Empirical Models of Spatial Interdependence

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

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

3 May 2007; corrections inserted 8 August 2008

[Forthcoming in *Oxford Handbook of Political Methodology*, J. Box-Steffensmeier, H. Brady, and D. Collier, eds., OUP]

I. Spatial Interdependence in Political Science

A. The Substantive Range of Spatial Interdependence

Until recently, empirical analyses of spatial interdependence in the social sciences remained largely confined to specialized areas of applied economics (e.g., environmental, urban/regional, real-estate economics) and sociology (i.e., network analysis). However, social-scientific interest in and applications of spatial modeling have burgeoned lately, due partly to advances in theory that imply interdependence and in methodology for addressing it, partly to global substantive changes that have raised at least the perception of and attention to interconnectivity, and likely the actual degree and extent of it, at all levels, from micro/personal to macro/international, and partly to advances in technology for obtaining and working with spatial data. In political science, too, spatial empirical analyses have grown increasingly common: a very welcome development as many phenomena that political scientists study entail substantively important spatial interdependence.

Perhaps the most extensive classical and contemporary interest in spatial interdependence in political science surrounds the diffusion of policy and/or institutions across national or sub-national governments. The study of policy-innovation diffusion among U.S. States in particular has deep roots and much contemporary interest (e.g., Crain 1966, Walker 1969, 1973, Gray 1973 and Bailey and Rom 2004, Boehmke and Witmer 2004, Daley and Garand 2004, Grossback et al. 2004, Shipan and Volden 2006, Volden 2006).[1] Similar policy-learning mechanisms underlie some comparative studies of policy

[1] And between, e.g., Knoke 1982, Caldiera 1985, Lutz 1987, Berry & Berry 1990, 1999, Case et al. 1993, Berry 1994, Rogers

diffusion (e.g., Schneider and Ingram 1988, Rose 1993, Meseguer 2004, 2005, Gilardi 2005). Interest in institutional or even regime diffusion is likewise long-standing and recently much reinvigorated in comparative and international politics. Dahl's (1971) classic *Polyarchy*, for instance, (implicitly) references international diffusion among the eight causes of democracy he lists; Starr's (1991) "Democratic Dominoes" and Huntington's (1991) *Third Wave* accord it a central role; and, finally, O'Loughlin et al. (1998) and Gleditsch and Ward (2006, 2007) empirically estimate the international diffusion of democracy. Simmons and Elkins (2004), Elkins et al. (2006), and Simmons et al. (2006) similarly stress international diffusion as the force behind recent economic liberalizations, as do Eising (2002), Brune et al. (2004), Brooks (2005), and many others in recent years.

The substantive range of important spatial-interdependence effects extends well beyond these more-obvious contexts of intergovernmental diffusion, however, spanning the subfields and substance of political science. Inside democratic legislatures, e.g., representatives' votes certainly depend on others' votes or expected votes; in electoral studies, election outcomes or candidate qualities or strategies in some contests surely depend on those in others. Outside legislative and electoral arenas, the probabilities and outcomes of coups (Li and Thompson 1975), riots (Govea and West 1981), and/or revolutions (Brinks and Coppedge 2006) in one unit depend in substantively crucial ways on those in others. In micro-behavioral work, too, some of the recently surging interest in *contextual effects* surrounds the effects on each respondent's behaviors or opinions of aggregates of others' behaviors and opinions—e.g., those of the respondent's region, community, or social network. Within the mammoth literature on contextual effects in political behavior (Huckfeldt and Sprague (1993) review), recent contributions that stress interdependence include Braybeck and Huckfeldt (2002ab), Cho (2003), Huckfeldt et al. (2005), Cho and Gimpel (2007), Cho and Rudolph (2007), Lin et al (2006). In international relations, meanwhile, the interdependence of states' actions essentially defines the subject. States' entry decisions in wars, alliances, or international organizations, e.g., heavily depend on how many and who (are expected to)

1995, Mintrom 1997ab, Mintrom & Vergari 1998, Mossberger 1999, Godwin & Schroedel 2000, Balla 2001, Mooney 2001.

enter.[2] Empirical attention to the inherent spatial interdependence of international relations has been greatest in the work of Ward, Gleditsch, and colleagues (e.g., Shin and Ward 1999, Gleditsch and Ward 2000, Gleditsch 2002, Ward and Gleditsch 2002, Hoff and Ward 2004, Gartzke and Gleditsch 2006, Salehyan and Gleditsch 2006, Gleditsch 2007). In comparative and international political economy also, interdependence is perhaps especially frequently substantively large and central. Simmons and Elkins (2004), Elkins et al. (2006), and Simmons et al. (2006) stress cross-national diffusion as the main force behind recent economic liberalizations, for examples, as do Eising (2002), Brune et al. (2004), Brooks (2005), and many others. In fact, globalization and international economic integration, arguably today's most-notable (and indisputably most-noted) political-economic phenomena, imply strategic and/or non-strategic interdependence of domestic politics, policymakers, and policies. Empirical work emphasizing such globalization-induced interdependencies includes Genschel (2002), Basinger and Hallerberg (2004), Knill (2005), Jahn (2006), Swank (2006), Franzese and Hays (2006b, 2007abc), and Kayser (2007).

B. Mechanisms of Spatial Interdependence

Spatial interdependence is, in summary, ubiquitous and often quite central throughout the substance of political science. Geographer Waldo Tobler (1930-) puts it simply: *everything is related to everything else, but near things are more related than distant things*. Moreover, as Beck et al. (2006) pithily stress titularly: *space is more than geography*. I.e., the substantive content of *proximity* in *Tobler's Law*, and so the pathways along which interdependence between units may operate, extend well beyond simple physical distance and bordering (as several examples above illustrate). Elkins and Simmons (2005) and Simmons et al. (2006), e.g., define and discuss four mechanisms by which interdependence may arise: coercion, competition, learning, and emulation. *Coercion*, which may be direct or indirect and hard (force) or soft (suasion), encompasses a generally "vertical" pathway by which the powerful induce actions among the weaker. *Competition* refers to interdependence stemming from economic pressures that the actions of each unit place upon others in competition with it or as substitutes for or complements

[2] Signorino (1999,2002) and Signorino & Tarar (2006) also stress the strategic interdependence of international relations in specifying empirical models, as do Signorino and Yilmaz (2003) and Signorino (2003) regarding strategic choice broadly.

to it. *Learning* entails situations where actors learn from others' actions, in rational-Bayesian or other fashion, something regarding the net appeal of their own alternatives.[3] *Emulation*, lastly, is ritualistic (i.e., neither coerced nor responsive to competition or to learning) following or doing oppositely of others (e.g., leaders, co-ethnics, co-partisans). Although enumerated specifically for contexts of cross-national diffusion, these four categories nicely span most possible channels of spatial interdependence across its broader substantive domain. We would add a fifth channel, *migration*, wherein components of some units move directly into others, the most obvious example being human im-/emigration, which will tend to generate a direct, mechanical interdependence in addition to strategic ones, only some of which pathways *competition* or *emulation* could cover in Simmons et al.'s (2006) schema.

C. A General Theoretical Model of Interdependence

More general-theoretically, one can show that strategic interdependence arises whenever the actions of some unit(s) affect the marginal utility of alternative actions for some other unit(s). (We follow Brueckner 2003 here; see also Braun and Gilardi 2006.) Consider two units (i,j) with (indirect) utilities, (U^i,U^j), from their alternative actions or policies, (p_i,p_j). Due to externalities, i's utility depends on its policy and that of j. E.g., imagine two countries with (homogenous) population preferences regarding, say, the economy and environment. Due to environmental externalities (e.g., pollution spillovers) and economic ones (e.g., regulatory-cost competition), domestic welfare (i.e., net political-economic benefits/utilities to policymakers) in each country will depend on both countries' actions:

$$U^i \equiv U^i(p_i, p_j) \;\; ; \;\; U^j \equiv U^j(p_j, p_i) \qquad (1).$$

When unit i chooses its policy, p_i, to maximize its own welfare, this alters the optimal policy in j, and *vice versa*. For example, i implementing more (less) effective anti-pollution policy reduces (increases) the need for effective anti-pollution policy in j due to environmental spillovers. We can express such strategic interdependence between i and j with a pair of best-response functions that give i's optimal policies, p_i^*, as a function of j's chosen policies, and *vice versa*:

[3] For interdependence to arise, what is learned must affect the utilities of actors' choices, but it may be objective or subjective, true/correct or false/incorrect, and may regard the politics, economics, sociology, or any other aspect of those choices.

$$p_i^* \equiv \text{Argmax}_{p_i} U^i(p_i, p_j) \equiv R^i(p_j) \;\; ; \;\; p_j^* \equiv \text{Argmax}_{p_j} U^j(p_j, p_i) \equiv R^j(p_i) \qquad (2).$$

The slopes of these best-response functions indicate whether actions by *i* induce *j* to move in the same direction, making *i* and *j strategic complements*, or in opposite directions as *strategic substitutes*. For instance, anti-pollution policies are strategic substitutes in their environmental effects as described above. The best-response functions' slopes depend on these ratios of second cross-partial derivatives:

$$\frac{\partial p_i^*}{\partial p_j} = -U^i_{p_i p_j} \Big/ U^i_{p_i p_i} \;\; ; \;\; \frac{\partial p_j^*}{\partial p_i} = -U^j_{p_j p_i} \Big/ U^j_{p_j p_j} \qquad (3).$$

If the units maximize their utilities, the second-order conditions imply negative denominators in (3). Thus, the slopes depend directly on the signs of the second cross-partial derivatives (i.e., the numerators). If $U^{i,j}_{p_i p_j} > 0$, i.e., if policies are strategic complements, reaction functions slope upward. Regarding any competitive costs of anti-pollution regulation, e.g., increased (reduced) regulation in *i* lowers (raises) the costs of regulation in competitors *j*, and so spurs *j* to increase (reduce) regulation too. If $U^{i,j}_{p_i p_j} < 0$, policies are strategic substitutes, so reaction functions slope downward, as noted regarding in the environmental benefits of anti-pollution regulation. If the second cross-partial derivative is zero, strategic interdependence does not materialize and best-response functions are flat.

Generally speaking, then, positive externalities induce strategic-substitute relations, with policies moving in opposite directions as *free-rider* dynamics obtain. Franzese and Hays (2006b) argue and find such free-riding dynamics in EU active-labor-market policies, for instance. Notice, furthermore, that free-rider advantages also confer late-mover advantages and so war-of-attrition (strategic delay and inaction) dynamics are likely. Conversely, negative externalities induce strategic complementarity, with policies moving in the same direction. The common example of tax-competition has these features. Tax cuts in one jurisdiction have negative externalities for competitors, thereby spurring them to cut taxes as well. These situations advantage early movers, so *competitive races* can unfold.[4] Other good examples

[4] We eschew the labels *races to the bottom* (or *top*) and *convergence* because these competitive races need not foster convergence on any top, bottom, or mean, and could further divergence (see, e.g., Plümper and Schneider 2006).

are competitive currency-devaluations or trade-protection. Early movers in these contexts reap greater economic benefits, so races to move first or earlier are likely. Thus, positive and negative externalities induce strategic-complement and -substitute relations, respectively, which spur competitive-races and free-riding, respectively, with their corresponding early- and late-mover advantages that foster strategic rush to go first on the one hand and delays and inaction on the other.

C. The Empirical-Methodological Challenges of Spatial Interdependence

A crucial challenge for empirical research, known as *Galton's Problem*,[5] is the great difficulty distinguishing true *interdependence* of units' actions, on the one hand, from the impacts of spatially correlated unit-level factors, of common or spatially correlated exogenous-external factors, and of context-conditional factors involving interactions of unit-level and exogenous-external explanators on the other (call these latter, non-spatial components of the model, this complex of correlated responses to correlated unit-level, contextual, or context-conditional factors: *common shocks*). On the one hand, ignoring or inadequately modeling interdependence processes tends to induce overestimation of the importance of common shocks, thereby privileging unit-level/domestic, contextual/exogenous-external, or context-conditional explanations. On the other hand, if the inherent simultaneity of interdependence is insufficiently redressed, then spatial-lag models (see below) yield misestimates (usually overestimates) of interdependence at the expense of common shocks, especially insofar as such common shocks are inadequately modeled. In other words, summarizing analyses in Franzese and Hays (2004,2006a,2007b), obtaining good estimates (unbiased, consistent, efficient) in substantive contexts having any appreciable

[5] Sir Francis Galton originally raised the issue thus: "[F]ull information should be given as to the degree in which the customs of the tribes and races which are compared together are independent. It might be that some of the tribes had derived them from a common source, so that they were duplicate copies of the same original...It would give a useful idea of the distribution of the several customs and of their relative prevalence in the world, if a map were so marked by shadings and colour as to present a picture of their geographical ranges" (*The Journal of the Anthropological Institute of Great Britain and Ireland* 18:270, quoted in Darmofal (2007).) In http://en.wikipedia.org/wiki/Galton's_problem, we find further historical context: "In [1888], Galton was present when Sir Edward Tylor presented a paper at the Royal Anthropological Institute. Tylor had compiled information on institutions of marriage and descent for 350 cultures and examined the correlations between these institutions and measures of societal complexity. Tylor interpreted his results as indications of a general evolutionary sequence, in which institutions change focus from the maternal line to the paternal line as societies become increasingly complex. Galton disagreed, pointing out that similarity between cultures could be due to borrowing, could be due to common descent, or could be due to evolutionary development; he maintained that without controlling for borrowing and common descent one cannot make valid inferences regarding evolutionary development. Galton's critique has become the eponymous *Galton's Problem*, as named by Raoul Naroll (1961, 1965), who proposed [some of] the first statistical solutions."

interdependence of coefficients and standard errors and, *a fortiori*, distinguishing domestic/unit-level, exogenous-external/contextual, and context-conditional factor explanations from interdependence ones, *by any empirical-methodological means*, whether qualitative or quantitative, is *not* straightforward.

