Spatial Econometrics and Political Science*

David Darmofal
Department of Political Science
University of South Carolina
350 Gambrell Hall
Columbia, SC 29208
Phone: (803) 777-5440
Email: darmofal@gwm.sc.edu

*A previous version of this paper was presented at the Annual Meeting of the Southern Political Science Association, Atlanta, GA, January 5-7, 2006. I thank Brady Baybeck, Jan Box-Steffensmeier, Sarah Brooks, Wendy Tam Cho, Tim Frye, Marcus Kurtz, Scott McClurg, and Chris Zorn for helpful comments and discussions on this project. Any remaining errors are my own.

Abstract

Many theories in political science predict the spatial clustering of similar behaviors among neighboring units of observation. This spatial autocorrelation poses implications for both inference and modeling that are distinct from the more familiar serial dependence in time series analysis. In this paper, I examine how political scientists can diagnose and model the spatial dependence that our theories predict. This diagnosis and modeling entails three simple sequential steps. First, univariate spatial autocorrelation is diagnosed via global and local measures of spatial autocorrelation. Next, diagnostics are applied to a model with covariates to determine whether any spatial dependence diagnosed in the first step persists after the behavior has been modeled. If so, the researcher simply chooses the spatial econometric specification indicated by the diagnostics. I present Monte Carlo results that demonstrate the importance of diagnosing and modeling spatial dependence in our data. I conclude by discussing how researchers can easily apply spatial econometric models in their research.

"[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."

Sir Francis Galton at The Royal Anthropological Institute, 1888
The Journal of the Anthropological Institute of Great Britain and Ireland 18: 270.

1 Introduction

Political science data are spatial data: the political behaviors, processes, and events we seek to understand occur at specific geographic locations. These locations, moreover, are often central to our understanding of these phenomena. Across a broad range of substantive concerns, from political communication to political conflict, democratization to dependency, policy diffusion to party mobilization, our theories posit that spatially proximate units are more likely to behave similarly than spatially distant units (Huckfeldt 1986; Vasquez 1995; Cardoso and Faletto 1979; Berry and Berry 1990). These theories, in short, predict positive spatial autocorrelation, the spatial clustering of similar behaviors among neighboring observations. Political science data are, in fact, particularly predisposed toward this positive spatial dependence.[1] Unlike more individuated concerns such as microeconomics, politics is, by nature, a collective concern. Shared political concerns combine with spatial proximity to promote familiarity. This familiarity in turn breeds both contempt and political conflict and interaction and political interdependence.

Until recently our ability to incorporate the spatial dimension of our theories in our models was quite limited, relying primarily on dummy variables to capture differences in behavior across geographically disparate units. Such an approach is suboptimal, as it is unable to address some of the central issues posed by spatially dependent data. Consider, for example, Sir Francis Galton's comment in the epigraph to this paper. Sir Galton's comment in response to Edward Tylor's presentation at The Royal Anthropological Institute in November 1888 clearly ranks among the most influential comments ever expressed at an academic presentation, remembered as it is more than a century later. Sir Galton's critique, which has since come to be known as Galton's problem, focuses on the critical substantive distinction between two alternative explanations for spatially dependent behavior.

On the one hand, spatial dependence may be produced by the diffusion of behavior between neighboring units. If so, the behavior is likely to be highly social in nature, and understanding the interactions between interdependent units is critical to understanding the behavior in question. Alternatively, neighboring units may share similar behaviors due simply to the units' independent adoptions of the behavior. If so, the spatial dependence observed in our data does not reflect a truly spatial process, but merely the geographic clustering of the sources of the behavior in question. Such dependence can be termed attributional dependence, as neighboring units have shared attributes that produce the clustering of behaviors. Clearly, determining which process is producing spatial dependence is critical to our substantive understanding of the behavior of interest.

A dummy variable approach is unable to distinguish between the two quite different explanations for spatial dependence. As proxies for our ignorance of the sources of spatial dependence,

[1] Our interest in spatial econometrics is not on the full joint density, as spatial dependence would suggest, but instead on spatial autocorrelation as a moment of the joint density (Anselin and Bera 1998, 240). However, for ease of exposition, I follow common practice in spatial econometrics and use both terms to refer to spatial autocorrelation.

statistically significant parameters on dummy variables for geographic areas merely tell us that behaviors differ for units in these particular areas in contrast to the reference category. Such an approach cannot tell us whether the spatial dependence is consistent with diffusion or with the spatial clustering of the behavior's sources.

Recent advances in spatial econometrics, however, now allow us to address Galton's problem econometrically. While spatial econometric models come in a variety of forms, at their most basic level they share a common feature that distinguishes them from standard econometric models: they explicitly model spatial autocorrelation.[2] Spatial econometric models allow us to address Galton's problem because each of the two alternative sources of spatial dependence posed by Galton presents its own distinct spatial econometric specification.

Spatial diffusion occurs because units' behavior is directly influenced by the behavior of "neighboring units."[3] This diffusion effect corresponds to a positive and significant parameter on a spatially lagged dependent variable capturing the direct influence between neighbors.[4] Conversely, the geographic clustering of the sources of the behavior implies an alternative specification. Assuming that we are unable to model fully the sources of spatial dependence in the data generating process (DGP), these sources will produce spatial dependence in the error terms between neighboring locations. This spatial error dependence can be modeled via a spatially lagged error term.[5] We can also extend Galton's critique and hypothesize that spatial dependence is produced both by diffusion and by the independent adoption of behaviors by neighbors. This joint spatial lag, spatial error dependence can also be modeled by incorporating both a spatially lagged dependent variable and a spatially lagged error term, with proper identifying restrictions imposed.

The substantive implications of properly modeling spatial dependence are intimately linked with methodological implications. Ignoring either form of spatial dependence in our empirical models poses its own distinct implications for inference. For example, estimating an OLS model that ignores a diffusion effect in the DGP produces biased and inconsistent parameter estimates. Estimating an OLS model that ignores spatial clustering in the sources of the behavior produces inefficient parameter estimates, standard error estimates that are biased downward, and Type I errors. Happily, the diagnosis and modeling of spatial dependence is a straightforward process that can be easily adapted by applied researchers. Global and local measures of spatial autocorrelation are estimated to determine whether the data exhibit spatial dependence. If they do, the researcher simply applies diagnostics to an OLS specification to determine whether the variables in the model

[2]The intellectual lineage of spatial econometric models can be traced to Isard's (1956) call for a regional science incorporating the spatial relationships between units of observation. As a field, spatial econometrics includes a broad range of models in both frequentist and Bayesian perspectives. Spatial econometric models range from linear models for continuous dependent variables to models for categorical and limited dependent variables (Bolduc, Fortin, and Gordon 1997; LeSage 2000; Fleming 2004), generalized method of moment (GMM) estimators (Kelejian and Prucha 1999), panel data models (Baltagi, Song, and Koh 2003), survival models (Banerjee and Carlin 2004), and more. The most widely applied spatial econometric model is the cross-sectional model for continuous dependent variables. Given political scientists' extensive use of and familiarity with its non-spatial alternative, ordinary least squares, this paper focuses on this spatial econometric model. The discussion proceeds from a frequentist perspective. Scholars interested in a Bayesian perspective will also be interested in LeSage 1997, 2000, Holloway, Shankar, and Rahman 2002, and MacNab 2003.

[3]As will be discussed, the definition of neighbors, those units hypothesized to exhibit spatial autocorrelation, is a critical decision when modeling spatial dependence. The definition of neighbors is generalizable and need not imply contiguity.

[4]Although a diffusion process implies a spatially lagged dependent variable, diagnostics may also indicate such a specification even in the absence of diffusion. A scale mismatch between the units of theoretical interest and the units that are available empirically may produce an artificial lag dependence, as is implied by the modifiable areal unit problem discussed in Section 10.1. Thus, we can state that a positive and significant autoregressive parameter on a spatially lagged dependent variable is consistent with, but is not definitive with regard to the existence of a diffusion process.

[5]Such spatial error dependence is also consistent with spatial clustering in measurement errors.

sufficiently capture the spatial dependence in the data. If the variables do not fully model the dependence, the diagnostics indicate whether the researcher should estimate a model with a spatially lagged dependent variable, a spatially lagged error term, or both.

The past decade has witnessed an increased use of spatial econometrics in a variety of disciplines, including economics, sociology, biostatistics, and urban planning. In contrast, despite the prominence of the spatial dimension in many of our theories, the application of spatial econometrics within political science remains quite limited.[6] By employing spatial econometrics, scholars in each of political science's empirical subfields can gain leverage on the spatial dependence that is inherent in our theories.

The paper is structured as follows. After examining the standard treatment of autocorrelation in econometrics texts, I discuss how spatial dependence differs from the more familiar serial dependence in time series analysis. Next, I examine the three sequential steps that researchers can easily adopt to diagnose and model the spatial dependence that is implied by our theories. I next present Monte Carlo results on the effects of omitted spatial dependence on OLS estimates. I conclude by discussing how political scientists can apply spatial econometric models in their research.

2 Standard Treatments of Spatial Autocorrelation

It is surprising that despite the increased recognition of the implications of spatial autocorrelation and the increased application of spatial econometrics in a variety of disciplines, there is actually very little discussion of spatial autocorrelation in most standard econometrics texts. As Anselin and Bera (1998, 237) note, there is no mention of spatial autocorrelation in standard texts such as Judge et al. (1982, 1985), Greene (1993), Poirier (1995), Fomby et al. (1984), Amemiya (1985), or Davidson and McKinnon (1993). We can add to this list Kmenta (1997), Long (1997), Agresti (2002), and Greene (2003).

Moreover, the brief mentions of non-spatial cross-sectional autocorrelation in these texts are incomplete and somewhat misleading. These texts, for example, argue that cross-sectional autocorrelation is not a concern; instead the analyst should focus on heteroskedasticity and leave the consideration of autocorrelation to the time domain. Thus, for example, Kmenta argues (1997, 299):

> . . . nonautocorrelation is more frequently violated in the case of relations estimated from time series data than in the case of relations estimated from cross-sectional data.. . . one might suspect that the effect of these factors operating in one period would, in part, carry over to the following periods. This seems more likely than that the effect would carry over from one family, firm, or other similar unit to another.

Yet these are precisely the effects that many political science theories predict will exist in our DGPs.

Greene (2003, 192) likewise argues that autocorrelation is primarily a time series issue and heteroskedasticity a cross-sectional issue. Elsewhere he emphasizes incorrect functional form as a source of cross-sectional autocorrelation. Although it is true that incorrect functional form can produce spatial error dependence, this is not the principal source of such dependence, as Galton's

[6]Spatial econometric models have seen some limited use within political science. Perhaps the most visible application of spatial econometrics has occurred in international relations, where examples include Gleditsch and Ward 2000, Gleditsch 2002, and Ward and Gleditsch 2002. American politics applications include Busch and Reinhardt 2000, Cho 2003, Gimpel and Cho 2004, and Darmofal 2006. Comparative politics applications include O'Loughlin, Flint, and Anselin 1994, O'Loughlin 2002, and Shin and Agnew 2002. Starr (1991, 2001), Starr and Most (1976, 1978, 1983), and Most and Starr (1980, 1982, 1983) also use a set of innovative approaches to examine the effects of spatial proximity on international conflict and democratization.

problem indicates. Moreover, Greene also promotes the use of the Durbin-Watson statistic as a diagnostic for autocorrelated errors in cross-sectional data (126). However, the spatial analog of the Durbin-Watson, the Moran's I test for spatial error dependence, is actually a highly unreliable test as it picks up several different misspecification errors (Anselin and Rey 1991). In fact, cross-sectional spatial autocorrelation is both a concern that merits political scientists' attention and one that is not diagnosable or treatable using methods developed for serial correlation over time.

3 Modeling Spatial Autocorrelation

> "There is an abundance of social psychological evidence demonstrating, in a variety of settings, the nexus between spatial proximity, interpersonal friendship, and similarity of attitudes."
>
> Samuel C. Patterson (1972, 361)

The statement by Patterson highlights the potential for spatial autocorrelation in many political phenomena. Formally, spatial autocorrelation implies a non-zero covariance between the values on a random variable for neighboring locations:

$$Cov(y_i, y_j) = E(y_i y_j) - E(y_i)E(y_j) \neq 0 \;\; \forall\, i \neq j, \tag{1}$$

where the i, j locations have a spatial interpretation (Anselin and Bera 1998, 241-242). The null hypothesis on a test of spatial autocorrelation is that the values on the random variable are distributed randomly in relation to space. That is, knowledge of units' spatial locations provides no leverage in predicting the units' values on the random variable. Positive spatial autocorrelation exists if neighboring units share similar values on the random variable. Positive spatial autocorrelation, for example, exists when neighboring countries have similar economic policies or neighboring voters favor the same candidate. Negative spatial autocorrelation exists if neighboring units have dissimilar values on the random variable. Negative spatial autocorrelation, for example, exists when neighboring countries have dissimilar economic policies or neighboring voters favor different candidates. As the Patterson quote would suggest, positive spatial autocorrelation will be more likely than negative spatial autocorrelation in most political science applications.

What implications does spatial autocorrelation pose for inference? At first glance, spatial autocorrelation exhibits a surface similarity with the more familiar temporal dependence in time series analysis. Both are instances of dependent data. And in both the cross-sectional spatial case and the longitudinal time series case, the dependence can be modeled either via a lagged dependent variable or via the error term.

Spatial dependence, however, is not the cross-sectional analog of serial dependence. The critical distinction between the two forms of dependence arises from the dimensionality of the dependence. In the time series case, dependence is unidimensional: the past influences the present. Cross-sectional spatial dependence, in contrast, is multidimensional.[7] In the diffusion case, neighbors influence the behavior of their neighbors *and vice versa*. In the case of attributional dependence, errors for neighboring observations exhibit simultaneous dependence. The simultaneous, multidimensional nature of spatial dependence leads to implications for inference that are distinct from the time series case. This also significantly complicates estimation of spatial econometric models incorporating this spatial dependence.

[7] In practice, the number of dimensions is typically fixed at two.

3.1 Spatial vs. Temporal Autocorrelation

To understand the differing implications of spatial and temporal autocorrelation, it is helpful to consider Anselin's (1988, 34) general spatial model for cross-sectional data:

$$y = \rho \mathbf{W}_1 y + \varepsilon$$

$$\varepsilon = \lambda \mathbf{W}_2 \varepsilon + \xi, \tag{2}$$

where y is an N by 1 vector of observations on the dependent variable, $\mathbf{W}_1 y$ is a spatially lagged dependent variable with spatial weights matrix $\mathbf{W}_1$, ρ is the spatial autoregressive parameter for the spatially lagged dependent variable, ε is an N by 1 vector of error terms, $\mathbf{W}_2\varepsilon$ is a spatially lagged error term with spatial weights matrix $\mathbf{W}_2$, λ is the spatial autoregressive parameter for the spatially lagged error term, and $\xi \sim N(0,\Omega)$, where $\Omega_{ii} = h_i(\mathbf{z}\alpha)$. When $\alpha = 0$, $h = \sigma^2$, and the errors are homoskedastic. The spatial lag model consistent with a diffusion process in the DGP results from setting λ equal to zero. The spatial error model consistent with attributional dependence results from setting ρ equal to zero.

