



The Review of Regional Studies

The Official Journal of the Southern Regional Science Association



*2014 SRSA Fellows Address: San Antonio, Texas March 28, 2014
53rd Annual Meetings of the Southern Regional Science Association*

What Regional Scientists Need to Know about Spatial Econometrics*

James P. LeSage

Department of Finance and Economics, Texas State University-San Marcos, USA

Abstract: Regional scientists frequently work with regression relationships involving sample data that is spatial in nature. For example, hedonic house-price regressions relate selling prices of houses located at points in space to characteristics of the homes as well as neighborhood characteristics. Migration, commodity, and transportation flow models relate the size of flows between origin and destination regions to the distance between origin and destination as well as characteristics of both origin and destination regions. Regional growth regressions relate growth rates of a region to past period own- and nearby-region resource inputs used in production. Spatial data typically violates the assumption that each observation is independent of other observations made by ordinary regression methods. This has econometric implications for the quality of estimates and inferences drawn from nonspatial regression models. Alternative methods for producing point estimates and drawing inferences for relationships involving spatial data samples comprise the broad topic covered by spatial econometrics. Like any subdiscipline, spatial econometrics has its quirks, many of which reflect influential past literature that has gained attention in both theoretical and applied work. This article asks the question: “What should regional scientists who wish to use regression relationships involving spatial data in an effort to shed light on questions of interest in regional science know about spatial econometric methods?”

Keywords: spatial regression models, local versus global spatial spillovers, spatial Durbin and spatial Durbin error model specifications, spatial weight matrices, model comparison

JEL Codes: C11, C23, O47, O52

1. INTRODUCTION

There has been an increasing number of spatial regression estimation methodologies proposed in the spatial econometrics literature, maximum likelihood, quasi-maximum likelihood, GMM, IV, Bayesian, maximum entropy, robust GMM, robust Bayesian, semiparametric, along with static and dynamic panel extensions, and variants for dealing with count data, truncated and limited dependent variables to mention a few. In addition to different estimation methods, there are a host of newly proposed specifications (the matrix exponential, Shiller smooth priors, hierarchical linear models, smooth transition autoregressive), which in conjunction with the plethora of estimation procedures provides a great deal of flexibility for applied modeling

*The author thanks R. Kelley Pace and Manfred M. Fischer for insightful comments, and Mark L. Burkey for thoughtful editorial comments that improved the manuscript.

LeSage is Fields Endowed Chair in the Department of Finance and Economics at Texas State University-San Marcos, in San Marcos, TX. E-mail: jlesage@spatial-econometrics.com

situations. There is also a growing literature on statistical testing in spatial econometrics, with tests of alternative specifications, Monte Carlo studies that compare performance of alternative specifications and estimation procedures, panel data model tests, tests for alternative spatial weight matrices, and tests for spatial dependence in disturbances, dependent and explanatory variables.

For the typical regional science practitioner the task of sorting through this literature is a daunting prospect. The thesis of this presentation is that there are a few principles that can guide regional science practitioners through the haze of spatial econometrics, and shine light on estimation methods and specifications that should be most useful for regional science research. The remainder of the paper is organized into sections, each discussing one of the guiding principles.

2. PRINCIPLE 1: LOCAL VERSUS GLOBAL SPILLOVER SPECIFICATIONS

An oft-quoted statement is that “all politics is local,” which is meant to imply that national congressional representatives react most strongly to issues of concern to local constituents.¹ A similar statement could be applied to spatial regression models, taking the form: “most spatial spillovers are local.” A major focus of regional science is spatial spillovers. The good news is that spatial regression models can be used to: 1) formally define the concept of a spatial spillover, 2) provide estimates of the quantitative magnitude of spillovers, and 3) test for the statistical significance of these (e.g., are the spillovers statistically significantly different from zero). The bad news is that for a host of historical reasons, most regional science applications of spatial regression models have not: 1) used an appropriate spatial regression specification to produce valid estimates of spatial spillovers, 2) correctly interpreted estimates of spillovers, or 3) produced valid inferences regarding the statistical significance of spillovers.

A (spatial) spillover arises when a causal relationship between the r^{th} characteristic/action of the i th entity/agent (X_i^r) located at position i in space exerts a significant influence on the outcomes/decisions/actions (y_j) of an agent/entity located at position j . In the context of a spatial regression relationship where $y_j, j = 1, \dots, n$ is a vector of outcomes/decisions/actions of an agent/entity located in region/location j , and X is a matrix of k characteristics/actions of all n regions/entities/agents, a formal definition would be: $\partial y_j / \partial X_i^r \neq 0$, which implies a spillover/impact from the r th characteristic/action of region/agent/entity i that impacts the outcome/decision/action in region j .

If the nonzero *cross-partial* derivative implies an impact on *neighboring* locations/regions that do not involve endogenous feedback effects, then we have a *local* spatial spillover. We note that *neighboring* regions can be (geographically) located nearby or far away, and that region j can have more than one neighbor, so $i \in \mathcal{N}$, where \mathcal{N} denotes the set of neighbors. In most applied regional modeling situations we know from substantive (or theoretical) aspects of the problem that a *local* spillover specification is the appropriate one. Examples include: state border crossing by social assistance program participants, cross-border shopping by cigarette smokers to avoid higher state cigarette taxes, and crossing school district boundaries by homeowners. A key facet of *local* spillovers is that endogenous interaction and feedback effects are *not* present. Endogenous interaction leads to a scenario where changes in

¹According to Wikipedia, this statement has been attributed to Tip O’Neil, former speaker of the U.S. House of Representatives.

one region/agent/entity set in motion a sequence of adjustments in (potentially) all regions in the sample such that a new long-run steady state equilibrium arises. Local spillovers do not produce this type of endogenous interaction outcome. For the examples given, local spillovers seem a more plausible specification if we believe that someone receiving public assistance might move from Ohio to a bordering state to maximize social assistance payments, or that smokers could regularly shop in bordering states to avoid higher state cigarette taxes. These actions do not cause further spillovers from these bordering states to *their* neighbors, however. We are often interested in the magnitude and significance of local spillovers as they pertain to the relationships we study, and fortunately these are easy to model and estimates of the spillover magnitudes are simple to interpret. Nonspatial regression specifications that exclude these from a model specification lead to estimates that suffer from omitted variables bias.