The first and primary consideration, as we have previously shown analytically for simple cases and via simulations in more realistic ones, are the relative and absolute theoretical and empirical precisions of the spatial and non-spatial parts of the model, i.e., of the *interdependence* part(s) and the *common-shocks* part(s). To elaborate: the relative and absolute accuracy and power with which the empirical specification of the spatial interdependence reflects and can gain leverage in the data upon the actual interdependence mechanisms operating and with which the domestic, exogenous-external, and/or context-conditional parts of the model reflect and gain leverage upon *common-shocks* alternatives crucially affect the empirical attempt to distinguish and evaluate their relative strength because the two mechanisms produce similar effects. This is the crux of *Galton's Problem.* Inadequacies or omissions in specifying the non-interdependence components of the model tend, intuitively, to induce overestimates of the importance of interdependence and *vice versa*. Secondarily, even if the *common-shocks* and *interdependence* mechanisms are specified properly into the spatial-lag, that (those) regressor(s) will be endogenous (i.e., will covary with residuals), so regression estimates of interdependence-strength (or, equally for that matter, attempts to distinguish interdependence from common shocks qualitatively) will suffer simultaneity biases. Conversely to the primary omitted-variable/misspecification biases described first, these secondary simultaneity biases favor overestimating interdependence-strength, which induces biases in the other direction for, i.e., under-estimation of, non-spatial factors' effects (*common shocks*).

Methodologically, one can discern two approaches to spatial analysis: spatial statistics and spatial econometrics (Anselin 2006). The distinction regards the relative emphasis in spatial-econometric approaches to theoretical models of interdependence processes (e.g., Brueckner 2003, Braun and Gilardi 2006, Franzese and Hays 2007abc) wherein space may often have broad meaning (Beck et al. 2006), well beyond geography and geometry across all manner of social, economic, or political connection that induces effects from outcomes in some units on outcomes in others. The spatial-lag regression model plays a starring role in that tradition (Hordijk 1974; Paelinck and Klaassen 1979; Anselin 1980, 1988, 1992; Haining 1990; LeSage 1999). Anselin (2002) notes that such theory driven models deal centrally with *substantive* spatial correlation, which suggests a corresponding approach to model specification and estimation wherein the importance of spatial interdependence is explored primarily by Wald tests upon the unrestricted spatial-lag model. Spatial-error models, analysis of spatial-correlation patterns, spatial kriging, spatial smoothing, and the like, characterize the more-exclusively data-driven approach and the typically narrower conception of space in solely geographic/geometric terms in the longer spatial-statistics tradition (initially inspired by Sir Galton's famous comments (see note 5), and reaching crucial methodological milestones in Whittle 1954; Naroll 1965, 1970; Cliff and Ord 1973, 1981; Besag 1974; Ord 1975; Ripley 1981; Cressie 1993). Anselin (2002) notes that this approach is often more driven by data problems like measurement error, with spatial correlation often seen as a *nuisance*, which suggests a different approach to model specification and estimation wherein the restricted spatial-error model and Lagrange-multiplier tests are dominant. However, the distinctions are subtle, with considerable and often fruitful cross-fertilization, and both approaches stress the dangers of ignoring spatial interdependence, namely overconfidence and bias, even for those interested primarily or even solely in domestic/unit-level or exogenous-external/contextual matters. Minimally, one should test for spatial interdependence and not proceed non-spatially unless it truly is negligible; otherwise, estimates of domestic/unit-level, exogenous-external/contextual, and/or context-conditional phenomena will be exaggerated. Finally, the most important task in any empirical spatial analysis, by either approach, is the pre-specification of the $N \times N$ spatial-weighting matrix, $\mathbf{W}$,[6] whose elements, w_{ij}, reflect the relative connectivity from unit j to i. As just emphasized: the relative and absolute accuracy and power with which the spatial-lag weights, w_{ij}, reflect and can gain leverage upon the interdependence mechanisms actually operating empirically and with which the domestic, exogenous-external, and/or context-

[6] Strategies for parameterizing $\mathbf{W}$ and estimating such models are of great interest but as yet mostly remain for future work.

conditional parts of the model can reflect and gain leverage upon the *common-shocks* alternatives critically affect the empirical attempt to distinguish and evaluate their relative strength because the two mechanisms produce similar effects.

II. Spatial Autoregressive Models

There are two workhorse regression models in empirical spatial analysis: spatial lag and spatial error models.

A. Spatial Lag Models

Spatial lag models imply spatial externalities in both modeled and unmodeled effects (i.e., the systematic and stochastic components) and are typically motivated by a theoretical model.

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (4);$$

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon} \qquad (5).$$

Note that the multipliers are restricted to be the same. This restriction can be relaxed (discussion below). In the cross-sectional context, the dependent variable, $\mathbf{y}$, is an $N \times 1$ vector observations; ρ is the spatial autoregressive coefficient, reflecting the overall or average strength of interdependence; and $\mathbf{W}$ is an $N{\times}N$ spatial-weighting matrix, with the elements w_{ij} reflecting the relative connectivity from unit *j* to *i*. $\mathbf{Wy}$ is a *spatial lag*; i.e., for each observation y_i, $\mathbf{Wy}$ gives a weighted sum of the y_j, with weights w_{ij}.

B. Spatial Error Models

Spatial error models imply that the pattern of spatial dependence is attributable to unmeasured covariates (i.e., the stochastic component) only. Spatial error specifications are rarely theory-driven.

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{u} \end{aligned} \qquad (6);$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda\mathbf{W})^{-1}\mathbf{u} \qquad (7).$$

Note the spatial moving-average (S-MA) alternative to this spatial autorgressive (S-AR) model, $\boldsymbol{\varepsilon} = \gamma \mathbf{W}\mathbf{u} + \mathbf{u}$, implies local autocorrelation or "pockets" of spatial interdependence because the reduced form does not contain and inverse (Anselin 1995, 2003).

C. Combined Lag and Error Models

A third model combines the two. Different externalities in modeled and unmodeled effects (a.k.a. systematic and stochastic components, which relaxes the previously noted constraint in the spatial-lag model. The resulting mixed SAR model is:

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda \mathbf{W}_2 \boldsymbol{\varepsilon} + \mathbf{u} \end{aligned} \qquad (8).$$

Analogously, a mixed SARMA model would be:

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda \mathbf{W}_2 \mathbf{u} + \mathbf{u} \end{aligned} \qquad (9).$$

III. Model Specification and Estimation

In this section we consider ways to specify and estimate spatial autoregressive models with continuous dependent variables and cross-sectional data (we consider models for time-series-cross-section and binary-choice contexts below). We begin with OLS estimation and specification testing under the null hypothesis of no spatial dependence. We then turn to the topic of estimating spatial lag models, and finish the section with a discussion of spatial error models. To illustrate the methods, we estimate a model of state-level welfare policy generosity in the US using cross-sectional data from Berry, Fording, and Hanson (2003) on the contiguous 48 states.

A. OLS with Specification Testing Under the Null

One approach to model specification is to estimate a non-spatial model using OLS and then conduct a battery of diagnostic tests on the residuals. This strategy makes sense when one does not have a theoretical model of spatial interdependence, and the spatial dependence in the data (if there is any) is seen primarily as a statistical nuisance. One could also perhaps argue for it on grounds of conservatism in that the approach tests upward from spatial-error models that by nature lack spatial dynamics in the systematic component, wherein one's core theoretical and substantive propositions are usually specified. The diagnostic tests can help identify whether the data generating process is spatial autoregressive and, in some cases, even detect the nature of the underlying spatial process (i.e., spatial lag vs. spatial error).

One of the most widely known and frequently used diagnostics for spatial correlation is Moran's I:

$$I = \frac{N}{S}\frac{\boldsymbol{\varepsilon}'\mathbf{W}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}, \text{ where } S = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}, \qquad (10).$$

When **W** is row-standardized (so row elements sum to one), the expression simplifies to:

$$I = \frac{\boldsymbol{\varepsilon}'\mathbf{W}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}} \qquad (11).$$

To test a null of no spatial correlation (in patterns given by **W)**, one can compare a properly standardized Moran's I to the standard normal distribution (Cliff and Ord 1973, Burridge 1980, Kelejian and Prucha 2001).

In addition to Moran's I, several Lagrange multiplier (LM) tests based on OLS residuals exist. The standard LM tests assume that the spatial autoregressive process is either a spatial lag or spatial error model. More precisely, in terms of (8), the standard LM test for the null hypothesis $\rho = 0$ against the spatial lag alternative *assumes* $\lambda = 0$. Likewise, the LM test for $\lambda = 0$ *assumes* $\rho = 0$. The standard, one-directional test against spatial lag alternative is calculated as

$$LM_{\rho} = \frac{\hat{\sigma}_{\varepsilon}^{2}\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\mathbf{y} / \hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{G + T\hat{\sigma}_{\varepsilon}^{2}}, \qquad (12),$$

where $G = (\mathbf{WX}\hat{\boldsymbol{\beta}})'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{WX}\hat{\boldsymbol{\beta}})$ and $T = \text{tr}[(\mathbf{W}' + \mathbf{W})\mathbf{W}]$. The standard, one-directional test against spatial error alternative is

$$LM_{\lambda} = \frac{\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\hat{\boldsymbol{\varepsilon}} / \hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{T} \qquad (13).$$

The drawback with these tests is that they have power against the incorrect alternative, which means they are usually not helpful for making specification choices. Regardless of whether the true spatial autoregressive process is a lag or error process, both tests are likely to reject the null hypothesis. Anselin et al. (1996) present robust LM tests for spatial dependence that are less problematic in this regard. The robust, one-directional test against spatial error alternative treats ρ in the mixed SAR model, (8), as a nuisance parameter and controls for its effect on the likelihood. The statistic is then calculated as

$$LM_{\lambda}^{*}=\frac{\left(\hat{\varepsilon}'W\hat{\varepsilon}/\hat{\sigma}_{\varepsilon}^{2}-\left[T\hat{\sigma}_{\varepsilon}^{2}\left(G+T\hat{\sigma}_{\varepsilon}^{2}\right)^{-1}\right]\hat{\varepsilon}'Wy/\hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{T\left[1-\dfrac{T\hat{\sigma}_{\varepsilon}^{2}}{G+T\hat{\sigma}_{\varepsilon}^{2}}\right]}. \qquad (14).$$

The robust, one-directional test against spatial lag alternative is

$$LM_{\rho}^{*}=G^{-1}\hat{\sigma}_{\varepsilon}^{2}\left(\hat{\varepsilon}'Wy/\hat{\sigma}_{\varepsilon}^{2}-\hat{\varepsilon}'W\hat{\varepsilon}/\hat{\sigma}_{\varepsilon}^{2}\right)^{2}. \qquad (15).$$

The two-directional LM test, finally, can be decomposed into the robust LM test for one alternative (lag or error) and the standard LM test for the other:

$$LM_{\rho\lambda}=LM_{\lambda}+LM_{\rho}^{*}=LM_{\rho}+LM_{\lambda}^{*} \qquad (16).$$

The one-directional test statistics are distributed χ_1^2 while the two-directional statistic is distributed χ_2^2. Using Monte Carlo simulations, Anselin et al. (1996) show that all five tests have the correct size in small samples. I.e., they all reject the null hypothesis at the stated rate when the null is true. The robust LM tests have lower power compared with the standard ones against the correct alternative, but the loss is relatively small and the robust tests are less likely to reject the null against the wrong alternative.

So, for example, when the true data generating process is a spatial AR error model ($\lambda \neq 0, \rho = 0$), rejection rates for LM_{λ} are about 5 percentage points higher on average across the range of λ than for LM_{λ}^{*}. The robustness of LM_{ρ}^{*} relative to LM_{ρ} is clear in this experiment. At $\lambda = .9$, LM_{ρ} rejects in favor of the incorrect alternative 89.9% of the time whereas LM_{ρ}^{*} rejects 17.1% of the time. The power advantage of the standard LM test is smaller when the true data generating process is a spatial AR lag model ($\lambda = 0, \rho \neq 0$). Rejection rates for LM_{ρ} are less than 2 percentage points higher on average than for LM_{ρ}^{*} across the full range of ρ. At $\rho = .9$, LM_{λ} rejects in favor of the incorrect alternative 100% of the time whereas LM_{λ}^{*} rejects 0.6% of the time. It seems the reduced power for increased robustness tradeoff strongly favors that the robust LM tests be included in the set of diagnostics.[7]

To help illustrate how these tests can be used in empirical research, we present OLS estimates for a non-spatial model of welfare policy generosity in column 1 of Table 1. All variables in our illustrative analysis are states' averages over the five years 1986-1990. The dependent variable is the maximum monthly AFDC benefit, and the independent variables are the state's poverty rate, average monthly wage in the retail sector, government ideology (ranging from 0=conservative to 100=liberal), degree of interparty competition (ranging from .5=competitive to 1.0=non-competitive, tax effort (revenues as a percentage of tax "capacity"), and the portion of AFDC benefits paid by the federal government. We use a standardized binary contiguity-weights matrix, which begins by coding w_{ij} = 1 for states i and j that share a border and w_{ij} = 0 for states that do not border. Then, we row-standardize (as commonly done in spatial-econometrics) the resulting matrix by dividing each cell in a row by that row's sum. This gives the unweighted average of the dependent variable in "neighboring" (so-defined) states.

