JOURNAL OF REGIONAL SCIENCE

JOURNAL OF REGIONAL SCIENCE, VOL. 52, NO. 2, 2012, pp. 192–209

PERSPECTIVES ON SPATIAL ECONOMETRICS: LINEAR SMOOTHING WITH STRUCTURED MODELS

Daniel P. McMillen

Department of Economics and Institute of Government and Public Affairs, University of Illinois at Urbana-Champaign, 1007 W. Nevada St. (MC-037), Urbana, IL 61801. E-mail: mcmillen@illinois.edu

ABSTRACT. Though standard spatial econometric models may be useful for specification testing, they rely heavily on a parametric structure that is highly sensitive to model misspecification. The commonly used spatial AR model is a form of spatial smoothing with a structure that closely resembles a semiparametric model. Nonparametric and semiparametric models are generally a preferable approach for more descriptive spatial analysis. Estimated population density functions illustrate the differences between the spatial AR model and nonparametric approaches to data smoothing. A series of Monte Carlo experiments demonstrates that nonparametric predicted values and marginal effect estimates are much more accurate then spatial AR models when the contiguity matrix is misspecified.

1. INTRODUCTION

The term "spatial econometrics" is now most closely associated with two specific parametric model structures. The first model specification is the *spatial AR* model, which adds a weighted average of nearby values of the dependent variable to the base set of explanatory variables: $\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{u} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\beta + \mathbf{u})$, where $\mathbf{W}$ is an nxn spatial weight matrix. The second specification, the *spatial error* model, uses a similar structure to directly model spatial relationships among the errors: $\mathbf{y} = \mathbf{XB} + \mathbf{u}$, with $\mathbf{u} = \theta\mathbf{W}\mathbf{u} + \mathbf{e} = (\mathbf{I} - \theta\mathbf{W})^{-1}\mathbf{e}$. These models were once difficult to estimate because they were not part of common statistical software packages and working with large, nxn matrices was extremely slow and required lots of memory. Now they have become routine and over-used.

The models are over-used precisely because they have become routine. As a result, they have become a quick fix for nearly any model misspecification issue related to space. It does not appear to be recognized commonly that these models are just another form of spatial smoothing. It should not be surprising that a weighted average of nearby values of the dependent variable or the error term adds significant explanatory power when nearly all data in typical spatial sets are highly correlated over space. As a result, the standard models are likely to produce statistically significant estimates whenever there are missing variables or problems with the functional form specification (in other words, virtually always).

In many applications, standard spatial econometric models are just another way of accounting for unexplained spatial effects. What is ironic is that the common estimation methods are derived from parametric specifications and log-likelihood functions that rely heavily on a known model structure. If the true model is unknown, why rely on cumbersome models that require *a priori* specification of the spatial relationships among all observations and the manipulation of enormous matrices to account for omitted spatial effects?

Received: August 2010; revised: August 2011; accepted: August 2011.

DOI: 10.1111/j.1467-9787.2011.00746.x

Despite this generally negative view of spatial econometric models, they do have a time and place. First, they provide convenient model specification tests that indicate when a base model does not adequately account for spatial relationships. Second, they provide convenient robustness checks that can provide some confidence in crucial statistical results. Finally, they may be useful for an important class of models in which the primary objective is to estimate the causal relationship of neighboring values of the dependent variable on itself, although even this apparent advantage is questionable since the models are identified only by functional form restrictions—a point emphasized forcefully by Gibbons and Overman (forthcoming). As the functional form restrictions are unlikely to be true and better procedures are available for descriptive data analysis, standard spatial models should be viewed as an additional statistical tool rather than as the primary means of analyzing spatial data.

In this paper, I argue that standard spatial econometric models are simply another form of spatial smoothing. Although my emphasis will be the spatial AR model, the analysis here applies to the spatial error model also. I begin by establishing the close similarity between the spatial AR model and a semiparametric model whose nonparametric components include the data set's geographic coordinates. Next, I compare spatial AR and nonparametric estimates of population density functions using data on census tracts in Chicago. The results demonstrate that both the spatial AR model and locally weighted regression accomplish the same purpose: they reveal that the rate of decline in population density is not symmetric across the city. Despite being roughly equivalent to a locally weighted regression with a very small window size, the spatial AR model's predictions are less accurate than the nonparametric model. A series of Monte Carlo experiments also demonstrates the advantages of nonparametric models over parametric spatial models. The nonparametric model provides more accurate predictions of the dependent variable and more accurate marginal effect estimates than a spatial AR model when the population density gradient varies spatially within the sample area.