If the nonzero *cross-partial* derivative implies an impact on *neighboring* regions, plus neighbors to the neighboring regions, neighbors to the neighbors, and so on, then we have a *global* spatial spillover scenario. Formally, if locations j include not only neighbors to i , but neighbors to the neighbors of i , neighbors to the neighbors of the neighbors to i , and so on, we have a global spillover scenario. A key facet of *global* spillovers is that endogenous interaction and feedback effects *are* present. Endogenous interaction leads to a scenario where changes in one region/agent/entity set in motion a sequence of adjustments in (potentially) all regions in the sample such that a new long-run steady state equilibrium arises. For example, if it seems plausible that changes in levels of public assistance (cigarette taxes) in state A would lead to a reaction by neighboring states B to change their levels of assistance (taxes), which in turn produce a game-theoretic (feedback) response of state A , and also responses of states C who are neighbors to neighboring states B , and so on, then a global spillover specification is most appropriate. A resource shared by numerous regions such as a highway (or river) can be one cause of global spillovers. Congestion (or pollution) on a highway/river segment passing through one region can produce impacts on all other regions through which the highway/river passes. These impacts might evoke feedback or endogenous interaction effects whereby commuters change their behavior in response to congestion, regional authorities react to pollution from neighbors, and so on. These feedback reactions will have an impact on the expected long-run steady state for the relationship being studied. Networks (airline, computer, trade, transportation, academic collaboration, etc.) represent one situation that may involve global spillovers, since impacts fall on many nodes of the network. Over time, actors located in network nodes adjust to this global nature of spillovers which likely changes the expected long-run steady state relationship.

Global spillover phenomena should be rarer than local spillovers in applied regional science modeling situations, hence the statement: “most spatial spillovers are local.” Despite the fact that global spillover situations are likely rare, the spatial regression specifications most commonly used in the applied regional science literature are those associated with global spillovers, not local. Global spillover specifications are more difficult to estimate and correct interpretation of estimates from these specifications is more difficult. Because of the interesting theoretical econometric aspects of these specifications, they represent those most studied in the spatial econometrics literature. Still, this does not mean they should be those most frequently used by regional science practitioners.

2.1 Local spillover specifications

Two local spatial spillover specifications are the *spatial lag of X model* (SLX) in (1) and *spatial Durbin error model* (SDEM) shown in (2).

$$(1) \quad y = X\beta_1 + WX\beta_2 + \varepsilon$$

$$(2) \quad \begin{aligned} y &= X\beta_1 + WX\beta_2 + u \\ u &= \lambda W u + \varepsilon \\ \varepsilon &\sim N(0, \sigma_\varepsilon^2 I_N) \end{aligned}$$

These models allow for local spillovers to neighboring observations through spatial lag terms for the explanatory variables: WX . A spatial lag consists of a matrix product such as WX , Wy , which forms a linear combination of values from the matrix X or vector y , reflecting neighboring region values. In the simplest case, an average of neighboring independent or dependent variable values could be used. The matrix W is of dimension $n \times n$, where n is the number of observations, and each observation represents a region (or location). Non-zero elements in the i, j row and column positions of the matrix W indicate that region/observation j is a neighbor to i . Main diagonal elements are zero, and rows are normalized so that elements of each row sum to unity.

If we consider the partial derivatives, own-region partial derivatives $\partial y_i / \partial X_i^k = \beta_1$, while cross-partial derivatives that reflect the local nature of spatial spillovers to *only* neighboring regions are $\partial y_i / \partial X_j^k = W\beta_2$. Since the main diagonal elements of W are zero and the row-sums are unity, we can interpret the coefficient β_2 as the (cumulative) cross-partial derivative (or indirect effect). By cumulative we mean that the coefficient β_2 denotes the sum of spillovers falling on all neighbors. Like all regression coefficients, β_2 reflects average or typical spillovers, where averaging takes place over all observations (regions). This makes these models easy to interpret relative to the global spillover models. For the case of the SLX model, least-squares coefficient estimates for β_1, β_2 along with measures of dispersion such as t -statistics can be used to produce inferences regarding the magnitude and significance of direct (own-region, β_1) and indirect (other-region, spillover, β_2) impacts, so standard regression software can be used to estimate the SLX model. LeSage and Pace (2009) argue that cross-sectional versions of these models have received too little attention in applied work by regional scientists.

For the case of the SDEM model, a spatial autoregressive specification is used for the disturbances, which allows for global diffusion of shocks to the model disturbances. To avoid confusion of terminology, we do not refer to these as spillovers. To see that we have global impacts arising from shocks to the disturbances, note that: $u = (I_n - \lambda W)^{-1} \varepsilon$, which can be expressed as: $u = (I_n + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 + \dots) \varepsilon$. A change in the disturbance of a single region can produce impacts on disturbances of neighboring regions $\lambda W \varepsilon$, neighbors to the neighboring regions, $\lambda^2 W^2 \varepsilon$, and so on. Because the scalar parameter $\lambda < 1$, impacts decay with order of neighbors, so higher-order neighbors (e.g., neighbors to neighbors) receive less impact.

An econometric point is that regression estimates of β_1, β_2 from the SLX model should be unbiased even when the true model is SDEM, since spatial dependence in the disturbances represents only an efficiency problem. A related point is that the partial derivatives with respect to the explanatory variables are the same for both models, but specialized software would be required to produce estimates (and valid t -statistics) for the SDEM specification.

It should be noted that neighboring regions/observations might be defined as those located far away in terms of geographical space. An example that uses the SDEM model is LeSage and Ha (2012), who study the impact of migration on county-level social capital. Their SDEM model takes the form in (3), where W_n and W_f represent migration-weighted spatial weight matrices. The matrix W_n identifies neighboring counties within 40 miles and assigns relative weights to these based on in-migration magnitudes. The matrix W_f identifies neighboring counties more than 40 miles away from each county that provide in-migration to each county i in the sample, and weights these according to in-migration magnitudes. The matrix V used to model dependence in the model disturbances was a spatial contiguity weight matrix, with equal weights assigned to all contiguous counties.

$$(3) \quad \begin{aligned} y &= X\beta + W_n X\theta + W_f X\gamma + u \\ u &= \rho V u + \varepsilon \end{aligned}$$

Pace and Zhu (2012) point out that a desirable aspect of the model in (3) is that dependence in the disturbances is modeled separately from spillovers. This means that misspecification of the spatial dependence structure in the model disturbances will not contaminate estimates (β, θ, γ) from the mean part of the model. The SDEM model in (3) allows *separation* of the (local) spillover impacts on county-level social capital levels arising from changes in population characteristics of nearby counties (providing in-migrants to each county in the sample) versus that arising from changes in population characteristics of far away (outside the region) counties (providing in-migrants to each county). For example, how do changes in educational attainment levels of population in counties within the region versus counties outside the region (providing in-migrants to each county) impact levels of social capital in the typical county? Are there important differences in the magnitude of impact associated with in-migration from within and outside the region? Are some changes in characteristics of in-migrants from nearby counties significant/insignificant while the same characteristics of in-migrants from outside the region are insignificant/significant?

By way of conclusion, we note that although the term *local* spillovers could be used to characterize the model in (3), this does not necessarily rule out consideration of spillover impacts involving great distances, since practitioners can try variants of individual W with different bandwidths to capture longer range dependencies. Therefore, practitioners of spatial regression models should spend time thinking about whether the phenomena being modeled are likely to produce local or global spillovers.