<Table 1 About Here>

The results for our non-spatial model suggest that high tax effort and low party competition are associated with more generous AFDC benefit payments. This seems reasonable. However, if the data exhibit spatial dependence, we need to worry about validity of these inferences. To check this possibility, we implement the diagnostic tests outlined above starting with Moran's I. The value of the standardized Moran-I test statistic is 3.312, which is statistically significant. We can reject the null hypothesis of no spatial dependence. We also include the LM tests. The result of the two-directional test leads to the same conclusion. Both the standard one-directional tests seem, predictably, statistically significant, which, unfortunately, gives us little guidance for specification. As expected, the robust one-directional tests are more helpful in this regard. The robust test against the spatial lag alternative is statistically significant

[7] See Anselin et al. (1996) Tables 3-6. These results are for the N=40 experiments.

while the robust test against the spatial error alternative is not. This suggests a spatial lag specification. We conclude with a warning. Ignoring evidence of spatial dependence can be extremely problematic, especially if the data suggest the true source of dependence is a spatial-lag process. In this case, simple OLS is likely to provide inaccurate coefficient estimates, particularly for variables that happen to cluster spatially (e.g., Franzese and Hays 2004, 2006a, 2007b).

B. Estimating Lag Models

The spatial lag model has become a very popular specification in social science research. One might arrive to this model via batteries of diagnostic tests or directly from theory. The theory-driven approach starts by estimating the spatial model, and then uses Wald, LR, and related tests to refine the specification. We begin with OLS estimation of spatial lag models, which we label spatial OLS (S-OLS).

1. Spatial OLS

Spatial OLS is inconsistent. To see this, we start by rewriting the spatial lag model as

$$\mathbf{y}=\mathbf{Z}\boldsymbol{\delta}+\boldsymbol{\varepsilon},\ \text{ where }\ \mathbf{Z}=\begin{bmatrix}\mathbf{Wy} & \mathbf{X}\end{bmatrix}\ \text{ and }\ \boldsymbol{\delta}=\begin{bmatrix}\rho & \boldsymbol{\beta}\end{bmatrix}' \qquad (17).$$

The matrices $\mathbf{Z}$ and $\boldsymbol{\delta}$ have dimensions Nx(k+1) and (k+1)x1 respectively. The asymptotic simultaneity bias for the S-OLS estimator is given by

$$\operatorname{plim}\hat{\boldsymbol{\delta}}=\boldsymbol{\delta}+\operatorname{plim}\left[\left(\frac{\mathbf{Z'Z}}{n}\right)^{-1}\frac{\mathbf{Z'}\boldsymbol{\varepsilon}}{n}\right]. \qquad (18).$$

In the case where $\mathbf{Z}$ is a single exogenous regressor, $\mathbf{x}$, $\left(k=1,\operatorname{cov}(\boldsymbol{\varepsilon},\mathbf{x})=0\right)$, we can rewrite (18) as

$$\operatorname{plim}\hat{\boldsymbol{\delta}}=\begin{bmatrix}\rho\\ \beta\end{bmatrix}+\begin{bmatrix}\dfrac{\operatorname{var}(\mathbf{x})}{\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{Wy})-\left[\operatorname{cov}(\mathbf{x},\mathbf{Wy})\right]^2} & \dfrac{-\operatorname{cov}(\mathbf{x},\mathbf{Wy})}{\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{Wy})-\left[\operatorname{cov}(\mathbf{x},\mathbf{Wy})\right]^2}\\ \dfrac{-\operatorname{cov}(\mathbf{x},\mathbf{Wy})}{\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{Wy})-\left[\operatorname{cov}(\mathbf{x},\mathbf{Wy})\right]^2} & \dfrac{\operatorname{var}(\mathbf{Wy})}{\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{Wy})-\left[\operatorname{cov}(\mathbf{x},\mathbf{Wy})\right]^2}\end{bmatrix}\begin{bmatrix}\operatorname{cov}(\boldsymbol{\varepsilon},\mathbf{Wy})\\ \operatorname{cov}(\boldsymbol{\varepsilon},\mathbf{x})\end{bmatrix} \qquad (19).$$

If we define $\boldsymbol{\Psi}=\operatorname{plim}\left(\dfrac{\mathbf{Z'Z}}{n}\right)$ and $\boldsymbol{\Gamma}=\operatorname{plim}\left(\dfrac{\mathbf{Z'}\boldsymbol{\varepsilon}}{n}\right)$, and do the matrix multiplication, (19) simplifies to

$$\operatorname{plim} \hat{\boldsymbol{\delta}} = \begin{bmatrix} \rho \\ \beta \end{bmatrix} + \begin{bmatrix} \dfrac{\Psi_{22}\Gamma_{11}}{|\boldsymbol{\Psi}|} \\ -\dfrac{\Psi_{12}\Gamma_{11}}{|\boldsymbol{\Psi}|} \end{bmatrix} \quad (20).$$

Since $\boldsymbol{\Psi}$ is a variance-covariance matrix, its determinant is strictly positive. With positive (negative) spatial dependence in the data, the covariances Ψ_{12} and Γ_{11} are positive (negative), and S-OLS will overestimate (underestimate) ρ and underestimate (overestimate) β. This is one of the analytic results we stressed repeatedly earlier: the simultaneity biases of S-OLS tend to induce exaggerated estimates of interdependence strength and correspondingly deflated estimates of the importance of non-spatial factors.

The S-OLS estimates are provided in column 2 of Table 1. Consistent with the results from our diagnostic tests, the estimated coefficient on the spatial lag is large, positive and statistically significant. The OLS estimates most affected by the switch to a spatial-lag specification are the party-competition and tax-effort coefficients, which become statistically insignificant. Conversely to S-OLS's simultaneity biases, the OLS coefficient estimates on these two variables may, because they cluster spatially, have suffered from omitted variable bias that would have inflated those estimates.

Franzese and Hays (2004, 2006a, 2007b) conclude that spatial OLS, despite its simultaneity, can perform acceptably under low-to-moderate interdependence-strength and reasonable sample-dimensions. Given our results, S-OLS is clearly preferable to OLS. In this particular case, however, both the size of the spatial-lag coefficient and the fact that no other coefficients are statistically significant should raise concern about simultaneity bias. We have advised using some consistent estimator under conditions like these. We discuss three consistent estimators below, starting with spatial-2SLS and spatial-GMM.

2. Spatial 2SLS and Spatial GMM

Spatial-2SLS and spatial-GMM provide consistent estimates for the coefficient on the spatial lag and use spatially weighted values of the exogenous variables in other units as instruments. The latter extends the former to account the heteroscedasticity in the quadratic form of the sample orthogonality conditions. If this particular form of heteroscedasticity is present, the S-GMM estimator yields smaller asymptotic variance than the spatial-2SLS estimator. If it is absent, the two estimators are equivalent.[8] Note that a mixed spatial autoregressive model of the form in (8) would suffer from heteroscedasticity, making S-GMM more efficient for estimating $\boldsymbol{\delta}$. (In this particular case, a generalized S-2SLS estimator using a Cochrane-Orcutt like transformation of the data is also available; see Kelejian and Prucha 1998, 1999).

To see how we estimate the spatial lag model (17) using S-2SLS, define the linear prediction of $\mathbf{Wy}$:

$$\widehat{\mathbf{Wy}} = \boldsymbol{\Pi}[\boldsymbol{\Pi})'(\boldsymbol{\Pi})]^{-1}\boldsymbol{\Pi}'\mathbf{Wy} \qquad (21),$$

where $\boldsymbol{\Pi}$ is the full set of *exogenous* variables including, at least, $\mathbf{X}$ and $\mathbf{WX}$. $\mathbf{WX}$ provides *spatial*-instruments.[9] Thus, $\boldsymbol{\Pi}$ is an N x L matrix, where $L \geq 2k$. The orthogonality condition for the 2SLS estimator is formally written as $E\left[\boldsymbol{\Pi\varepsilon}\right] = 0$. Next, define $\hat{\mathbf{Z}}$ as an Nx(k+1) matrix of the predicted values of $\mathbf{Wy}$ and $\mathbf{X}$,

$$\hat{\mathbf{Z}} = \left[\widehat{\mathbf{Wy}} \quad \mathbf{X}\right]. \qquad (22).$$

Using this definition, the spatial-2SLS estimator is

$$\hat{\boldsymbol{\delta}}_{\text{S2SLS}} = (\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}'\mathbf{y} \qquad (23);$$

$$\widehat{\text{var}}\left(\hat{\boldsymbol{\delta}}_{\text{S2SLS}}\right) = s^2(\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1} \qquad (24).$$

where s^2 is calculated from residuals in the original structural model, (17), with $\hat{\boldsymbol{\delta}}_{\text{S2SLS}}$ substituted for $\boldsymbol{\delta}$.

The GMM estimator minimizes a weighted quadratic form of the sample moment conditions derived from the orthogonality assumptions. More specifically, this criterion is

$$q = E[\boldsymbol{\mu}(\boldsymbol{\delta})\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}(\boldsymbol{\delta})'], \qquad (25),$$

with the corresponding moment conditions:

[8] When the number of excluded exogenous variables exactly equals the number of endogenous variables, the GMM, 2SLS, and ILS estimators are equivalent. Therefore, we could more accurately say that GMM improves on 2SLS when the coefficients in the system/equation are overidentified and heteroscedasticity exists. In this case, we have one endogenous regressor, the spatial lag, in one equation. Provided the number of exogenous variables in $\mathbf{X}$ exceeds one, the number of spatial instruments will exceed one, making the coefficient on the spatial lag overidentified.

[9] One can also include higher order spatial instruments in $\boldsymbol{\Pi}$ --that is, $\left\{\mathbf{W}^2\mathbf{X}, \mathbf{W}^3\mathbf{X}, \mathbf{W}^4\mathbf{X}, \ldots\right\}$.

$$\boldsymbol{\mu}(\boldsymbol{\delta}) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\pi}_i(y_i - \mathbf{z}_i'\boldsymbol{\delta}) \tag{26;}$$

$$\boldsymbol{\Sigma} = E\left[\boldsymbol{\mu}(\delta)\boldsymbol{\mu}(\delta)'\right] = \frac{1}{N}E\left[\sum_{i=1}^{N}\boldsymbol{\pi}_i\boldsymbol{\pi}_i'(y_i - \mathbf{z}_i'\boldsymbol{\delta})^2\right] = \frac{1}{N}\sum_{i=1}^{N}\omega\boldsymbol{\pi}_i\boldsymbol{\pi}_i = \frac{1}{N}\left(\boldsymbol{\Pi}'\boldsymbol{\Omega}\boldsymbol{\Pi}\right) \tag{27.}$$

In these equations, $\boldsymbol{\pi}_i$ is a column vector (lx1) that is the transpose of the i^{th} row of $\boldsymbol{\Pi}$ (representing the i^{th} observation) and, similarly, $\mathbf{z}_i$ is a (k+1)x1 vector that is the transpose of the i^{th} row of $\mathbf{Z}$. The GMM weighting matrix is calculated by inverting a consistent estimate of the variance-covariance matrix of the moment conditions.[10] White's estimator provides a consistent non-parametric estimate of $\boldsymbol{\Sigma}$ provided we have a consistent estimator of $\boldsymbol{\delta}$ (Anselin 2006). Fortunately, spatial-2SLS can provide these. Thus, the estimate for $\boldsymbol{\Sigma}$ is

$$\mathbf{S}_0 = \sum_{i=1}^{N}\boldsymbol{\pi}_i\boldsymbol{\pi}_i'\left(y_i - \mathbf{z}_i'\hat{\boldsymbol{\delta}}_{\text{S2SLS}}\right)^2 \tag{28,}$$

and the GMM estimator for $\boldsymbol{\delta}$ is

$$\hat{\boldsymbol{\delta}}_{\text{SGMM}} = \left[\mathbf{Z}'\boldsymbol{\Pi}(\mathbf{S}_0)^{-1}\boldsymbol{\Pi}'\mathbf{Z}\right]^{-1}\left[\mathbf{Z}'\boldsymbol{\Pi}(\mathbf{S}_0)^{-1}\boldsymbol{\Pi}'\mathbf{y}\right] \tag{29;}$$

$$\widehat{\text{var}\left(\hat{\boldsymbol{\delta}}_{\text{SGMM}}\right)} = \left[\mathbf{Z}'\boldsymbol{\Pi}\left(\hat{\mathbf{S}}_0^{-1}\right)\boldsymbol{\Pi}'\mathbf{Z}\right]^{-1} \tag{30.}$$

We present the S-2SLS and S-GMM estimates for the spatial-lag model of welfare policy generosity in columns 3 and 4 of Table 1. The S-2SLS estimates for this particular specification and dataset are troubling as the spatial-lag coefficient estimate exceeds one, giving a non-stationary spatial process. This is a bit surprising when compared with the smaller S-OLS result, given that the S-OLS estimator has likely-inflationary simultaneity biases and S-2SLS likely does not. Of course, this can happen with a single sample and/or if the exogeneity of the instruments is violated.[11] The S-GMM estimates are better. The spatial-lag coefficient estimate is well below one (though it is still large) and the standard errors are

[10] The logic here is similar to that behind the WLS estimator. OLS is consistent in the presence of heteroscedasticity, but WLS is more efficient. Likewise, 2SLS is consistent under heteroscedasticty, but GMM is asymptotically more efficient.