Consider, first, the spatial lag model, where $\lambda = 0$. This bears some similarity to a time series model with a lagged dependent variable:

$$y_t = \rho y_{t-1} + \varepsilon_t, \tag{3}$$

where the lag in (3) is temporal rather than spatial, as it is in (2). In the time series case, OLS is a biased but consistent estimator of ρ in the absence of serial correlation and other misspecification errors. Thus, although the OLS estimator should not be relied on in small samples, it can still be employed for asymptotic inference.

In contrast to the time series case, OLS estimates of the autoregressive parameter ρ in a spatial lag model will be biased and inconsistent, regardless of whether the errors exhibit serial correlation. The distinction between the spatial and temporal cases exists because of the multidimensional nature of spatial dependence. Consider first, the bias of the OLS estimator of the spatial autoregressive parameter ρ. The expected value of the OLS estimator, $\hat{\rho}$, is:

$$E(\hat{\rho}) = (y'\mathbf{W}_1'\mathbf{W}_1 y)^{-1} y'\mathbf{W}_1'(\rho \mathbf{W}_1 y + \varepsilon)$$

$$= \rho + (y'\mathbf{W}_1'\mathbf{W}_1 y)^{-1} y'\mathbf{W}_1'\varepsilon. \tag{4}$$

As in the time series case, the expected value of the estimator will not equal the true value of ρ. In the spatial case, the sources of bias are twofold. First, as Anselin (1988, 78) notes, just as in the time series case, the complex nature of the matrix inverse induces a non-zero correlation with the error term. Unique to the spatial case, however, the expected value of $y'\mathbf{W}_1'\varepsilon$ is also non-zero whenever $\rho \neq 0$ due to the multidimensional nature of spatial, as opposed to temporal, dependence.

The consistency of the OLS estimator in the time series case exists only because y_{t-1} is uncorrelated with ε_t when there is no serial correlation in the errors. This does not hold in the spatial case, due to the multidimensional nature of spatial dependence. We can reformulate the spatial lag model in (2) as:

$$y = (\mathbf{I} - \rho \mathbf{W}_1)^{-1}\varepsilon, \tag{5}$$

where $(\mathbf{I} - \rho \mathbf{W}_1)$ is the Leontief inverse, which acts as a spatial multiplier, linking the spatially lagged dependent variable to the errors at all locations. As Anselin and Bera (1998, 246-47) show, in contrast to the unidimensional time-series case where the matrix inverse is triangular, in the multidimensional spatial case the matrix inverse is a full matrix, producing an infinite series, $(\mathbf{I} + \rho\mathbf{W}_1 + \rho^2\mathbf{W}_1^2 + \rho^3\mathbf{W}_1^3 + \dots)\varepsilon$. As a result, $\mathbf{W}_1 y_i$ is correlated not only with ε_i, but also with the errors at all other locations.

Because of the simultaneous nature of the spatial dependence, the OLS estimator $\hat{\rho}$ is inconsistent, regardless of whether there is dependence in the error term or not. This can be seen via the probability limit:

$$\text{plim}\, N^{-1}(y'\mathbf{W}_1'\varepsilon) = \text{plim}\, N^{-1}\varepsilon'\mathbf{W}_1(\mathbf{I} - \rho\mathbf{W}_1)^{-1}\varepsilon, \tag{6}$$

where the error term takes a quadratic form and only when $\rho = 0$ does the probability limit equal zero (Anselin 1988, 58).

As Anselin (1988) shows, if a diffusion process exists in the DGP and a spatially lagged dependent variable is omitted from the model altogether, the result is biased and inconsistent parameter estimates for the covariates in the model, reflecting omitted variable bias. Estimation of the spatial lag model incorporating the spatially lagged dependent variable must proceed via either a maximum likelihood specification or an instrumental variables specification. These two approaches are taken up in Section 9, which discusses spatial econometric specifications for modeling spatially dependent DGPs.

The second principal spatial model is a spatial error model where now the dependence pertains to the error term, rather than to a spatially lagged dependent variable. The spatial error model bears a resemblance to a time series model with serially correlated errors. The implications of spatial error dependence are similar, though not identical, to those of serial correlation in time series. As in the time series case, OLS parameter estimates remain unbiased, but are no longer efficient. In the presence of spatial error dependence, standard error estimates will be biased downward, producing Type I errors (Anselin 1988). The loss of information implicit in this spatial error dependence must be accounted for in estimation in order to produce unbiased standard error estimates.

Similar to the case of a spatially lagged dependent variable, the simultaneous error dependence produces a non-zero covariance between the error terms at all locations, via the spatial multiplier:

$$E[\varepsilon\varepsilon'] = \sigma^2(\mathbf{I} - \lambda\mathbf{W}_2)^{-1}(\mathbf{I} - \lambda\mathbf{W}_2')^{-1}, \tag{7}$$

with this covariance declining as the order of contiguity increases (Anselin 1988). Moreover, the spatial multiplier induces heteroskedasticity in the errors, which must be accounted for in estimation (Anselin and Bera 1998, 248).[8]

In the unidimensional serial correlation case, iterative FGLS estimators such as the Cochrane-Orcutt and Durbin estimators may be applied, producing consistent estimates of the autoregressive parameter for the serial dependence. These approaches are not applicable in the case of multidimensional spatial dependence. A spatial analog of the Cochrane-Orcutt estimator does not produce consistent estimates of the autoregressive parameter, λ (Anselin 1988, 110). And as Kelejian and Prucha (1997, 108) demonstrate, λ is unidentified in a spatial analog of the Durbin two-step method. As a result, estimation of the autoregressive spatial error parameter, λ, must proceed via maximum likelihood estimation.

[8] Alternatively, the spatial error dependence may reflect a moving average process. The moving average specification differs from the autoregressive specification in implying a more localized error dependence, in which only first-order and second-order neighbors exhibit error dependence (Anselin and Bera 1998, 250). In contrast to the case in time series, spatial moving average error processes have seen only limited application. As a result, I do not consider them further in this paper.

4 Imposing Constraints on Spatial Dependence

The researcher's first step in modeling spatial autocorrelation is the diagnosis of this autocorrelation in the absence of covariates. When diagnosing this spatial autocorrelation in cross-sectional data, constraints must be imposed on the covariances between observations, since the parameters of the complete covariance matrix are unidentified (Anselin 2002, 256). Assuming that observations cannot influence themselves, there are $N(N-1)$ potential spatial correlations in a cross-sectional data set. Clearly, there is insufficient information in cross-sectional data to estimate each of these separate covariances. Nor can one rely on asymptotics, since there is an incidental parameter problem: the number of potential spatial correlations in the data increases exponentially as the number of observations increases (Anselin and Bera 1998, 242). Instead, the spatial autocorrelation must be parameterized in a limited number of autoregressive or moving average parameters. In most applied work, a single autoregressive or moving average parameter is estimated to capture this spatial dependence.[9] The parameterization of spatial dependence is significantly influenced by the nature of one's spatial data. Here it is useful to consider a taxonomy of forms of spatial data.

First, consider the following spatial stochastic process:

$$Y(\mathbf{s}) : \mathbf{s} \in D, \tag{8}$$

where $\mathbf{s} \in \Re^d$ is a data location in d-dimensional Euclidean space, $D \subset \Re^d$, and $Y(\mathbf{s})$ is a random vector at location $\mathbf{s}$ in D.[10] A particular realization of (8) can be denoted as $y(\mathbf{s}) : \mathbf{s} \in D$. There are three basic forms of data in $\Re^d$: lattice data, geostatistical data, and point pattern data. Each presents its own implications for how spatial dependence is conceptualized and modeled.

In the case of lattice data, D is a fixed subset of $\Re^d$, partitioned into a finite number of areal objects, e.g., polygons (Banerjee, Carlin, and Gelfand 2004, 2). Regular lattice data take the form of a regular grid of square or rectangular objects. Irregular lattice data, which are more common in political science, occur when the shapes of the lattice objects differ from unit to unit. The observed values for the areal units are conceived of as a stochastic draw from a superpopulation, rather than samples from a continuous underlying surface. Given political scientists' interest in irregular lattice objects such as counties, states, and nations, lattice data are the principal data likely to be employed in political science applications of spatial econometrics, and thus this paper's discussion of spatial econometric models and its Monte Carlo applications emphasize a lattice data perspective.

As with lattice data, in geostatistical data, D is again a fixed subset of $\Re^d$. In contrast to lattice data, however, $\mathbf{s}$ varies continuously over D. The observed data are, in turn, sample data from this continuous underlying surface. Generally, the researcher's principal interest in geostatistical data involves inferring information about unobserved observations on the spatial plane from the sampled observations. Kriging, the spatial interpolation of values for these unsampled observations based on the values of the geo-referenced sampled observations, is a critical concern of geostatistical analysis (Cressie 1993). A geostatistical perspective is unlikely to be applicable for most political science applications since, unlike sample data in the natural sciences, most political science samples are not drawn from a continuous spatial field. Geostatistical data will, however, be examined later in this paper as an econometric solution for the identification problem in models with spatial autoregressive lag and spatial autoregressive error parameters applied to identical spatial weights matrices for lattice data (see, e.g., Dubin 2003).

[9]The spatial econometric literature has seen some limited application of "horserace" models in which multiple autocorrelation parameters are estimated simultaneously to determine which potential spatial interactions exert the strongest influence in the DGP (see Selb 2004; Ertur and Koch 2005).

[10]This section draws significantly on the notation and discussion in Banerjee, Carlin, and Gelfand 2004 and the discussion in Cressie 1993.

In contrast to lattice and geostatistical data, in the case of point pattern data, D is a random subset of $\Re^d$. Here the observed locations are the locations of discrete events. Where the central concern with lattice data is modeling any spatial autocorrelation or spatial heterogeneity in the behaviors of the observed units, the central concern with point pattern data is the locations of the events themselves. Thus, the critical question is whether the observed locations of events exhibit spatial clustering beyond what we would expect given the underlying environmental heterogeneity in event occurrence attributable to demographics and other factors (Gatrell et al. 1996, 262). Point pattern data have natural applicability within epidemiology, where spatial patterns in disease contraction, e.g., cancer clusters, are a principal substantive concern.[11] A natural extension of point pattern analysis to political science would be event history models of spatial clustering in events such as international conflicts, policy adoptions, and democratic transitions. As such spatial survival models are beyond the scope of this paper, I do not address point pattern data further.

4.1 Lattice Data Approach

As stated earlier, the particular strategy for imposing constraints on spatial autocorrelation is shaped significantly by the nature of one's spatial data. If the data are lattice data, the constraint follows from the definition of the relevant neighbors for each observation via a spatial weights matrix. Assuming that one is modeling first-order spatial autocorrelation, the j neighbors of unit i are those units for whom such first-order spatial autocorrelation with i is permissable.[12]

The standard spatial weights matrix is an n x n matrix, $\mathbf{W}$, in which all of the j neighbors of i have non-zero values, ($\mathbf{w}_{ij} \neq 0$), the k non-neighbors of i have zero values, ($\mathbf{w}_{ik} = 0$), and i is not a neighbor of itself ($\mathbf{w}_{ii} = 0$). Generally, spatial weights matrices are row-standardized so that the sum of the weights for each observation equals 1. As a result, the spatial influence from neighbors is a weighted average of this influence across the j neighbors.

Clearly, the definition of neighbors is a critical decision in the modeling of spatial autocorrelation. If one misspecifies the spatial dependence in the DGP by treating non-neighbors as neighbors, or vice versa, subsequent spatial autocorrelation estimates will be biased. Closely related is the form and extent of spatial dependence between neighbors. Do all of the j neighbors of i exert the same influence on i? Or is this influence greater for neighbors closer to i? In defining neighbors and the form of spatial dependence between these neighbors, the constraints on potential spatial dependence incorporated in the weights matrix should reflect a priori theoretical expectations.

The simplest definition of neighbors is the first-order contiguity case. Here, there are three principal possibilities. A rook contiguity definition defines objects sharing a common edge with unit i as neighbors of i. A bishop contiguity definition defines objects sharing a common vertex with i as neighbors of i. A queen contiguity definition incorporates both the rook and bishop definitions as any object sharing either a common edge or vertex with i is defined as a neighbor of i. Thus, in the queen contiguity definition, all lattice objects contiguous to i are neighbors of i (Anselin 1988, 18).

Alternatively, one may wish to relax the strict contiguity neighbor definition, but still retain a nearness conception of spatial influence. A k-nearest neighbor definition retains the nearness

[11] As Gatrell et al. (1996, 257) note, the development of modern epidemiology can be traced to Dr. John Snow's early, primitive use of point pattern analysis. Snow's finding that cholera deaths during the London cholera epidemic of 1854 were clustered near the Broad Street pump in Soho led to the removal of the pump's handle, the conclusion that cholera was waterborne rather than airborne, and the subsequent development of modern epidemiology (Centers for Disease Control and Prevention 2004, 783).

[12] First-order spatial autocorrelation exists when neighbors exhibit significant spatial dependence with each other. Higher-order spatial autocorrelation exists when spatial autocorrelation between units exists through dependencies between shared neighbors. Thus, for example, second-order spatial autocorrelation exists if neighbors of neighbors exhibit spatial dependence.

conception while not assuming that there is any substantive importance to the Euclidean distance between units. In a k-nearest neighbor definition, all units among the k nearest neighbors of unit i are treated as neighbors of i, while the $k + 1, \ldots, k + n$ units are treated as non-neighbors. Clearly the value k should be theoretically informed.

Often, the researcher's conception of spatial dependence will incorporate the Euclidean distances between observations. If the researcher believes that all observations within a critical distance of unit i exhibit the same dependence with i and that dependence is absent beyond this threshold, then the researcher may employ a distance band definition of neighbors. All observations within the specified threshold distance of i are treated as neighbors of i while all observations beyond this threshold are treated as non-neighbors of i. The definition of the distance band should reflect a theoretical conception of the potential for spatial dependence.