2.2 Global spillover specifications

The *spatial Durbin model* (SDM) is a global spillover specification, taking the form shown in (4).

$$(4) \quad y = \rho W y + \alpha \iota_n + X\beta_1 + WX\beta_2 + \varepsilon$$

This model includes a spatial lag vector $W y$ representing a linear combination of values of the dependent variable vector from neighboring observations, as well as a matrix of own-region characteristics X , and a matrix of characteristics of neighboring regions (WX) as additional explanatory variables. An example would be a model where y represents a cross-section of metropolitan area census tract commuting times, and X population characteristics of the tracts. The motivation for global spillovers in this setting is that commuting time congestion spillovers arise from shared roadways that traverse census tracts. Commuters are likely to react

to congestion over time and change their behavior, reflecting endogenous interaction effects. Direct and indirect effects for the SDM model shown in (4) for the r^{th} explanatory variable in the matrix X , are given by the matrix partial derivative expression in (5).

$$(5) \quad \partial y / \partial X_r' = (I_n - \rho W)^{-1} (I_n \beta_1^r + W \beta_2^r)$$

The presence of global spillovers can be seen by recognizing that: $(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \dots$. This means that we have an $n \times n$ matrix of partial derivatives associated with a change in each of the explanatory variables. LeSage and Pace (2009) proposed using an average of the main diagonal elements of this matrix as a scalar summary measure of the own-partial derivatives ($\partial y_i / \partial X_i^k$) which they label *direct effects*. An average of the cumulative sum of off-diagonal elements reflecting cross-partial derivatives ($\partial y_j / \partial X_i^k$) provides a scalar summary measure of spillovers which they label *indirect effects*.

A literal interpretation of the partial derivatives from a cross-sectional model such as the SDM would be that the cross-partial derivative impacts on neighboring regions (indirect effects or spillovers) arise simultaneously. Intuitively, such impact should take time, yet there is no explicit role for time in a cross-sectional setting. Because of this, LeSage and Pace (2009) argue that spillovers in the context of (cross-sectional) spatial regression models should be interpreted as comparative static changes that will arise in the dependent variable (as a result of changes in the explanatory variables) as the relationship under study moves to a new steady-state equilibrium. Cross-sectional observations could be viewed as reflecting a (comparative static) slice at one point in time of a long-run steady-state equilibrium relationship, and the partial derivatives viewed as reflecting a comparative static analysis of changes that represent new steady-state relationships that would arise.

In contrast to the case of the SLX local spillover specification, the SDM and SDEM specifications require specialized software to produce valid estimates. Software algorithms developed for estimating spatial error dependence models (commonly labeled SEM) can be used with the addition of spatial lags of the explanatory variables (WX) for the case of the SDEM specification. Estimates of the direct and indirect effects can be based on coefficients β_1, β_2 from (2), and associated t -statistics can be used for inference.

In the case of the SDM specification, the direct and indirect effects associated with each explanatory variable and valid t -statistics are produced by numerous software programs available for estimating spatial regression models involving spatial lags of the dependent variable. These routines calculate and report direct and indirect effects estimates along with t -statistics that allow for statistical tests of significance of the direct and indirect or spillover effects associated with each explanatory variable. MATLAB and R-language programs are available as well as a Stata module.²

2.3 Between local and global spillover specifications

The distinction drawn between local and global spillover specifications is somewhat artificial, since it suggests that local spillovers involve *only* neighbors, but not higher-order neighbors (neighbors to the neighbors, neighbors to the neighbors of the neighbors, and so on

² MATLAB functions for spatial econometric analysis are at: www.spatial-econometrics.com, an R-language package by Roger Bivand (spdep), a Stata module, SPPACK for cross-section spatial-autoregressive models, by David M. Drukker, Hua Peng, Ingmar Prucha, and Rafal Raciborski.

represent higher-order neighbors).³ What about a situation where first- and second-order neighbors are involved, that is neighbors plus neighbors to the neighbors? Is this a local or global specification? According to the definition used here, it is neither!

Making a distinction that rules out a host of specifications that differ slightly in terms of the order of included neighbors may serve a useful purpose. Without large samples containing sample data information collected at a very fine spatial scale, it should be very difficult to statistically distinguish between slightly different specifications. It should also be difficult to draw inferences regarding the magnitude of decay of influence over space or distance without data containing such detail. One advantage of maintaining the somewhat artificial distinction is that it makes empirical tests for local versus global specifications easier (as we will see later). LeSage and Pace (2010) show that in the case of a global SDM specification, the model produces similar estimates, inferences and predictions for similar (e.g., 6 vs. 8 nearest neighbors) weight matrices. An implication of this is that empirical tests of alternative specifications will not work well, since these are based on measures of model fit involving the (similar) predictions. In applied work, we simply have to recognize the limitations of our models.

Given situations where sample data information is weak regarding slightly different spatial weight matrix specifications, a Bayesian approach provides one way to introduce subjective prior information regarding this aspect of the problem. Han and Lee (2013) set forth a Bayesian approach to estimating SDEM models that include varying orders of neighbors, but rely on Shiller smoothness priors to impose a stochastic structure of decay on the coefficients associated higher-order neighbors. They also set forth formal Bayesian model comparison methods that calculate posterior model probabilities based on log-marginal likelihoods.

3. PRINCIPLE 2: THERE ARE ONLY TWO MODEL SPECIFICATIONS WORTH CONSIDERING FOR APPLIED WORK

One obstacle confronting regional scientists who attempt to use spatial econometric methods is the plethora of alternative model specifications discussed in the literature. The family of models that have been labeled SAR, SDM, SEM, and SAC were popularized by the Anselin (1988) text, and a great deal of the literature on statistical testing of alternative model specifications has built on this family of models.

If one can narrow down the relationship being investigated as reflecting a local spillover situation, then the SDEM model is the only model one needs to estimate. This specification subsumes the SLX and SEM specifications as special cases. To see this, consider that the SDEM specification in (6) collapses to the SLX when $\lambda = 0$ and there is no spatial dependence in the disturbances. The SDEM specification collapses to the SEM specification when $\beta_2 = 0$ and there are no (local) spatial spillovers. When both $\lambda = 0$ and $\beta_2 = 0$, we have an ordinary non-spatial regression model.

$$(6) \quad \begin{aligned} y &= X\beta_1 + WX\beta_2 + u \\ u &= \lambda Wu + \varepsilon \end{aligned}$$

³Remember that neighbors need not be defined as only regions in close physical proximity, as noted in the discussion surrounding Equation (3).

For the case where a global spillover specification is implied by theoretical or substantive aspects of the problem, one need only estimate an SDM specification. It is not difficult to argue that the SDM specification in (7) subsumes the SAR as a special case when $\beta_2 = 0$.

$$(7) \quad y = \rho W y + \alpha \iota_n + X \beta + W X \beta_2 + \varepsilon$$

Nevertheless, a specification that seems particularly appealing to the novice spatial econometrician is the SAC (spatial autoregressive *combination*) model shown in (8). This specification has been explored in numerous *theoretical* spatial econometrics articles. From a theoretical viewpoint, it poses the greatest problems for estimation and inference, making it an attractive subject for theoretical econometric research. An interesting question is: what is the intuitive appeal of this specification to novice practitioners?

$$(8) \quad \begin{aligned} y &= \rho W y + \alpha \iota_n + X \beta + u \\ u &= \lambda W u + \varepsilon \end{aligned}$$

I believe the answer lies in *confusion* about what the specification represents. A first point to note is that this specification involves two spatial dependence parameters, ρ and λ . The additional spatial dependence parameter adds to the *costs* and complexity of estimation for this type of model. We should ask: do the marginal *benefits* associated with this specification justify the marginal *costs*?