[11] Franzese and Hays (2004) show that the exogeneity at issue here is that the y_i must not cause the x_j, a condition we call (no) cross-spatial endogeneity. Such reverse "diagonal" causality seems unlikely to arise in many substantive contexts, although we do note also that spatial correlation among the other regressors plus the typical endogeneity from y to x would create it.

about 5% smaller than the S-2SLS standard errors on average, as expected given the likely efficiency of the GMM estimator. The coefficients on government ideology and on party competition are statistically significant. The results suggest that, *ceteris paribus*, welfare benefits are highest in states with non-competitive party systems and liberal governments.

3. Spatial Maximum Likelihood

Implementing S-ML is not complicated, although the spatial-lag model adds a slight wrinkle to the standard linear additive case, and the maximization can be computationally intense for large samples. To see the minor complication, start by isolating the stochastic component of the spatial-lag model:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \Rightarrow \boldsymbol{\varepsilon} = \left(\mathbf{I} - \rho\mathbf{W}\right)\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \equiv \mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \qquad (31).$$

Assuming *i.i.d.* normality, the likelihood function for $\boldsymbol{\varepsilon}$ is then just the typical linear one:

$$L(\boldsymbol{\varepsilon}) = \left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{N}{2}} \exp\left(-\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^2}\right) \qquad (32),$$

which, in this case, will produce a likelihood in terms of $\mathbf{y}$ as follows:

$$L(\mathbf{y}) = |\mathbf{A}|\left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right) \qquad (33),$$

and the log-likelihood takes the form

$$\ln L(\mathbf{y}) = \ln|\mathbf{A}| - \left(\frac{N}{2}\right)\ln\left(2\pi\right) - \left(\frac{N}{2}\right)\ln\sigma^2 - \left(\frac{1}{2\sigma^2}\left(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right) \qquad (34).$$

This still resembles the typical linear-normal likelihood, except that the transformation from $\boldsymbol{\varepsilon}$ to $\mathbf{y}$, is not by the usual factor of 1, but by $|\mathbf{A}|=|\mathbf{I}-\rho\mathbf{W}|$. Since $|\mathbf{A}|$ depends on ρ, each recalculation of the likelihood in maximization routine must recalculate this determinant for the updated ρ-values. Ord's (1975) solution to this computational-intensity issue was to approximate $|\mathbf{W}|$ by $\boldsymbol{\Pi}_i \lambda_i$ because the eigenvector $\boldsymbol{\lambda}$ in this approximation does not depend on ρ. Then $|\mathbf{I}-\rho\mathbf{W}| = \boldsymbol{\Pi}_i\left(1-\lambda_i\right)$, which requires the estimation routine only to recalculate a product, not a determinant, as it updates.[12] The estimated variance-covariances of parameter estimates follow the usual ML formula (negative the inverse of Hessian of the likelihood) and so are also functions of $|\mathbf{A}|$. The same approximation serves there.

Typically, estimation proceeds by maximizing a concentrated-likelihood. Given an estimate of the spatial-lag coefficient, ρ, an analytic optimum estimate of the non-spatial coefficients can be found as:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \rho(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} = \hat{\boldsymbol{\beta}}_{\mathbf{O}} - \rho.\hat{\boldsymbol{\beta}}_{\mathrm{L}} \qquad (35).$$

Note that the first term in the second two expressions of (35) is just the OLS regression of $\mathbf{y}$ on $\mathbf{X}$, and the second term is just ρ times the OLS regression of $\mathbf{Wy}$ on $\mathbf{X}$. Both of these rely only on observables, (except for ρ), and so are readily calculable given some ρ (estimate). Next, define these terms:

$$\hat{\boldsymbol{\varepsilon}}_{\mathbf{O}} = \mathbf{y}\text{ - }\mathbf{X}\boldsymbol{\beta}_{\mathbf{O}} \text{ and } \hat{\boldsymbol{\varepsilon}}_{\mathrm{L}} = \mathbf{W}\mathbf{y}\text{ - }\mathbf{X}\boldsymbol{\beta}_{\mathrm{L}} \qquad (36).$$

It then follows that

$$\hat{\sigma}^2 = (1/N)(\hat{\boldsymbol{\varepsilon}}_{\mathbf{O}} - \rho\hat{\boldsymbol{\varepsilon}}_{\mathrm{L}})'(\hat{\boldsymbol{\varepsilon}}_{\mathbf{O}} - \rho\hat{\boldsymbol{\varepsilon}}_{\mathrm{L}}) \qquad (37)$$

is the S-ML estimate of the standard-error of the regression, and

$$\ln L_C(\mathbf{y}) = -\left(\frac{N}{2}\right)\ln\pi + \ln|\mathbf{A}| - \frac{N}{2}\ln\left(\frac{1}{N}(\boldsymbol{\varepsilon}_{\mathbf{O}} - \rho\boldsymbol{\varepsilon}_{\mathrm{L}})'(\boldsymbol{\varepsilon}_{\mathbf{O}} - \rho\boldsymbol{\varepsilon}_{\mathrm{L}})\right) \qquad (38)$$

yields the S-ML estimate of ρ which is substituted into (35) to get $\hat{\boldsymbol{\beta}}$. The procedure may be iterated, and estimated variance-covariances of parameter estimates derive from the information matrix as usual, although they could also be bootstrapped.

The S-ML estimates for our spatial lag model of welfare policy generosity are provided in column 5 of Table 1. These estimates are mostly similar to the S-GMM estimates. The most noteable difference is in the estimate of ρ. The S-ML coefficient is approximately 36% smaller than the S-GMM coefficient, and it is estimated much more precisely, the standard error being about half the size of the S-GMM

[12] Unfortunately, the approximation may be numerically unstable (Anselin 1988, 2001; Kelejian and Prucha 1998). On the other hand, S-ML may enjoy a practical advantage over S-2SLS in multiple-W models in that S-ML does not require differentiated instrumentation for each **W** to gain distinct leverage on its corresponding ρ. The instruments, **WX**, would differ by virtue of **W** differing for the alternative interdependence processes, so S-2SLS is estimable for multiple-**W** models even with identical **X** in the **WX** instruments, but we harbor doubts about the practical identification leverage obtainable thereby.

standard error. Three of the coefficients in this model are statistically significant including the tax effort coefficient. The S-ML estimates imply welfare benefits are systematically larger, all else equal, in states with high taxes, liberal governments, non-competitive party systems. Franzese and Hays (2004, 2006a, 2007b) find that S-ML generally outperforms S-2SLS on mean squared error grounds. S-GMM lessens the efficiency advantage for S-ML over the IV class of estimators.

C. Estimating Error Models

If specification tests indicate that spatial dependence is of the form in (6), OLS coefficient estimates are consistent, but standard-error estimates will be biased. One could combine OLS coefficient estimates with robust standard errors (e.g., PCSE's: Beck and Katz 1995, 1996). Another option is to estimate a spatial-error model. In this section we consider the maximum-likelihood estimator for this model. Again, we start by isolating the stochastic component, which, in this case, is

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\equiv\mathbf{X}\boldsymbol{\beta}+\left(\mathbf{I}-\lambda\mathbf{W}\right)^{-1}\mathbf{u}\Rightarrow\mathbf{u}=\left(\mathbf{I}-\lambda\mathbf{W}\right)\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\equiv\mathbf{B}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right) \quad (39).$$

The likelihood for a spatial error process is

$$L(\mathbf{y})=|\mathbf{B}|\left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{N}{2}}\exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\mathbf{B}'\mathbf{B}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right) \quad (40),$$

where $|\mathbf{B}|=|\mathbf{I}-\lambda\mathbf{W}|$, and the log-likelihood takes the form

$$\ln L(\mathbf{y})=\ln|\mathbf{B}|-\left(\frac{N}{2}\right)\ln\left(2\pi\right)-\left(\frac{N}{2}\right)\ln\sigma^2-\left(\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\mathbf{B}'\mathbf{B}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right) \quad (41).$$

We first calculate OLS residuals, and then estimate λ by maximizing the concentrated likelihood:

$$\ln L_C(\mathbf{y})=-\left(\frac{N}{2}\right)\ln\left(2\pi\right)+\ln|\mathbf{B}|-\frac{N}{2}\ln\left(\frac{1}{N}(\hat{\boldsymbol{\varepsilon}}'\mathbf{B}'\mathbf{B}\hat{\boldsymbol{\varepsilon}})\right) \quad (42).$$

Given λ, the ML estimates for $\boldsymbol{\beta}$ are calculated using FGLS

$$\hat{\boldsymbol{\beta}}_{ML}=(\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{y} \quad (43).$$

The asymptotic variance covariance matrix for $\hat{\boldsymbol{\beta}}_{ML}$ is

$$\text{var}\left(\hat{\boldsymbol{\beta}}_{ML}\right) = \hat{\sigma}^2 \left(\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X}\right)^{-1} \quad (44).$$

where $\hat{\sigma}^2 = (1/N)(\hat{\boldsymbol{\varepsilon}}'\mathbf{B}'\mathbf{B}\,\hat{\boldsymbol{\varepsilon}})$ and $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ML}$. The asymptotic variance for λ is

$$\text{var}\left(\hat{\lambda}\right) = 2\text{tr}\left(\mathbf{W}\mathbf{B}^{-1}\right)^2 \quad (45).$$

The S-ML estimates for the spatial error model of welfare policy generosity are provided in the last column of Table 1. We note only that the log-likelihood value for the error model is less than the log-likelihood for the lag model, and this is consistent with the robust LM specification test results.

IV. Calculating and Presenting Spatial Effects

Calculation and presentation of *effects* in empirical models with spatial interdependence, as in any model beyond the purely linear-additive, involve more than simply considering coefficient estimates. In empirical models containing *spatial dynamics*, as in those with only temporal dynamics, coefficients on explanatory variables give only the *pre-dynamic* impetuses to the outcome variable from increases in those variables. This represents the pre-interdependence impetus, which, incidentally, is unobservable if spatial dynamics are instantaneous (i.e., incur within observation period). This section discusses calculation of spatial multipliers, which allow expression of the *effects* of counterfactual shocks across units, and it applies the delta-method to compute standard errors for these *effects*.[13]

$$\begin{aligned}
\mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\
&= (\mathbf{I_N} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= \begin{bmatrix} 1 & -\rho w_{1,2} & \cdots & \cdots & -\rho w_{1,\mathrm{N}} \\ -\rho w_{2,1} & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 1 & -\rho w_{(\mathrm{N}-1),\mathrm{N}} \\ -\rho w_{\mathrm{N},1} & \cdots & \cdots & -\rho w_{\mathrm{N},(\mathrm{N}-1)} & 1 \end{bmatrix}^{-1} \left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \\
&= \mathbf{M}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)
\end{aligned} \quad (46)$$

Denote the i^{th} column of $\mathbf{M}$ as $\mathbf{m}_i$ and its estimate as $\hat{\mathbf{m}}_i$. If we are interested in the spatial effects of a

[13] For an excellent discussion of spatial multipliers, see Anselin (2003). For an application (without standard errors), see Kim, Phipps, and Anselin (2003).

one-unit increase in explanatory variable k in country i, we calculate $\frac{dx_{i,k}\beta_{i,k}}{dx_{i,k}}$. The effect of this change on country i's neighbors is $\mathbf{M}\frac{dx_{i,k}\beta_{i,k}}{dx_{i,k}}$ or simply, $\mathbf{m}_i\beta_k$.