Often, researchers will wish to posit that the spatial autocorrelation between observations will decline as the distance between observations increases. This implies a distance-decay function and is consistent with Tobler's (1970, 236) first law of geography in which "everything is related to everything else, but near things are more related than distant things." Two common specifications of distance-decay functions that expand upon Tobler's law to potentially incorporate size effects independent of distance are an inverse distance function, as in a gravity model, and a negative exponential function, as in an entropy model. A gravity model takes the following general form:

$$\mathbf{w}_{ij} = \frac{S_j}{d_{ij}^{\alpha}}, \tag{9}$$

where $\mathbf{w}_{ij}$ is the weight for i,j, S_j is a size measure, allowing neighbors to have differing impacts independent of distance, d_{ij} is the distance between i and j, and α is a parameter for the particular form of the distance decay. An entropy model takes the form:

$$\mathbf{w}_{ij} = S_j e^{-\beta d_{ij}}, \tag{10}$$

where β is a distance decay parameter, e is the exponential function, and the remaining notation is as in (9). The size measure, S_j, in both (9) and (10) may reflect population size, GDP, length of border, or other factors posited to influence autocorrelation independent of the distance between observations.

Gravity and entropy models highlight an important consideration in modeling autocorrelation. Although the stochastic process in (8) is conceived of in Euclidean space, we need not limit ourselves to a spatial conception of autocorrelation. The lattice data approach to dependence between observations is sufficiently generalizable that we need not incorporate a spatial component in this dependence at all. We can, for example, posit a social network effect, in which dependence is not a function of the spatial dependence between units, but instead, is a function of these units' membership in a common social network, no matter how spatially dispersed. Alternatively, Beck, Gleditsch, and Beardsley (2006) model dependence as a function of trade flows between countries.

Two particular concerns in the lattice data approach to constraining spatial autocorrelation merit attention. The spatial weights matrix must be exogenous to the spatial econometric model estimated. If covariates from the model are also incorporated in the neighbor definition in the weights matrix, the resulting spatial econometric model will become highly non-linear, and the endogeneity must be removed via instruments (Anselin 2002, 259). Second, significant redundancy may exist in a weights matrix if the researcher is interested in higher-order spatial dependence. Higher-order matrices may incorporate dependencies between observations that also exist at lower orders (e.g., if one of the j neighbors of i is both a second- and first-order neighbor of i). In order to validly estimate the distinct effects of higher-order relationships, such redundancies must be removed before estimation (Anselin 1988, 24).

4.2 Geostatistical Approach

In contrast to the case of lattice data, in geostatistical data, the autocorrelation between observations is typically assumed to be a function of the distance between paired observations. Assuming that the spatial stochastic process in (8) is intrinsically stationary, the variance between observed values is a function only of the distance, $\mathbf{h}$, between the sampled observations (Banerjee, Carlin, and Gelfand 2004, 22). This produces the following function for the variogram of an intrinsically stationary spatial stochastic process:

$$E[y(\mathbf{s}+\mathbf{h}) - y(\mathbf{s})]^2 = Var(y(\mathbf{s}+\mathbf{h}) - y(\mathbf{s})) = 2\gamma(\mathbf{h}), \tag{11}$$

where $2\gamma(\mathbf{h})$ is the variogram and the function $\gamma(\mathbf{h})$ is the semivariogram. Note that paired observations at identical distances exhibit the same variance in an intrinsically stationary spatial process. As a result, the variances between the observed values can be plotted as a function of the distances between paired observations. Geostatistical analysis thus proceeds by plotting an empirical semivariogram of these variances for paired observations at specific distances:

$$\hat{\gamma}(t) = \frac{1}{2N(t)} \sum_{(\mathbf{s}_i,\mathbf{s}_j)\in N(t)} [y(\mathbf{s}_i) - y(\mathbf{s}_j)]^2, \tag{12}$$

where *N(t)* is the set of pairs of points in which $||\mathbf{s}_i - \mathbf{s}_j|| = t$, the length of the separation vector between the points, and $|N(t)|$ is the number of pairs in the set (Banerjee, Carlin, and Gelfand 2004, 29). Geostatistical analysis then consists of estimating an empirical semivariogram to model the variances between pairs of observations at identical distances. In practice, since distances across pairs of sampled observations in D are highly unlikely to be identical, the interunit distances are typically binned so that several pairs of observations are included within each bin. As Banerjee, Carlin, and Gelfand (2004, 29) show, gridding up the t-space into intervals $I_1 = (0, t_1)$, $I_2 = (t_1, t_2), \ldots, I_K = (t_{K-1}, t_K)$ produces the revised definition of $N(t)$ in practice:

$$N(t_k) = (\mathbf{s}_i, \mathbf{s}_j) : ||\mathbf{s}_i - \mathbf{s}_j|| \in I_k,\ k = 1, \ldots, K. \tag{13}$$

The empirical semivariogram is then plotted, and a theoretical, parametric semivariogram is then fit to the plot. This parametric semivariogram is employed for kriging purposes; because of the noise in any empirical semivariogram, a parametric distribution is seen as providing better guidance for interpolation of unsampled values. As Banerjee, Carlin, and Gelfand (2004, 30) note, in practice this fitting of a parametric distribution to the empirical semivariogram has largely been an art, focused more on visual inspection of the empirical graph, rather than the use of goodness of fit measures, though the latter have been developed. A common parametric distribution for the semivariogram, the Gaussian, is discussed later in the consideration of the geostatistical approach as a solution to the identification problem in a lattice model with identical weights matrices for the spatially lagged dependent variable and spatial autoregressive error dependence.

5 Global Measures of Spatial Autocorrelation

Tests for spatial autocorrelation in lattice data can proceed at either the global or local levels. Tests for global spatial autocorrelation examine whether the data as a whole exhibit spatial autocorrelation (against a null of spatial randomness) as well as the strength and direction (positive or negative) of any spatial autocorrelation. Tests for local spatial autocorrelation (again, against a null of spatial randomness) identify particular observations that are autocorrelated with neighboring observations on the random variable of interest and also determine the strength and, depending upon the statistic, also the direction of this spatial autocorrelation. Typically, researchers first employ global tests of spatial autocorrelation and subsequently employ local tests to decompose the global result.

Tests for either global or local spatial autocorrelation in lattice data proceed through the use of a Γ index. A Γ index consists of the sum of the cross products of the corresponding elements $\mathbf{w}_{ij}, \mathbf{v}_{ij}$ in two matrices, $\mathbf{W}$ and $\mathbf{V}$:

$$\Gamma = \Sigma_i \Sigma_j \mathbf{W}_{ij} \mathbf{V}_{ij}, \tag{14}$$

where $\mathbf{W}$ and $\mathbf{V}$ are, respectively, matrices of spatial (dis)similarity (a spatial weights matrix) and value (dis)similarity (Anselin 1995, 98). Measures of spatial autocorrelation are variants of this Γ index, with the $\mathbf{v}_{ij}$ elements in $\mathbf{V}$ reflecting how value (dis)similarity is conceptualized in the particular form of the Γ index.

Global join count statistics are applicable when the random variable is dichotomous. By convention, units with a value of 1 on the binary variable are denoted as Black (B) and units with a value of 0 are denoted as White (W). There are then three possible types of joins: BB, BW, and WW. Assuming that the weights matrix, $\mathbf{W}$, is symmetric, the number of BB, BW, and WW joins is, respectively,

$$BB = \frac{1}{2} \Sigma_i \Sigma_j \mathbf{w}_{ij} y_i y_j, \tag{15}$$

$$BW = \frac{1}{2} \Sigma_i \Sigma_j \mathbf{w}_{ij} (y_i - y_j)^2, \tag{16}$$

and

$$WW = A - (BB + WW), \tag{17}$$

where y_i and y_j are the observed values on the random variable at locations i and j and A is the number of joins in the data (Cliff and Ord 1973, 4).[13] The observed frequencies of BB, BW, and WW joins are compared to the expected frequencies under the null of spatial randomness to determine whether the data are spatially autocorrelated (Cliff and Ord 1981).

In contrast to the dichotomous case, the two principal global spatial autocorrelation measures for interval-level variables are the Moran's I and Geary's c. Again, both are Γ indices, with the two measures differing in their conception of value (dis)similarity. For the Moran's I, the value (dis)similarity in the $\mathbf{v}_{ij}$ elements of the $\mathbf{V}_{ij}$ matrix is measured as deviations from the mean. The global Moran's I is thus:

$$I = \frac{N}{S} \frac{\Sigma_i \Sigma_j \mathbf{w}_{ij} (y_i - \overline{y})(y_j - \overline{y})}{\Sigma_i (y_i - \overline{y})^2}, \tag{18}$$

where N is the number of observations, S is the sum of the weights, y_i and y_j are the values on the random variable at locations i and j, and $\overline{y}$ is the mean on y.

A positive global Moran's I that differs significantly from the expected value under the null indicates positive spatial autocorrelation – the clustering of similar values on the random variable among neighboring observations. A negative global Moran's I that differs significantly from the expected value under the null indicates negative spatial autocorrelation – the clustering of dissimilar values on the random variable among neighboring observations.

Although both this interpretation of the global Moran's I and the form of the measure suggest

[13] In contrast to the standard case for other spatial autocorrelation measures, $\mathbf{W}$ is unstandardized in join count analysis. The joins in a join-count analysis are symmetric and binary (Moran 1948). In contrast, row-standardization of a weights matrix produces a non-symmetric, non-binary matrix.

a correlation coefficient, Moran's I differs from a correlation coefficient in two key respects. The expected value of Moran's I under the null hypothesis is not zero, but instead, is $\frac{-1}{N-1}$, and thus is a function of the number of observations (Anselin 1992, 133). As a consequence, the expected value of Moran's I under the null is negative, though it approaches zero asymptotically. Moreover, unlike a correlation coefficient, Moran's I is not bounded at $\pm$ 1. Instead, the bounds are a function of the data and will generally be narrower than $\pm$ 1 (Cliff and Ord 1981, 21).

Where the global Moran's I defines value (dis)similarity as deviations from the mean, the global Geary's c defines value (dis)similarity as the squared difference in values between neighboring observations:

$$c = \frac{N-1}{2S} \frac{\Sigma_i \Sigma_j \mathbf{w}_{ij}(y_i - y_j)^2}{\Sigma_i (y_i - \overline{y})^2}, \tag{19}$$

where the notation is as in (18). Interpretation of Geary's c differs significantly from interpretation of Moran's I. The expected value of Geary's c under the null is 1. A Geary's c that is significantly larger than 1 indicates negative spatial autocorrelation, while a Geary's c that is significantly smaller than 1 indicates positive spatial autocorrelation (Anselin 1992, 133). Due to the squared term in the numerator in (19), Geary's c gives greater weight to extreme values than does Moran's I (Cliff and Ord 1981, 14-15). As a consequence, the global Moran's I is generally preferred in practice.

Inference on a Γ index of spatial autocorrelation takes either of two approaches. If the random variable is normally distributed with a constant variance, then the Γ indices are asymptotically normally distributed under the null. Inference then proceeds by comparing the observed z-value to its probability given the normal distribution.

Alternatively, one can apply a randomization approach. In the randomization approach for a Γ measure, the observed values on the random variable are randomly permuted across all locations. Note that this randomization without regard to spatial location corresponds to the null of spatial randomness. The appropriate Γ index is then calculated for each permutation to form an empirical reference distribution. The observed global spatial autocorrelation measure is then compared to the reference distribution to determine pseudo-statistical significance.

6 Local Measures of Spatial Autocorrelation

Often our interest lies not in determining whether the data as a whole exhibit spatial autocorrelation, but instead, in identifying the specific observations that exhibit spatial autocorrelation with their neighbors. Here, the researcher can examine local measures of spatial autocorrelation. There are two classes of local spatial autocorrelation measures. The first, such as Ord and Getis' (1995) (see also Getis and Ord 1992) G_i and G_i^* statistics, have no global spatial autocorrelation analog. The second, local indicators of spatial association (LISA statistics), have the attractive feature that they are proportional to a corresponding global measure. As will be discussed, this relationship to a global measure makes LISA statistics particularly helpful in disaggregating global spatial autocorrelation.

6.1 Ord and Getis Statistics

As originally developed (Getis and Ord 1992), the G_i and G_i^* statistics were expressly for the case of a distance-based neighbor definition and a symmetric, binary weights matrix. In a significant revision of the measures, Ord and Getis (1995) extended the statistics to the case of non-distance-based definitions of neighbors, and non-symmetric, non-binary weights matrices. I focus my discussion in this section on these revised G_i and G_i^* statistics. The G_i and G_i^* statistics, respectively, take the forms:

$$G_i = \frac{\Sigma_j \mathbf{w}_{ij} y_j - \Sigma_j \mathbf{w}_{ij} \overline{y}(i)}{s(i)([(n-1)S_{1i}) - (\Sigma_j \mathbf{w}_{ij})^2]/(n-2))^{\frac{1}{2}}}, \tag{20}$$

and

$$G_i^* = \frac{\Sigma_j \mathbf{w}_{ij} y_j - (\Sigma_j \mathbf{w}_{ij} + \mathbf{w}_{ii}) \overline{y}}{s([(n)S_{1i}^*) - (\Sigma_j \mathbf{w}_{ij} + \mathbf{w}_{ii})^2]/(n-1))^{\frac{1}{2}}}, \tag{21}$$

where $\mathbf{w}_{ij}$ are the elements of the spatial weights matrix $\mathbf{W}$ corresponding to the j units in the neighborhood set for unit i, $\mathbf{w}_{ii}$ is a non-zero weight in the case in which i is in its own neighborhood set, $\overline{y}(i)$ is the mean of the values on the random variable for the j neighbors of i, $\overline{y}$ is the mean of the values on the random variable when i is in its own neighborhood set, $S_{1i} = \Sigma_j \mathbf{w}_{ij}^2$, $S_{1i}^* = \Sigma_j \mathbf{w}_{ij}^2 + \mathbf{w}_{ii}^2$, and s^2 is the sample variance (Ord and Getis 1995, 289). G_i and G_i^* are thus standard normal variates, where G_i^* includes i in its own neighborhood set (thus incorporating a non-zero weight for $\mathbf{w}_{ii}$) and G_i does not.

Note that common to other local spatial autocorrelation measures and unlike global measures, the Ord and Getis statistics incorporate only the neighborhood set for each observation. However, the incorporation of observations in their own neighborhood sets in the G_i^* statistic is non-standard, running counter to the construction of most spatial weights matrices. In practice, political scientists are unlikely to wish to include observations as their own neighbors in most applications, and thus, if wishing to employ one of the Ord and Getis statistics, will likely favor the G_i statistic.

The interpretation of values on the G_i and G_i^* statistics is quite distinct from interpretation of Moran's I. Positive values on Moran's I indicate positive spatial autocorrelation. This includes both cases where high values on the random variable spatially cluster with other high values and cases where low values on the variable spatially cluster with other low values. Negative values on Moran's I, in contrast, indicate that observations with higher values on the random variable are neighbors of observations with lower values on the random variable, and vice versa. In contrast, positive values on the G_i and G_i^* statistics indicate that high values are spatially clustered with other high values on the random variable. Negative values on the G_i and G_i^* statistics indicate that low values are spatially clustered with other low values on the random variable. Negatively autocorrelated cases (those with negative values on Moran's I) will exhibit only weak, negative G_i and G_i^* values (Getis and Ord 1992, 198). Note that a consequence of this is that the G_i and G_i^* statistics, unlike Moran's I, cannot distinguish cases of positive spatial autocorrelation from cases of negative spatial autocorrelation.