One possible benefit from the SAC specification would be added precision (increased efficiency) that would arise in cases where the model disturbances exhibit spatial dependence. Ignoring spatial dependence in the disturbances has consequences similar to those arising from serial correlation in the errors. Adjusting for spatial dependence in the disturbances will produce more accurate *t*-statistics used for inference regarding the effects on the dependent variable *y* arising from changes in model explanatory variables, *X*. However, the SAC specification imposes a severe restriction on the partial derivatives, which take the form: $\partial y / \partial X'_r = (I_n - \rho W)^{-1} I_n \beta'_1$. The scalar summary measures for *all* explanatory variables ($r = 1, \dots, R$) based on the own- and cross-partial derivatives reflect the same *relative* relationship. That is, if explanatory variable X_1 has a direct effect of 1.0 and indirect effect of 0.5, then explanatory variable X_2 must also have a direct effect that is twice that of the indirect effect, as do all other explanatory variables (see Elhorst, 2010). Needless to say, this is an extremely restrictive aspect of the both the SAR as well as SAC specifications, which should make them relatively unattractive compared to the SDM and SDEM specifications for use in applied situations.⁴

A second *perceived* benefit derives from some confusion regarding the SAC specification. At first glance, the specification appears to represent a mixture or combination of both spatial dependence in the dependent variable reflected in $W y$ and spatial dependence in the disturbances represented by $W u$. This leads to the *misconception* that the SAC reflects a more general model specification than SAR or SEM, and, generality is often perceived to have advantages in econometrics. A more formal examination of what specification arises from a mixture of specifications reflecting spatial dependence in the dependent variable and dependence in the disturbances is provided by LeSage and Pace (2009, p. 31). They show what arises if we form a linear combination of the SAR and SEM specifications reflecting spatial dependence in

⁴ The SDM and SDEM are more general because a separate β is estimated for the own and lagged values for each independent variable.

the dependent variable and disturbances, respectively. Let $\pi_a + \pi_b = 1$, and use these weights to form a linear combination of the SAR and SEM data generating processes as shown in (9).

$$(9) \quad y_c = \pi_a y_a + \pi_b y_b$$

$$(10) \quad y_a = (I_n - \rho W)^{-1}(X\beta + \varepsilon)$$

$$(11) \quad y_b = X\beta + (I_n - \rho W)^{-1}\varepsilon$$

$$(12) \quad y_c = \rho W y_c + X\beta_1 + WX\beta_2 + \varepsilon$$

The resulting specification shown in (12) is that of the SDM. In addition to this misconception regarding what the SAC specification represents, there are a number of other drawbacks to its use in applied practice that are often overlooked.

Another cost associated with using a two-parameter specification for spatial dependence is that misspecification in one part of the model can contaminate estimates from the other part. Pace and Zhu (2012) describe models that *separate* the spatial process assigned to govern the mean part of the model from that used in the disturbance specification. They argue that this approach provides some protection against misspecification in one part of the model contaminating estimates from other parts. The data-generating processes for the SAC model represents a situation where the *same* spatial autoregressive process $(I_n - \rho W)^{-1}$ is assigned to both the mean model ($X\beta$) as well as the disturbances ε , based on $(I_n - \lambda W)^{-1}$. This aspect of these models means that misspecification of the disturbances can produce undesirable impacts on estimates for the mean model which is likely the focus of inference.

By way of conclusion, one should consider the costs versus benefits from adopting a two-parameter specification such as the SAC over alternative one-parameter specifications such as SDM, SDEM, SLX, SAR and SEM. Practitioners should also be aware that alternative two-parameter specifications exist that may have better benefit-cost ratios (Pace and Zhu 2012).

3.1 Testing the two model specifications

There are relationships we explore in regional science where the distinction between global and local spillovers is not obvious. Consider the case of state-level cigarette demand, with a well-know model from the second edition of Baltagi's (2001) textbook on panel data econometrics. State-level demand for cigarettes is modeled using a panel of information on: a dependent variable measuring demand for cigarettes (real per capita sales of cigarettes by persons of smoking age (14 years and older), and explanatory variables: the average retail price of a pack of cigarettes measured in real terms, real per capita disposable income, and the minimum real price of cigarettes in any neighboring state. The minimum price in neighboring states is a proxy for local spillover effects that arise when consumers engage in *casual smuggling* (shop across state borders to take advantage of lower taxes in neighboring states). This variable (say P_{\min}) can be written as WP_{\min} , where the matrix W contains a value of one (in each row) for the neighboring state with the minimum price. (We could extend the model specification using a matrix W that forms a linear combination of all neighboring states whose cigarette taxes are below those of state i , or we could include a linear combination of real per capita disposable income of residents from neighboring states, retail prices of cigarettes from neighboring states, etc.)

This would seem to be an example of *local spatial spillovers*, since it is an argument that explanatory variable characteristics of only a single neighboring region (that with contiguous

borders and the minimum tax on cigarettes) should exert an impact on the outcome variable (demand for cigarettes). It might seem difficult to envision a global spillover specification as appropriate for this type of phenomena.

Still, *commercial smuggling network* activity could produce *global* spatial spillovers. Low tax states such as North Carolina and Kentucky may be a source of illegal exports to many states as a result of deliberate endogenous interaction by smugglers. Commercial smugglers would obviously react to state-level excise taxes. LaFaive and Nesbit (2010, 2013a, 2013b) claim that (over the 1990 to 2006 period) the states with the top five average smuggling import rates as a percentage of their total estimated in-state cigarette consumption, including both legally and illegally purchased cigarettes, were California (24.5 percent of the state's total cigarette consumption), New York (20.9 percent), Arizona (20.6 percent), Washington state (20.1 percent), and Michigan (16.0 percent). They claim that commercial smuggling import rates were highest in New Jersey (13.8 percent), Massachusetts (12.7 percent), and Rhode Island (12.7 percent) and that casual smuggling import rates were highest in New York (9.9 percent), Washington (8.9 percent), and Michigan (6.0 percent).

The possibility that both casual and commercial smuggling are at work suggests that this may be a case where formal model comparison tests would be useful in distinguishing between a local versus global specification. There are likely many other situations where regional scientists are interested in testing whether endogenous interaction is present or absent. For example, are local governments engaged in tax or services provision competition with neighboring governments? Or, do differences in tax and service provision by local governments simply result in a local spillover phenomena? The former situation suggests a global spillover specification (SDM), while the latter implies a local spillover (SDEM) specification. Effective statistical tests should be useful in these circumstances.

