/T$ in the linear case, because $\mathbf{Ly}*$ includes observations' dependence only back to the first observation (assuming that the true process extends back indefinitely). From this perspective, we are using the conditional (on the first period) likelihood, which has this familiar small-sample (in T) bias (and a related small-sample inefficiency), rather than the unconditional likelihood, which would treat the first observation as stochastic and be unbiased as well as consistent and asymptotically efficient.

[12] The other comprehensive strategy currently in use is Bayesian MCMC by Metropolis-Hastings-within-Gibbs sampling. Appendix III briefly introduces and explains that estimation strategy.

by the interdependence in spatial, temporal or spatiotemporal probit.[13] To approximate a cumulative multivariate-normal distribution such as the following, for example,

$$p = \int_{-\infty}^{\mathbf{x}_0} f_n(\mathbf{x})d\mathbf{x}, \tag{12}$$

where $f_n(\mathbf{x})$ is the density and $[-\infty, \mathbf{x}_0]$ the interval over which we want to integrate, we first choose an 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*. Multiplying and dividing the right-hand side of the integral we wish to calculate, Equation 12, by the density of this importance distribution, $g_n^c(\mathbf{x})$, simply restates Equation 12 as:

$$p = \int_{-\infty}^{\mathbf{x}_0} \frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})} g_n^c(\mathbf{x})d\mathbf{x}. \tag{13}$$

By definition, because $g_n^c(\mathbf{x})$ is a valid *pdf* over the integral's range, this integral is a mean. Thus Equation 13 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:

$$p = E\left[\frac{f_n(\mathbf{x})}{g_n^c(\mathbf{x})}\right] \approx \frac{1}{R}\sum_{r=1}^{R} \frac{f_n(\tilde{\mathbf{x}}_r)}{g_n^c(\tilde{\mathbf{x}}_r)} \equiv \hat{p}. \tag{14}$$

To implement the RIS estimator, we draw $\mathbf{x}$ from the importance distribution, for which we will use a truncated multivariate normal[14] R times and calculate the mean as $f_n(\mathbf{x})/g_n^c(\mathbf{x})$. Given that the densities for $f_n(\mathbf{x})$ and $g_n^c(\mathbf{x})$ are $\phi_n(\mathbf{x})$ and $\phi_n(\mathbf{x})/\Phi_n(\mathbf{x})$, where $\phi_n(\cdot)$ and $\Phi_n(\cdot)$ are the density and distribution functions for the multivariate normal, the ratio simplifies to $\Phi_n(\mathbf{x})$.

Again, in the standard probit model with independence, the log-likelihood would simply sum the log of n univariate cumulative-standard-normal distributions, which is manageable. In spatial probit, with its interdependence, however, the likelihood is one n-dimensional cumulative normal. To calculate this integral, we first define $\mathbf{Q}$ as a diagonal matrix with $q_{ii} = 1 - 2y_i$, $\boldsymbol{\upsilon} = \mathbf{Qu}$ and $\mathbf{v} = -\mathbf{Q}(\mathbf{I} - \phi\mathbf{L} - \rho\mathbf{W})^{-1}$.[15] Specifically, we want to calculate

$$p(\boldsymbol{\upsilon} < \mathbf{v}), \tag{15}$$

where $\boldsymbol{\upsilon} \sim MVN(\mathbf{0}, \Omega)$ is the $n \times 1$ vector of reduced-form spatiotemporally interdependent errors, with $\Omega = \mathbf{Q}\Sigma\mathbf{Q}'$, and $\mathbf{v}$ is the vector of spatially interdependent cutpoints. The RIS estimator exploits that $\boldsymbol{\Omega}$, as a variance-covariance matrix, is positive definite, and so has a Cholesky decomposition, $\Omega^{-1} = \mathbf{C}'\mathbf{C}$, with $\mathbf{C}$ being upper triangular and $\boldsymbol{\eta} \equiv \mathbf{C}\boldsymbol{\upsilon}$ giving

[13] We introduce RIS following Vijverberg's (1997) and Beron and Vijverberg's (2004) notation.

[14] 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.

[15] Note that $q_{ii} = 1 - 2y_i$ is -1 for $y_i = 1$ and 1 for $y_i = 0$; thus, pre-multiplying $\mathbf{X}\boldsymbol{\beta}$ by $-\mathbf{Q}$ serves merely to correctly sign the systematic component up to which to integrate the distribution of the stochastic component; pre-multiplying $\mathbf{u}$ by $\mathbf{Q}$ ensures that the covariances of the stochastic component are correctly signed.

n independent standard normals, $\boldsymbol{\eta}$. (This is familiar, as the same exploitation also applied in GLS.) Then, letting $\mathbf{B} \equiv \mathbf{C}^{-1}$ and substituting $\boldsymbol{\upsilon} = \mathbf{C}^{-1}\boldsymbol{\eta} \equiv \mathbf{B}\boldsymbol{\eta}$ into Equation 15 gives:

$$\Pr(\mathbf{B}\boldsymbol{\eta} < \mathbf{v}) = \Pr\left(\begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & \cdots & b_{1,n} \\ 0 & b_{2,2} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_{n-1,n-1} & b_{n-1,n} \\ 0 & \cdots & 0 & 0 & b_{n,n} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \vdots \\ \vdots \\ \eta_{n-1} \\ \eta_n \end{bmatrix} < \begin{bmatrix} v_1 \\ \vdots \\ \vdots \\ v_{n-1} \\ v_n \end{bmatrix}\right). \tag{16}$$

The n elements of $\boldsymbol{\eta}$ are independent, so the probability in Equation 16 can be calculated by evaluating the univariate cumulative-normal distributions row by row at the implied upper bounds, which are determined recursively starting with the last row, and then multiplying these probabilities.[16] To determine these upper bounds, start by solving the inequalities in Equation 16 for the vector $\boldsymbol{\eta}$:

$$\Pr\left(\begin{bmatrix} \sum_{i=1}^{n} b_{1,i}\eta_i \\ \vdots \\ b_{n-1,n-1}\eta_{n-1} + b_{n-1,n}\eta_n \\ b_{n,n}\eta_n \end{bmatrix} < \begin{bmatrix} v_1 \\ \vdots \\ v_{n-1} \\ v_n \end{bmatrix}\right) = \Pr\left(\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_{n-1} \\ \eta_n \end{bmatrix} < \begin{bmatrix} b_{1,1}^{-1}\left(v_1 - \sum_{i=2}^{n} b_{1,i}\eta_i\right) \\ \vdots \\ \vdots \\ b_{n-1,n-1}^{-1}\left(v_{n-1} - b_{n-1,n}\eta_n\right) \\ b_{n,n}^{-1}v_n \end{bmatrix}\right). \tag{17}$$

First, calculate the upper bound for the truncated-normal distribution of the nth row, 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 $(n-1)$ th row conditional on the nth as $b_{n-1,n-1}^{-1}\left[v_{n-1} - b_{n-1,n}\tilde{\eta}_n\right]$. Evaluate the cumulative standard normal at this upper bound and call it p_{n-1}. Use the first two draws to calculate the $(n-2)$th upper bound, and calculate p_{n-2} and proceed analogously through all n rows. Formally, this recursive process can be depicted as:

$$\left.\begin{array}{r} \eta_n < b_{n,n}^{-1}v_n \equiv \psi_n \\ \eta_{n-1} < b_{n-1,n-1}^{-1}\left[v_{n-1} - b_{n-1,n}\tilde{\eta}_n\right] \equiv \psi_{n-1} \\ \eta_{n-2} < b_{n-2,n-1}^{-1}\left[v_{n-2} - b_{n-2,n-1}\tilde{\eta}_{n-1} - b_{n-2,n}\tilde{\eta}_n\right] \equiv \psi_{n-2} \\ \vdots \end{array}\right\} \Rightarrow \eta_j < b_{j,j}^{-1}\left[v_j - \sum_{i=j+1}^{n} b_{j,i}\tilde{\eta}_i\right] \equiv \psi_j. \tag{18}$$

The likelihood of observing a given sample of 1's and 0's can now be found by evaluating the *univariate* cumulative-normal distribution at each of these bounds, p_j, and *multiplying*

[16] Importantly, the marginal probabilities in these rows do not correspond to those of the corresponding observations; the RIS procedure relies only on the fact that their product is the correct joint probability.

those probabilities: $\prod_{j=1}^{n} p_j = \prod_{j=1}^{n} \Phi(\psi_j)$.[17] Repeating the entire process R times and averaging gives the RIS estimate of the joint probability (that is, the simulated likelihood, $\hat{l}$) as this simple mean:

$$\hat{l} = (1/R) \sum_{r=1}^{R} \left[\prod_{j=1}^{n} \Phi(\psi_{j,r}) \right]. \tag{19}$$

One can then maximize this simulated likelihood by any standard optimization routine to estimate parameters and apply the standard maximum likelihood (ML) estimator for the variance covariance $(-[H(\hat{l})]^{-1})$.[18] Like ML, STP-MSL is BANC: best (asymptotically efficient) asymptotically normal and consistent.