The standard errors calculation, using the delta method, is

$$\operatorname{var}\left(\hat{\mathbf{m}}_i\hat{\beta}_k\right)=\left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]\operatorname{var}\left(\hat{\boldsymbol{\theta}}\right)\left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]', \text{ where } \hat{\boldsymbol{\theta}}=\begin{bmatrix}\hat{\rho}\\ \hat{\beta}_k\end{bmatrix} \text{ and } \left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]=\begin{bmatrix}\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\rho}} & \mathbf{m}_i\end{bmatrix} \quad (47)$$

The vector $\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\rho}}$ is the i^{th} column of $\beta_k\frac{\partial\hat{\mathbf{M}}}{\partial\hat{\rho}}$. Since $\mathbf{M}$ is an inverse matrix, the derivative in equation (47) is calculated as $\frac{\partial\hat{\mathbf{M}}}{\partial\hat{\rho}}=-\hat{\mathbf{M}}\frac{\partial\hat{\mathbf{M}}^{-1}}{\partial\hat{\rho}}\hat{\mathbf{M}}=-\hat{\mathbf{M}}\left(\mathbf{I}-\rho\mathbf{W}\right)\hat{\mathbf{M}}=-\hat{\mathbf{M}}(-\mathbf{W})\hat{\mathbf{M}}=\hat{\mathbf{M}}\mathbf{W}\hat{\mathbf{M}}$. We do not calculate and present the spatial effects implied by the models in Table 1. Instead, we concentrate on calculating spatio-temporal effects using one of the panel models in the next section. These spatio-temporal calculations are slightly more complicated than the purely spatial ones.

V. Extensions

In this section we consider several newer applications of spatial techniques in empirical political-science research: SAR models with multiple lags, SAR models for binary dependent variables, and STAR models for panel data.

A. Spatial Autoregressive Models with Multiple Lags

One innovation in the booming literature on policy and institutional diffusion in recent years is the use of spatial autoregressive models with multiple lags to evaluate distinct diffusion mechanisms (Simmons and Elkins 2004; Elkins, Guzman and Simmons 2006; Lee and Strang 2006). This section briefly highlights some of the difficulties involved in estimating these models, focusing on the linear additive case. Again, there are two main versions of this model. Brandsma and Ketellapper (1979) and Dow (1984) estimate biparametric error models of the form

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda_1 \mathbf{W}_1 \boldsymbol{\varepsilon} + \lambda_2 \mathbf{W}_2 \boldsymbol{\varepsilon} + \mathbf{u}\end{aligned}, \qquad (48).$$

using the maximum likelihood technique described in IV.C. In this case, the concentrated likelihood contains $|\mathbf{B}|=|\mathbf{I}-\lambda_1\mathbf{W}_1-\lambda_2\mathbf{W}_2|$ (see (42)). Again, the OLS estimator for $\boldsymbol{\beta}$ is consistent but inefficient when the spatial dependence takes the form in (48). Lacombe (2004) estimates a biparametric lag model:

$$\mathbf{y} = \rho_1 \mathbf{W}_1 \mathbf{y} + \rho_2 \mathbf{W}_2 \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (49).$$

As with the single-spatial-lag model, S-OLS estimation of the biparametric model suffers simultaneity bias. However, the problem is potentially worse in the case of multiple spatial lags with its two or more endogenous variables rather than one. To see this, first rewrite the model (without exogenous regressors):

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\rho} + \boldsymbol{\varepsilon} \text{ where } \mathbf{Z} = \begin{bmatrix} \mathbf{W}_1\mathbf{y} & \mathbf{W}_2\mathbf{y} \end{bmatrix} \text{ and } \boldsymbol{\rho} = \begin{bmatrix} \rho_1 & \rho_2 \end{bmatrix}' \qquad (50).$$

The *asymptotic* simultaneity bias for the S-OLS estimator is given by

$$\text{plim } \hat{\boldsymbol{\rho}} = \boldsymbol{\rho} + \text{plim} \left[\left(\frac{\mathbf{Z'Z}}{n} \right)^{-1} \frac{\mathbf{Z'}\boldsymbol{\varepsilon}}{n} \right] \qquad (51),$$

which can be written as

$$\text{plim } \hat{\boldsymbol{\rho}} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} + \begin{bmatrix} \dfrac{\text{var}(\mathbf{W}_2\mathbf{y})}{\text{var}(\mathbf{W}_2\mathbf{y})\,\text{var}(\mathbf{W}_1\mathbf{y}) - [\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})]^2} & \dfrac{-\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})}{\text{var}(\mathbf{W}_2\mathbf{y})\,\text{var}(\mathbf{W}_1\mathbf{y}) - [\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})]^2} \\ \dfrac{-\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})}{\text{var}(\mathbf{W}_2\mathbf{y})\,\text{var}(\mathbf{W}_1\mathbf{y}) - [\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})]^2} & \dfrac{\text{var}(\mathbf{W}_1\mathbf{y})}{\text{var}(\mathbf{W}_2\mathbf{y})\,\text{var}(\mathbf{W}_1\mathbf{y}) - [\text{cov}(\mathbf{W}_1\mathbf{y},\mathbf{W}_2\mathbf{y})]^2} \end{bmatrix} \begin{bmatrix} \text{cov}(\boldsymbol{\varepsilon},\mathbf{W}_1\mathbf{y}) \\ \text{cov}(\boldsymbol{\varepsilon},\mathbf{W}_2\mathbf{y}) \end{bmatrix} \qquad (52).$$

If we define $\boldsymbol{\Psi} = \text{plim}\left(\frac{\mathbf{Z'Z}}{n} \right)$ and $\boldsymbol{\Gamma} = \text{plim}\left(\frac{\mathbf{Z'}\boldsymbol{\varepsilon}}{n} \right)$, and carry out the matrix multiplication, equation (52) simplifies to

$$\text{plim } \hat{\boldsymbol{\rho}} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} + \begin{bmatrix} \dfrac{\Psi_{22}\Gamma_{11} - \Psi_{12}\Gamma_{21}}{|\boldsymbol{\Psi}|} \\ \dfrac{\Psi_{11}\Gamma_{21} - \Psi_{12}\Gamma_{11}}{|\boldsymbol{\Psi}|} \end{bmatrix} \qquad (53).$$

Since $\boldsymbol{\Psi}$ is a variance-covariance matrix, its determinant is strictly positive $\left(\left[\Psi_{11}\Psi_{22} - \Psi_{12}^2 \right] > 0 \right)$.

Therefore, if we assume that 1) there is positive spatial dependence $(\boldsymbol{\rho}, \Psi_{12}, \Gamma_{11} > 0)$, 2) the spatial lags have the same degree of endogeneity $(\Gamma_{11} = \Gamma_{21})$, and 3) the variance of $\mathbf{W}_2\mathbf{y}$ is less than the variance of $\mathbf{W}_1\mathbf{y}$ $(\Psi_{22} < \Psi_{11})$, it follows that S-OLS will underestimate ρ_1 and overestimate ρ_2 *asymptotically*.

Fortunately, the maximum likelihood estimator can be implemented in almost the same manner described in IV.B.3. In the biparametric case, the error term is

$$\boldsymbol{\varepsilon} = (\mathbf{I}_N - \rho_1\mathbf{W}_1 - \rho_2\mathbf{W}_2)\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \equiv \mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \qquad (54).$$

With this change, the likelihood function in (38) can be used for estimation. The main practical difficulty in maximizing the concentrated likelihood is calculating the log-determinant of **A**. Lacombe (2004) addresses this difficulty by calculating $\log|\mathbf{A}|$ over a grid of values for ρ_1 and ρ_2 prior to estimation. His routine calls values from this table during the optimization process.

B. Spatial Models for Binary Outcomes

The methods for estimating and analyzing spatial latent variable models for categorical data have received significant attention in the literature recently. Much of the methodological research has focused on the spatial probit model (e.g., McMillen 1992, LeSage 2000). This is also one of the most frequently used models in the applied research (Beron et al. 2003, Simmons and Elkins 2004). In this section we consider spatial models for binary outcomes, starting with the spatial lag probit model.

1. Spatial Lag Probit Models

The structural model for the spatial probit takes the form

$$\mathbf{y}^* = \rho\mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (55)$$

which can be written in its reduced form as

$$\mathbf{y}^* = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad (56)$$

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

$$y_i = \begin{cases} 1 \text{ if } y_i^* > 0 \\ 0 \text{ if } y_i^* \le 0 \end{cases}. \tag{57}$$

The marginal probabilities are calculated as follows

$$\Pr(y_i = 1 \mid \mathbf{x}_i) = \Pr\left((\mathbf{I} - \rho\mathbf{W})_i^{-1} \mathbf{x}_i'\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})_i^{-1} \varepsilon_i > 0 \right)$$

or

$$\Pr(y_i = 1 \mid \mathbf{x}_i) = \Pr\left(u_i < \frac{(\mathbf{I} - \rho\mathbf{W})_i^{-1} \mathbf{x}_i'\boldsymbol{\beta}}{\sigma_i} \right) \tag{58}$$

using the marginal distribution from a multivariate normal with variance-covariance matrix $[(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})]^{-1}$. The denominator in equation (58), which is the square root of the variance for unit *i*, is attributable to the heteroscedasticity induced by the spatial dependence. This heteroscedasticity distinguishes the spatial probit from the conventional probit and makes the estimator for the latter inconsistent for the spatial case. The fact that the u_i are interdependent also makes the standard probit estimator inappropriate for the spatial model. One does not sum the log of *n* one-dimensional probabilities to estimate the model, but rather calculates the log of one *n*-dimensional normal probability.

Beron et al. (2003) proposed estimation by simulation using recursive importance sampling (RIS), which is discussed extensively in Vijverberg (1997). LeSage (2000) has suggested using Bayesian Markov Chain Monte Carlo (MCMC) methods. The MCMC approach is mostly straightforward. The full conditional distributions are standard except one, and therefore the Gibbs sampler can be used. The parameter ρ has a non-standard conditional distribution. Metropolis-Hastings sampling is used to draw values from this posterior.

We estimate several spatial lag probits in Table 2 using both standard ML and MCMC methods. In keeping with our state welfare spending example, we switch the dependent variable from maximum AFDC benefits to whether or not a state's CHIP (Children's Health Insurance Program) includes a monthly premium payment (Volden 2006). We keep the same independent variables since this dependent variable also reflects the generosity of the welfare program.

<Table 2 About Here>

In the first two columns, the models are estimated assuming the spatial lags are exogenous. The model in the first column is estimated using standard ML techniques. The parentheses in this column contain estimated standard errors and the hypothesis tests assume that the asymptotic t-statistics are normally distributed. The models in columns two and three are estimated using MCMC methods with diffuse zero-mean priors. The reported coefficient estimate is the mean of the posterior distribution based on 10,000 observations after a 1000 observation burn-in period. The number in parentheses is the standard deviation of the posterior distribution. The p-values are also calculated using the posterior. The results in columns two and three are very similar, as they should be given our diffuse priors. Because the estimator used in column two incorrectly treats the spatial lag as exogenous (i.e., like any other right-hand-side variable) the likelihood is misspecified and the sampler draws from the wrong posterior distribution for the spatial coefficient ρ. This specification error has serious consequences for drawing inferences about the importance of spatial interdependence.

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

2. Spatial Error Probit Models

The spatial error version of the probit model takes the form

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\ , \tag{59}$$

where $\mathbf{u} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon}$. In this case, the marginal probabilities are calculated as

$$\Pr(y_i = 1 \mid \mathbf{x}_i) = \Pr\left(u_i < \frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma_i}\right) \tag{60}$$

using the marginal distribution from a multivariate normal with variance-covariance matrix $[(\mathbf{I}-\lambda\mathbf{W})'(\mathbf{I}-\lambda\mathbf{W})]^{-1}$. The same estimation techniques used for the spatial lag model can be used for the spatial error model. We present estimates for the spatial error model in column four of Table 2. In this case, none of the 10,000 sampled spatial AR coefficients were negative. We do not discuss specification tests (*lag* vs. *error*) for the spatial probit, but not that they are covered in Anselin (2006).

C. Spatio-temporal Models for Panel Data

The spatio-temporal autoregressive (STAR) lag model can write in matrix notation as

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \phi\mathbf{V}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{61}$$

where $\mathbf{y}$, the dependent variable, is an $NT \times 1$ vector of cross sections stacked by periods (i.e., the N first-period observations, then the N second-period ones, and so on to the N in the last period, T).[14] The parameter ρ is the spatial autoregressive coefficient and $\mathbf{W}$ is an $NT\times NT$ block-diagonal spatial-weighting matrix. More specifically, we can express this $\mathbf{W}$ matrix as the Kronecker product of a $T\times T$ identity matrix and an $N\times N$ weights matrix $\left(\mathbf{I}_T \otimes \mathbf{W}_N\right)$, with elements w_{ij} of $\mathbf{W}_N$ reflecting the relative degree of connection from unit j to i. $\mathbf{Wy}$ is thus the spatial lag; i.e., for each observation y_{it}, $\mathbf{Wy}$ gives a weighted sum of the y_{jt}, with weights, w_{ij}, given by the relative connectivity from j to i. Notice how $\mathbf{Wy}$ thus directly and straightforwardly reflects the dependence of each unit i's policy dependence on unit j's policy, exactly as in the formal model and theoretical arguments reviewed above. The parameter ϕ is the temporal autoregressive coefficient, and $\mathbf{V}$ is an $NT\times NT$ matrix with ones on the minor diagonal, i.e., at coordinates $(N+1,1), (N+2,2), \ldots, (NT, NT\text{-}N)$, and zeros elsewhere, so $\mathbf{Vy}$ is the (first-order) temporal lag. The matrix $\mathbf{X}$ contains $NT\times k$ observations on k independent variables, and $\boldsymbol{\beta}$ is a $k\times 1$ vector of coefficients on them. The final term in equation (61), $\boldsymbol{\varepsilon}$, is an $NT\times 1$ vector of disturbances, assumed to be independent and identically distributed.[15]

[14] With some work, nonrectangular panels and/or missing data are manageable, but we assume rectangularity and completeness for simplicity of exposition.