My emphasis throughout the paper is on an econometric model's accuracy in predicting the dependent variable and for estimating the marginal effects of exogenous explanatory variables. Gibbons and Overman (2011) argue persuasively that standard spatial econometric models are unlikely to identify "neighborhood effects" accurately regardless of whether the neighboring variable of interest is the dependent variable or an exogenous explanatory variable. But standard spatial econometric models are not the best approach even when the primary objective is descriptive data analysis or to control for spatial effects that may influence parameter estimates.

2. SPATIAL SMOOTHING

Some simple examples clearly demonstrate the similarity between the spatial AR or spatial error model and other forms of smoothing. Assume to start that the base model is $y_i = \mathbf{X}_\mathrm{i}\beta + \mathrm{u}_\mathrm{i}$ and that the spatial arrangement of the data is one-dimensional, with the observations ordered from smallest to largest according to a single variable, z_i. Spatial weight matrices are commonly based on a contiguity matrix with entries that equal one if observation i borders observations j and zero otherwise. To form the spatial weight matrix, the contiguity matrix is then row-standardized such that each row sums to one. In this simple one-dimensional example, the spatial weight matrix has the following form: $\mathbf{W}_{1,2} = \mathbf{W}_{n,n-1} = 1$; $\mathbf{W}_{i,i+1} = \mathbf{W}_{i,i-1} = 0.5$ for $2 \leq i \leq n{-}1$; and $\mathbf{W}_{ij} = 0$ otherwise. The spatial AR version of the base model would add $\mathbf{Wy}$ as an explanatory variable to the base model, while the spatial error model would write the error vector as $\mathbf{u} = \theta\mathbf{Wu} + \mathbf{e}$.

Linear interpolation is the simplest version of a smoothing estimator. Replacing each y_i with a value interpolated from its nearest neighbors produces the following estimate of y_i: $\hat{y}_i = [(z_{i+1} - z_i)y_{i-1} + (z_i - z_{i-1})y_{i+1}]/(z_{i+1} - z_{i-1})$. The only difference between this interpolation pattern and the spatial AR model is that **Wy** imposes equal weights of 0.5 on each of the neighboring values rather than placing more weight on the closer observation.

Interpolation is similar for two-dimensional spatial frameworks. Changing notation slightly, let $\mathbf{z}_i = (z_{1i}, z_{2i})$ represent the geographic coordinates for a given observation i and assume that $\mathbf{z}_i$ falls within a cell bounded by four other coordinate pairs, ν_0, ν_1, ν_2, and ν_3. Standard interpolation procedures use weighted averages of the values of y at each of the four coordinates to estimate the value at z_i, i.e., $\hat{y}_i = \psi_0 y(\nu_0) + \psi_1 y(\nu_1) + \psi_2 y(\nu_2) + \psi_3 y(\nu_3)$, where $\psi_0 + \psi_1 + \psi_2 + \psi_3 = 1$. Details can be found in Loader (1999, pp. 215–217). The point is that linear interpolation again produces a predicted value of y that is a weighted average of four neighboring values, with more weight placed on observations that are closer to the target point. This pattern corresponds to one of the most commonly used spatial weight patterns, a "rook" pattern of "first-order spatial contiguity." For data drawn from a regular lattice, this pattern places equal weights of 1/4 on values of **y** drawn from the points due north, south, east, and west of the target point, i.e, $W_{ij} = 1/4$ for these four points and $W_{ij} = 0$ otherwise. (Values may equal 1/2 or 1/3 at the borders, but the general pattern is the same.) Again, the only difference between the spatial weight variable, **Wy**, and linear interpolation is that the most commonly used patterns for **W** place equal weight on a small number of neighboring observations, whereas linear interpolation places more weight on closer data points.