ESTIMATED SPATIAL/SPATIOTEMPORAL EFFECTS AND RESPONSE PATHS WITH CERTAINTY ESTIMATES

Properly estimating *parameters* like *coefficients* and their certainties is an essential start to valid inference, but the ultimate aim is to estimate and draw inferences regarding *effects* (ideally: causal ones), that is, *outcome* responses associated with (ideally: caused by) changes in explanatory factors or other counterfactual shocks. We estimate *coefficients* like ρ, ϕ and $\boldsymbol{\beta}$ for the purposes of estimating *effects* like $\frac{\Delta y_i^*}{\Delta x_i}$, that is, effects of changes in some x_i on the latent variable, y_i* or, better, on the outcome probability, $\frac{\Delta p(y_i=1)}{\Delta x_i}$.[19] Given interdependence, even these sorts of *within-unit* counterfactuals (effects of $x_{\underline{i}}$ on $y_{\underline{i}}$) involve feedback from i through other units j back to i. The effect of economic development in Senegal on its own probability of internal conflict, for instance, includes the reverberating feedback from Senegal to its neighbors and back. In contexts involving interdependence, however, interest usually focuses on cross-unit feedback effects, like $\frac{\Delta y_j^*}{\Delta x_i}$ or $\frac{\Delta p(y_j=1)}{\Delta x_i}$ (the reverberating effects of Senegalese development across Sub-Saharan Africa).[20] For interdependent binary outcomes, perhaps most interesting of all would be effects on outcome probabilities in some unit(s) i of hypothetical *outcomes* in other units, j ($y_j = 1$ or $y_j = 0$), which we denote $\frac{\Delta p(\mathbf{y}_i=1)}{\Delta \mathbf{y}_j}$: the effect on the probability of a Senegalese civil war on internal conflict versus peace in Guinea-Bissau, for instance. In spatiotemporal contexts, finally, we will want to estimate the response paths and long-run-steady-state effects of all of these sorts of hypotheticals.

Because they involve spatial, temporal or spatiotemporal feedback dynamics, none of these substantive *effects* or *response paths* is simple to estimate. The feedback multipliers $(\mathbf{I} - \rho\mathbf{W} - \phi\mathbf{L})^{-1}$

[17] The italicized is possible because the Cholesky-transformed η are independent, and subscript j substituted for subscript i to emphasize that these are probabilities for the rows $j = 1 \ldots N$ and not the observations $i = 1 \ldots N$.

[18] When estimating temporal or spatiotemporal models, we drop the first observation, thereby omitting the first period from the likelihood and treating it as deterministic. This is similar to using the conditional likelihood in the continuous outcome case, except that, since y* is unobserved, we condition the joint likelihood for observations $2 \ldots T$ on the first period's vector of binary outcomes. If the true autoregressive process extends back indefinitely, this likelihood treating the first observation as deterministic is misspecified. However, given that observations $2 \ldots T$ are correctly incorporated into the likelihood, the importance of this first-period misspecification for estimation declines as T increases. Our Monte Carlo experiments below, which condition on the first period's outcomes in this way, yield estimates broadly consistent with small-sample bias of order $1/T$ as just described.

[19] Actually, more fully and precisely, the aims are/should be to estimate the structural model from which we can, among other things, calculate such estimated responses to substantive hypotheticals of all sorts, carrying around the empirically estimated parameterized model as useful analytical understanding of the empirical data-generation process.

[20] For responses to underlying *propensities*, y_i*, instead of specific explanators, x_i, imagine shocking the intercept.

imply that these various spatiotemporal effects and responses involve (nonlinear) combinations of coefficients and variables, so we could not read *effects* directly from a table of *coefficients* even if we confined discussion to responses in latent variables, $\hat{y}^*$, rather than to the probabilities of actual interest, $\hat{\mathbf{p}}$. Since the latent-variable model is a spatial linear regression, estimated effects in terms of $\hat{y}^*$ and their certainties would derive exactly as in that case, which we have discussed elsewhere (Franzese and Hays 2004, 2007, 2008a, 2008b; Hays et al. 2010; Appendix IV).

Beron and Vijverberg (2004) cover the effect of changes in x_i on the outcome probability in unit i. Simply applying the chain rule, the immediate effect, including all spatial feedback, is:[21]

$$\frac{\partial p(y_i = 1\,|\,\mathbf{X},\mathbf{W})}{\partial x_i} = \phi_{pdf}\left(\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]_i\Big/\sigma_{u_i}\right)\left[(\mathbf{I}-\rho\mathbf{W})^{-1}_{ii}\boldsymbol{\beta}\right]\Big/\sigma_{u_i}, \tag{20}$$

where ϕ_{pdf} is the univariate standard-normal density function. This calculation can be extended straightforwardly to include steady-state effects and response paths as well as cross-unit effects.

We focus below on perhaps the most substantively interesting type of effect, which, to the best of our knowledge, has not been considered elsewhere: the response path in outcome probabilities for some unit(s) i from hypothetical changes to *outcomes* in other unit(s) j. We start with the response path in outcome probabilities for a single unit i to a change at time t in the outcome for a single unit j. Using the definition of conditional probability, these effects are calculated as:

$$\begin{aligned}&p[y_{i,t+s}{=}1\,|\,y_{j,t}{=}1;\mathbf{X},\mathbf{W},\mathbf{L}]-p[y_{i,t+s}{=}1\,|\,y_{j,t}{=}0;\mathbf{X},\mathbf{W},\mathbf{L}]\\ &=\frac{p[y_{i,t+s}{=}1,y_{j,t}{=}1\,|\,\mathbf{X},\mathbf{W},\mathbf{L}]}{p[y_{j,t}{=}1\,|\,\mathbf{X},\mathbf{W},\mathbf{L}]}-\frac{p[y_{i,t+s}{=}1,y_{j,t}{=}0\,|\,\mathbf{X},\mathbf{W},\mathbf{L}]}{p[y_{j,t}{=}0\,|\,\mathbf{X},\mathbf{W},\mathbf{L}]},\ \text{for } s{=}0,\ \ldots\ .\end{aligned} \tag{21}$$ [22]

Thus, given estimates of ρ, ϕ and $\boldsymbol{\beta}$, we calculate the cumulative distribution functions for two univariate and two multivariate-normal distributions, one for each side of the counterfactual, to produce an estimated spatiotemporal response path (or spatial or spatiotemporal long-run steady-state effect).[23] The joint probabilities in the numerators depend on the strength of the

[21] For ∂x_j, for example, we have: $\frac{\partial p(y_i=1\,|\,\mathbf{X},\mathbf{W})}{\partial x_j} = \phi_{pdf}\left(\left[(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]_i\Big/\sigma_{u_i}\right)\times\left[(\mathbf{I}-\rho\mathbf{W})^{-1}_{ij}\beta\Big/\sigma_{u_i}\right]$, and first-difference effects could be calculated in the usual fashion by $\mathbf{P}\left\{(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}_1\boldsymbol{\beta}\right\}-\mathbf{P}\left\{(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}_0\boldsymbol{\beta}\right\}$, where $\mathbf{P}$ is the cumulative nonstandard normal (having variance-covariance as in Equations 5, 8 or 11 as appropriate, and $\mathbf{X}_1$ and $\mathbf{X}_0$ indicating the two sides of the hypothetical of interest).