3.2 Testing even two specifications can be tricky

Given only the SDEM versus SDM specifications, consider a case where the true model that generated the data is the SDM in (14), but the SDEM specification in (13) is estimated.

$$(13) \quad \begin{aligned} y &= X\beta_1 + WX\beta_2 + u \\ u &= \lambda Wu + \varepsilon \end{aligned}$$

$$(14) \quad y = \rho Wy + \alpha \iota_n + X\beta + WX\beta_2 + \varepsilon$$

In the absence of a spatial lag of the dependent variable in the estimated SDEM specification, spatial dependence from the dependent variable will land in the disturbances, likely producing a significant coefficient estimate for λ that might be similar in magnitude to the true value of ρ . Similarly, in the converse scenario where the SDEM in (13) is the true data-generating process, but the SDM specification is estimated, since the SDM specification does not accommodate dependence in the disturbances, we might expect to see estimates of ρ similar in magnitude to the true value of λ from the DGP.

A practical implication is that practitioners cannot simply estimate an SDM and SDEM model and draw meaningful conclusions about the correct model specification from point estimates of ρ or λ . This has raised a serious debate in the spatial econometrics literature about *identification* of model parameters. In our case, we have seen that the SDM specification implies global spatial spillovers, while the SDEM is consistent with local spillovers. These have very different policy implications in applied work. LeSage (2014) shows that despite the difficult

circumstances facing comparison of SDM versus SDEM specifications, formal Bayesian model comparison methods can provide accurate answers regarding the appropriate specification.

3.3 Why does Bayesian model comparison work well?

Distinguishing between even SDM and SDEM specifications is complicated by the fact that both specifications collapse to the SLX specification: $y = X\beta_1 + WX\beta_2 + \varepsilon$, when ρ, λ approach zero. Most of the spatial econometrics literature on comparing model specifications relies on likelihood ratio or Lagrange multiplier statistics. These involve evaluating the (log) likelihood function at maximum likelihood point estimates which produces a single scalar magnitude for two model specifications. A statistical test is then carried out to see whether one of the two scalars reflects a significantly better fit to the sample data for one of the two specifications.

Of course, inferences regarding comparison of the two model specifications from this approach are *conditional* on the maximum likelihood estimates from the two specifications. And they assume one of the two specifications is true. When spatial dependence is weak however, the best fitting specification might be the SLX, which makes neither the SDM nor SDEM the true model.

Bayesian model comparison takes a different approach, integrating out all model parameters to produce a log-marginal likelihood for each model specification. These are used to calculate model probabilities in favor of each specification. We can easily include three specifications, SLX, SDM and SDEM and produce model probabilities for all three. In cases of weak dependence evidenced by small values of ρ or λ , posterior probabilities should approach $\frac{1}{3}$ for each model, making it easy to interpret results from the model comparison exercise.

One could (loosely) view likelihood ratio statistics as comparing the single modal values of the densities shown for both models, attempting to discern (statistically significant) differences between these scalar magnitudes. In contrast, Bayesian model comparison takes into account the entire profile of the density by (numerically) integrating over this final spatial dependence parameter for both models, and all other parameters are integrated out analytically. It should be clear that there is a great deal of information to be found regarding the curvature of the log-marginal likelihood from the two models by comparing the area under these two densities. This is in stark contrast to a model comparison method that focuses only on the single points reflecting the modal values of the two densities shown.

As an illustration, Table 1 shows model probabilities for the SLX, SDM, and SDEM specifications produced using a panel data Monte Carlo experiment that systematically decreases the spatial dependence parameter ρ towards zero. The panel data set contained 49 regions and 30 time periods. It is relatively easy to interpret the model probabilities that favor the (true) SDM specification for levels of dependence greater than 0.1 in absolute value. For lower values of dependence, it becomes clear that one cannot draw a meaningful distinction between the three specifications. It would be far more difficult to arrive at this type of conclusion using likelihood ratio tests that rely on two-way comparisons of SDM versus SLX, SDEM versus SDM and SDEM versus SLX specifications.

4. PRINCIPLE 3: CONSIDER PERFORMANCE OF ESTIMATION PROCEDURES, NOT SIMPLY POINT ESTIMATES

A spatial regression *estimation procedure* (e.g., Bayesian, IV, GMM, maximum likelihood, robust Bayesian, GMM-HAC, etc.) needs to produce not only *point estimates* of the underlying parameters (β, ρ, λ), but other information as well. Specifically, 1) scalar summaries of the partial derivatives (labeled direct and indirect/spillover *effects estimates*) that show how changes in the explanatory variables impact the dependent variable, 2) estimates of *dispersion* for the underlying parameters that can be used for *inference* regarding the statistical significance of (underlying) parameters such as ρ, λ measuring spatial dependence, 3) estimates of *dispersion* for the direct and indirect effects parameters that can be used for inference regarding the statistical significance of spillovers associated with different explanatory variables, and perhaps 4) predictions (in- or out-of-sample) as well as measures of dispersion for these.

4.1 Should we believe Monte Carlo studies?

Surprisingly, almost all of the spatial econometric literature contains Monte Carlo experiments that explore performance of alternative estimation procedures focusing only on the mean-variance relationships of the point estimates for β, ρ, λ . That is, they ignore performance of scalar summary *effects estimates* of the partial derivatives produced by alternative procedures. Since these effects estimates are those that regional scientists will use to learn about the relationship under study, the spatial econometric literature has little to offer practitioners.

4.2 Point estimates versus estimation procedures

An *estimation procedure* (such as robust Bayes, robust GMM, maximum likelihood) needs to produce not only *point estimates*, but also estimates of *dispersion* that can be used for *inference* regarding the point estimates. For example, if a proposed robust GMM *procedure* produces *slightly* better (less biased) point estimates versus *slightly* (more biased) robust Bayes, but *dramatically* inferior variance-covariance estimates (which are used for inference), then the *proposed procedure* might be considered inferior as an *estimation procedure* for producing estimates *and* inference in applied settings.

Marginal or *effects estimates* and inferences for spatial regression models containing spatial lags of the dependent variable are produced by simulating thousands of parameters β, ρ from the estimated mean and variance-covariance matrix. Then using these in the nonlinear partial derivative relationship produces an empirical estimate of the mean and dispersion of the scalar summary effects estimates.

Table 1. Monte Carlo results for $\rho, \lambda \rightarrow 0, N = 49, T = 30$

True SDM Operator, ρ	Prob(SLX)	Prob(SDM)	Prob(SDEM)
-0.1	0.0601	0.5451	0.3948
-0.05	0.2290	0.3999	0.3711
-0.01	0.3266	0.3470	0.3264
-0.001	0.3268	0.3326	0.3406
0	0.3389	0.3374	0.3237
0.001	0.3485	0.3145	0.3370
0.01	0.3436	0.3236	0.3328
0.05	0.2209	0.3766	0.4025
0.1	0.0633	0.5176	0.4191

4.3 Spatial econometric Monte Carlo studies: Raising the bar

Monte Carlo studies typically perform some number of replications using the data generating process to produce say, 1,000 vectors y based on randomly generated disturbances. Mean estimates of the parameters β, ρ are reported from the 1,000 runs along with estimated standard deviations for these parameters. The means and standard deviations are compared to the true values for the parameters β, ρ along with the standard deviation in these outcomes, where the true standard deviation is sometimes known and other times unknown. A superior estimation procedure is judged to be the one with the least bias (gap between the true β, ρ and mean of the 1,000 estimates) and smallest variance in the 1,000 estimates of β, ρ .