[15] Alternative distributions of $\boldsymbol{\varepsilon}$ are possible but add complication without illumination.

The likelihood for the spatio-temporal model is a straightforward extension of this spatial-lag likelihood. Written in ($N\times 1$) vector notation, spatio-temporal-model conditional-likelihood is mostly conveniently separable into parts, as seen here:

$$\text{Log} f_{\mathbf{y}_t,\mathbf{y}_{t-1},\ldots,\mathbf{y}_2|\mathbf{y}_1} = -\frac{1}{2}N(T-1)\log\left(2\pi\sigma^2\right) + (T-1)\log\left|\mathbf{I}-\rho\mathbf{W}\right| - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t \tag{62}$$

where $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \rho\mathbf{W}_N\mathbf{y}_t - \phi\mathbf{I}_N\mathbf{y}_{t-1} - \mathbf{X}_t\boldsymbol{\beta}$.

The issue of stationarity arises in more-complicated fashion in spatio-temporal dynamic models than in purely temporally dynamic ones. Nonetheless, the conditions and issues arising in the former are reminiscent although not identical to those arising in the latter. Define $\mathbf{A} = \phi\mathbf{I}$, $\mathbf{B} = \mathbf{I} - \rho\mathbf{W}$, and ω as a characteristic root of $\mathbf{W}$, the statio-temporal process generating the data is covariance stationary if

$$\left|\mathbf{A}\mathbf{B}^{-1}\right| < 1$$

or, equivalently, if

$$\begin{cases} |\phi| < 1-\rho\omega_{\max}, \ \text{if}\ \rho \geq 0 \\ |\phi| < 1-\rho\omega_{\min}, \ \text{if}\ \rho < 0 \end{cases} \tag{63}$$

If W is row-standardized and both the temporal and spatial dependence are positive ($\rho > 0$ and $\phi > 0$), stationarity requires simply that $\phi + \rho < 1$.

Finally, we note that he unconditional (exact) likelihood function, the one that retains the first time-period observations as non-predetermined, is more complicated (Elhorst 2001, 2003, 2005).

$$\begin{aligned} \text{Log} f_{\mathbf{y}_t,\ldots,\mathbf{y}_1} = & -\frac{1}{2}NT\log\left(2\pi\sigma^2\right) + \frac{1}{2}\sum_{i=1}^{N}\log\left(\left(1-\rho\omega_i\right)^2 - \phi^2\right) + (T-1)\sum_{i=1}^{N}\log\left(1-\rho\omega_i\right) \\ & -\frac{1}{2\sigma^2}\sum_{t=2}^{T}\varepsilon_t'\varepsilon_t - \frac{1}{2\sigma^2}\varepsilon_1'\left(\left(\mathbf{B}-\mathbf{A}\right)'\right)^{-1}\left(\mathbf{B}'\mathbf{B}-\mathbf{B}'\mathbf{A}\mathbf{B}^{-1}\left(\mathbf{B}'\mathbf{A}\mathbf{B}^{-1}\right)'\right)^{-1}\left(\mathbf{B}-\mathbf{A}\right)^{-1}\varepsilon_1 \end{aligned} \tag{64}$$

where $\boldsymbol{\varepsilon}_1 = \mathbf{y}_1 - \rho\mathbf{W}_N\mathbf{y}_1 - \phi\mathbf{I}_N\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}$. When T is small, the first observation contributes greatly to the overall likelihood, and the unconditional likelihood should be used to estimate the model. In other cases, the more compact conditional likelihood is acceptable for estimation purposes.

Note that the same condition that complicates ML estimation of the spatio-temporal lag model, namely the first set of observations is stochastic, also invalidates the use of OLS to estimate a model with a temporally lagged spatial lag. The spatio-temporal model with time-lagged dependent variable and time-lagged spatial-lag is

$$\mathbf{y}_t = \eta \mathbf{W} \mathbf{y}_{t-1} + \phi \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t . \tag{65}$$

If the first set of observations is stochastic, the unconditional (exact) log-likelihood is

$$\begin{aligned} \operatorname{Log} f_{\mathbf{y}_t,\ldots,\mathbf{y}_1} = & -\frac{1}{2} NT \log\left(2\pi\sigma^2\right) + \frac{1}{2}\sum_{i=1}^{N} \log\left(1-\left(\phi+\eta\omega_i\right)^2\right) - \frac{1}{2\sigma^2}\sum_{t=2}^{T} \varepsilon_t' \varepsilon_t \\ & - \frac{1}{2\sigma^2} \varepsilon_1' \left(\left(I-A\right)'\right)^{-1} \left(I - AA'\right)^{-1} \left(I-A\right)^{-1} \varepsilon_1 \end{aligned} \tag{66}$$

where $\boldsymbol{\varepsilon}_1 = \mathbf{y}_1 - \left(\phi + \eta \mathbf{W}_N\right)\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}$, $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \eta \mathbf{W}_N \mathbf{y}_{t-1} - \phi \mathbf{y}_{t-1} - \mathbf{X}_t \boldsymbol{\beta}$, and $\mathbf{A} = \phi \mathbf{I} + \eta \mathbf{W}$. For the derivation of this likelihood function, see Elhorst (2001, 126-130). Note that the second term in the likelihood function causes the OLS estimator to be biased. Asymptotically $\left(T \to \infty\right)$, this bias goes to zero.

We present estimates for a panel model of welfare policy generosity in Table 3. The data are annual observations from 1981-1990 on the contiguous 48 states. The dependent variable is the maximum AFDC benefit, and the independent variables remain unchanged. All the regressions include fixed state effects. The first column contains a non-spatial model estimated with OLS. Clearly, from Moran's I statistic and the two-directional LM statistics, there is spatial dependence in the dataset. The diagnostics do not provide clear evidence in favor of a spatial lag or error specification, however. We estimate both with contemporaneous spatial lags. The second column contains a spatio-temporal lag model, and the third column contains a combined temporal lag and spatial error model. Interestingly, the retail wage variable is statistically significant and positive in all three regressions. Once again, the tax effort coefficient becomes statistically insignificant with the change from a non-spatial to spatial specification.

<Table 3 About Here>

To calculate marginal spatio-temporal effects (non-cumulative) or plot the over-time path of the effect of a permanent one-unit change in an explanatory variable (cumulative), and their standard errors, simply solve for $\mathbf{y}$ in (61):

$$\begin{aligned}\mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \phi\mathbf{V}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \left(\rho\mathbf{W} + \phi\mathbf{V}\right)\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \left[\mathbf{I}_{\mathrm{NT}} - \rho\mathbf{W} - \phi\mathbf{V}\right]^{-1}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \\ &\equiv \mathbf{M}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)\end{aligned} \tag{67}$$

Denote the i^{th} column of $\mathbf{M}$ as $\mathbf{m}_i$ and its estimate as $\hat{\mathbf{m}}_i$. The spatial effects of a one-unit increase in explanatory variable k in country i are $\mathbf{m}_i\beta_k$ with delta method standard errors calculated as

$$\operatorname{var}\left(\hat{\mathbf{m}}_i\hat{\beta}_k\right) = \left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]\operatorname{var}\left(\hat{\boldsymbol{\theta}}\right)\left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right]' \quad , \tag{68}$$

where $\hat{\boldsymbol{\theta}} = \begin{bmatrix}\hat{\rho} \\ \hat{\phi} \\ \hat{\beta}_k\end{bmatrix}$, $\left[\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\boldsymbol{\theta}}}\right] = \begin{bmatrix}\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\rho}} & \frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\phi}} & \hat{\mathbf{m}}_i\end{bmatrix}$, and the vectors $\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\rho}}$ and $\frac{\partial\hat{\mathbf{m}}_i\hat{\beta}_k}{\partial\hat{\phi}}$ are the i^{th} columns of $\hat{\beta}_k\hat{\mathbf{M}}\mathbf{W}\hat{\mathbf{M}}$ and $\hat{\beta}_k\hat{\mathbf{M}}\hat{\mathbf{M}}$ respectively. In Table 4, we present the immediate and long-run (steady-state) spatial effects on regional AFDC benefits from a permanent \$100 increase to monthly retail wages in Missouri using the calculations in equations (67) and (68). The immediate (steady-state) effects range from a low of \$0.44 (\$3.68) in Kentucky to a high of \$0.77 (\$6.38) in Kansas.

<Table 4 About Here>

In Figure 1, we present the spatio-temporal effects on AFDC benefits in Missouri from a permanent \$100 increase to monthly retail wages in Missouri (with 95% C.I.). The marginal effects decay rapidly with most of the total effect experienced within the first 2 years after the shock. The cumulative 10-year effect is approximately \$55.75. In Figure 2, we present the spatio-temporal effects on AFDC benefits in Nebraska from a \$100 counterfactual increase to monthly retail wages in Missouri (with 95% C.I.). The cumulative 10-year effect is about \$4.11. Interestingly, the maximum effects in

Nebraska are not experienced until one or two years after the initial shock. This serves to highlight an important point, namely the contemporaneous spatial lag specification does not imply all (or even most) of the spatial effects are instantaneous.

<Figures 1 and 2 About Here>

VI. Conclusions

Spatial analysis has become much more common in empirical political science research recently. New theories, data, and technology have all contributed to what is likely to be a lasting trend in the study of politics. In our view, the incorporation of spatial models into political science represents a very positive development. After all, spatial interdependence is an important part of the politics that political scientists aim to understand. If there is a concern, it is that the applied research on diffusion and other sources of spatial interdependence is approaching the limits of our methodological knowledge about best practices. This partly reflects the time it takes for new methods from other disciplines to become standard tools in the political science toolkit, as seems to be the case, for example, with simulation based estimation, but it is also due the fact that political scientists ask unique questions and have distinct methodological needs. In this chapter, we have surveyed some developments with respect to diagnosing spatial interdependence, specifying and estimating spatial models, and presenting spatial (and spatio-temporal) effects. It is incumbent on applied researchers to familiarize themselves with these techniques, but it is equally important for political methodologists to accept the challenge of developing spatial methods for political science.

Table 1. State Welfare Policy (Maximum AFDC Benefit)

Independent Variables	OLS	Spatial AR Lag (S-OLS)	Spatial AR Lag (S-2SLS)	Spatial AR Lag (S-GMM)	Spatial AR Lag (S-MLE)	Spatial AR Error (S-MLE)
Constant	54.519	-246.76	-422.09	-500.05	-156.282	676.120
	(531.830)	(450.75)	(437.74)	(413.02)	(429.130)	(471.965)
Poverty Rate	-6.560	8.04	13.205	7.29	3.657	3.239
	(11.262)	(10.022)	(9.977)	(8.452)	(8.917)	(10.062)
Retail Wage	-.121	.016	.089	-.008	-.025	-.344
	(.226)	(.193)	(.187)	(.201)	(.181)	(.243)
Government Ideology	1.513	1.397	1.359*	1.655**	1.432*	1.696**
	(1.030)	(.863)	(.825)	(.761)	(.806)	(.822)
Inter-party Competition	621.799**	368.65	286.98	438.9**	444.677*	263.887
	(290.871)	(250.55)	(243.72)	(197.47)	(226.911)	(238.419)
Tax Effort	3.357**	2.022	1.553	2.397	2.423*	2.936**
	(1.587)	(1.364)	(1.328)	(1.493)	(1.262)	(1.213)
Federal Share	-4.405	-5.818	-6.012	-3.654	-5.393	-6.882*
	(5.001)	(4.20)	(4.014)	(3.415)	(3.901)	(4.099)
Spatial AR		.767***	1.069***	.840***	.537***	.565***
		(.178)	(.232)	(.237)	(.122)	(.131)
Moran I-statistic	3.312***					
$LM_{\rho\lambda}$	12.322***					
LM_{ρ}	11.606***					
LM_{ρ}^{*}	6.477**					
LM_{λ}	5.845***					
LM_{λ}^{*}	.716					
Log-likelihood					-270.763	-272.728
Adj.-R^2	.461	.622	.595	.606	.510	.588
Obs.	48	48	48	48	48	48

Notes: The spatial lags are generated with a binary contiguity weighting matrix. All the spatial weights matrices are row-standardized. ***Significant at the 1% Level; **Significant at the 5% Level; *Significant at the 10% Level.

Table 2. State Welfare Policy (Monthly CHIP Premium)

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

Notes: In the first two columns, the models are estimated assuming the spatial lags are exogenous. The model in the first column is estimated using standard ML techniques. The parentheses in this column contain estimated standard errors and the hypothesis tests assume that the asymptotic t-statistics are normally distributed. The models in columns two through four are estimated using MCMC methods with diffuse zero-mean priors. The reported coefficient estimate is the mean of the posterior density based on 10,000 observations after a 1000 observation burn-in period. The number in parentheses is the standard deviation of the posterior density. The p-values are also calculated using the posterior density. The last two models are estimated with true spatial estimators described in the text. In third column, 30 of the 10,000 spatial AR coefficients sampled from the posterior distribution were negative. In the fourth column, none of the 10,000 sampled spatial AR coefficients were negative. ***p-value <.01, **p-value<.05, *p-value <.10.