[22] Notice that $\mathbf{X}$ and $\mathbf{W}$ and $\mathbf{L}$ are taken as data (exogenously given and nonstochastic) in the usual manner.

[23] A more general statement from which we could, in principle, calculate the outcome-probability responses across all units over time to counterfactual outcomes across all units and histories, $\mathbf{y}^1_{j\neq i}$versus $\mathbf{y}^0_{j\neq i}$ and $\mathbf{y}^1_{i,t-r}$versus $\mathbf{y}^0_{i,t-r}$, would be:

$$\begin{aligned}&p[y_{i,t+s}=1\,|\,(\mathbf{y}_{j\neq i,t}=\mathbf{y}^1_{j\neq i,t},\mathbf{y}_{i,t-r}=\mathbf{y}^1_{i,t-r});\mathbf{X},\mathbf{W},\mathbf{L}]-p[y_{i,t+s}=1\,|\,(\mathbf{y}_{j\neq i,t}=\mathbf{y}^0_{j\neq i,t},\mathbf{y}_{i,t-r}=\mathbf{y}^0_{i,t-r});\mathbf{X},\mathbf{W},\mathbf{L}]\\ &=\frac{p[y_{i,t+s}=1,(\mathbf{y}_{j\neq i,t}=\mathbf{y}^1_{j\neq i,t},\mathbf{y}_{i,t-r}=\mathbf{y}^1_{i,t-r});\mathbf{X},\mathbf{W},\mathbf{L}]}{p[(\mathbf{y}_{j\neq i,t}=\mathbf{y}^1_{j\neq i,t},\mathbf{y}_{i,t-r}=\mathbf{y}^1_{i,t-r});\mathbf{X},\mathbf{W},\mathbf{L}]}-\frac{p[y_{i,t+s}=1,(\mathbf{y}_{j\neq i,t}=\mathbf{y}^0_{j\neq i,t},\mathbf{y}_{i,t-r}=\mathbf{y}^0_{i,t-r});\mathbf{X},\mathbf{W},\mathbf{L}]}{p[(\mathbf{y}_{j\neq i,t}=\mathbf{y}^0_{j\neq i,t},\mathbf{y}_{i,t-r}=\mathbf{y}^0_{i,t-r});\mathbf{X},\mathbf{W},\mathbf{L}]}.\end{aligned}$$

$$\text{for } s=0,\ \ldots \text{ and } r=1,\ \ldots\ ,(T-1);\text{ and }\forall i.$$

Both numerators and denominators of these conditional probabilities would be computed by multivariate non-standard normal (with the estimated $V(\mathbf{u})$) cumulative distribution functions at the estimated spatiotemporally transformed cutpoints. As the number of units in $\mathbf{y}_{j\neq i,t}$ and $\mathbf{y}_{i,t-r}$ grows, the joint probabilities become more spatiotemporal connections between $y_{i,t+s}$ and $y_{j,t}$. When s is small, the spatial connection largely determines the size of the effect. As s gets large, the time between events grows and $p[y_{i,t+s} = 1,0 \,|\, y_{j,t} = 1,0]$ converges to $p[y_{i,t+s} = 1,0] \times p[y_{j,t} = 1,0]$, driving the effect to 0. We estimate uncertainties for these estimated spatiotemporal responses by parametric simulation: repeat the procedure for each of many draws of parameter estimates from their estimated joint distribution, and the average and standard deviation across draws give the estimate and its standard error.[24] For those who study interdependence in binary outcomes, these and related quantities are substantively important. In the Monte Carlo experiments and illustration that follow, we evaluate our ability to estimate this conditional effect accurately and then use this calculation for analytical purposes.

MONTE CARLO ANALYSES

We explore the small-sample properties of the STP-MSL estimator for the spatiotemporal-lag probit model using a data-generation process that follows as closely as possible Beron and Vijverberg's (2004) purely cross-sectional spatial-probit Monte Carlo experiments.[25] We evaluate the quality of the STP-MSL parameter estimates, and of the spatial and spatiotemporal effect estimates described in the Estimated Spatial/Spatiotemporal Effects and Response Paths with Certainty Estimates section. Regarding the effect estimates, we compare the small-sample performance of our suggested strategy for calculating first differences in probabilities for y_i conditional on outcomes $y_{j\neq i}$ with effect estimates from naïve spatiotemporal-probit models that include spatial lags (but treat them as exogenous regressors) and that address temporal dynamics using either a first-order Markov regime-switching or discrete-time hazard framework. The former includes time-lagged observed outcome, **Ly**, not **Ly***, as a regressor, and the latter a polynomial in time-since-event, s, (we use s, s^2 and s^3). With all regressors assumed to be exogenous, both naïve models apply simple ML as the estimator. Our spatio-temporal data-generating process takes the form:

$$\mathbf{y}^* = (\mathbf{I}_n - \rho\mathbf{W} - \phi\mathbf{L})^{-1}(\beta_0 + \beta_1\mathbf{x} + \boldsymbol{\varepsilon}),\ \boldsymbol{\varepsilon} \sim N(0,1). \tag{22}$$

We apply Equation 3 to generate **y** from these **y***. For **W**, we use a row-standardized binary-contiguity matrix with the 50 US states as in Beron and Vivjerberg (2004). Data for each unit are generated for 20 periods, giving us a sample size of 1,000 in each of our Monte Carlo experiments.[26] In each of our experiments we set $\beta_0 = -1.5$, and $\beta_1 = 3.0$, yet vary $\rho = \{0.10, 0.25\}$ and $\phi = \{0.30, 0.5\}$, giving us four experiments with different levels of spatial and temporal dependence. Finally, **x** is drawn from a standard uniform distribution on the interval

(F'note continued)
complicated to calculate. At the limit, this effect estimate is as difficult to compute as the likelihood itself, although the task is simpler since the unknown parameters have already been estimated. Either brute-force or importance sampling could be used for this purpose. Alternatively, these types of conditional counterfactual effects can be estimated via brute-force simulation as in Hays (2009). We include a brief explanation of this technique in Appendix IV as it produces accurate results and can be easily generalized to higher dimensions, up to and including those for which computing cumulative multivariate normal probabilities as shown here is either computationally inefficient or unachievable (due to from limits within existing commands/packages).

[24] MSL, like ML, is BANC, so this asymptotically normal with the mean of the estimated parameter vector and variance/covariance of the parameter estimates' estimated variance/covariance.

[25] Appendix V gives simulation results for the Bayesian MCMC estimator and for experiments with varying ρ.

[26] This is similar to that found in many political science applications in both American and Comparative politics.

TABLE 1 *Simulation Results for STP-MSL by RIS Parameter Estimates (100 Trials)*

	$\beta_0 = -1.5$	$\beta_1 = 3.0$	ρ	ϕ
Experiment #1: $\rho = 0.10$, $\phi = 0.30$				
Mean coefficient estimate	−1.467	2.962	0.092	0.276
Root mean-squared error	0.104	0.173	0.043	0.032
Actual SD of estimates	0.101	0.169	0.042	0.021
Mean of reported SE	0.129	0.231	0.045	0.025
Overconfidence	0.784	0.730	0.938	0.840
Experiment #2: $\rho = 0.10$, $\phi = 0.50$				
Mean coefficient estimate	−1.385	2.797	0.097	0.464
Root mean-squared error	0.150	0.273	0.034	0.042
Actual SD of estimates	0.095	0.183	0.034	0.021
Mean of reported SE	0.105	0.199	0.042	0.020
Overconfidence	0.906	0.920	0.802	1.022
Experiment #3: $\rho = 0.25$, $\phi = 0.30$				
Mean coefficient estimate	−1.450	2.922	0.224	0.280
Root mean-squared error	0.117	0.197	0.047	0.028
Actual SD of estimates	0.106	0.181	0.038	0.020
Mean of reported SE	0.119	0.212	0.046	0.024
Overconfidence	0.884	0.854	0.834	0.833
Experiment #4: $\rho = 0.25$, $\phi = 0.50$				
Mean coefficient estimate	−1.363	2.752	0.241	0.471
Root mean-squared error	0.172	0.322	0.034	0.035
Actual SD of estimates	0.104	0.205	0.033	0.019
Mean of reported SE	0.107	0.202	0.031	0.020
Overconfidence	0.969	1.013	1.043	0.946

Note: $N = 50$ and $T = 20$ for all experiments reported in text.