One problem with these procedures is that using only a few data points to interpolate the value for a base observation can lead to "under-smoothing," i.e., a high variance. Kernel smoothers are one way to avoid this problem. The general form of a kernel smoother is

$$\hat{y}_i = \frac{\sum_{j=1}^{n} K\left(\frac{z_j - z_i}{h}\right) y_j}{\sum_{j=1}^{n} K\left(\frac{z_j - z_i}{h}\right)}$$

where **z** may represent either a single variable or a vector of geographic coordinates and h is the bandwidth. Kernel smoothers again use weighted averages of nearby observations to construct an estimate value for y_i, with weights that decline with distance. The same pattern of weights is sometimes used to construct spatial weight matrices. Whether interpreted as a kernel density regression or as a specification of the spatial weight matrix, the resulting smoother can be written as $\hat{\mathbf{y}} = \mathbf{Wy}$. Each value of **y** is estimated as a weighted average of nearby values of **y**.

Viewing the spatially lagged dependent variable, **Wy**, as just another form of linear smoothing leads to a much different view of the spatial AR model than is commonly adopted in the literature. Rather than viewing **Wy** as a direct causal effect of neighboring values of the dependent variable on **y**, the spatial lag variable can just as readily be viewed as a set of predicted values of the dependent variable. Either way, the spatial AR model is subject to a severe identification problem that is only overcome by imposing a simple structure on W, despite the fact that this structure is never known *a priori*.

Viewed in this light, the model becomes either a tautology or a paradox: how can predicted values of the dependent variable *not* prove to be statistically significant as an explanatory variable? One possibility is that location is not a good predictor of **y**. This is clearly not what most researchers have mind given Tobler's (1970) oft-quoted first law of geography—"everything is related to everything else, but near things are more related than distant things"—and the fact that spatial AR and spatial error models are typically employed because some factors related to location are not adequately captured by **X** alone. Also, as Pinkse and Slade (2010) emphasize, a model in which **Wy** is the intended regressor rather than the substitute for an expectation term with the form E(Y|•) is likely to be at least weakly identified. However, a model that is formally identified when correctly specified may still produce biased results when misspecified, and the spatial lag variable, **Wy**, is very likely to be correlated with the source of the misspecification.

One of the most vexing problems faced when analyzing spatial data is that, ultimately, virtually every variable is correlated with location. For example, house prices vary by location, but so do common explanatory variables in a hedonic price function such as lot size, structural characteristics, and neighborhood characteristics. Moreover, there may be good reasons to expect that the coefficients themselves vary over space. Thus, a model like $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$ could just as easily be written as a function of the geographic coordinates alone: $\mathbf{y} = \mathbf{X}(\mathbf{z}_1, \mathbf{z}_2)\beta(\mathbf{z}_1, \mathbf{z}_2) + \mathbf{u} = f(\mathbf{z}_1, \mathbf{z}_2) + \mathbf{u}$. If the function is specified correctly, it may be possible to estimate $\beta(\mathbf{z}_1, \mathbf{z}_2)$, precisely enough to account for how the coefficients vary over space. But the combination of missing variables and incorrect functional forms are likely to produce a model with the form $\mathbf{y} = \mathbf{X}_1(\mathbf{z}_1, \mathbf{z}_2)\beta_1(\mathbf{z}_1, \mathbf{z}_2) + (\mathbf{u} + \mathbf{X}_2(\mathbf{z}_1, \mathbf{z}_2)\beta_2(\mathbf{z}_1, \mathbf{z}_2))$, where $\mathbf{X}_1$ is a set of correctly specified variables and $\mathbf{X}_2$ is a form of model misspecification. Since $\mathbf{X}_2$ and β_2 are both likely to be functions of $\mathbf{z}_1$ and $\mathbf{z}_2$, smoothing over space by adding $\hat{\mathbf{y}} = \mathbf{Wy}$ adds a variable that is correlated with $\mathbf{X}_2(\mathbf{z}_1, \mathbf{z}_2)\beta_2(\mathbf{z}_1, \mathbf{z}_2)$ as an explanatory variable that will clearly add significant explanatory power. The fundamental point is the same for both the spatial lag and the spatial error model: model misspecification produces significant estimates of ρ or θ even when the true model has $\rho = \theta = 0$.