The focus of Monte Carlo studies on the estimates of β, ρ is flawed, because the inferences of practitioners will be based upon how the explanatory variables impact outcomes on their own and neighboring regions. Because the scalar summaries of the $n \times n$ matrix of partial derivatives for each explanatory variable is a nonlinear function of the parameters β, ρ , the bias and precision of the *impacts* can be very different from the bias and precision of the *parameters* β, ρ . We will demonstrate this below.

For *each* of the 1,000 trials, estimated *scalar summary direct, indirect and total effects* (based on the $n \times n$ matrix of nonlinear partial derivatives) should be stored, along with estimated empirical standard deviations based on the simulated parameters β, ρ arising from the point estimates and estimated variance-covariance matrix. Comparing the mean of the set of 1,000 scalar summary effects estimates from alternative estimation procedures to the true effects estimates provides a measure of bias associated with the various procedures. Comparing the mean of the set of 1,000 empirical standard deviations across alternative estimation procedures provides a measure of relative efficiency of the alternative approaches.

There are likely to be trade-offs between bias and efficiency. Decisions regarding these trade-offs require a loss function, with the mean squared error (MSE) criterion arising from a quadratic loss function. The MSE assesses the quality of an estimator in terms of its variation and degree of bias. $MSE = \text{variance}(\text{effects}) + \text{bias}(\text{effects})^2$. Table 2 shows bias and MSE results from a Monte Carlo study of three estimation procedures for both the β, ρ estimates alongside bias and MSE results for the total effects estimates (direct + indirect effects = total effects). Of course, practitioners would draw conclusions about the relationship between y and x –variables using the total effects estimates, not estimates of β, ρ .

From results for parameters ρ, β_1, β_2 in the table (those typically reported in Monte Carlo studies), Method 2 seems superior since it exhibits *much smaller bias* in estimates for these parameters, with a *slightly larger* MSE than Method 3. However, results based on the total effects estimates that would be used in applied practice make it clear that both Methods 1 and 3 are far superior to method 2 in terms of total effects estimates.

This should make it clear that past Monte Carlo studies comparing alternative estimation procedures are not very useful regarding performance that the typical practitioner will experience when using alternative methods. Spatial heteroscedastic autocorrelation consistent (SHAC) procedures provide one possible illustration of this issue (Kelejian and Prucha, 2007, Doğan and Taspınar, 2014). Excessive focus on (possibly small) bias arising from the presence of

Table 2. An Illustration of Trade-offs between Bias and Efficiency

	Method 1		Method 2		Method 3	
	bias	MSE	bias	MSE	bias	MSE
ρ	-0.0350	0.0507	-0.0008	0.0502	-0.0333	0.0482
β_1	0.0599	1.5914	0.0854	1.5602	0.0799	1.4401
β_2	-0.0801	1.5977	-0.0869	1.5685	-0.1470	1.4269
Total Effects						
$Total_1$	-0.7455	14.6243	6.5850	82.8110	-0.7656	12.6486
$Total_2$	-2.5610	14.1155	-5.9543	51.6293	-2.4481	12.2358

heteroscedastic disturbances has led to proposed alternative estimation procedures. These alternatives involve adjustments to the estimation procedure including formulas used to calculate variance-covariance estimates. The impact of adjustments to the variance-covariance matrix estimates on simulated parameter values used to construct empirical measures of dispersion for the direct, indirect and total effects estimates has gone unexplored. If changes in the estimated variance-covariance matrix lead to much larger dispersion in the simulated parameters and empirical estimates of dispersion, then the ability to draw meaningful inferences regarding statistical significance of direct and indirect effects estimates arising from changes in explanatory variables on the dependent variable outcomes may be dramatically altered by these alternative SHAC estimation procedures. Given that past Monte Carlo studies have not explored this issue by producing empirical estimates of dispersion for the effects estimates based on simulating parameters from the point estimates and (SHAC) variance-covariance matrix, we really have little idea about this important issue.⁵

In fact, Table 2 results compare maximum likelihood (Method 1) to a SHAC (Method 2) and robust Bayesian (Method 3) for the case of heteroskedastic disturbances. As indicated in the discussion surrounding Table 2, results based on the conventionally reported estimates for the parameters ρ, β appear to support benefits provided by the SHAC procedure, while results based on the total effects estimates clearly contradict these conclusions. It should be noted that even conclusions regarding bias of the parameters β, ρ cannot be trusted when it comes to bias of the effects estimates, since upward (downward) bias in ρ could be offset by downward (upward) bias in β producing smaller bias in the effects estimates. This appears to be the case for the Table 2 results.

5. PRINCIPLE 4: AVOID THE TEMPTATION FOR OBSERVATION-LEVEL INFERENCE

Given that partial derivatives for spatial econometric models containing a spatial lag of the dependent variable take the form of an $n \times n$ matrix, one might suppose there is no need for scalar summary estimates. The matrix of partial derivatives literally shows how every observation in the sample data responds to changes in every observation of the explanatory variables. While observation-level inference is possible, it is not advisable. Intuitively, can we draw realistic and useful inferences regarding an $n \times n$ matrix of parameters based on n sample

⁵ Doğan and Taspınar (2014) do not calculate empirical standard deviations for each Monte Carlo iteration and explore the dispersion of these standard deviation outcomes. They simply calculate total effects for each Monte Carlo iteration and assess the standard deviation of these total effects over all iterations.

observations? Models that provide *observation-level* inference appear to be the holy grail for many practitioners, which may account for the popularity of geographically weight regression methods as well as other semi-parametric approaches that produce estimates that can be mapped for every region/observation. A practical aspect of this desire is that clients of regional economic studies are frequently officials from a single region/observation used in the sample data set. These clients are mostly interested in how outcomes in their region respond to changes in characteristics of their (and perhaps neighboring) regions.

We have some responsibility to inform clients that regression methods (including spatial regression) provide estimates and inferences based on averaging dependent variable responses to explanatory variable changes over all sample observations. Inferences regarding any single observation are far less likely to be indicative of the general influences of changes in regional characteristics/policies on outcomes that regression methods focus on producing. Observation-level inferences are also not likely to possess much precision. This can lead to the possibility that perceived differences between point estimates for each region are important, when in fact they are not. For example, if these differences in point estimates were considered alongside confidence intervals for each point estimate, they would not be (statistically) significantly different from other observations.

As an example of observation-level inference Kelejian and Mukerji (2011) construct effects they label *emanating effects* that represent spillovers from a single observation to all other observations in the sample. These are essentially rows of the $n \times n$ matrix of partial derivatives: $\partial y / \partial X'_k = (I_n - \rho W)^{-1} \beta_k$.⁶ An applied illustration reports these observation-level effects for GDP per capita growth response of individual countries (observations) to changes taking place in other countries (observations). A great deal of discussion and analysis of differences between various observations/countries responses is provided, but no (statistical) indication of whether these differences are statistically different across observations.