[−1, 2], resulting in an expected value of 0.5 and a variance of roughly 2. These choices produce, on average, a balanced sample of 0's and 1's. Table 1 presents results for 100 trials using the STP-MSL-RIS spatiotemporal-probit estimator (with $R = 100$).[27] We do not compare these parameter estimates with those of the naïve spatiotemporal-probit, because the forms of temporal dependence they assume (regime-switching, discrete-time hazard) do not generate the same parameters as Equation 22. Comparing the implied first/same-period spatial effects is appropriate, though, and we do so below.

We begin by noting that our estimates are reasonably accurate across all of the experiments. The mean coefficient estimates are always within 10 percent of the true parameter values. This performance compares favorably to the results in Beron and Vijverberg (2004) and can be explained by our sample sizes, which are larger than theirs, as well as by the fact that the covariate, **x**, in our data-generating process has a larger variance. Despite this accuracy, our results exhibit the well-known small-sample bias in ML estimates of autoregressive-lag models: a tendency to underestimate, on average, both the spatial coefficient ρ as well as the time-dependence coefficient ϕ. The bias in the mean estimate for ρ seems to grow slightly in percentage terms as we increase its true value from 0.10 to 0.25, whereas the bias in the mean estimate for ϕ drops in percentage terms as we increase the degree of temporal dependence from 0.30 to 0.50. However, given the relatively small number of trials in our experiments and the small changes observed in the biases, we do not want to attach too much significance to these

[27] Given the computational time to estimate the model, 100 is common for Monte Carlo trials in spatial latent-variable contexts (e.g., Beron and Vijverberg 2004).

TABLE 2 *Simulation Results for Effect Estimates (100 Trials)*

	STP-MSL	RS-ML	DTH-ML
Experiment #1: ρ = 0.10, ϕ = 0.30, *TrueEffect* = 0.055			
Mean effect estimate	0.049	0.034	0.038
Actual SD of estimates	0.023	0.022	0.022
Root mean-squared error	0.024	0.030	0.028
Experiment #2: ρ = 0.10, ϕ = 0.50, *TrueEffect* = 0.063			
Mean effect estimate	0.056	0.030	0.040
Actual SD of estimates	0.019	0.019	0.022
Root mean-squared error	0.020	0.038	0.032
Experiment #3: ρ = 0.25, ϕ = 0.30, *TrueEffect* = 0.135			
Mean effect estimate	0.117	0.087	0.095
Actual SD of estimates	0.022	0.037	0.024
Root mean-squared error	0.028	0.061	0.047
Experiment #4: ρ = 0.25, ϕ = 0.50, *TrueEffect* = 0.151			
Mean effect estimate	0.136	0.076	0.107
Actual SD of estimates	0.018	0.038	0.023
Root mean-squared error	0.027	0.084	0.050

patterns. We also see attenuation biases in the mean estimates for β_0 and β_1. These biases are quite small when the degree of spatiotemporal dependence is low, approximately 2 percent, but increase to around 8–10 percent when both ϕ and ρ are large.

Ultimately, we are not interested in the parameter estimates *per se*, but rather in the effects that they imply. We start with the first difference $\Pr[y_i = 1 \,|\, \mathbf{x}, \mathbf{W}, \mathbf{L}, y_j = 1] - \Pr[y_i = 1 \,|\, \mathbf{x}, \mathbf{W}, \mathbf{L}, y_j = 0]$ for the immediate (first/same-period) spatial effect (that is, the effect of a time t change in the outcome for unit j on the time t probability that we observe a particular outcome for unit i). In Table 2, we compare the first/same-period spatial effect estimates implied by the spatiotemporal-lag model with comparable estimates from two naïve spatial models that address temporal dependence alternatively via regime switching (RS) or discrete-time hazard (DTH) functions. To illustrate, we select a specific pair of units i and j directly connected in the spatial-weights matrix (Alabama and Mississippi) and calculate the true effect, $\Pr[y_i = 1 \,|\, \mathbf{x}, \mathbf{W}, \mathbf{L}, y_j = 1] - \Pr[y_i = 1 \,|\, \mathbf{x}, \mathbf{W}, \mathbf{L}, y_j = 0]$, assuming $\mathbf{x}$ equals its last sample values. In our example, the true effect size ranges from 0.055 to 0.151 given different levels of spatial and temporal dependence (as indicated in Table 2). To calculate these effects using the spatiotemporal-lag probit model, we use the parameter estimates noted above and the method outlined in the Estimated Spatial/Spatiotemporal Effects and Response Paths with Certainty Estimates section. The naïve spatial models replace the spatial-lag latent variable in Equation 22 with a spatial lag in observed outcomes, $\mathbf{Wy}$, treating it as exogenous. The temporal-lag latent variable is replaced with a temporally lagged observed outcome in the RS model and with the last sample value of the peace years counter in the DTH model. To estimate the first/same-period spatial effect, we set $y_j = 0$ and calculate $\Pr[y_i = 1]$, change to $y_j = 1$ and recalculate $\Pr[y_i = 1]$, and then difference these probabilities.

It is possible that the naïve alternatives do well enough in terms of effect estimates that they provide reasonable substitutes for the more complicated STP-MSL estimator, even when the spatiotemporal-lag probit model is the true data-generating process. In Table 2, we show that this is not the case. As noted in Table 1, STP-MSL slightly underestimates, on average, all of the model's parameters. Consequently, it will also underestimate the first difference $\Pr[y_i = 1 \,|\, \mathbf{x}, \mathbf{W}, \mathbf{L}, y_j = 1] - \Pr[y_i = 1 \,|\, \mathbf{x}, \mathbf{W}, \mathbf{L}, y_j = 0]$. Across the four experiments, the

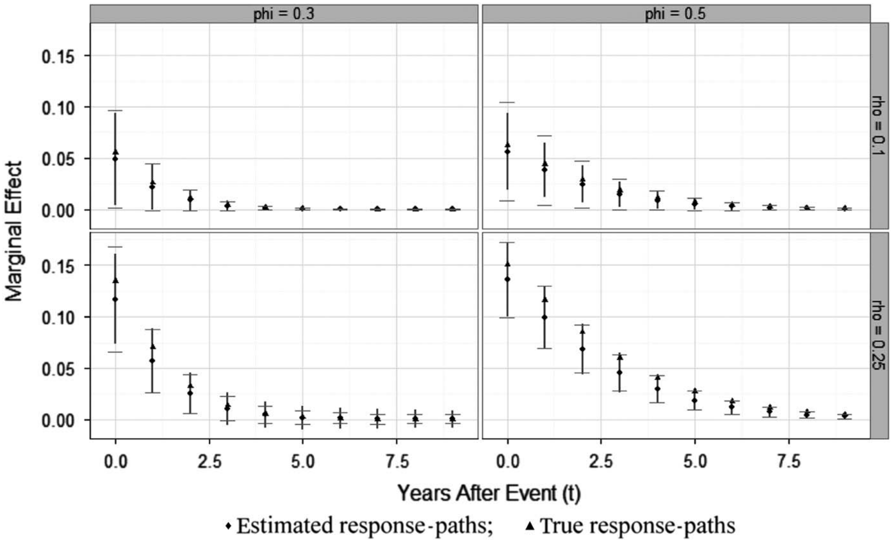

Fig. 1. Accuracy of STP-MSL estimates of spatiotemporal response-path estimates.
Note: horizontal bars indicate 95 percent confidence intervals using the estimated standard errors (within trials); vertical lines indicate 95 percent intervals using the actual standard deviation of the effect estimates (across trials). ◆ = estimated response paths; ▲ = true response paths

effect estimate by STP-MSL is ~89 percent of the true first difference. This compares with average effect estimates across the experiments of 68 and 56 percent for the DTH-ML and RS-ML models, respectively. In our experiments, the performance gap in terms of bias is largest when the degree of temporal dependence is high. This is true when performance is gauged in root-mean-square error (RMSE) terms as well. When ρ is 0.25 and ϕ is 0.5, the RMSE for the DTH-ML effect estimate is 85 percent larger than the STP-MSL estimator, and the RMSE for the RS-ML effect estimate is 211 percent larger. Thus, we do not find evidence to suggest that the simpler DTH-ML and RS-ML models provide reasonably approximate effect estimates when the spatiotemporal-lag probit is the true data-generating process.