A key contribution of the G_i and G_i^* statistics, as with other measures of local spatial autocorrelation, is that they aid in identification of local pockets of spatial clustering. Such clustering, moreover, may occur even in the absence of global spatial autocorrelation. In fact, in an application of the initial, distance-based forms of the G_i and G_i^* statistics, Getis and Ord identified five counties in North Carolina that exhibited significant local spatial autocorrelation in Sudden Infant Death Syndrome (SIDS) rates despite the lack of a global structure to such death rates in the state as a whole (Getis and Ord 1992, 200). Local spatial autocorrelation can exist in the absence of global autocorrelation when the clustering at the local level is limited as a proportion of the overall number of observations or when local patterns are off-setting, producing no global pattern as a consequence. Clearly, a critical implication of Getis and Ord's analysis is that scholars should not limit their analyses to the global level, for they risk overlooking potentially important local clustering in phenomena if they do so.

6.2 Local Indicators of Spatial Association (LISA Statistics)

The second form of local spatial autocorrelation measure is a local indicator of spatial association (LISA statistic). Anselin (1995, 94) defines a LISA statistic as any statistic satisfying the following two conditions: the LISA for each observation measures the extent of significant spatial clustering of similar values around the observation, and the sum of LISAs for all observations is proportional to a corresponding global indicator of spatial association.[14] Formally, the second condition implies:

$$\sum_i L_i = \gamma \Lambda, \tag{22}$$

where L_i are the LISA statistics for each observation, γ is a scale factor, and Λ is a corresponding global spatial autocorrelation measure. The sum of the LISAs is thus proportional to a global analog up to a scaling factor.

Although the G_i and G_i^* statistics meet the first requirement of a LISA statistic, they do not meet the second. Instead, the principal LISA statistics are the local Moran's I and the local Geary's c. The forms of the local Moran and local Geary are, respectively:

$$I_i = \frac{\Sigma_j \mathbf{w}_{ij}(y_i - \overline{y})(y_j - \overline{y})}{(y_i - \overline{y})^2}, \tag{23}$$

and

$$c_i = \frac{\Sigma_j \mathbf{w}_{ij}(y_i - y_j)^2}{(y_i - \overline{y})^2}, \tag{24}$$

where the notation is as in their global analogs in (18) and (19). Again, only the j neighbors of i are incorporated in the LISA for i. The interpretation of values of the local Moran's I and local Geary's c is analogous to their global counterparts.

As with the G_i and G_i^* statistics, LISAs aid considerably in identifying local clustering of similar or dissimilar values on the random variable. This is additionally aided through the use of a Moran scatterplot. A Moran scatterplot plots observed values on the random variable (along the x-axis) and the weighted average of the values in each observation's neighborhood set (along the y-axis) as standardized values. The result is a plot in four quadrants. Significant LISAs in the upper right quadrant denote positive local spatial autocorrelation above the mean on the variable. Significant LISAs in the lower left quadrant denote positive local spatial autocorrelation below the mean. Significant LISAs in the upper left quadrant indicate negative local spatial autocorrelation in which observations have lower values than their neighbors' values. Significant LISAs in the lower right quadrant indicate negative local spatial autocorrelation in which observations have higher values than their neighbors' values. The observations' locations in the Moran scatterplot can then be combined with their significance on the local autocorrelation measure to map areas of positively autocorrelated values above or below the mean as well as observations exhibiting particular forms of negative local spatial autocorrelation.

In addition to the identification of local spatial clustering, the correspondence between LISA statistics and global spatial autocorrelation measures carries significant additional advantage in decomposing these global measures. Through the estimation of LISA statistics, the researcher can identify which observations are consistent with the global pattern of positive or negative spatial autocorrelation and which observations run counter to this global pattern. Moreover, high leverage observations can also be identified. Often this follows a two-sigma rule: observations with LISAs

[14] Although Anselin defines a LISA in terms of the clustering of similar values, his definition is unnecessarily restrictive here. LISA statistics, like their global analogs, identify both positive and negative spatial autocorrelation.

that are more than two standard deviations from the mean can be examined to determine whether they are unduly influencing the global measure (Anselin 1995, 97).

Inference on the G_i and G_i^* statistics and LISA statistics proceeds in the same manner as for the global measures of spatial autocorrelation. On the one hand, the researcher can rely on asymptotics and assume a normal distribution. Alternatively, the researcher can employ a randomization approach in which the values on the variable are randomly permuted across all locations. In contrast to the global approach, however, only as many observations as are in each observation's neighborhood set need be resampled from these permuted values. Repeated sampling with replacement produces an empirical reference distribution and the observed local measure of spatial autocorrelation for each observation is compared to the distribution of spatially randomized values to determine statistical significance.

7 Spatial Heterogeneity

The diagnosis of univariate spatial autocorrelation via global and local measures of spatial autocorrelation is a prerequisite for modeling this spatial dependence. In section 8, I examine diagnostics for spatial lag and spatial error dependence in OLS specifications. These diagnostics are critical for determining whether a spatial econometric specification is required to model the spatial dependence, and if so, whether the specification should reflect diffusion-based dependence or attribute-based dependence. First, however, I consider an alternative source of the univariate spatial dependence identified by global and local measures of spatial autocorrelation – spatial heterogeneity.

The spatial autocorrelation diagnosed by measures of spatial autocorrelation such as Moran's I and Geary's c may be produced not by a diffusion process nor by an attributional process in the DGP. Instead, it may be produced by behavioral heterogeneity. Within a modeling context, this will take the form of spatial heterogeneity in parameters. If this parameter heterogeneity is not modeled, spatial dependence will persist in the presence of covariates. Residual spatial dependence in a multivariate model may also be produced by another undiagnosed form of spatial heterogeneity – functional form heterogeneity. Depending upon the modeling strategy employed to account for spatial heterogeneity in parameters or functional form, neither a spatial econometric specification nor an instrumental variables specification may be required to model the spatial heterogeneity. Instead, the researcher may choose to model the spatial heterogeneity via standard econometric approaches.

7.1 Spatial Heterogeneity in Parameters

The intuition of how spatial heterogeneity in parameters may produce univariate spatial autocorrelation is straightforward. If a covariate differs in its effects on the political phenomenon of interest across observations, and if the effects are similar among neighboring observations, this may produce similar values on the dependent variable. I examine three sets of approaches for modeling spatial heterogeneity in parameters – spatial random coefficients models, spatial switching regressions models for discrete parameter heterogeneity, and spatially varying coefficients models for continuous parameter heterogeneity – in turn next.

7.1.1 Spatial Random Coefficients Models

Random coefficients models have received extensive use in econometrics as an approach for modeling heterogeneity in parameters (Swamy 1970; Hsiao 1975). The standard, non-spatial random coefficients specification takes the following form:

$$y_i = \mathbf{X}_i \beta_i + \varepsilon_i$$

$$\beta_i = \beta + \mu_i, \tag{25}$$

where β_i is no longer assumed constant, but instead is allowed to vary across observations as a function of a mean, β, and a stochastic term, μ_i. The random coefficients model induces heteroskedasticity, which is modeled via FGLS.

In the standard random coefficients model, the stochastic variation, μ_i, around the common mean, β, is assumed to be random with regard to the spatial locations of observations. In contrast, in the spatial random coefficients approach, there is spatial dependence in this variation around the mean. As a result, the spatial random coefficients specification takes the form:

$$y_i = \mathbf{X}_i \beta_i + \varepsilon_i$$

$$\mu_i = \beta_i - \beta \tag{26}$$

with

$$\mu_i = \lambda \Sigma_j \mathbf{w}_{ij} (\beta_j - \beta) + \xi_i \tag{27}$$

or

$$\mu_i = \lambda \Sigma_j \mathbf{w}_{ij} \mu_j + \xi_i, \tag{28}$$

where μ_i are the stochastic variations around the common mean, ξ are i.i.d. error terms, λ is the spatial autoregressive parameter for the stochastic dependence around the common mean, and $\mathbf{w}_{ij}$ are the elements of the spatial weights matrix, $\mathbf{W}$, for the neighborhood set of i (Anselin and Cho 2002, 283). The simultaneous spatial dependence in (28) induces heteroskedasticity and a nonzero covariance between the errors for all observations. Ignoring this heteroskedasticity and spatial dependence will produce inefficient estimates of the common mean, β, and biased estimates of the variance (Anselin and Cho 2002). Thus, the spatial dependence in the error terms must be accounted for in estimation of spatial random coefficients models and the explicitly spatial econometric form of these models differs from other approaches for modeling spatial heterogeneity in parameters or functional form.

7.1.2 Spatial Switching Regressions

Discrete spatial heterogeneity in parameters occurs when a parameter is homogeneous within spatial subsets of the data and heterogeneous across these subsets. The concept of discrete spatial heterogeneity in parameters can be found in many political science theories. Consider, for example, studies of American voting behavior, policy diffusion, or legislative studies where behavioral parameters are hypothesized to operate differently in the South vs. the non-South. Or consider the comparative politics concepts of the global North and South where relationships such as those between financial weakness and capital account liberalization or between globalization and welfare state spending differ in discrete spatial subsets of the data (Brooks 2004; Rudra 2002).

Tests on such structural instability in parameters are well-known in the time series literature where Chow tests are often employed to diagnose structural breaks in parameters over time. We can adapt the Chow test to the case of spatial structural instability via a spatial switching regressions specification, also known as a spatial regimes model. Here, in contrast to the standard time series approach, where the subsets of the data reflect discrete temporal periods, the subsets of the data are instead spatially indexed:

$$\begin{bmatrix} y_i \\ y_j \end{bmatrix} = \begin{bmatrix} \mathbf{X}_i & 0 \\ 0 & \mathbf{X}_j \end{bmatrix} \begin{bmatrix} \beta_i \\ \beta_j \end{bmatrix} + \begin{bmatrix} \varepsilon_i \\ \varepsilon_j \end{bmatrix}, \tag{29}$$

where i,j index discrete spatial subsets of the data, D_i, D_j in $\Re^d$. The spatial Chow test then proceeds as a test of the null that $\beta_i = \beta_j$ via an F test:

$$\mathbf{C} = [(\mathbf{e}_R^{'}\mathbf{e}_R - \mathbf{e}_U^{'}\mathbf{e}_U)/K][\mathbf{e}_U^{'}\mathbf{e}_U/(N-2K)] \sim F_{K,N-2K}, \tag{30}$$

where $\mathbf{e}_R$ and $\mathbf{e}_U$ are the OLS residuals from a restricted model (with the equality restriction imposed) and from an unrestricted model, N is the number of discrete spatial subsets in the data, and K is the number of regressors (Anselin 1990, 192).

If the switching regressions specification models fully the sources of univariate spatial autocorrelation, a spatial econometric specification is not required. If, however, spatial lag or spatial error dependence remains after the spatial heterogeneity in parameters has been modeled, the researcher will wish to employ either an instrumental variables specification or a spatial econometric specification incorporating both spatial heterogeneity and spatial autocorrelation.

A critical related concern is the effect of spatial dependence on Chow tests. Anselin (1990) demonstrates that spatial error dependence can invalidate spatial Chow tests for spatial structural instability in parameters. Consider the spatial autoregressive error process:

$$\varepsilon = (\mathbf{I} - \lambda\mathbf{W})^{-1}\varepsilon. \tag{31}$$

In this case, a common spatial autoregressive process extends across the spatial regimes, allowing for the use of a single spatial weights matrix, $\mathbf{W}$. When such an autoregressive error process exists, the correct form for the spatial Chow test becomes:

$$\mathbf{C}_{err} = [\mathbf{e}_R^{'}(\mathbf{I}-\lambda\mathbf{W})^{'}(\mathbf{I}-\lambda\mathbf{W})\mathbf{e}_R - \mathbf{e}_U^{'}(\mathbf{I}-\lambda\mathbf{W})^{'}(\mathbf{I}-\lambda\mathbf{W})\mathbf{e}_U]/\sigma^2 \sim \chi^2_K. \tag{32}$$

In a series of Monte Carlo experiments, Anselin demonstrates that a spatial autoregressive error process can inflate both the size and power of standard spatial Chow tests. There is, however, a noteworthy asymmetry to these effects. Overrejection of the null increases markedly as λ reaches .5 for certain spatial regime layouts. In contrast, negative spatial autocorrelation has little effect on the size of the standard spatial Chow. Negative spatial autocorrelation does tend to produce a modest reduction in the power of the standard spatial Chow, but this reduction in power is much less severe than the inflation of the power of the test in the case of positive spatial autocorrelation.

Given that positive, rather than negative spatial autocorrelation is likely to predominate in political science data, scholars should be keenly aware of the implications of positive spatial error dependence for measures of spatial structural instability in parameters. Scholars should thus first estimate spatial dependence via global and local measures of spatial autocorrelation. Given a finding of spatial dependence, scholars should then incorporate this dependence in their spatial Chow tests. This will produce more valid estimates of both forms of spatial effects than considering only spatial heterogeneity or spatial dependence in isolation.

7.1.3 Spatial Expansion and Geographically Weighted Regression Models

In contrast to the discrete spatial heterogeneity implicit in a spatial regimes approach, we may posit instead continuous spatial heterogeneity in parameters. Here, continuous parameter heterogeneity is typically conceived of as reflecting a continuous spatial drift in a parameter across spatial locations, where the parameter takes on differing values at each spatial location and there is a continuous drift in the parameter values as one moves across one of the d dimensions in $\Re^d$. The principal approaches for modeling this continuous spatial heterogeneity are spatial expansion models and geographically weighted regression models.

In the spatial expansion model, the parameters of an initial model are themselves modeled as functions of a set of covariates in an expansion equation, producing a combined terminal model. As in a standard hierarchical modeling approach, lower level areal units could be conceived of as nested within higher level areal units, with lower-level parameters modeled as a function of higher-level effects. This approach, however, will generally not prove useful in modeling continous parameter heterogeneity. Instead, we will typically model the parameters in the initial equation as functions of the x,y coordinates of the observations. Thus, the spatial expansion model for continuous parameter heterogeneity may take the form:

$$y_i = \beta_0 + \mathbf{x}_i \beta_{1i} + \varepsilon_i \tag{33}$$

with

$$\beta_i = \gamma_0 + \gamma_1 \mathbf{z}_{1i} + \gamma_2 \mathbf{z}_{2i} + \mu_i \tag{34}$$

producing the combined model

$$y_i = \beta_0 + \gamma_0 \mathbf{x}_i + \gamma_1 \mathbf{z}_{1i} \mathbf{x}_i + \gamma_2 \mathbf{z}_{2i} \mathbf{x}_i + \mu_i \mathbf{x}_i + \varepsilon_i, \tag{35}$$

where $\mathbf{z}_1$ and $\mathbf{z}_2$ are covariates measuring the x,y coordinates, ε_i is the error term in the initial equation, and μ_i is the error term in the expansion equation. The simple linear expansion in coordinates can be modified to reflect more realistic trend surfaces, such as a second degree polynomial allowing for curvature of the surface or higher-degree polynomials to incorporate multiple inflection points on the surface. Clearly, the inclusion of the error term μ_i in the expansion equation induces heteroskedasticity. The error variance will include a constant variance σ^2_ε from the initial model as well as a sum of squares of the expanded variables weighted by the error variance in the expansion equation $(\sigma^2_\mu)(\mathbf{x}^2)$ (Anselin 1992, 341). This heteroskedasticity is modeled via FGLS.