Table 3. State Welfare Policy (Maximum AFDC Benefit, 1981-1990)

Independent Variables	OLS	Spatial AR Lag (MLE)	Spatial AR Error (MLE)
Poverty Rate	-.855	-.911	-.903
	(1.130)	(1.050)	(1.198)
Retail Wage	.217***	.204***	.197***
	(.036)	(.034)	(.037)
Government Ideology	.053	.059	.027
	(.087)	(.081)	(.083)
Inter-party Competition	18.960	25.540	18.633
	(24.046)	(22.442)	(22.382)
Tax Effort	.388*	.322	.349
	(.223)	(.208)	(.218)
Federal Share	.483	.859*	.750
	(.521)	(.491)	(.510)
Temporal AR	.663***	.628***	.666***
	(.030)	(.031)	(.030)
Spatial AR		.143***	.200***
		(.044)	(.058)
Moran I-statistic	3.296***		
$LM_{\rho\lambda}$	11.896***		
LM_{ρ}	9.976***		
LM_{ρ}^{*}	1.446		
LM_{λ}	10.450***		
LM_{λ}^{*}	1.921		
Log-likelihood		-1991.357	-1991.290
Adj.-R^2	.981	.981	.982
Obs.	480	480	480

Notes: All regressions include fixed period and unit effects; those coefficient-estimates suppressed to conserve space. The spatial lags are generated with a binary contiguity weighting matrix. All the spatial weights matrices are row-standardized. ***Significant at the 1% Level; **Significant at the 5% Level; *Significant at the 10% Level.

Table 4. Spatial Effects on AFDC Benefits from a $100 Counterfactual Shock to Monthly Retail Wages in Missouri

Neighbor	Immediate Spatial Effect	Long-Run Steady State Effect
	.51	4.26
Arkansas	[.16,.87]	[1.01,7.52]
	.62	5.11
Illinois	[.19,1.04]	[1.25,8.97]
	0.52	4.37
Iowa	[.15,.88]	[.99, 7.75]
	0.77	6.38
Kansas	[.23,1.31]	[1.60,11.17]
	0.44	3.68
Kentucky	[.13,.75]	[.87,6.50]
	0.52	4.44
Nebraska	[.15,.89]	[.99,7.90]
	0.52	4.47
Oklahoma	[.15,.89]	[.96,7.98]
	0.38	3.21
Tennessee	[.12,.65]	[.75,5.67]

Notes: Effects calculated using estimates from the spatial AR lag model in Table 3. Brackets contain a 95% confidence interval.

Figure 1. Spatio-Temporal Effects on AFDC Benefits in Missouri from a $100 Counterfactual Shock to Monthly Retail Wages in Missouri (with 95% C.I.)

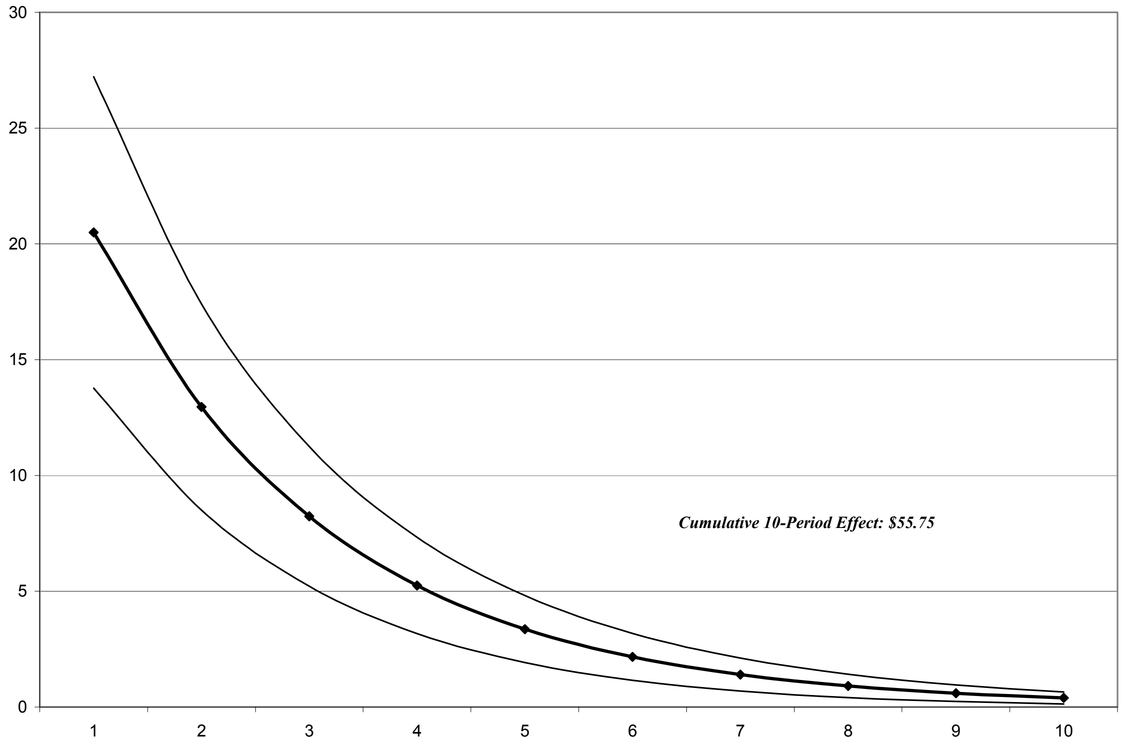
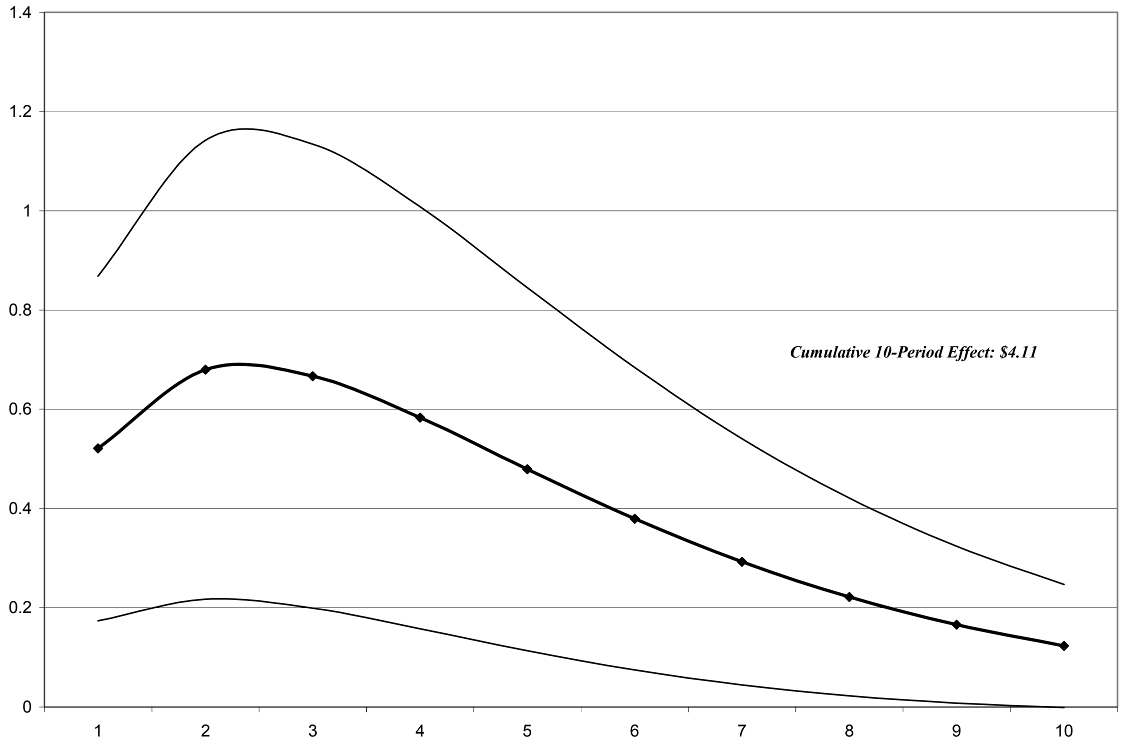


Figure 2. Spatio-Temporal Effects on AFDC Benefits in Nebraska from a $100 Counterfactual Shock to Monthly Retail Wages in Missouri (with 95% C.I.)



References

Anselin, L. 2006. Spatial Econometrics. In T.C. Mills and K. Patterson, eds., *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*. Basingstoke: Palgrave Macmillan, pp. 901-941.

Anselin, L. 2003. "Spatial externalities, spatial multipliers and spatial econometrics," *International Regional Science Review*, 26(2):153–166.

Anselin, L. 2002. "Under the hood. Issues in the specification and interpretation of spatial regression models." *Agricultural Economics*, 27(3):247–267.

Anselin L. 1995. "Local Indicators of Spatial Association–LISA," *Geographical Analysis* 27:93-115.

Anselin, L. 1992. "Space and applied econometrics. Introduction." *Regional Science and Urban Economics*, 22:307–316.

Anselin, L. 1980. "Estimation Methods for Spatial Autoregressive Structures." *Regional Science Dissertation and Monograph Series*, Cornell University, Ithaca, NY.

Anselin, L., Bera, A., Florax, R. J., Yoon, M. 1996. "Simple diagnostic tests for spatial dependence," *Regional Science and Urban Economics* 26:77–104.

Bailey, M., Rom, M. 2004. "A Wider Race? Interstate Competition across Health and Welfare Programs," *Journal of Politics* 66(2):326–47.

Balla, S. 2001. "Interstate Professional Associations and the Diffusion of Policy Innovations," *American Politics Research* 29(3):221–45.

Basinger, S., Hallerberg, M. 2004. "Remodeling the Competition for Capital: How Domestic Politics Erases the Race-to-the-Bottom," *American Political Science Review* 98(2):261-76.

Beck, N., Gleditsch, K. S., Beardsley, K. 2006. "Space is more than geography: Using spatial econometrics in the study of political economy." *International Studies Quarterly* 50:27–44.

Beck, N., Katz, J. 1995. "What to do (and Not to Do) with Time Series Cross Section Data in Comparative Politics." *American Political Science Review* 89:634-47.

Beck, N., Katz, J. 1996. "Nuisance vs. Substance: Specifying and Estimating Time Series Cross Section Models." *Political Analysis* 6:1-34.

Beck, P., Dalton, R., Greene, S., Huckfeldt, R. 2002. "The Social Calculus of Voting: Interpersonal, Media, and Organizational Influences on Presidential Choices," *American Political Science Review* 96: 57-74.

Beron, K. J., Murdoch, J. C., Vijverberg, W. P. 2003. "Why cooperate? Public goods, economic power, and the Montreal Protocol." *The Review of Economics and Statistics*, 85(2):286–297.

Berry, F.S. 1994. "Sizing Up State Policy Innovation Research," *Policy Studies Journal* 22(3):442–56.

Berry, F.S., Berry, W. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis," *American Political Science Review* 84(2):395–415.

Berry, F.S., Berry, W. 1999. "Innovation and Diffusion Models in Policy Research," in P. Sabatier, ed., *Theories of the Policy Process*. Boulder: Westview.

Besag, J. 1974. "Spatial interaction and the statistical analysis of lattice systems." *Journal of the Royal Statistical Society B*, 36:192–225.

Boehmke, F., Witmer, R. 2004. "Disentangling Diffusion: The Effects of Social Learning and Economic Competition on State Policy Innovation and Expansion," *Political Research Quarterly* 57(1):39–51.

Brandsma, A. Ketellapper, R. H. 1979. "A biparametric approach to spatial autocorrelation." *Environment and Planning A*, 11:51–58.

Braun, D., Gilardi, F. 2006. "Taking 'Galton's Problem' Seriously: Towards a Theory of Policy Diffusion," *Journal of Theoretical Politics* 18(3):298–322.

Braybeck, B., Huckfeldt, R. 2002a. "Spatially Dispersed Ties Among Interdependent Citizens: Connecting Individuals and Aggregates," *Political Analysis* 10: 261-275.

Braybeck, B., Huckfeldt, R. 2002b. "Urban Contexts, Spatially Dispersed Networks, and the Diffusion of Political Information," *Political Geography* 21: 195-220.

Brinks, D., Coppedge, M. 2006. "Diffusion Is No Illusion: Neighbor Emulation in the Third Wave of Democracy," *Comparative Political Studies* 39(4):463-89.

Brooks, S. 2005. "Interdependent and Domestic Foundations of Policy Change: The Diffusion of Pension Privatization Around the World," *International Studies Quarterly* 49(2):273–94.

Brueckner, J. K. 2003. "Strategic interaction among governments: An overview of empirical studies."