Moreover, note that the consequences of the naïve models' misspecifications, both assuming the spatial lag is exogenous and getting the temporal dynamics wrong, are minimized in these first/same-period spatial-effect estimates. Misspecifying temporal dynamics has relatively little impact on *immediate*-effect estimates, and the coefficient-estimate bias from assuming that the spatial-lag is exogenous arises in part because it accounts for both direct and indirect effects, both of which these spatial-effect estimates incorporate to some degree. Thus Table 2 represents the best case for the naïve spatial models. Most other effect estimates of interest would exhibit larger, likely much larger, performance gaps.

Figure 1 graphs the spatiotemporal response paths for these Monte Carlo experiments, using the spatially and temporally lagged latent variable probit model.[28] The graph provides information about (1) the bias in our response-path estimates, comparing the truth triangles and average-estimate diamonds, (2) the efficiency of these estimates, by the 95-percentile range across trials in the estimates and (3) the accuracy of our uncertainty estimates, by the horizontal

[28] The naïve models assume different temporal-dynamic processes, complicating direct comparisons of the implied response-path estimates.

check marks giving the average 95 percent confidence interval estimates.[29] As expected, these estimated response paths, which begin with the first/same-period spatial effects from Tables 1 and 2, decay over time, and we tend to slightly underestimate their magnitudes due to the underestimation of ρ and ϕ. Certainty estimates tend to be quite accurate as well, marginally under-confident in (many of) the first periods, and occasionally over-confident thereafter. In totality, the estimated response paths reflect the true paths accurately.

ILLUSTRATION: CONTAGION OF CONFLICT IN AFRICA

To illustrate, we model the (inter)dependence in civil conflict episodes, estimating spatial and spatiotemporal-lag models of civil war incidences in Sub-Saharan Africa, covering all countries from 1966 to 2001.[30] Anecdotally, signs of spatiotemporal dependence in the region's civil conflict dynamics seem strong. Several countries in the Great Lakes region of (East) Africa (for example, DRC, Uganda, Rwanda, Burundi, Tanzania) have seen numerous civil wars that appear related. O'Loughlin and Raleigh (2008, 10) argue that these conflicts "highlight how many current civil wars are not state-specific, but related and supported by a host of external conditions."[31] The West Africa region (for example, Sierra Leone, Liberia, Guinea, Senegal) seems to exhibit similar conflict dynamics. In 2003, *The Economist* reported: "West Africa's civil wars are usually reported as tragedies befalling individual states…In fact, all these wars are intertwined, and it is impossible to understand one without reference to the others." These accounts highlight the need to explicitly account for spatial interdependence in analyses of civil war.

Theoretically, conflict studies recognizing spatial dependence in civil wars are hardly new. A large and long-standing literature concludes that civil conflicts diffuse across international borders, with violence in one country making civil war more likely in adjacent ones (for example, Most and Starr 1980; Starr and Most 1983; Diehl 1991; Lake and Rothchild 1998; Gleditsch 2002; Ward and Gleditsch 2002).[32] Refinements have attempted to discern and distinguish the specific mechanisms by which civil war diffusion occurs (Murdoch and Sandler 2002; Saleyhan and Gleditsch 2006; Buhaug and Gleditsch 2008), to understand the ability of some states to resist transmission (Raleigh 2004; Braithwaite 2010) and to explore how expectations of diffusion may influence third-party intervention (Kathman 2010). This research has generally inferred support for such *neighborhood effects* from the significance of some spatial-parameter estimate, usually the coefficient on a spatial lag of conflict (occasionally, time-lagged spatial lag) entered in a standard logit or probit estimation. Usually the simultaneity of these spatial lags goes unaddressed, and the likely high levels of interdependence suggest the resultant biases may be substantial.[33] Moreover, to our knowledge, no work has interpreted the effects of the estimated interdependence in terms of the spatial or spatiotemporal responses in

[29] Alternatively, the numeric values for all information reported in Figure 1 are available in tabular form in Appendix VI.

[30] The panel is unbalanced, given the emergence of new countries during the sample period, ranging from 31 countries in the earliest period to 42 in the final year.

[31] Additionally, see O'Loughlin and Raleigh (2008) for an example of how a particular conflict (in this case Uganda) draws in other actors in surrounding states and, in turn, increases the risk of civil conflict in adjacent countries.

[32] That contagion is the cause is not entirely consensual, however; *inter alia*, Hegre et al. (2001) contend that, instead, the spatial clustering of domestic correlates of civil war causes the observed clustering of conflicts.

[33] As noted, time lagging the spatial lag is a possible, though often inadequate, redress. Most current research has preferred to include the spatial lag simultaneously, however (e.g., Salehyan and Gleditsch 2006; Buhaug and Gleditsch 2008).

TABLE 3 *Spatial-Autoregressive Models of Conflict in Africa*

	Model 1 (Naive Spatial)	Model 2 (STP-MSL)
Constant	–4.5713 (1.0619)***	–4.0136 (1.063)***
Spatial Lag	0.7801 (0.1620)***	0.1294 (0.0489)***
Neighborhood Dem	–0.0298 (0.0208)	–0.0309 (0.0207)
Neighborhood Dem sq.'d	0.0015 (0.0031)	–0.0009 (0.0031)
Neighborhood GDP	0.0248 (0.1181)	0.0326 (0.1250)
GDP per capita (ln)	0.0543 (0.0755)	0.0325 (0.0725)
Population (ln)	0.3678 (0.0380)***	0.3546 (0.0353)***
Democracy	0.0042 (0.0103)	–0.0007 (0.0098)
Democracy sq.'d	–0.0103 (0.0020)***	–0.0094 (0.0018)***
Post-Cold War	0.1116 (0.0991)	0.0946 (0.0953)
N (states)	1403 (42)	1403 (42)

Note: dependent variable: civil-war incidence (0 = No, 1 = Yes). All spatial lags use row-standardized binary-contiguity spatial-weights matrices. Significant at ***1, **5, *10%.

civil conflict *probabilities* across countries to hypothetical shocks to civil conflict risks in some country or countries, as we suggest here.

Therefore, we apply the STP-MSL estimator to a spatial-probit model of civil war incidence. In Models 1 and 2, we focus exclusively on (cross-sectional) spatial dependence, modeled by a simultaneous spatial lag (See Table 3). The spatial relationships between states are given by a row-standardized binary-contiguity **W** matrix, with the sample countries averaging 4.0167 neighbors (ranging from 1 to 8).[34] Our data come from Buhaug and Gleditsch (2008), who explore the extent to which regional conflict clustering stems from the spatial clustering of country-specific characteristics (that is, in **X**'s) rather than direct *contagion* or *spillover* of conflict. As they do, we measure conflict incidence using the UCDP/PRIO Armed Conflict Database, which codes civil conflicts as occurring if violent incidents between a state government and organized opposition result in at least 25 deaths.[35] Our models include all of the regressors in Buhaug and Gleditsch's full models, including measures for the most common predictors of civil war in peace studies (for example, *GDP, Population, Regime Type*).[36]

Though both models estimate a significant spatial-lag coefficient, the estimates from the naïve model would seem to radically overstate the extent of interdependence, as the estimate is nearly *seven times* what the STP-MSL estimates seem to indicate. However, as noted above, parameter estimates may belie the ability of either model to recover accurate estimates of spatial effects, which are of greater importance. That is, we are ultimately concerned about the extent to which neighboring conflict (peace) makes conflict more (less) likely in various interesting counterfactual questions we could pose regarding conflict spatial and/or spatiotemporal dynamics in Sub-Saharan Africa. Even just a cursory review of recent political coverage uncovers the tragic frequency with which such questions are (implicitly) posed: "Fears mount that Côte d'Ivoire conflict could spill to Liberia,"[37] "Zambia Concerned About DRC conflict spillover,"[38]

[34] A table summarizing the information contained in the weights matrix is available in Appendix VII.