Geographically weighted regression presents an alternative strategy for modeling continuous spatial heterogeneity in parameters. Where the spatial expansion method models the parameters as functions of the observations' coordinates, geographically weighted regression employs distance weights to give more spatially proximate observations greater weight in the calculation of the spatially varying parameters. The concept is straightforward. An unweighted OLS model gives each observation equal weight in calculation of the common parameter, β. In a geographically weighted regression, this approach is modified to weight observations' influences on the spatially varying parameter, β_i, by their proximity to i (see Fotheringham, Charlton, and Brunsdon 1998). The weights may reflect any of the approaches discussed in Section 4.1.

7.2 Spatial Heterogeneity in Functional Form

Greene (2003, 192) emphasized incorrect functional form as a source of cross-sectional (though not explicitly spatial) autocorrelation in errors. The intuition of how an incorrect functional form can induce spatial error dependence is straightforward. Assume, for example, that the response variable is modeled as a linear function of the covariates. However, perhaps the correct functional form is an alternative function, such as a quadratic. If so, the misspecification is likely to produce correlated errors, and this is likely to particularly be the case at neighboring locations.

Closely related is the issue of spatial heterogeneity in functional form. Here, different functional forms are valid in different spatially indexed subsets of the data. Thus, for example, where D_i and

D_j are subsets of $\Re^d$, a linear functional form may be valid in D_i while a quadratic form is valid in D_j. Fitting a linear functional form to all observations will produce spatial clustering in the errors in D_j.

The issues of spatial dependence and functional form must be considered in tandem. Standard tests of functional form, such as the Box-Cox, typically assume no spatial autocorrelation in the DGP. Thus, it is important to examine how spatial dependence may affect tests for functional form.

In a series of Monte Carlo experiments, Baltagi and Li (2001, 2005) examine the performance of four sets of test statistics – joint LM tests, unidirectional LM tests, modified LM tests, and conditional LM tests – for functional form in the presence of either spatial lag or spatial error dependence. The joint Lagrange multiplier tests test the joint null of no spatial lag (error) dependence and a particular functional form (a linear or a loglinear form, depending on the null). The simple unidirectional LM tests test the null of a linear or loglinear functional form, assuming no spatial lag or error dependence. Such simple LM tests, however, are not robust against the presence of the alternative form of spatial dependence (the presence of spatial lag dependence when error dependence is assumed to be absent, and vice versa). Under such misspecification of spatial dependence, the simple unidirectional LM test converges to a noncentral chi-square. The modified LM test developed by Bera and Yoon (1993) accounts for the noncentrality parameter, producing a test statistic that is robust to misspecification of the spatial autocorrelation (Baltagi and Li 2001, 200-202). Finally, Baltagi and Li also examine conditional LM tests of both linear and loglinear functional forms, given unknown autoregressive parameters, ρ and λ, for spatial lag and error dependence.

Baltagi and Li's Monte Carlos produce similar results whether spatial lag or spatial error dependence exists in the DGP. The joint LM tests for functional form overreject the null as both ρ and λ depart from zero. Similarly, overrejection increases on the modified LM tests as spatial lag or spatial error dependence increase. Perhaps somewhat surprisingly, the simple unidirectional LM tests are not sensitive to departures from the null of spatial randomness. Finally, the conditional LM tests both overreject the null as ρ and λ depart from zero. Clearly one must be sensitive to spatial dependence in the DGP when intrepreting tests for functional form.

Recently, Beck and Jackman (1998) have argued for the use of generalized additive models (GAMs) to model nonlinear functional forms. GAMs are more flexible than standard approaches such as the Box-Cox transformation, as they allow for local rather than global departures from linearity and model these departures nonparametrically. Spatial dependence, however, is typically not incorporated in standard GAMs. This is problematic, as spatial dependence in the DGP can produce concurvity (additive dependence) between the covariates in a GAM. Ramsay, Burnett, and Krewski (2003) demonstrate via a set of Monte Carlo experiments that this concurvity induced by spatial dependence can bias downward standard error estimates for linear parameters in a semiparametric GAM, producing Type I errors. The bias increases as spatial dependence and concurvity increase. Thus, though often overlooked, spatial dependence poses critical implications for inference, even when scholars are sensitive to questions of functional form.

7.3 Spatial Heterogeneity in Error Variance

A third form of spatial heterogeneity is spatial heterogeneity in error variance. As the discussion in Section 2 demonstrates, heteroskedasticity is a common concern in cross-sectional data. Spatial heteroskedasticity – non-constant error variance that is related to units' spatial locations – is particularly likely given the aggregate areal data employed in most spatial econometric applications. Often spatial data take the form of rates based on differing population sizes. Rates based on smaller population bases will tend to have larger error variances than rates based on larger population bases. Units with similar population sizes will often be located near each other (e.g., neighboring sparsely populated rural counties), inducing spatial heteroskedasticity. Spatial heteroskedasticity may also be induced by variation in measurement error across aggregate units, or by other well-known sources of heteroskedasticity, such as variation in behavioral processes across geographically located populations. The implications of heteroskedasticity for OLS estimates are well-known. OLS parameter estimates remain unbiased, but are inefficient and standard error estimates are biased.

There are two interrelated concerns regarding heteroskedasticity. The first is the effect of heteroskedasticity on diagnostics for spatial dependence. I defer this discussion until Section 8. The second is the effect of spatial autocorrelation on standard tests for heteroskedasticity.

As with tests of functional form, standard tests against heteroskedasticity such as the White and Breusch-Pagan tests do not incorporate spatial dependence. In a series of Monte Carlo experiments, Anselin and Griffith (1988) examine the performance of both tests in the presence of both positive and negative spatial error dependence. Consistent with the common criticism of White's general test, they find that the test has consistently low power, in this case regardless of the degree of heteroskedasticity or error dependence. They find that the power of the Breusch-Pagan is more variable in response to levels of spatial error dependence. Specifically, they find that the power of the test declines as positive spatial error dependence increases in value, with the effects most pronounced as λ increases to .5 and above. To avoid the possibility of Type II errors, Anselin and Griffith propose a sequential test for heteroskedasticity and spatial autocorrelation. First, they suggest a joint test for spherical errors (against the joint existence of heteroskedastic and autocorrelated disturbances). This test is equivalent to the sum of a Breusch-Pagan LM test and an LM test against spatial error autocorrelation. A rejection of the joint null is then followed by separate tests against heteroskedasticity and spatial error autocorrelation.

8 Diagnostics for Spatial Dependence in OLS Models

Assuming that one has identified significant spatial dependence via global and local tests of spatial autocorrelation and that spatial heterogeneity in parameters does not account fully for this univariate spatial dependence, the next step is to model this spatial autocorrelation via covariates. The researcher thus specifies and estimates an OLS model and applies diagnostics for spatial lag and spatial error dependence to determine whether the covariates fully model the spatial dependence. If the diagnostics indicate spatial lag dependence in the presence of covariates, evidence consistent with a diffusion process exists. At this point, the researcher will wish to estimate either a maximum likelihood mixed regressive, spatial autoregressive (spatial lag) specification or an instrumental variables specification. If the diagnostics indicate residual spatial error dependence in the OLS specification, the researcher may either wish to estimate a more fully specified OLS model or estimate a spatial error model. I examine the principal diagnostics for spatial lag and spatial error dependence in OLS models next.

8.1 The Moran's *I* Diagnostic for Spatial Error Dependence in an OLS Model

Given the wide familiarity with the Moran's I as a diagnostic for univariate spatial autocorrelation, it is not surprising that the test has been extended to the diagnosis of spatial dependence in the presence of covariates. The Moran's I test for spatial error dependence in OLS regression residuals takes the form:

$$I = \frac{N}{S}\frac{\mathbf{e}'\mathbf{W}\mathbf{e}}{\mathbf{e}'\mathbf{e}}, \tag{36}$$

where N is the number of observations, S is the sum of the weights, $\mathbf{e}$ are the residuals from an OLS regression, and $\mathbf{W}$ is the spatial weights matrix. Assuming that $\mathbf{W}$ is row-standardized,

$\frac{N}{S}$ simplifies to 1. As can be seen from the form of (36), Moran's I is the spatial analog of the Durbin-Watson test for serial correlation in residuals. Although Greene (2003, 192) recommends such a test as a diagnostic for cross-sectional autocorrelation, the Moran's I is actually a highly unreliable diagnostic for spatial error dependence in OLS residuals.

Anselin and Rey (1991) examine the performance of the Moran's I diagnostic for spatial error dependence in a set of Monte Carlo experiments. The Moran's I is found to have low power in the presence of non-normally distributed errors, resulting in a significant underrejection of the null. In contrast, the power of Moran's I is inflated in the presence of heteroskedasticity, leading to an overrejection of the null. Moreover, the Moran's I diagnostic for spatial error dependence is found to have power against both error and lag dependence, rejecting the null even in the latter case in the absence of error dependence. As a result, it should not be relied on for guidance for the proper alternative specification.

8.2 Lagrange Multiplier Diagnostics for Spatial Lag and Spatial Error Dependence in an OLS Model

Given the poor performance of the Moran's I diagnostic, Anselin and Rey argue for the use instead of Lagrange Multiplier diagnostics in OLS specifications. Here, there are two basic LM diagnostics. The first is a Lagrange Multiplier diagnostic for spatial lag dependence in the presence of covariates in an OLS model. Following Anselin and Rey's (1991, 119) notation, the LM diagnostic for lag dependence takes the form:

$$LM_{Lag} = [N\mathbf{e}'\mathbf{W}_1 y/\mathbf{e}'\mathbf{e}]^2[N(\mathbf{W}_1\mathbf{X}\hat{\beta})'\mathbf{M}(\mathbf{W}_1\mathbf{X}\hat{\beta})/\mathbf{e}'\mathbf{e} + \text{tr}(\mathbf{W}_1'\mathbf{W}_1 + \mathbf{W}_1^2)]^{-1}, \tag{37}$$

where N is the number of observations, $\mathbf{e}$ are the OLS residuals, $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\hat{\beta}$ is the OLS estimate of β, tr is the matrix trace operator, and $\mathbf{W}_1$ is the spatial weights matrix for the spatially lagged dependent variable. The Lagrange Multiplier diagnostic for spatial error dependence in the presence of covariates in an OLS model takes the form:

$$LM_{Error} = [N\mathbf{e}'\mathbf{W}_2\mathbf{e}/\mathbf{e}'\mathbf{e}]^2[\text{tr}(\mathbf{W}_2'\mathbf{W}_2 + \mathbf{W}_2^2)]^{-1}, \tag{38}$$

where $\mathbf{W}_2$ is the spatial weights matrix for the spatially lagged error term, and the remaining notation is as in (37) (Anselin and Rey 1991, 119).[15]

The application of both tests is straightforward. The researcher specifies an OLS model and runs the diagnostics on this model. For each diagnostic, the null hypothesis is the absence of the particular form of spatial dependence. If the null cannot be rejected on either diagnostic, the OLS specification is sufficient for modeling the spatial dependence estimated via the global and local measures of spatial autocorrelation. If the null is rejected on the Lagrange Multiplier diagnostic for spatial lag dependence, but is not rejected on the Lagrange Multiplier diagnostic for spatial error dependence, the researcher should proceed by estimating either a mixed regressive, spatial autoregressive (spatial lag) model or an instrumental variables model with an instrument for the spatially lagged dependent variable. If, alternatively, the null is rejected on the Lagrange Multiplier diagnostic for spatial error dependence but is not rejected on the Lagrange Multiplier diagnostic for spatial lag dependence, the researcher can either proceed with a more fully specified OLS model or a maximum likelihood spatial error specification.

Anselin and Rey's (1991) Monte Carlos indicate that the Lagrange Multiplier diagnostic for

[15] A Lagrange Multiplier diagnostic for the joint presence of spatial lag dependence and spatial moving average error dependence has also been developed. However, in practice, the LM SARMA test is quite limited in usefulnesss as it has power against each of the two alternative forms of spatial dependence and thus will reject the null even in the absence of the joint SARMA dependence (Anselin 2005a, 197).

spatial lag dependence is robust both to heteroskedasticity and to deviations from normality. The basic Lagrange Multiplier diagnostics, however, are not without their limitations. The Lagrange Multiplier diagnostic for spatial error dependence underrejects the null in the presence of non-normally distributed errors and over-rejects the null in the presence of heteroskedasticity. Another limitation of these (and other) diagnostics for spatial lag and error dependence is their non-nested nature. Faced with diagnostics that point toward both spatial lag and spatial error dependence, the researcher typically chooses the larger test statistic, and estimates the corresponding spatial model (see, e.g., Gimpel and Cho 2004, 1002). However, if both spatial lag and spatial error dependence are indicated, the researcher may also wish to estimate a mixed lattice/geostatistical model incorporating both forms of spatial dependence.

Perhaps most problematic, however, the standard Lagrange Multiplier diagnostics are not valid in the presence of local misspecification (i.e., in the presence of the alternative form of spatial dependence). The Lagrange Multiplier diagnostic for spatial lag dependence will reject the null even if $\rho = 0$, if spatial error dependence is present. Similarly, the Lagrange Multiplier diagnostic for spatial error dependence will reject the null even if $\lambda = 0$, if spatial lag dependence is present (Anselin and Bera 1998, 274). As a result, researchers may wish to apply Lagrange Multiplier diagnostics that are robust to the alternative form of dependence.