Another aspect of observation-level inference is that results are likely to be very sensitive to the spatial weight matrix specification used. Intuitively, the same two scalar estimates for the parameters ρ and β_k are used to produce $1 \times n$ vectors of responses in y to changes in each observation of the explanatory variable x_{ik} . The major difference between these response vectors arises from differences in the number of spatial neighbors to each observation, making observation-level inferences sensitive to changes in specification of the weight matrix. In contrast, the scalar summary estimates for the *indirect/spillover effects* have been shown to be robust with respect to changes in the spatial weight matrix specification (LeSage and Pace, 2010).

6. PRINCIPLE 5: KEEP THE WEIGHT MATRIX SIMPLE

The role played by the spatial weight matrix has long been a controversial aspect of spatial regression methods. Much of the controversy stems from the mistaken belief that minor changes in specification of the weight matrix produce major changes in spatial regression model estimates and inferences. This is simply not true, if one produces valid estimates and inferences reflecting the partial derivative changes for the SDM using: $\partial y / \partial X'_r = (I_n - \rho W)^{-1} (I_n \beta_1 + W \beta_2)$. LeSage and Pace (2010) discuss how historically incorrect interpretation of estimates from spatial regression models involving spatial lags of the dependent variable may have created

⁶They also construct *vulnerability effects* reflecting columns of the same matrix of partial derivatives.

the impression of sensitivity of estimates and inferences to the weight matrix specification. Some guiding principles regarding weight matrix specification are provided here.

6.1 Sparse connectivity structures work best

As already motivated, the ability to distinguish between local and global spillover specifications is important for the task of producing plausible estimates of spatial spillovers. A common practice is to specify W based on distance between centroids of regions with some type of decay imposed so more distant regions receive smaller weights than nearby regions. Unless the decay imposed includes a cut-off distance beyond which weights are restricted to zero values, a distance-decay based weight matrix will result in a nonsparse weight matrix.⁷ Since distance between centroids of all regions is positive, a distance-based weight matrix specification (without a cut-off distance) implies all regions are connected to all others. You should ask: does this type of weight matrix make sense in the context of a local spillover specification? In many applied situations, the intention behind using a local spillover specification would be to limit the spatial extent of spillover influences to regions nearby (in geographical space), so allowing for impacts falling on all other regions in the sample (even if these are small due to distance decay) is counterintuitive at best.

As already noted, maintaining a clear distinction between the local and global model specifications simplifies the task of empirically testing for differences between these two models. Use of a distance-based weight matrix (without a cut-off that produces sparseness) blurs the distinction between local and global specifications.

A related point regards interpretation, which calculates *cumulative* spatial spillovers adding up off-diagonal elements from each row of the matrix inverse (for the case of the SDM): $(I_n - \rho W)^{-1}(I_n \beta_1 + W \beta_2)$, and then taking an average of these row-sums to produce a scalar summary measure. Even if off-diagonal elements associated with distant regions are small due to distance decay, summing up numerous small elements may have a nontrivial impact on the scalar summary measure of spillovers.

6.2 Identification problems for parameterized decay

Often, distance decay is specified as a parameter to be estimated, e.g., $W_{ij} = 1/(d_{ij})^\gamma$. Some technical problems arise here as the decay parameter γ is not defined for values of the spatial dependence parameter ρ equal to zero, producing a point discontinuity in the likelihood. Aside from this issue, since slightly different weight matrices produce similar estimates and inferences, attempts to statistically test different values for the decay parameter are likely hindered by ridges in the likelihood function values associated with similar decay magnitudes.

Sparse weight matrices arise from use of some number m of nearest neighboring regions, or simply the set $m_i, i = 1, \dots, n$ of regions contiguous to each observation i . One could also weight contiguous regions by lengths of borders in common, or assign differing weights to each of the m nearest neighbors based on distance between centroids of the regions. The latter is equivalent to imposing a cut-off number of neighbors beyond which weight elements equal zero. Of course, a physical distance of z miles from the centroid of each region could be used to identify a varying number of neighboring regions to include for each observation/region.

⁷Sparse matrices are those containing a large percentage of zero-valued elements.

LeSage and Pace (2010) show that the correlation between spatial lags of a standard independent normal $n \times 1$ vector u , $W_a u$ and $W_b u$ is: $corr(W_a u, W_b u) = (m_a/m_b)^{0.5}$, where $m_a \leq m_b$, and m_a, m_b represents the number of neighbors used in constructing W_a, W_b . For example, if W_a is based on 15 nearest neighbors and W_b 16 nearest neighbors, the correlation between $W_a u, W_b u$ is 0.97.

Spatial regression models also involve use of vectors involving higher-order spatial lags such as $W_a^j u, W_b^j u$ for $j > 1$. LeSage and Pace (2010) show that for doubly stochastic symmetric weight matrices (having both row- and column-sums of unity), whether based on nearest neighbors, inverse distance, contiguity, common border lengths, or any other method of construction, the limiting values (as j becomes large) of $W_a^j u, W_b^j u$ are identical. LeSage and Pace (2010) argue that other weight matrices (those that are not doubly stochastic) share the general result that we would expect the exact form of W to become less important for many of the higher order neighboring relations that play a role in spatial regression models.

They also show that predicted values as well as effects estimates (true partial derivatives $\partial y/\partial X_r'$) are highly correlated for weight matrices based on similar numbers of nearest neighbors, for example 10 nearest neighbors versus 8 or 12 neighbors. Practitioners can take advantage of robustness for estimates and inferences derived from model specifications based on similar sparse weight matrices, by avoiding non-sparse weight structures.

6.3 Avoid the pitfall of multiple weight matrices

A temptation facing practitioners new to spatial regression methods involves specification of weight matrices that reflect complicated connectivity structures between regions. One seemingly intuitive approach is to rely on more than a single weight matrix, with each weight matrix representing a different kind of regional connectivity. For example, it seems natural to specify a model such as that in (15), where W_S reflects spatial proximity of regions and W_T technological proximity.

$$(15) \quad y = \rho_S W_S y + \rho_T W_T y + X\beta + \varepsilon$$

Technological proximity could be specified using information on industry structure of the regional observations, such that say m regions exhibiting the highest correlation with the industry structure of each region are labeled technological neighbors. The model purportedly posits that both spatial neighbors as well as technological neighbors exert an influence on outcomes y .

While this seems a simple and intuitively appealing extension of the standard spatial regression model, LeSage and Pace (2011) point to a number of problems with this type of specification. One problem is that the model does not allow for covariance between W_S, W_T , which would involve inclusion of a third term: $W_S W_T y$. This would reflect the fact that if spatial proximity as well as technological proximity are deemed important, then regions that are spatial neighbors as well as technological neighbors should exert a different impact than those that are spatial but not technological neighbors, or technological but not spatial neighbors. One can arrive at this more general model result by viewing the model specification as involving both spatial and technological filters/smoothers, as shown in (16).

$$(16) \quad (I_n - \rho_S W_S)(I_n - \rho_T W_T)y = X\beta + \varepsilon$$

$$(17) \quad y = \rho_S W_S y + \rho_T W_T y - \rho_S \rho_T W_S W_T y + X\beta + \varepsilon$$

$$(18) \quad y = [(I_n - \rho_S W_S)(I_n - \rho_T W_T)]^{-1}(X\beta + \varepsilon)$$

Taking this view of the model, the expression in (17) indicates that exclusion of the covariance term implies a restriction that: $-\rho_S \rho_T = 0$, which is inconsistent with the model posited in (15), where both ρ_S, ρ_T are assumed different from zero.