[35] This occurs in 307 observations, roughly 22 percent of our sample.

[36] Our estimation models differ greatly from those in Buhaug and Gleditsch (2008)—notably, they estimate logistic models in a global sample, whereas ours are probit models in an African sample—making direct comparisons of parameter estimates inappropriate.

[37] http://www.unhcr.org/4d8b61839.html.

[38] http://www.voanews.com/content/zambia-concerned-about-drc-conflict-spillover/1553526.html.

"The conflict in Mali could be creating a 'ticking time bomb' for neighbouring Western Sahara,"[39] etc. Such hypotheticals are examples of the *conditional* effects discussed in the Estimated Spatial/Spatiotemporal Effects and Response Paths with Certainty Estimates section. To illustrate how we can answer such substantively meaningful conditional probabilities, we explore the extent to which the outbreak of conflict in Guinea-Bissau in 1998 was a consequence of the ongoing conflict in neighboring Senegal, subject to the important caveat that we have not yet accounted for temporal dependence in conflict propensities.

Historically, the border between Guinea-Bissau and Senegal has been a breeding ground for instability. Guinea-Bissau served as the training ground for *Mouvement des Forces Democratiques de Casamance* (MFDC) fighters and a conduit to funnel arms into the decades-long Casamance conflict. Most view the 1998–99 civil war in Guinea-Bissau as an outgrowth of these same tensions, with Senegalese forces ultimately fighting on both sides of the conflict (Humphreys and Mohamed 2005).[40] We explore the extent to which the conflict in Senegal affected the onset of civil war in Guinea-Bissau, or, stated differently to highlight the conditional probability of the counterfactual: in the absence of the Casamance conflict, how likely was conflict in Guinea-Bissau? Utilizing simulations as described in the Estimated Spatial/Spatiotemporal Effects and Response Paths with Certainty Estimates and Monte Carlo Analyses sections, Guinea-Bissau has a civil war 8.5 percent of the time when Senegal is peaceful and 11.14 percent of the time when Senegal is at war. Thus the model estimates suggest that Senegal's conflict increased the risk of war in Guinea-Bissau by 2.64 percent. To calculate our uncertainty about these effects estimates, we sample parameter estimates from their estimated sampling distribution; doing so reveals that the effect sizes at the 5th and 95th percentiles are 0.90 and 4.89 percent, respectively.

Next, using the time-series cross-section of these data, we estimate models accounting for both spatial and temporal dependence in conflict episodes (Table 4). The naïve model again appears to overstate the extent of spatial dependence, as both the RS-ML (Model 3) and DTH-ML (Model 4) models using an exogenous spatial lag indicate a significant spatial effect. When we estimate the spatiotemporal-lag probit model, however, we find a smaller and non-significant spatial-lag coefficient. What, then, causes the discrepancy between the insignificant spatial effect found here and the significant finding in all other models? We know that naïve estimation strategies (Models 3 and 4) that treat the spatial lag as exogenous will tend to overstate the degree of spatial dependence. Similarly, even if we account for the endogeneity of the spatial lag, failure to allow for the possibility of temporal dependence (Model 2) could lead us to infer incorrectly from our models that spatial dependence is present when it is not.[41]

This can be the case when the observable and unobservable determinants of an event cluster in both space and time, making it difficult for misspecified models to distinguish between spatial and temporal dependence. For example, if the level of democracy in Guinea-Bissau at time $t-1$ correlates with the level of democracy in Senegal at $t-1$ (spatial clustering)—which, in turn, correlates with the level of democracy in Senegal at time t (temporal clustering)—the latent propensity for civil war in Guinea-Bissau at time $t-1$ will correlate with the latent propensity

[39] http://www.theguardian.com/world/2013/apr/09/mali-conflict-spread-western-sahara.

[40] Specifically, the Senegalese government sent troops to support the Vieira regime, while the MFDC sent forces to support the revolutionary Ansoumane Mané.

[41] We should also note that direct comparison of the spatial coefficients produced by Models 2 and 5 would be inappropriate. To make these estimates more comparable with the purely spatial Model 2, we would decompose the spatial dynamics here into a short-run component, given by the coefficient on the spatial lag, ρ, and the long-run accumulation of those dynamic feedbacks, given by that spatial-lag coefficient times the temporal long-run multiplier, $1/(1-\phi)$, so: $\rho/(1-\phi)$.

TABLE 4 *Spatiotemporal Models of Conflict in Africa (Buhaug and Gleditsch 2008)*

	Model 3 (Naïve Spatial w/RS)	Model 4 (Naïve Spatial w/DTH)	Model 5 (STP-MSL w/lagged-**y***)
Constant	–2.4153 (1.3764)*	–0.4272 (1.1352)	–1.2411 (0.7953)
Spatial Lag	0.4881 (0.2110)**	0.4599 (0.2203)**	0.0072 (0.0316)
Temporal Dependence	2.456 (0.1126)***	–0.5373 (0.0383)***	0.7217 (0.0247)***
Time-since-event^2	—	0.0294 (0.0029)***	—
Time-since-event^3	—	–0.0004 (0.0001)***	—
Neighborhood Dem	–0.0054 (0.0254)	0.0009 (0.0244)	–0.0261 (0.0169)
Neighborhood Dem sq.'d	0.0050 (0.0038)	0.0054 (0.0037)	0.0018 (0.0025)
Neighborhood GDP	–0.1670 (0.1559)	–0.2350 (0.1562)	–0.0829 (0.0929)
GDP per capita (ln)	0.0540 (0.0989)	0.0999 (0.0956)	0.0472 (0.0545)
Population (ln)	0.1731 (0.0483)***	0.2153 (0.0463)***	0.1403 (0.0272)***
Democracy	0.0222 (0.0131)*	0.0107 (0.0134)	0.0150 (0.0081)*
Democracy sq.'d	–0.0032 (0.0026)	–0.0038 (0.0026)	–0.0043 (0.0016)***
Post-Cold War	0.1346 (0.1295)	0.0082 (0.1290)	0.1738 (0.0811)**
N (states)	1403 (42)	1403 (42)	1403 (42)

Note: dependent variable: civil-war incidence (0 = No, 1 = Yes). All spatial lags use row-standardized binary-contiguity spatial-weights matrices. Significant at ***1, **5, *10%.

for civil war in Senegal at time *t*. If there is only temporal dependence in the true data-generating process for a country's latent propensity to experience civil war (and no spatial dependence), and we put Senegal's latent propensity for civil war at time *t* on the right-hand side of Guinea-Bissau's latent propensity at time *t* and exclude Guinea-Bissau's latent propensity at time $t-1$, omitted-variable bias will erroneously make it seem as though there are spatial spillovers.[42]

This appears to be what we observe with respect to civil conflict. While we do not want to make sweeping claims on the basis of our admittedly limited illustration, we do want to emphasize that it is possible to mistake spatial dependence for temporal dependence and vice versa. Given that much of the scholarly quantitative literature finds spatial dependence in civil war and the popular press treats these spillovers as a nearly undeniable reality, this is important to know. Our analysis suggests that the evidence of such a relationship may not be as strong as currently believed.

We do, however, find that temporal dependence is, unsurprisingly, high and highly significant, indicating great persistence in conflict behavior. To analyze these temporal dynamics more explicitly, we again turn from estimated *coefficients* to estimated *effects*. Continuing with the Casamance conflict in Senegal as an example, but shifting focus to the temporal dynamics, this time we ask: What effect did Senegal's prior conflict history have on the realization of conflict in 1998? In particular, beginning with 1992—when the conflict in Senegal escalated—we explore the extent to which fighting (or peace) from that point onward influenced the likelihood of conflict in 1998.