8.3 Robust Lagrange Multiplier Diagnostics for Spatial Lag and Spatial Error Dependence in an OLS Model

The robust Lagrange Multiplier diagnostics for OLS models apply Bera and Yoon's modified Lagrange Multiplier tests to the diagnosis of spatial lag and spatial error dependence in OLS specifications. As Anselin, Bera, Florax, and Yoon (1996) show, the robust LM diagnostics adjust the unidirectional LM diagnostics in Section 8.2 by accounting for the non-centrality parameter. In essence, the spatial lag dependence is estimated in the diagnostic for lag dependence by accounting for any spatial error dependence that may also exist. Likewise, spatial error dependence is estimated in the diagnostic for error dependence by accounting for any spatial lag dependence that may exist. The robust Lagrange Multiplier diagnostic for spatial lag dependence developed by Anselin, Bera, Florax, and Yoon (1996, 83) as an extension of the Bera and Yoon modified LM test takes the form:

$$LM^*_{Lag} = \frac{(\mathbf{e}'\mathbf{W}_1 y/s^2 - \mathbf{e}'\mathbf{W}_1\mathbf{e}/s^2)^2}{(N\tilde{J}_{1\rho\cdot\beta})^{-1} - t_1}, \tag{39}$$

where $s^2 = \frac{\mathbf{e}'\mathbf{e}}{N}$, and $(N\tilde{J}_{1\rho\cdot\beta})^{-1} = [t_1 + (\mathbf{W}_1\mathbf{X}\beta)'\mathbf{M}(\mathbf{W}_1\mathbf{X}\beta)/s^2]^{-1}$, with $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $t_1 = \text{tr}(\mathbf{W}_1'\mathbf{W}_1 + \mathbf{W}_1^2)$. The robust Lagrange Multiplier diagnostic for spatial error dependence in an OLS model takes the form:

$$LM^*_{Error} = [(\mathbf{e}'\mathbf{W}_2\mathbf{e}/s^2 - t_2) - (N\tilde{J}_{2\rho\cdot\beta})^{-1}(\mathbf{e}'\mathbf{W}_2 y/s^2)]^2/[t_2 - t_2^2(N\tilde{J}_{2\rho\cdot\beta})^{-1}], \tag{40}$$

where $(N\tilde{J}_{2\rho\cdot\beta})^{-1} = [t_2 + (\mathbf{W}_2\mathbf{X}\beta)'\mathbf{M}(\mathbf{W}_2\mathbf{X}\beta)/s^2]^{-1}$, with $t_2 = \text{tr}(\mathbf{W}_2'\mathbf{W}_2 + \mathbf{W}_2^2)$, and the remaining notation is as in (39) (Anselin, Bera, Florax, and Yoon 1996, 82). In a series of Monte Carlo experiments, Anselin, Bera, Florax, and Yoon find that the robust Lagrange Multiplier diagnostics are less prone to Type I errors than their non-robust counterparts in the presence of the alternative form of spatial dependence. As Anselin and Bera (1998, 277) note, however, the robustification of the Lagrange Multiplier diagnostics does not come without a price. The robust Lagrange Multiplier diagnostic for spatial lag (error) dependence has reduced power against spatial lag (error) dependence than the unidirectional Lagrange Multiplier diagnostic for spatial lag (error) dependence in the absence of spatial error (lag) dependence.

The researcher is thus advised to estimate both the robust and non-robust Lagrange Multiplier diagnostics. As the non-robust LM diagnostic is prone to Type I errors in the presence of the alternative form of dependence and the robust LM diagnostic is prone to Type II errors in the absence of the alternative form of dependence, a rejection of the null on both spatial lag (error) diagnostics indicates that a spatial econometric, IV, or more fully specified OLS model is in order. Null findings on both diagnostics indicate that the current OLS specification is sufficient. If the null is rejected on the non-robust LM diagnostic for spatial lag (error) dependence due to the presence of spatial error (lag) dependence, the null will also likely be rejected on the non-robust LM diagnostic for spatial error (lag) dependence. The researcher, however, is likely to accept the null on the robust LM diagnostic for lag (error) dependence and, due to the reduction in power, may also accept the null on the robust LM diagnostic for spatial error (lag) dependence.

Faced with these conflicting diagnostics, the researcher is advised to estimate both spatial lag and spatial error models. There is a significant cost to be paid in potential bias, inconsistency, and inefficiency from estimating an OLS model in response to conflicting diagnostics. Given the increased availability of spatial econometric estimators in packages such as R, GeoDa, and Stata, there is little cost in estimating both spatial econometric alternatives (or an IV model if one chooses for spatial lag dependence). Information criteria may be employed after estimation to choose the proper spatial econometric specification.

8.4 The Kelejian-Robinson Diagnostic for Spatial Error Dependence in an OLS Model

Thus far, each of the five diagnostics for spatial dependence in the presence of covariates in an OLS model has assumed normally distributed errors. If the population errors are believed to be non-normally distributed (for example, as indicated by a Kiefer-Salmon test on the regression residuals), the researcher has two options. The dependent variable can be transformed to induce normality in the errors. Alternatively, the researcher can employ a non-parametric diagnostic that does not require normally distributed errors. Here, the principal non-parametric diagnostic is the Kelejian-Robinson diagnostic for spatial error dependence.

The Kelejian-Robinson diagnostic takes the form:

$$KR = \frac{\hat{\gamma}'\mathbf{Z}'\mathbf{Z}\hat{\gamma}}{\hat{\sigma}^4} \sim \chi^2_K, \tag{41}$$

where $\hat{\gamma}$ is the parameter vector from an auxiliary regression of the cross-products of residuals and the cross-products of covariates in the matrix, $\mathbf{Z}$, in which the cross-products are all paired observations for which there is a non-zero correlation, and K is the number of regressors in the observation matrix, $\mathbf{Z}$ (Anselin 1992, 179). The term, $\hat{\sigma}^4$, is any consistent estimator of σ^4, such as $[\frac{\mathbf{e}'\mathbf{e}}{N}]^2$, where $\mathbf{e}$ are the residuals from the auxiliary regression and N is the number of observations (Anselin and Bera 1998, 268). Although the diagnostic has the advantage of not assuming normality in the errors, it has the disadvantage of being limited to the case of spatial error dependence.

9 Modeling Spatial Dependence

If the OLS diagnostics indicate the existence of spatial lag dependence, the researcher has two options. She can estimate a mixed regressive, spatial autoregressive (spatial lag) model via maximum likelihood estimation. Alternatively, she can estimate an instrumental variables specification incorporating an instrument for the spatially lagged dependent variable. I consider these two options in turn next.

9.1 Maximum Likelihood Spatial Lag Estimation

The mixed regressive, spatial autoregressive model extends the pure spatial autoregressive model in (2) to include a set of covariates and associated parameters:

$$y = \rho \mathbf{W}_1 y + \mathbf{X}\beta + \varepsilon, \tag{42}$$

where $\mathbf{X}$ is an N by K matrix of observations on the covariates, β is a K by 1 vector of parameters, and the remaining notation is as in (2). As discussed in Section 3, OLS estimates of the spatial autoregressive parameter, ρ, will be biased and inconsistent, regardless of whether spatial dependence exists in the error term or not. Instead, the researcher may wish to estimate ρ via maximum likelihood estimation. Here, the log-likelihood function takes the form:

$$L_{Lag} = \sum_i \ln(1 - \rho\omega_i) - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{(y - \rho\mathbf{W}_1 y - \mathbf{X}\beta)'(y - \rho\mathbf{W}_1 y - \mathbf{X}\beta)}{2\sigma^2}, \tag{43}$$

where ω_i are the eigenvalues of the spatial weights matrix. As Anselin (1988, 181-182) shows, the likelihood function in (43) can be expressed as a concentrated log-likelihood function of the spatial autoregressive parameter, ρ:

$$L^c_{Lag} = -\frac{N}{2}\ln\left[\frac{(\mathbf{e}_O - \rho\mathbf{e}_L)'(\mathbf{e}_O - \rho\mathbf{e}_L)}{N}\right] + \sum_i \ln(1 - \rho\omega_i), \tag{44}$$

where $\mathbf{e}_O$ and $\mathbf{e}_L$ are, respectively, the residuals from OLS regressions of y on $\mathbf{X}$ and from $\mathbf{W}_y$ on $\mathbf{X}$ (Anselin 1988, 181). The maximum likelihood estimate of ρ is then found from the optimization of this concentrated log-likelihood function (Anselin and Bera 1998, 256). Given the maximum likelihood estimate of ρ, the parameters, β, and the error variance, σ^2, are then easily computed.

9.2 Instrumental Variables Spatial Lag Estimation

The maximum likelihood estimation of the mixed regressive, spatial autoregressive model requires the assumption of normality in the errors. Alternatively, one may relax this assumption and employ an instrumental variables approach. Here, the researcher will employ instruments for the spatially lagged dependent variable that are asymptotically uncorrelated with the error term. Assuming that the proper instruments can be found, the instrumental variables estimator will be consistent (as the instruments will be uncorrelated with the error), but will not be the most efficient estimator. The relative efficiency will depend upon the choice of the instruments (Anselin 1988, 84; Anselin and Bera 1998, 259). Anselin (1988, 85) suggests two potential instruments for the spatially lagged dependent variable. The researcher may use spatially lagged predicted values of the dependent variable from a regression of this variable on non-spatial covariates. Alternatively, these covariates themselves may be spatially lagged, with the lagged versions employed as instruments for the spatially lagged dependent variable. This latter approach, however, is likely to suffer from multicollinearity. In practice, given the computational ease of maximum likelihood estimation of spatial lag models and the potential difficulty in finding proper instruments, researchers will often prefer the maximum likelihood approach for modeling spatially lagged dependent variables.

9.3 Maximum Likelihood Spatial Error Estimation

If the OLS diagnostics indicate the existence of spatial error dependence in the presence of covariates, the researcher has two options. She can estimate a more fully specified OLS model to model the spatial dependence. Or she can estimate a maximum likelihood model incorporating the spatial dependence in the errors. The model with spatial autoregressive error dependence takes the form:

$$y = \mathbf{X}\beta + \varepsilon$$

$$\varepsilon = \lambda \mathbf{W}_2 \varepsilon + \xi, \tag{45}$$

where the notation is as in (2) and (42). As stated in Section 3, OLS estimates of the autoregressive parameter, λ, will not be consistent. Instead, if the spatial dependence in the errors cannot be modeled via covariates, the autoregressive error dependence must be modeled via maximum likelihood estimation. Here, the log-likelihood takes the form:

$$L_{Error} = \sum_i \ln(1 - \lambda\omega_i) - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{(y - \mathbf{X}\beta)'(\mathbf{I} - \lambda\mathbf{W}_2)'(y - \mathbf{X}\beta)}{2\sigma^2}. \tag{46}$$

Similar to the spatial lag case, this log-likelihood can be expressed as a concentrated log-likelihood in the parameter, λ:

$$L^c_{Error} = -\frac{N}{2}\ln\left[\frac{\mathbf{e}'\mathbf{e}}{N}\right] + \sum_i \ln(1 - \lambda\omega_i), \tag{47}$$

where $\mathbf{e}'\mathbf{e}$ is the residual sum of squares from the regression of the spatially filtered variables, $y - \lambda\mathbf{W}_2 y$ and $\mathbf{X} - \lambda\mathbf{W}_2\mathbf{X}$ (Anselin 1992, 210). As in the spatial lag case, the parameters, β, and error variance, σ^2, are then computed, given the maximum likelihood estimate of λ.

9.4 Models for Joint Spatial Lag and Spatial Error Dependence

Thus far, I have considered only models for either spatial lag or spatial error dependence. As indicated in Section 8.2, in practice, rarely are specifications with both a spatially lagged dependent variable and spatial dependence in the error term estimated. In part, this reflects the identification problem inherent in models with identical weights matrices for both the spatially lagged dependent variable and the spatial autoregressive error dependence. However, there may be times when the researcher will wish to estimate a model with both lag and error dependence. I consider the complications posed by joint spatial lag and error dependence as well as an approach to such models employed by Dubin (2003).

Throughout this paper, I have employed differing notation for the weights matrix for the spatially lagged dependent variable ($\mathbf{W}_1$) and for the spatially lagged error term ($\mathbf{W}_2$). If the two weights matrices are distinct, estimation of the joint spatial lag, spatial error model is straightforward. Often, the researcher will wish to employ differing weights matrices for the lagged dependent variable and the error term on substantive grounds. For example, the researcher may posit that a diffusion process exists between contiguous neighbors, with additional attributional dependence for non-contiguous neighbors within a particular distance of each other. If so, the researcher will wish to include a weights matrix with a queen contiguity definition for the spatially lagged dependent variable, and a spatial weights matrix with a distance band conception of dependence (excluding contiguous observations as neighbors) for the spatial error dependence.

Estimation, however, becomes significantly more complex when the identical weights matrix is employed for both the spatially lagged dependent variable and the spatial error dependence. As

Anselin and Bera (1998, 252) show, here the model takes the form:

$$y = \rho \mathbf{W}_1 y + \mathbf{X}\beta + \varepsilon$$

$$\varepsilon = \lambda \mathbf{W}_1 \varepsilon + \xi, \tag{48}$$

or

$$y = (\rho + \lambda)\mathbf{W}_1 y - \lambda\rho \mathbf{W}_1^2 y + \mathbf{X}\beta - \lambda \mathbf{W}_1 \mathbf{X}\beta + \xi. \tag{49}$$

As can be seen from (49), λ and ρ are not identified in the absence of nonlinear constraints (Anselin and Bera 1998, 252). Enforcing the nonlinear constraints between β and $-\lambda\beta$ will produce an estimate of λ. However, this estimate of λ will also produce two estimates of ρ unless nonlinear constraints are strictly enforced (Anselin and Bera 1998, 252).

Dubin (2003) offers an alternative approach for jointly modeling spatial lag and spatial error dependence. Dubin's approach is to model the spatial lag dependence via a lattice data approach, while employing a geostatistical approach to model the spatial error dependence. As shown in Section 4.2, the geostatistical approach does not employ a spatial weights matrix; instead it models the spatial dependence as a function of the distance between paired observations. As a result, one can employ similar distance-based neighbor definitions for both the spatially lagged dependent variable and the spatial error dependence without encountering the identification problem inherent in employing an identical spatial weights matrix for both forms of dependence.

Dubin's (2003, 6) specification takes the following form:

$$y = \rho \mathbf{W}_1 y + \mathbf{X}\beta + \varepsilon, \tag{50}$$

with

$$\varepsilon \sim N(0, \sigma^2 K), \tag{51}$$

where K is the particular functional form specified for the error dependence based on separation distances. A commonly applied functional form in geostatistics is the Gaussian, where the spatial dependence is modeled as:

$$K(d) = \begin{cases} \alpha_1 \exp - \left(\frac{d}{\alpha_2}\right)^2 & \text{for } d > 0 \\ 1 & \text{for } d = 0 \end{cases} \tag{52}$$

where *K(d)* is the spatial autocorrelation between paired observations separated by distance *d*, and α_1 and α_2 are parameters to be estimated (Dubin 2003, 5). Although the joint lattice data/geostatistical approach offers an alternative approach for estimating models with both spatial lag and spatial error dependence, the treatment of political science data as geostatistical data generally reflects an artificial modeling choice. Political science data generally are not sample observations from a continuous underlying surface. Thus, for example, the concept of kriging will generally have little applicability to political science data. Political scientists, therefore, will generally not wish to follow Dubin's suggestion of extending his general spatial model by employing kriging to improve the predictive accuracy of the model.