The DGP for the model shown in (18) points to another issue raised by LeSage and Pace (2011), the inability to separately identify direct and indirect effects associated with spatial and technological proximity.⁸ This is because the impact of changes in outcomes y associated with changes in the r th regional characteristic X_r take the form: $\partial y / \partial X_r' = [(I_n - \rho_S W_S)(I_n - \rho_T W_T)]^{-1} \beta$. Since $(AB)^{-1} = B^{-1}A^{-1}$, and $A^{-1} = (I_n + \rho_S W_S + \rho_S^2 W_S^2 + \dots)$, while $B^{-1} = (I_n + \rho_T W_T + \rho_T^2 W_T^2 + \dots)$, an infinite number of cross-product terms involving matrix products such as: $W_S W_T^3, W_S^2 W_T$ will arise in the expressions used to interpret our model estimates.

It seems pointless to extend the simple model to include multiple types of connectivity if we cannot separately identify how these different types of connectivity work in conjunction with changes in regional characteristics (X) to impact regional outcomes (y).

6.4 Interpretation pitfalls with non-spatial weight matrices

Another issue that arises when specifying connectivity using non-spatial notions of proximity such as technology is that of endogeneity of the matrix W_T . Since industry structure of regions changes over time, how are we to interpret the partial derivatives from a model involving non-spatial lags? LeSage and Pace (2009) point out that we interpret the partial derivatives from models involving spatial lags of the dependent variable as the changes that arise when moving the system of regions in our sample from one steady-state equilibrium to a new steady-state equilibrium (comparative statics). Specifically, we view the cross-sectional relationship we estimate as reflecting a long-run steady-state equilibrium relationship between regional characteristics (X) and outcomes (y). We use estimates of the parameters of the model relationship to infer how changes in regional characteristics produce comparative static changes in long-run steady-state outcomes. Since spatial proximity of regions is fixed over time, there is no problem envisioning (global) spillover impacts that diffuse over space based on spatial proximity relationships during movement of the regions to a new steady-state equilibrium. This is of course not the case when we define connectivity of regions using notions such as technological proximity that change over time.

7. CONCLUSIONS

Regional scientists who wish to use spatial regression methods in applied work are likely to find the spatial econometrics literature confusing. It is argued here that much of the literature can be ignored as it does not provide relevant information pertinent to practitioners.

First, the literature places too much emphasis on the SAC specification because of its theoretical econometric interest. This specification can be safely ignored by practitioners as it has numerous drawbacks in applied use. In fact, it is argued that only two specifications, the SDM and SDEM, should be considered by regional science practitioners. This should greatly simplify the task of deciding on an appropriate model specification. A further simplification

⁸There may be special cases where these can be separately identified, but not in general.

arises when the modeling context makes it clear that *either* a local or global spillover specification is the only specification that is theoretically justified. In cases of doubt, Bayesian model comparison methods can provide accurate inferences regarding which of the two specifications is most appropriate. These model comparison methods can be used for simple cross-sectional problems as well as static and dynamic panel data problems. Unfortunately, these methods have not yet been developed for probit and tobit models.

Second, practitioners are likely to experience very different performance of alternative estimation methodologies than portrayed by spatial econometric Monte Carlo studies reported in the literature. Monte Carlo methodologies used in spatial econometric studies are in need of serious reform. Studies have overwhelmingly ignored evaluation of estimation procedures in terms of the estimates and inferences that typical applied users will rely on. Further, the focus of most studies is on model specifications other than SDM and SDEM, which it is argued here should be the mainstay models used by practitioners.

Third, the literature has perpetrated a myth that estimates and inferences from spatial regression models are sensitive to specification of the weight matrix. Estimates and inferences that are robust with respect to weight matrix specification can be achieved by following simple principles discussed in the body of this paper.

It is hoped that regional science practitioners with little experience with spatial regression methods can benefit from the discussion of simple principles for applied use of spatial regression models set forth here.

REFERENCES

- Anselin, Luc. (1988) *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers: Dordrecht, Germany.
- Baltagi, Badi. H. (2001) *Econometric Analysis of panel data*, 2nd ed. Wiley: Chichester.
- Doğan, Osman and Süleyman Taşpınar. (2014) “Spatial Autoregressive Models with Unknown Heteroskedasticity: A Comparison of Bayesian and Robust GMM Approach,” *Regional Science and Urban Economics*, 45, 1–21.
- Elhorst, J. Paul. (2010) “Applied Spatial Econometrics: Raising the Bar,” *Spatial Economic Analysis*, 5, 9–28.
- Han, Xiaoyi and Lung-fei Lee. (2013) “Bayesian Estimation and Model Selection for Spatial Durbin Error Models with Finite Distributed Lags,” *Regional Science and Urban Economics*, 43, 816–837.
- Kelejian, Harry H. (2007) “HAC Estimation in a Spatial Framework,” *Journal of Econometrics*, 140, 131–154.
- Kelejian, Harry H. and Purba Mukerji. (2011) “Important Dynamic Indices in Spatial Models,” *Papers in Regional Science*, 90, 693–702.
- LaFaive, Michael D. and Todd Nesbit. (2010) “Cigarette Taxes and Smuggling: An Update of Earlier Research,” *Mackinac Center for Public Policy*, www.mackinac.org/14210;
- _____. (2013a) “Cigarette Smuggling Remains High in Michigan, Elsewhere,” *Viewpoint on Public Issues*, No. 2013-05 ISSN 1093-2240 Available online at mackinac.org/v2013-05

- _____. (2013b) “Higher Cigarette Taxes Create Lucrative, Dangerous Black Market,” *Mackinac Center For Public Policy*, www.mackinac.org/18128;
- LeSage, James P. (2014) “Spatial Econometric Panel Data Model Specification: A Bayesian Approach,” *Spatial Statistics*. 9, 122–145.
- LeSage, James P. and Christina Ha. (2012) “The Impact of Migration on Social Capital—Do Migrants Take Their Bowling Balls with Them?,” *Growth and Change*, 43,1–26.
- _____. (2009) *Introduction to Spatial Econometrics*. CRC Press, Taylor & Francis Group: Boca Raton, FL.
- _____. (2010) “The Biggest Myth in Spatial Econometrics,” Available in October 2014 at <http://ssrn.com/abstract=1725503> or <http://dx.doi.org/10.2139/ssrn.1725503>.
- _____. (2011) “Pitfalls in Higher Order Model Extensions of Basic Spatial Regression Methodology,” *Review of Regional Studies*, 41, 13–26.
- Pace, R. Kelley and Shuang Zhu. (2012) “Separable Spatial Modeling of Spillovers and Disturbances,” *Journal of Geographical Systems*, 14, 75–90.