As before, we can use our simulation strategy to condition on a single counterfactual outcome, finding that conflict as opposed to peace in 1992 made Senegal 51.9 percent more likely to experience conflict in 1993.[43] To estimate counterfactual response paths for various

[42] Using a similar logic, Cook (2015) shows how neglecting spatial dependence limits our understanding of how, and how much, civil conflicts persist over time.

[43] As before, we can calculate our uncertainty about these estimates by sampling parameter estimates from their estimated distribution and repeating this whole procedure for each parameter-vector draw.

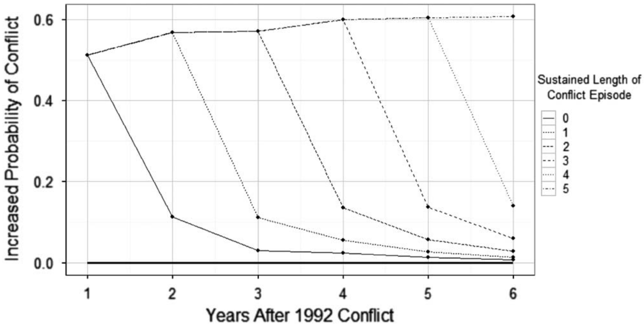

Fig. 2. Estimated response-paths to alternative counterfactual conflict histories in Senegal

alternative conflict histories, we simply increase the number of outcomes that are conditioned upon (that is, the number of dimensions in our cumulative probability). Figure 2 illustrates, depicting six (of the possible 64) conflict histories since 1992 that could have preceded 1998 in Senegal. We compare each of these histories—lengths of sustained conflict since 1992 of 0 to 5 years—against one in which Senegal was peaceful for all those years, which is given by the thick-black x-axis.

The results suggest some interesting features of conflict dynamics. First, the effects of conflict persistence (auto-dependence) are quite large immediately following a conflict, but dissipate rapidly once peace emerges. For instance, we estimate that, if the Senegalese conflict had ended after 1992, no discernible effect of previous conflict on future conflict probabilities would have persisted by 1998. (The thin solid line converges quickly to the steady state.) We also see how histories full of conflict can mire states in so-called *conflict traps*, as years of fighting produce high probabilities that conflict will persist into subsequent years. Notice, for instance, that at the end of the all-conflict counterfactual history, the likelihood of fighting in 1998 is 62 percent higher than it would be following the all-peace history.

Lastly, we compare the predictive accuracy of the spatiotemporal-lag probit model estimated by maximum simulated likelihood (STP-MSL), the regime-switching model estimated by maximum likelihood (RS-ML) and the discrete-time hazard model estimated by maximum likelihood (DTH-ML). We begin by focusing on the overall ability of the respective models to predict conflict and peace outcomes using the area under the receiver-operating characteristic (ROC) curve. In our illustration, ROC curves gauge the ability of each model to correctly predict instances of conflict or true positives without making false positive predictions across all possible prediction rules.[44] Area under the curve (AUC) scores are portions of the area of the unit square and therefore range between 0 and 1, with higher scores indicating a better job of predicting peace and conflict outcomes. (A random model-free prediction strategy generates an AUC score of 0.5.) In terms of overall predictive performance, there is not much separation between the models. STP-MSL has the highest score (0.9318), but the difference in comparison with DTH-ML is minimal (0.9316). The performance gap with RS-ML is larger (0.9254), but remains minor.

[44] A prediction rule in our example is a probability threshold that is used for predicting conflict. This rule reflects one's relative sensitivity to making false-positive and false-negative predictions. On the one hand, if sensitivity to false positives is high, one would set the probability threshold for predicting conflict close to 1. On the other hand, if sensitivity to false negatives (predicting peace when there is conflict) is high, one would set the probability threshold close to 0.

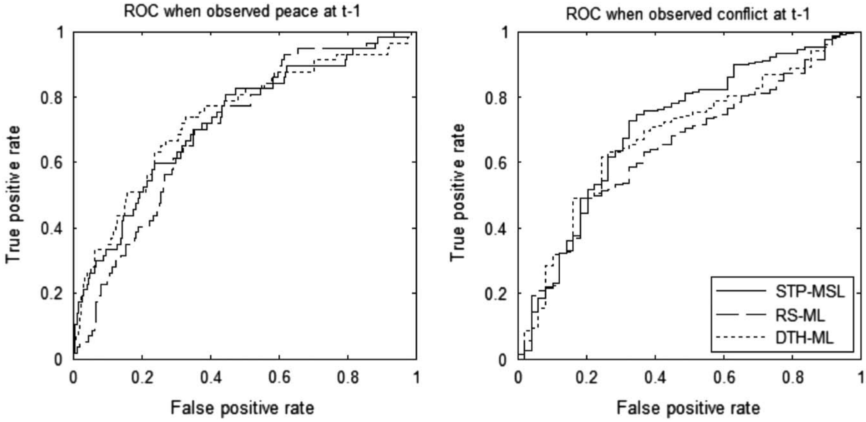

Fig. 3. ROC curves for onset and duration subsamples

In Figure 3, we divide the data into two sub-samples based on whether conflict was observed in the prior period, thereby allowing us to break down the ability of the models to accurately predict conflict onset (or peace duration) and conflict persistence (or peace onset). The AUC scores from the ROC curves indicate that DTH-ML (0.7342) and STP-MSL (0.7222) are more or less equivalent at predicting conflict onset (plot to the left in Figure 3), with DTH-ML offering a 1.7 percent improvement over STP-MSL and both besting RS-ML (0.6988). A clearer performance gap emerges when we focus on the prior conflict sub-sample (plot to the right in Figure 3), with STP-MSL outperforming both DTH-ML and RS-ML at most of the evaluated thresholds. Comparing AUC scores, STP-MSL (0.7090) offers a 9.6 percent improvement over RS-ML (0.6469) and a 4.2 percent improvement over DTH-ML (0.6801).

In sum, STP-MSL has a slight advantage over DTH-ML and RS-ML in overall predictive performance. When we separate the data based on the prior outcome, we see that the superior performance of STP-MSL is due to its ability to better predict conflict duration (peace onset), with no estimator offering as clear an advantage in predicting conflict onset. While the predictive advantages offered by STP-MSL are relatively minor here, we expect that the benefits would be clearer in applications exhibiting significant spatial dependence.[45] Furthermore, our unified autoregressive framework for spatial and temporal dependence is much more elegant than the alternatives. For this reason, and despite the added complexity surrounding parameter estimation, the spatiotemporal probit simplifies effect calculations both conceptually and computationally.

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, and are also misspecified if that

[45] Estimated ROC curves from the Monte Carlo experiments showed clearer support for our estimator, with MSL-RIS strictly dominating in each of the four possible outcome predictions (peace-peace, peace-conflict, conflict-conflict, conflict-peace).

interdependence is modeled by including endogenous spatial lags and/or temporally lagged outcomes (as opposed to lags of latent variables), treating these as exogenous regressors. Implementing consistent and efficient estimators for spatial- and spatiotemporal-lag probit models, such as MSL-by-RIS, can be challenging and is computationally demanding, but the benefits seem to outweigh these costs when latent propensities are autoregressive in time and space. The more-easily implemented naïve spatiotemporal alternatives yield inefficient and sometimes badly biased estimates of important *effects* when data are generated by spatial, temporal or spatiotemporal autoregressive processes (in latent variables). And latent propensities are likely autoregressive in time and space in many, if not most, substantive areas of research in the social sciences. In the civil conflict example pursued here, a country's latent propensity for civil conflict seems likely persistent over time and may spill across borders even if conflict does not actually erupt.[46] Most importantly, we have shown how to provide substantively important counterfactual effect estimates from spatial/spatiotemporal-autoregressive binary-outcome models, including spatiotemporal response paths, and naïve alternatives will not generally produce these naturally or estimate them accurately. In our view, these estimated substantive spatiotemporal effects and response paths alone more than justify the computational costs of estimating models that directly reflect the spatiotemporal autoregressive processes generating our data.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/psrm.2015.14

REFERENCES

Beck, Nathaniel, David Epstein, Simon Jackman, and Sharyn O'Halloran. 2001. 'Alternative Models of Dynamics in Binary Time-Series–Cross-Section Models: The Example of State Failure'. 2001 Annual Meeting of the Society for Political Methodology, Emory University, Atlanta, GA.