10 Issues in Diagnosing and Modeling Spatial Dependence

The diagnosis and modeling of spatial dependence is thus a straightforward process entailing three simple sequential steps. First, estimate global and local measures of spatial autocorrelation to diagnose univariate spatial autocorrelation in the absence of covariates. Next, estimate a standard econometric model and conduct diagnostics on it to determine whether the covariates sufficiently model the spatial dependence in the DGP. If they do not, estimate the spatial econometric or instrumental variables specification indicated by the diagnostic. While diagnosing and modeling spatial autocorrelation is a simple process, researchers should be attentive to two particular issues – the Modifiable Areal Unit Problem and the Boundary Value Problem – when diagnosing and modeling the spatial dependence in their DGPs.

10.1 Modifiable Areal Unit Problem

The choice of the proper areal unit of analysis is a critical concern in modeling spatial dependence. As Openshaw and Taylor (1979) note, different levels of areal units of analysis can produce fundamentally different measures of spatial autocorrelation, or, as they colorfully note, "a million or so correlation coefficients." This dependence of spatial autocorrelation estimates on the level of analysis is known as the modifiable areal unit problem and is analogous to the issue of temporal aggregation in time series analysis (see, e.g., Freeman 1989). The modifiable areal unit problem is particularly problematic because it implies that an improper choice of geographic scale can produce artificial spatial dependence. Consider, for example, an analysis of municipal-level tax policies. We might expect negative spatial autocorrelation in such policies as suburban municipalities pursue low tax strategies to attract economic development away from center cities with higher tax rates. To examine this possible negative spatial dependence, we need data collected at the municipal level. However, assume we can only collect data at the neighborhood level. An analysis proceeding at the (improper) neighborhood level of analysis may identify positive spatial autocorrelation in tax policies as neighborhoods nested within the same municipality share these policies decided at the municipal level of government. Clearly, in order to validly estimate spatial dependence in political phenomena, the choice of areal level of analysis must be guided by theoretical considerations and should approximate the areal level of theoretical concern as closely as possible.

10.2 Boundary Value Problem

A second problem inherent in many spatial econometric applications is the transcending of spatial dependence beyond the observed data. Assuming that observed units do not cover the full plane of possible spatial dependence, units at the boundaries of the observed data may be spatially autocorrelated with units outside of the observed data.[16] This boundary value problem, or edge effect, can produce biased estimates of spatial dependence, as unobserved units are influencing the spatial autocorrelation in the data. As Anselin (1988, 173) notes, this problem is not as severe as would appear at first glance, however. Maximum likelihood estimators of spatial dependence remain consistent even in the presence of edge effects.

The bias induced in finite samples, however, may remain a concern for applied researchers. To date, solutions for the boundary value problem are not completely satisfying. One solution proposed early in the literature is to wrap the observed data on a torus, so that units on opposite edges of the observed data are treated as neighbors (Griffith 1983). This "cure" of creating artificial neighbors, however, is often more problematic than the disease, as there is generally little substantive reason for assuming that observations on opposite edges of the spatial plane can be treated as first-order neighbors. More commonly, researchers specify a buffer zone, in which boundary observations are dropped from spatial econometric models, but are allowed to exert spatial influence on neighbors retained in the analysis. This approach, of course, is problematic if one has a particular substantive

[16]Note, if one has missing data on the interior of the plane, this same problem can arise away from boundary locations.

interest in the observations on the boundary. As Cressie (1993, 439) notes, additional research into solutions for edge effects remains warranted.

11 Monte Carlo Analysis

In this section, I assess the performance of the OLS estimator when spatial lag dependence or spatial error dependence is ignored through a series of Monte Carlo experiments.[17] Recall from Section 3.1 the expected performance of the OLS estimator if spatial dependence in the DGP is ignored. If spatial autocorrelation consistent with a diffusion process exists in the data generating process and a spatially lagged dependent variable is omitted from the model, OLS parameter estimates for the remaining covariates will be biased and inconsistent.[18] If, alternatively, the spatial autocorrelation pertains only to the errors, OLS will remain an unbiased estimator, but it will no longer be efficient. Standard error estimates will be biased downward, producing Type I errors.

In the Monte Carlos, the DGP for the case of spatial lag dependence takes the form:

$$y = \rho \mathbf{W}_1 y + \beta_0 + \beta_1 \mathbf{x} + \varepsilon, \tag{53}$$

where $\varepsilon \sim N(0, \sigma^2 \mathbf{I})$. The DGP for the case of spatial error dependence takes the form:

$$y = \beta_0 + \beta_1 \mathbf{x} + \varepsilon$$

$$\varepsilon = \lambda \mathbf{W}_2 y \varepsilon + \xi, \tag{54}$$

where $\xi \sim N(0, \sigma^2 \mathbf{I})$. In both cases, the independent variable, $\mathbf{x}$, is normally distributed with a mean of 0 and a standard deviation of 3. I set $\beta_0 = \beta_1 = 1$. The OLS estimates for each experiment reflect a standard OLS specification, ignoring the spatial lag or spatial error dependence.

I examine the bias of the OLS estimates of β_1 when spatial lag and error dependence in the DGP are omitted from the OLS specification. I also examine the OLS estimates of the standard error of β_1 when spatial error dependence is omitted from the OLS specification. I examine the performance of OLS varying both the number of observations (and the corresponding spatial weights matrices) and the degree of spatial autocorrelation, as reflected in the autoregressive parameters ρ and λ. For each set of experiments, the observations are arrayed in regular square lattices. Monte Carlos are performed for four different sizes of square lattice structures: a 5 by 5 lattice ($n = 25$), a 10 by 10 lattice ($n = 100$), a 20 by 20 lattice ($n = 400$), and a 30 by 30 lattice ($n = 900$). In each case, a queen contiguity definition of neighbors is employed. The performance of OLS is examined for ten values of both ρ and λ: -0.9, -0.7, -0.5, -0.3, -0.1, 0.1, 0.3, 0.5, 0.7, 0.9. For each combination of lattice size and ρ or λ value, 1000 replications were performed.

11.1 Monte Carlo Results for Omitted Spatial Lag Dependence

Table 1 reports the bias of the OLS estimator of β_1 when spatial dependence consistent with a diffusion process exists in the DGP but is omitted from the OLS specification. As can be seen from

[17] The Monte Carlos are based on modified versions of the **R** code for omitted spatial error dependence in Anselin (2005b).

[18] Of course, the OLS estimate of the spatial autoregressive parameter for the lagged dependent variable will be biased and inconsistent if this term is included in the OLS model; estimation should instead proceed via maximum likelihood estimation or an instrumental variables specification. However, given that a diffusion process is often implied by our theories and yet is often ignored in practice, it is critical to examine the performance of the OLS estimator for non-spatial covariates when this diffusion process is ignored.

the table, OLS performs well at low levels of both positive and negative spatial autocorrelation. The bias of the OLS estimator, however, increases appreciably as $|\rho|$ increases to .5 and beyond. Moreover, there is an asymmetric effect, as bias is markedly more problematic at high levels of positive spatial autocorrelation. With an n of 25, the OLS estimator overstates the true value of β_1 by 15% when $\rho = .7$ and by 35% when $\rho = .9$.

Table 1: Bias of the OLS Estimator with Omitted Spatially Lagged Dependent Variable

	ρ									
N	-.9	-.7	-.5	-.3	-.1	.1	.3	.5	.7	.9
25	.09	.05	.02	.01	.00	.01	.03	.07	.15	.35
100	.12	.07	.04	.01	.00	.00	.01	.04	.11	.29
400	.11	.06	.04	.02	.00	.00	.01	.04	.09	.25
900	.08	.04	.02	.00	.00	.00	.02	.06	.14	.36

This bias of the OLS estimator when there is an omitted spatial autoregressive process pertaining to the dependent variable can be seen graphically in Figure 1. This figure plots the bias in the OLS estimator for an n of 900 for each of the ten values of ρ. Here again is the pattern of negligible bias at low levels of ρ, increasing as both negative and positive spatial autocorrelation become more acute. Again, the bias in the OLS estimator is largest at high levels of positive spatial autocorrelation. These results are consistent with the expectations from Section 3.1.

11.2 Monte Carlo Results for Omitted Spatial Error Dependence

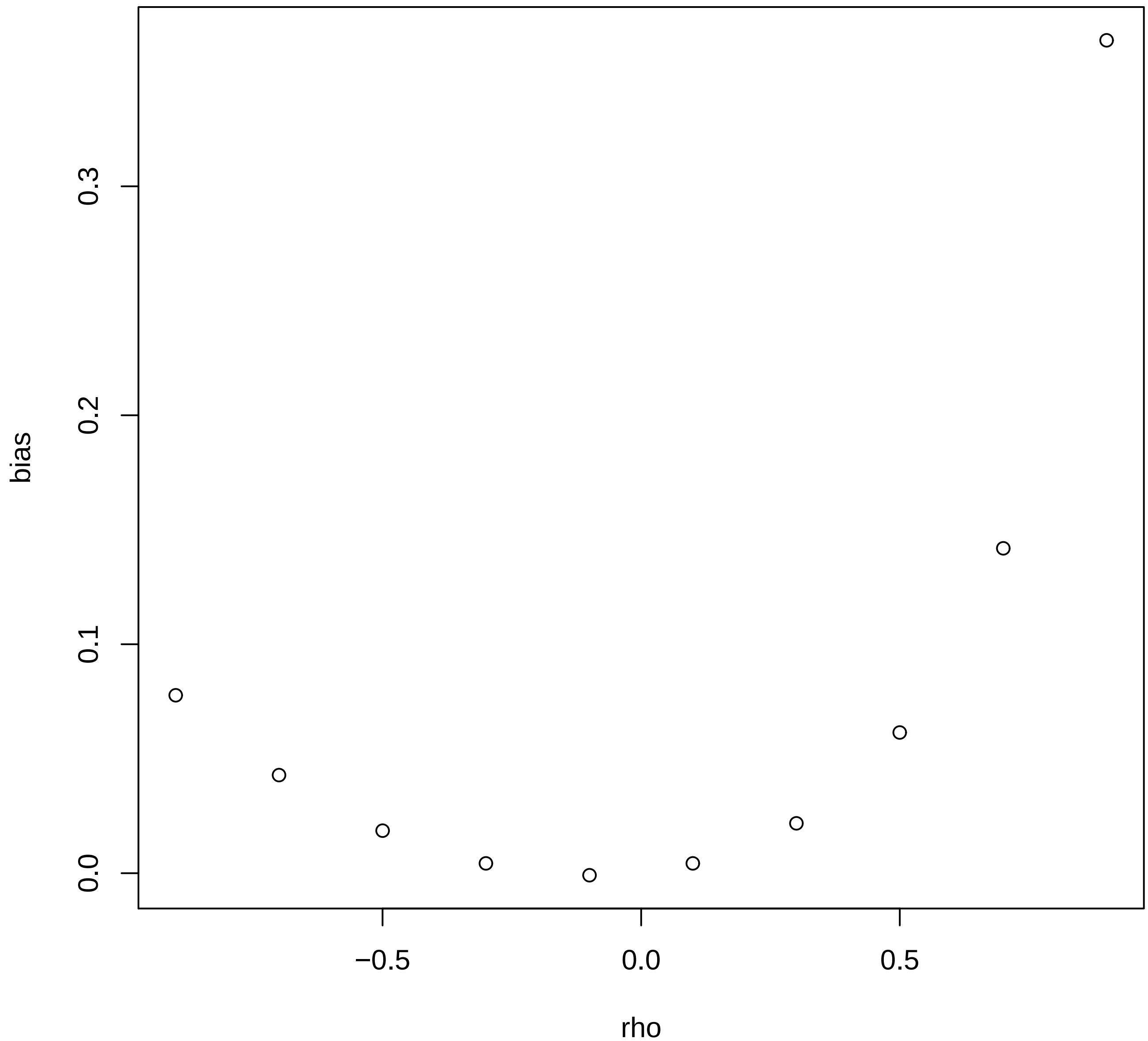
In contrast to the Monte Carlo results with omitted lag dependence, the Monte Carlo experiments with omitted spatial error dependence show no appreciable bias in the OLS estimator. Employing proportional measures of bias, as in Table 1, there is no bias to two decimal places across the four different lattice sizes and the ten values of λ. Consistent with expectations from Section 3.1, OLS remains an unbiased estimator of slope parameters even at high levels of positive and negative spatial error dependence.

As expected, however, OLS standard errors are unduly optimistic in the presence of spatial error dependence. This can be seen in Table 2, which reports the ratio of the OLS standard errors to the true standard errors. At low levels of spatial error dependence, OLS standard errors remain unbiased. However, as both positive and negative error dependence increase in size, the standard errors reported by OLS increasingly understate the true standard errors. Similar to the case of omitted spatial lag dependence, the problem is particularly acute for high levels of positive spatial error autocorrelation. For example, with an n of 900 and a λ of .9, the OLS standard error is only .57 the size of the true standard error. In the presence of both positive and negative spatial error dependence, inference based on OLS is likely to lead to Type I errors.

Table 2: Ratio of OLS Standard Error to True Standard Error with Spatial Error Dependence

	λ									
N	-.9	-.7	-.5	-.3	-.1	.1	.3	.5	.7	.9
25	.87	.89	.95	.97	1.03	1.01	.98	.96	.85	.71
100	.85	.93	.98	.96	1.00	.98	.96	.97	.86	.64
400	.86	.93	.95	1.02	1.00	.97	.98	.92	.85	.59
900	.87	.92	.93	1.02	.98	1.00	.96	.92	.83	.57

Figure 1: Bias of the OLS Estimator with Omitted Spatially Lagged Dependent Variable



12 Conclusion

Since the time of Galton, political scientists have been keenly aware that the spatial clustering in the phenomena of interest to them may be produced by two quite distinct processes: diffusion processes, or, alternatively, the independent occurrences of the phenomena among neighboring units. In more recent decades, scholars have been highly attuned to the methodological problems posed by this spatial dependence. Until recently, however, scholars lacked a methodological approach that could simultaneously speak to Galton's problem and address the hurdles to inference posed by spatially dependent data. Spatial econometrics now provides scholars a rigorous method for addressing the spatial dimension that is central to many of our theories of political behaviors, processes, and events.