International Regional Science Review, 26(2):175–188.
Brune, N., Garrett, G., Kogut, B. 2004. "The International Monetary Fund and the Global Spread of Privatization," *IMF Staff Papers* 51(2):195–219.
Burridge, P. 1980. "On the Cliff-Ord Test for Spatial Autocorrelation." *Journal of the Royal Statistical Society B*, 42:107-8.
Caldeira, G. 1985. "The Transmission of Legal Precedent: A Study of State Supreme Courts," *American Political Science Review* 79(1):178–94.
Case, A., Hines, J., Rosen, H. 1993. "Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States," *Journal of Public Economics* 52(3):285–307.
Cho, W.T. 2003. "Contagion Effects and Ethnic Contribution Networks," *American Journal of Political Science* 47(2):368-87.
Cho, W.T., Gimpel, J. 2007. "Spatial Dimensions of Arab American Voter Mobilization after September 11," *Political Geography* 26(3):330-51.
Cho, W.T., Rudolph, T. 2007. "Emanating Political Participation: Untangling the Spatial Structure behind Participation," *British Journal of Political Science* 37(1). Forthcoming.
Cliff, A., Ord, J. 1973. *Spatial Autocorrelation*, London: Pion.
Cliff, A., Ord, J. 1981. *Spatial Processes: Models and Applications*. London: Pion.
Crain, R. 1966. "Fluoridation—Diffusion of an Innovation among Cities," *Social Forces* 44(4):467–76.
Cressie, N. 1993. *Statistics for Spatial Data*. Wiley, New York.
Dahl, R. 1971. *Polyarchy: Participation and Opposition*. New Haven: Yale University Press.
Daley, D., Garand, J. 2005. "Horizontal Diffusion, Vertical Diffusion, and Internal Pressure in State Environmental Policymaking, 1989–1998," *American Politics Research* 33(5):615–44.
Darmofal, D. 2006. "Spatial Econometrics and Political Science," Society for Political Methodology Working Paper Archive: http://polmeth.wustl.edu/workingpapers.php.
Dow, M. "A Biparametric Approach to Network Autocorrelation: Galton's Problem." *Sociological Methods and Research*, 13(2): 201-217.
Eising, R. 2002. "Policy Learning in Embedded Negotiations: Explaining EU Electricity Liberalization," *International Organization* 56(1):85–120.
Elhorst, J.P. 2001. "Dynamic models in space and time." *Geographical Analysis* 33:119–140.
Elhorst, J.P. 2003. "Specification and estimation of spatial panel data models." *International Regional Science Review* 26:244–68.
Elhorst, J.P. 2005. "Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels." *Geographical Analysis* 37:85–106.
Elkins, Z., Simmons, B. 2005. "On Waves, Clusters, and Diffusion: A Conceptual Framework," *Annals of the American Academy of Political and Social Science* 598(1):33–51.
Elkins, Z., Guzman, A., Simmons, B. 2006. "Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960-2000." *International Organization*, 60(4): 811-846.
Franzese, R., Hays, J. 2004. Empirical modeling strategies for spatial interdependence: Omitted-variable vs. simultaneity biases. Summer meetings of the Political Methodology Society.
Franzese, R., Hays, J. 2006a. Spatio-temporal models for political-science panel and time-series-cross-section data. Summer meetings of the Political Methodology Society.
Franzese, R., Hays, J. 2006b. "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. "Empirical Models of International Capital-Tax Competition," in *International Taxation Handbook*, G. Gregoriou, C. Read, eds., Elsevier Press, 2007: 43-72.
Franzese, R., Hays, J. 2007b. "Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data," *Political Analysis* 15(2):140-64.
Franzese, R., Hays, J. 2007c. "Spatial Interdependence in Comparative Politics: Theoretical- and Empirical-Model Specification, Estimation, Interpretation, and Presentation," forthcoming in 40th Anniversary Issue of *Comparative Political Studies*.
Frieden, J., Rogowski, R. 1996. "The Impact of the International Economy on National Policies: An Analytical Overview," In R. Keohane, ed., *Internationalization and Domestic Politics*, Cambridge:

Cambridge University Press.
Gartzke, E., Gleditsch, K.S. 2006. "Identity and Conflict: Ties that Bind and Differences that Divide," *European Journal of International Relations* 12(1): 53–87.
Genschel, P. 2002. "Globalization, Tax Competition, and the Welfare State," *Politics and Society* 30(2):245–75.
Gilardi, F. 2005. "The Institutional Foundations of Regulatory Capitalism: The Diffusion of Independent Regulatory Agencies in Western Europe," *Annals of the American Academy of Political and Social Science* 598:84–101.
Gleditsch, K.S. 2002. *All International Politics is Local: The Diffusion of Conflict, Integration, and Democratization*. Ann Arbor: University of Michigan Press.
Gleditsch, K.S. 2007. "Civil War and its Spread," in J. Bercovitch, V. Kremenyuk, I.W. Zartman, eds., *Handbook on Conflict Resolution*. London: Sage. Forthcoming.
Gleditsch, K.S., Beardsley, K. 2004. "Nosy Neighbors: Third Party Actors in Central American Civil Conflicts," *Journal of Conflict Resolution* 48(3):379-402.
Gleditsch, K.S., Ward, M. 2000. "War and peace in space and time: The role of democratization," *International Studies Quarterly* 44(1):1–29.
Gleditsch, K.S., Ward, M. 2006. "Diffusion and the International Context of Democratization," *International Organization* 60(4): 911–33.
Gleditsch, K.S., Ward, M. 2007. "Diffusion and the Spread of Democratic Institutions," in F. Dobbins, G. Garret, B. Simmons, eds. *The Global Diffusion of Democracy and Markets*. Cambridge: Cambridge University Press.
Godwin, M., Schroedel, J. 2000. "Policy Diffusion and Strategies for Promoting Policy Change: Evidence from California Local Gun Control Ordinances," *Policy Studies Journal* 28(4):760–76.
Goldenberg, Edie N., Michael W. Traugott, and Frank K. Baumgartner. 1986. "Preemptive and Reactive Spending in U.S. House Races." *Political Behavior* 8:3-20.
Govea, R., West, G. 1981. "Riot Contagion in Latin America, 1949-1963," *Journal of Conflict Resolution* 25(2):349-368.
Gray, V. 1973. "Innovation in the States: A Diffusion Study," *American Political Science Review* 67(4):1174–85.
Grossback, L., Nicholson-Crotty, S., Peterson, D. 2004. "Ideology and Learning in Policy Diffusion," *American Politics Research* 32(5):521–45.
Haining, R. 1990. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge.
Hoff, P., Ward, M. 2004. "Modeling dependencies in international relations networks," *Political Analysis* 12(2):160–75.
Hordijk, L. 1974. "Spatial correlation in the disturbances of a linear interregional model." *Regional Science and Urban Economics*, 4:117–140.
Huckfeldt, R., Johnson, P.E., Sprague, J. 2005. " Individuals, Dyads and Networks: Autoregressive Patterns of Political Influence," in Alan S. Zuckerman (ed.), *The Social Logic of Politics: Personal Networks as Contexts for Political Behavior*. Philadelphia: Temple University Press.
Huntington, S. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. Norman: The University of Oklahoma Press.
Jahn, D. 2006. "Globalization as 'Galton's Problem': The Missing Link in the Analysis of Diffusion Patterns in Welfare State Development," *International Organization* 60:401-31.
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. 1998. "A generalized spatial two stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances." *Journal of Real Estate Finance and Economics*, 17:99–121.
Kelejian, H.H., Prucha, I. 1999. "A generalized moments estimator for the autoregressive parameter in a spatial model." *International Economic Review*, 40:509–533.
Kelejian, H.H., Prucha, I. 2001. "On the asymptotic distribution of the Moran I test statistic with applications," *Journal of Econometrics* 104:219–57.

Kim, C.-W., Phipps, T. T., Anselin, L. 2003. "Measuring the benefits of air quality improvement: A spatial hedonic approach," *Journal of Environmental Economics and Management*, 45:24–39.
Knill, C. 2005. "Introduction: Cross-National Policy Convergence: Concepts, Approaches and Explanatory Factors," *Journal of European Public Policy* 12(5):764–74.
Knoke, D. 1982. "The Spread of Municipal Reform: Temporal, Spatial, and Social Dynamics," *American Journal of Sociology* 87(6):1314–39.
Krasno, J., Green, D., Cowden, J. 1994. "The Dynamics of Campaign Fundraising in House Elections," *Journal of Politics* 56(2):459-474.
Lacombe, D.J. 2004. "Does Econometric Methodology Matter? An Analysis of Public Policy Using Spatial Econometric Techniques." *Geographical Analysis* 36(2): 105-118.
Lee, C-K., Strang, D. 2006. "The International Diffusion of Public-Sector Downsizing: Network Emulation and Theory-Driven Learning." *International Organization*, 60(4): 883-909.
LeSage, J.P. 2000. "Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models," *Geographical Analysis*, 32(1): 19-35.
LeSage, J.P. 1999. Spatial Econometrics. http://rri.wvu.edu/WebBook/LeSage/spatial/spatial.html.
Li, R., Thompson, W. 1975. "The 'Coup Contagion' Hypothesis," *Journal of Conflict Resolution* 19(1):63–88.
Lin, T. M., Wu, C. E., Lee, F. Y. 2006. "'Neighborhood' influence on the formation of national identity in Taiwan: Spatial regression with disjoint neighborhoods," *Political Research Quarterly* 59:35–46.
Lutz, J. 1987. "Regional Leadership Patterns in the Diffusion of Public Policies," *American Politics Quarterly* 15(3):387–98.
McMillen, D. P. 1992. "Probit with spatial autocorrelation." *Journal of Regional Science*, 32:335–348.
Meseguer, C. 2004. "What Role for Learning? The Diffusion of Privatisation in OECD and Latin American Countries," *Journal of Public Policy* 24(3):299–325.
Meseguer, C. 2005. "Policy Learning, Policy Diffusion, and the Making of a New Order, *The Annals of the American Academy of Political and Social Science* 598(1):67–82.
Mintrom, M. 1997a. "Policy Entrepreneurs and the Diffusion of Innovation," *American Journal of Political Science* 41(3):738–70.
Mintrom, M. 1997b. "The State-Local Nexus in Policy Innovation Diffusion: The Case of School Choice," *Publius: The Journal of Federalism* 27(3):41–59.
Mintrom, M., Vergari, S. 1998. "Policy Networks and Innovation Diffusion: The Case of State Education Reforms," *Journal of Politics* 60(1):126– 48.
Mooney, C. 2001. "Modeling Regional Effects on State Policy Diffusion," *Political Research Quarterly* 54(1):103–24.
Mossberger, K. 1999. "State-Federal Diffusion and Policy Learning: From Enterprise Zones to Empowerment Zones," *Publius: The Journal of Federalism* 29(3):31–50.
O'Loughlin, J., Ward, M., Lofdahl, C., Cohen, J., Brown, D., Reilly, D., Gleditsch, K., Shin, M. 1998. "The Diffusion of Democracy, 1946–1994," *Annals of the Association of American Geographers* 88(4):545–74.
Ord, J. K. 1975. "Estimation methods for models of spatial interaction." *Journal of the American Statistical Association*, 70:120–126.
Paelinck, J., Klaassen, L. 1979. *Spatial Econometrics*. Saxon House, Farnborough.
Ripley, B. D. 1981. *Spatial Statistics*. Wiley, New York.
Rogers, E. 1995. *Diffusion of Innovations*. New York: Free Press.
Rose, R. 1993. *Lesson-Drawing in Public Policy: A Guide to Learning across Time and Space*. Chatham: Chatham House.
Salehyan, I., Gleditsch, K.S. 2006. "Refugees and the Spread of Civil War," *International Organization* 60(2):335-66.
Schneider, A., Ingram, H. 1988. "Systematically 'Pinching' Ideas: A Comparative Approach to Policy Design," *Journal of Public Policy* 8(1):61–80.
Shin, M., Ward, M. 1999. "Lost in space: Political geography and the defense-growth trade-off.," *Journal of Conflict Resolution* 43:793–816.
Shipan, C., Volden, C. 2006. "Bottom-Up Federalism: The Diffusion of Antismoking Policies from U.S.

Cities to States," *American Journal of Political Science* 50(4), 825–843.
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.
Simmons, B., Dobbin, F., Garrett, G. 2006. "Introduction: The International Diffusion of Liberalism," *International Organization* 60(4):781–810.
Starr, H. 1991. "Democratic Dominoes: Diffusion Approaches to the Spread of Democracy in the International System," *Journal of Conflict Resolution* 35(2):356–81.
Swank, D. 2006. "Tax Policy in an Era of Internationalization: Explaining the Spread of Neoliberalism," *International Organization* 60: 847-82.
Vijverberg, W. P. 1997. "Monte Carlo evaluation of multivariate normal probabilities." Journal of Econometrics," 76:281–307.
Volden, C. 2006. "States as Policy Laboratories: Emulating Success in the Children's Health Insurance Program." *American Journal of Political Science* 50(2): 294-312.
Walker, J. 1969. "The Diffusion of Innovations among the American States." *American Political Science Review* 63(3):880–99.
Walker, J. 1973. "Problems in Research on Diffusion of Policy Innovations," *American Political Science Review* 67(4):1186–91.
Ward, M., Gleditsch, K.S. 2002. "Location, location, location: An MCMC approach to modeling the spatial context of war and peace," *Political Analysis* 10(3):244–60.
Whittle, P. 1954. "On stationary processes in the plane." *Biometrika*, 41:434–449.