Beck, Nathaniel, Johnathan Katz, and Richard Tucker. 1998. 'Taking Time Seriously: Time-Series–Cross-Section Analysis with a Binary Dependent Variables'. *American Journal of Political Science* 42(4):1260–88.

Beck, Nathaniel, Kristian S. Gleditsch, and Kyle Beardsley. 2006. 'Space is More than Geography: Using Spatial Econometrics in the Study of Political Economy'. *International Studies Quarterly* 50(1):27–44.

Beron, Kurt J., James C. Murdoch, and Wim P.M. Vijverberg. 2003. 'Why Cooperate? Public Goods, Economic Power, and the Montreal Protocol'. *Review of Economics and Statistics* 85(2):286–97.

Beron, Kurt J., and Wim P.M. Vijverberg. 2004. 'Probit in a Spatial Context: A Monte Carlo Analysis'. In Luc Anselin, Raymond Florax and Sergio J. Rey (eds), *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Berlin: Springer-Verlag, 169–96.

Braithwaite, Alex. 2010. 'Resisting Infection: How State Capacity Conditions Conflict Contagion'. *Journal of Peace Research* 47(3):311–9.

Buhaug, Halvard, and Kristian S. Gleditsch. 2008. 'Contagion or Confusion? Why Conflicts Cluster in Space'. *International Studies Quarterly* 52(2):215–33.

Carter, David B., and Curt S. Signorino. 2010. 'Back to the Future: Modeling Time Dependence in Binary Data'. *Political Analysis* 18(3):271–92.

[46] Given that this finding contrasts with the conventional wisdom (and expanding literature) on conflict contagion, we revisit this question in greater detail in additional substantive work (Cook 2015).

Chamberlain, Gary. 1993. 'Feedback in Panel Data Models'. Working Paper No. 1656. Cambridge, MA: Harvard Institute of Economic Research.

Cook, Scott J. 2015. 'The Echo of Conflict: Modeling the Dependence of Civil Conflict in Space and Time'. Working Paper, Texas A&M University.

Diehl, Paul F. 1991. 'Geography and War: A Review and Assessment of the Empirical Literature'. *International Interactions* 17:11–27.

Fleming, Mark M. 2004. 'Techniques for Estimating Spatially Dependent Discrete-Choice Models'. In Luc Anselin, Raymond Florax and Sergio J. Rey (eds), *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Berlin: Springer-Verlag, 145–68.

Franzese, Robert J., and Jude C. Hays. 2004. 'Empirical Modeling Strategies for Spatial Interdependence: Omitted-Variable Vs. Simultaneity Biases'. Presented at the 2004 Summer Meeting of the Society for Political Methodology, Stanford University, Stanford, CA. Available at http://www.umich.edu/~franzese/FranzeseHays.PolMeth.2004.pdf.

Franzese, Robert J., and Jude C. Hays. 2007. 'Spatial-Econometric Models of Cross-Sectional Interdependence in Political-Science Panel & Time-Series-Cross-Section Data'. *Political Analysis* 15 (2):140–64.

Franzese, Robert J., and Jude C. Hays. 2008a. 'Empirical Models of Spatial Interdependence'. In J. Box-Steffensmeier, H. Brady and D. Collier (eds), *Oxford Handbook of Political Methodology*, 570–604. Oxford: Oxford University Press.

Franzese, Robert J., and Jude C. Hays. 2008b. 'Interdependence in Comparative Politics: Substance, Theory, Empirics, Substance'. *Comparative Political Studies* 41(4/5):742–80.

Franzese, Robert J., Jude C. Hays, and Scott J. Cook. 2012. 'Spatial-, Temporal-, and Spatiotemporal-Autoregressive Probit Models of Interdependent Binary Outcomes: Estimation and Interpretation'. Presented at the 2012 Annual European Political Science Association, Berlin.

Hays, Jude C. 2009. 'Bucking the System: Using Simulation Methods to Estimate and Analyze Systems of Equations with Qualitative and Limited Dependent Variables'. Presented at the 2009 Annual St. Louis Area Methods Meeting (SLAMM), Washington University in St. Louis.

Heckman, James J. 1978. 'Dummy Endogenous Variables in a Simultaneous Equation System'. *Econometrica* 46:931–59.

Hegre, Havard, Tanja Ellingsen, Scott Gates, and Nils Petter Gleditsch. 2001. 'Towards a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816-1992'. *American Political Science Review* 95(1):33–48.

Honore, Bo E., and Ekaterini Kyriazidou. 2000. 'Panel Data Discrete Choice Models with Lagged Dependent Variables'. *Econometrica* 68(4):839–74.

Jackman, Simon. 2000. 'In and Out of War and Peace: Transitional Models of International Conflict'. Working paper. Stanford, CA: Stanford University. Available at http://jackman.stanford.edu/papers/inandout.pdf.

Kathman, Jacob D. 2010. 'Civil War Contagion and Neighborhood Interventions'. *International Studies Quarterly* 54:989–1012.

Klier, Thomas, and Daniel P. McMillen. 2005. 'Clustering of Auto Supplier Plants in the US: GMM Spatial Logit for Large Samples'. Working Paper Series No. WP-O5-18, Federal Reserve Bank of Chicago.

Lake, David A., and Donald Rothchild, eds. 1998. *The International Spread of Ethnic Conflict*. Princeton, NJ: Princeton University Press.

LeSage, James P. 1999. *Spatial Econometrics*. http://www.rri.wvu.edu/WebBook/LeSage/spatial/wbook.pdf, accessed 22 May 2015.

LeSage, James P. 2000. 'Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models'.
Geographical Analysis 32(1):19–35.

LeSage, James P., and Robert K. Pace. 2009. *Introduction to Spatial Econometrics*. Boca Rotan, FL: CRC Press.

McMillen, Daniel P. 1992. 'Probit with Spatial Autocorrelation'. *Journal of Regional Science* 32:335–48.

McMillen, Daniel P. 1995. 'Selection Bias in Spatial Econometric Models'. *Journal of Regional Science* 35(3):417–36.

Most, Benjamin, and Harvey Starr. 1980. 'Diffusion, Reinforcement, and Geopolitics and the Spread of War'. *The American Political Science Review* 74(4):932–46.

Murdoch, James, and Todd Sandler. 2002. 'Economic Growth, Civil Wars, and Spatial Spillovers'. *Journal of Conflict Resolution* 46(1):91–110.

O'Loughlin, John, and Clionadh Raleigh. 2008. 'Spatial Analysis of Civil War Violence'. In Kevin R. Cox, Murray M. Low and Jennifer Robinson (eds), *The Sage Handbook of Political Geography*. Thousand Oaks, CA: Sage, 493–508.

Pinkse, Joris, and Margaret E. Slade. 1998. 'Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models'. *Journal of Econometrics* 85:125–54.

Raleigh, Clionadh. 2004. 'Neighbours and Neighbourhoods: Understanding the Role of Context in Civil War'. Presented at the 5th Pan-European International Relations Conference. The Hague, The Netherlands, 9–11 September.

Salehyan, Idean, and Kristina S. Gledistch. 2006. 'Refugees and the Spread of Civil War'. *International Organization* 60(2):335–66.

Starr, Harvey, and Benjamin A. Most. 1983. 'Contagion and Border Effects on Contemporary African Conflict'. *Comparative Political Studies* 16(1):92–117.

Vijverberg, Wim P.M. 1997. 'Monte Carlo Evaluation of Multivariate Normal Probabilities'. *Journal of Econometrics* 76:281–307.

Ward, Michael, and Kristian S. Gledistch. 2002. 'Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace'. *Political Analysis* 10(3):244–60.