This paper has sought to highlight both the importance and the ease of modeling the spatial dependence in our data. As both the analytical results and the Monte Carlo results demonstrate, spatially dependent DGPs pose significant challenges to inference that are not easily addressed by techniques developed to handle the more familiar serial dependence in time series analysis. Instead, the researcher working with spatially dependent data will often wish to employ spatial econometrics to model this spatial autocorrelation. The diagnosis and modeling of this spatial autocorrelation, moreover, is a straightforward and simple three-step process. The researcher first diagnoses univariate spatial dependence in the phenomenon of interest by estimating global and local measures of spatial autocorrelation. Next, the researcher attempts to model this spatial dependence with covariates, and applies simple diagnostics to determine whether a standard econometric specification sufficiently captures the spatial dependence. If the standard econometric specification does not capture the spatial dependence, the researcher then estimates the spatial econometric or instrumental variables specification indicated by the diagnostic.

The application of spatial econometrics is also aided by two parallel developments that are occurring in the social sciences. First, the past decade has witnessed an explosion in the availability of geo-coded data in the social sciences. Today, institutions such as the Census Bureau and state governments provide free downloads of geo-coded data and corresponding ESRI shape files (which significantly ease the creation of the spatial weights matrices required in the lattice data approach to spatial econometric modeling). Additionally, spatial econometric estimators are increasingly being incorporated in statistical software. This includes packages dedicated to spatial econometrics, such as GeoDa. It also includes general purpose packages with which political scientists are already familiar, such as R, Stata, and WinBUGS.

As researchers, we routinely test for violations of modeling assumptions such as non-normality or heteroskedasticity. But while time series analysts routinely test for dependence in their data, testing for spatial dependence in cross-sectional political science research is rare despite the spatial autocorrelation predicted by many of our theories. Happily, researchers can now employ easily applied methods and tools to diagnose and model the spatial dependence that is implied by our theories. In doing so, researchers are likely to draw renewed attention to the interactions and interdependence that are at the core of many of the political phenomena we examine.

References

[1] Agresti, Alan. 2002. *Categorical Data Analysis*, 2nd Ed. Hoboken, NJ: John Wiley & Sons.

[2] Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

[3] Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Kluwer: Dordrecht.

[4] Anselin, Luc. 1990. "Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis." *Journal of Regional Science* 30(2): 185-207.

[5] Anselin, Luc. 1992. "SpaceStat Tutorial: A Workbook for Using SpaceStat in the Analysis of Spatial Data." Typescript. University of Illinois at Urbana-Champaign.

[6] Anselin, Luc. 1995. "Local Indicators of Spatial Association – LISA." *Geographical Analysis* 27(2): 93-115.

[7] Anselin, Luc. 2002. "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models." *Agricultural Economics* 27: 247-267.

[8] Anselin, Luc. 2005a. "Exploring Spatial Data with GeoDa™: A Workbook." Typescript. University of Illinois at Urbana-Champaign and Center for Spatially Integrated Social Science.

[9] Anselin, Luc. 2005b. "Spatial Regression Analysis in R: A Workbook." Typescript. University of Illinois at Urbana-Champaign and Center for Spatially Integrated Social Science.

[10] Anselin, Luc, and Anil K. Bera. 1998. "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." In *Handbook of Applied Economic Statistics*, eds. Aman Ullah and David E.A. Giles. New York: Marcel Dekker.

[11] Anselin, Luc, Anil K. Bera, Raymond Florax, and Mann J. Yoon. 1996. "Simple Diagnostic Tests for Spatial Dependence." *Regional Science and Urban Economics* 26: 77-104.

[12] Anselin, Luc, and Wendy K. Tam Cho. 2002. "Spatial Effects and Ecological Inference." *Political Analysis* 10(3): 276-297.

[13] Anselin, Luc, and Daniel Griffith. 1988. "Do Spatial Effects Really Matter in Regression Analysis?" *Papers of the Regional Science Association* 65: 11-34.

[14] Anselin, Luc, and Serge Rey. 1991. "Properties of Tests for Spatial Dependence in Linear Regression Models." *Geographical Analysis* 23(2): 112-131.

[15] Baltagi, Badi H., and Dong Li. 2001. "LM Tests for Functional Form and Spatial Error Correlation." *International Regional Science Review* 24: 194-225.

[16] Baltagi, Badi H., and Dong Li. 2005. "Testing for Linear and Log-Linear Models Against Box-Cox Alternatives with Spatial Lag Dependence." In *Advances in Econometrics, Volume 18: Spatial and Spatiotemporal Econometrics*, eds. James P. LeSage and R. Kelley Pace. Oxford: Elsevier.

[17] Baltagi, Badi H., Seuck Heun Song, and Won Koh. 2003. "Testing Panel Data Regression Models with Spatial Error Correlation." *Journal of Econometrics* 117: 123-150.

[18] Banerjee, Sudipto, and Bradley P. Carlin. 2004. "Parametric Spatial Cure Rate Models for Interval-Censored Time-to-Relapse Data." *Biometrics* 60: 268-275.

[19] Beck, Nathaniel, Kristian Gleditsch, and Kyle Beardsley. 2006. "Space is More than Geography: Using Spatial Econometrics in the Study of Political Economy." Forthcoming. *International Studies Quarterly*

[20] Beck, Nathaniel, and Simon Jackman. 1998. "Beyond Linearity by Default: Generalized Additive Models." *American Journal of Political Science* 42(2): 596-627.

[21] Bera, Anil K., and Mann J. Yoon. 1993. "Specification testing with Locally Misspecified Alternatives." *Econometric Theory* 9: 649-658.

[22] Berry, Frances Stokes, and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *American Political Science Review* 84(2): 395-415.

[23] Bolduc, Denis, Bernard Fortin, and Stephen Gordon. 1997. "Multinomial Probit Estimation of Spatially Interdependent Choices: An Empirical Comparison of Two New Techniques." *International Regional Science Review* 20(1 & 2): 77-101.

[24] Brooks, Sarah M. 2004. "Explaining Capital Account Liberalization in Latin America: A Transitional Cost Approach." *World Politics* 56(3): 389-430.

[25] Busch, Marc L., and Eric Reinhardt. 2000. "Geography, International Trade, and Political Mobilization in U.S. Industries." *American Journal of Political Science* 44(4): 703-719.

[26] Cardoso, Fernando Henrique, and Enzo Faletto. 1979. *Dependency and Development in Latin America*. Berkeley, CA: University of California Press.

[27] Centers for Disease Control and Prevention. 2004. "150th Anniversary of John Snow and the Pump Handle." *Morbidity and Mortality Weekly Report* 53(34): 783-806.

[28] Cho, Wendy K. Tam. 2003. "Contagion Effects and Ethnic Contribution Networks." *American Journal of Political Science* 47(2): 368-387.

[29] Cliff, A.D., and J.K. Ord. 1973. *Spatial Autocorrelation*. London: Pion.

[30] Cliff, A.D., and J.K. Ord. 1981. *Spatial Processes: Models and Applications*. London: Pion.

[31] Cressie, Noel A.C. 1993. *Statistics for Spatial Data*. New York: John Wiley & Sons.

[32] Darmofal, David. 2006. "The Political Geography of Macro-Level Turnout in American Political Development." Forthcoming. *Political Geography*.

[33] Dubin, Robin. 2003. "Spatial Lags and Spatial Errors Revisited: Some Monte Carlo Evidence." Typescript. Case Western Reserve University.

[34] Ertur, Cem, and Wilfried Koch. 2005. "Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence." Manuscript. Available at: http://eswc2005.econ.ucl.ac.uk/papers/ESWC/2005/2209/ErturKoch.pdf. Accessed 11-9-05.

[35] Fleming, Mark M. 2004. "Techniques for Estimating Spatially Dependent Discrete Choice Models." In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, eds. Luc Anselin, R.J.G.M. Florax, and S.J. Rey. Berlin: Springer.

[36] Fomby, Thomas B., R. Carter Hill, and Stanley R. Johnson. 1984. *Advanced Econometric Methods*. New York: Springer-Verlag.

[37] Fotheringham, A.S., M.E. Charlton, and C. Brunsdon. 1998. "Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis." *Environment and Planning A* 30: 1905-1927.

[38] Freeman, John R. 1990. "Systematic Sampling, Temporal Aggregation and the Study of Political Relationships." *Political Analysis* 1: 61-98.

[39] Gatrell, Anthony C., Trevor C. Bailey, Peter J. Diggle, and Barry S. Rowlingson. 1996. "Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology." *Transactions of the Institute of British Geographers* 21: 256-274.

[40] Getis, Arthur, and J.K. Ord. 1992. "The Analysis of Spatial Association by Use of Distance Statistics." *Geographical Analysis* 24(3): 189-206.

[41] Gimpel, James G., and Wendy K. Tam Cho. 2004. "The Persistence of White Ethnicity in New England Politics." *Political Geography* 23: 987-1008.

[42] Gleditsch, Kristian Skrede. *All International Politics is Local: The Diffusion of Conflict, Integration, and Democratization*. Ann Arbor, MI: University of Michigan Press.

[43] Gleditsch, Kristian Skrede, and Michael D. Ward. 2000. "Peace and War in Time and Space: The Role of Democratization." *International Studies Quarterly* 43: 1-29.

[44] Greene, William H. 1993. *Econometric Analysis*, 2nd Ed. Upper Saddle River, NJ: Prentice Hall.

[45] Greene, William H. 2003. *Econometric Analysis*, 5th Ed. Upper Saddle River, NJ: Prentice Hall.

[46] Griffith, Daniel A. 1983. "The Boundary Value Problem in Spatial Statistical Analysis." *Journal of Regional Science* 23(3): 377-387.

[47] Holloway, Garth, Bhavani Shankar, and Sanzidur Rahman. 2002. "Bayesian Spatial Probit Estimation: A Primer and an Application to HYV Rice Adoption." *Agricultural Economics* 27: 383-402.

[48] Hsiao, Cheng. 1975. "Some Estimation Methods for a Random Coefficient Model." *Econometrica* 43(2): 305-326.

[49] Huckfeldt, Robert. 1986. *Politics in Context: Assimilation and Conflict in Urban Neighborhoods*. New York: Agathon Press:

[50] Isard, Walter. 1956. *Location and Space-Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure*. New York: Wiley.

[51] Judge, George G., W.E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee. 1985. *The Theory and Practice of Econometrics*, 2nd Ed. New York: John Wiley & Sons.

[52] Kelejian, Harry H., and Ingmar R. Prucha. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review* 40(2): 509-533.

[53] Kmenta, Jan. 1997. *Elements of Econometrics*, 2nd Ed. Ann Arbor, MI: University of Michigan Press.

[54] LeSage, James P. 1997. "Bayesian Estimation of Spatial Autoregressive Models." *International Regional Science Review* 20: 113-129.

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

[56] Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.

[57] MacNab, Ying C. 2003. "Hierarchical Bayesian Modeling of Spatially Correlated Health Service Outcome and Utilization Rates." *Biometrics* 59: 305-316.

[58] Moran, P.A.P. 1948. "The Interpretation of Statistical Maps." *Journal of the Royal Statistical Society. Series B (Methodological)* 10(2): 243-251.

[59] Most, Benjamin A., and Harvey Starr. 1980. "Diffusion, Reinforcement, Geopolitics, and the Spread of War." *American Political Science Review* 74(4): 932-46.

[60] Most, Benjamin A., and Harvey Starr. 1982. "Case Selection, Conceptualizations and Basic Logic in the Study of War." *American Journal of Political Science* 26(4): 834-856.

[61] Most, Benjamin A., and Harvey Starr. 1983. "Conceptualizing 'War': Consequences for Theory and Research." *Journal of Conflict Resolution* 27(1): 137-159.

[62] O'Loughlin, John. 2002. "The Electoral Geography of Weimar Germany: Exploratory Spatial Data Analyses (ESDA) of Protestant Support for the Nazi Party." *Political Analysis* 10: 217-243.

[63] O'Loughlin, John, Colin Flint, and Luc Anselin. 1994. "The Geography of the Nazi Vote: Context, Confession and Class in the Reichstag Election of 1930." *Annals of the Association of American Geographers* 84: 351-380.

[64] Openshaw, Stan, and Peter Taylor. 1979. "A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem." In *Statistical Applications in the Spatial Sciences*, ed. N. Wrigley. London: Pion.

[65] Ord, J.K., and Arthur Getis. 1995. "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application." *Geographical Analysis* 27(4): 286-306.

[66] Patterson, Samuel C. 1972. "Party Opposition in the Legislature: The Ecology of Legislative Institutionalization." *Polity* 4(3): 344-366.

[67] Poirier, Dale J. 1995. *Intermediate Statistics and Econometrics: A Comparative Approach*. Cambridge, MA: MIT Press.

[68] Ramsay, Timothy, Richard Burnett, and Daniel Krewski. 2003. "Exploring Bias in a Generalized Additive Model for Spatial Air Pollution Data." *Environmental Health Perspectives* 111(1): 1283-1288.

[69] Rudra, Nita. 2002. "Globalization and the Decline of the Welfare State in Less-Developed Countries." *International Organization* 56(2): 411-445.

[70] Selb, Peter. 2004. "A Spatial Approach to the Dynamics of Electoral Success." Manuscript. Available at: http://www.uni-lueneburg.de/fb2/zdemo/dvpw/dateien/selb.pdf. Accessed 11-9-05.

[71] Shin, Michael, and John Agnew. 2002. "The Geography of Party Replacement in Italy, 1987-1996." *Political Geography* 21: 221-242.

[72] Starr, Harvey. 1991. "Democratic Dominoes: Diffusion Approaches to the Spread of Democracy in the International System." *Journal of Conflict Resolution* 35(2): 356-381.

[73] Starr, Harvey. 2001. "Using Geographic Information Systems to Revisit Enduring Rivalries: The Case of Israel." *Geopolitics* 5(Summer): 37-56.

[74] Starr, Harvey, and Benjamin A. Most. 1976. "The Substance and Study of Borders in International Relations Research." *International Studies Quarterly* 20(4): 581-620.

[75] Starr, Harvey, and Benjamin A. Most. 1978. "A Return Journey: Richardson, 'Frontiers' and Wars in the 1946-1965 Era." *Journal of Conflict Resolution* 22(3): 441-467.

[76] Starr, Harvey, and Benjamin A. Most. 1983. "Contagion and Border Effects on Contemporary African Conflict." *Comparative Political Studies* 16(February): 92-117.

[77] Swamy, P.A.V.B. 1970. "Efficient Inference in a Random Coefficient Regression Model." *Econometrica* 38(2): 311-323.

[78] Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46: 234-240.

[79] Tylor, Edward B. 1889. "On a Method of Investigating the Development of Institutions; Applied to Laws of Marriage and Descent." *The Journal of the Anthropological Institute of Great Britain and Ireland* 18: 245-272.

[80] Vasquez, John A. 1995. "Why Do Neighbors Fight? Proximity, Interaction, or Territoriality." *Journal of Peace Research* 32(3): 277-293.

[81] Ward, Michael D., and Kristian Skrede Gleditsch. 2002. "Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace." *Political Analysis* 10: 244-260.