3. ESTIMATION PROCEURES FOR THE SPATIAL AR MODEL

When the spatial lag variable **Wy** is viewed as a form of spatial smoothing, it may seem surprising that its coefficient does not hover near unity in most applications. The log-likelihood function for this model is

$$(1)\quad \ln L = -\frac{n}{2}\ln(\pi) - \frac{n}{2}\ln(\sigma^2) - (\mathbf{y} - \rho\mathbf{Wy} - \mathbf{X}\beta)'(\mathbf{y} - \rho\mathbf{Wy} - \mathbf{X}\beta) + \sum_{i=1}^{n}\ln(1 - \rho\omega_i).$$

The last term is the log of the determinant of the Jacobian matrix, $|\partial\mathbf{u}/\partial\mathbf{y}|$, which Ord (1975) showed can be simplified to a function of the eigenvalues of **W**, ω_i. Apart from this term, equation (1) is simply the standard maximum likelihood version of a linear regression model with normally distributed, homoskedastic errors, which means that the model could be estimated by a simple regression of **y** on **X** and **Wy**. Since the Jacobian term reaches a peak at $\rho = 0$ and declines smoothly as $|\rho|$ approaches one, it serves as a penalty term, pulling the estimate back toward zero.

The GMM estimator of Kelejian and Robinson (1993) and Kelejian and Prucha (1999) replaces **Wy** with an instrumental variable created by regressing **Wy** on **X** and terms such as **WX** and $\mathbf{W}^2\mathbf{X}$. This approach also breaks the direct link between **Wy** and spatial smoothing. However, the GMM approach sometimes produces estimates of ρ that greatly exceed unity, which is often viewed as an upper bound for acceptability since **Wy** closely resembles a lagged dependent variable from a time series model.

The GMM estimator has significant advantages over the maximum likelihood version of the model. Though maximum likelihood provides consistent and efficient estimates when the model is correctly specified and the errors truly follow the assumed normal distribution, there is no reason to expect in practice that the errors are actually normally distributed or homoskedastic, and it certainly is not the case that the model is known to be correctly specified since the very reason for adopting the spatial model is that the base specification does not account adequately for spatial effects. Moreover, maximum likelihood estimation may require large sample sizes to be accurate, yet equation (1) requires at least some manipulation of large, nxn matrices.[1] GMM estimation does not require any distributional assumptions and it often turns out to be less sensitive to model misspecification.

4. ALTERNATIVE APPROACHES

Although I have extolled the advantages of nonparametric and semiparametric approaches before (McMillen, 2010), it is worth emphasizing some of the differences in modeling strategy between these approaches and the spatial lag and spatial error models. Spatial econometric models are fundamentally unidentified. There are $n(n-1)$ potential relationships among the observations, but only n data observations are available. The spatial lag and spatial error models attempt to solve this problem by imposing model structure at the start in the form of the spatial weight matrix. The $\rho\mathbf{Wy}$ or $\theta\mathbf{Wu}$ term reduces the impossible task of estimating $n(n-1)$ relationships to the estimation of a single parameter (or perhaps two) by imposing an extremely simple form on $\mathbf{W}$. Alternative specifications are sometimes tried for $\mathbf{W}$—first order versus second order contiguity, weights based on distance versus weights based on some other variables, etc.—and the results sometimes prove sensitive to the specification.

The core of the standard spatial approach is a classical statistical model in which the true model is known and specified beforehand; only the values of the parameters are unknown. In fact, the true model structure is unknown and it takes far more hubris to believe that an $n(n-1)$ set of spatial relationships can be specified correctly beforehand than to assume a functional form and a set of explanatory variables that might influence the dependent variable. Nonparametric and semiparametric approaches admit at the start that the true model structure is unknown. They can easily be applied to quite large data sets, and as McMillen and Redfearn (2010) emphasize, they are not necessarily profligate consumers of degrees of freedom. Most importantly, they are more apt to be viewed explicitly as a form of specification and robustness testing rather than as a direct estimate of a true underlying model form.

Consider the standard linear regression model $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$. After regressing $\mathbf{y}$ on $\mathbf{X}$, a spatial lag or spatial AR model might be used to determine whether the base model is adequate. If it turns out that the estimated value of ρ is statistically significant while θ turns out to be insignificant, there is a strong temptation to simply report the spatial AR results as the final of estimate of the true model. The additional explanatory variable $\mathbf{Wy}$ is viewed as controlling for omitted spatial effects while the estimated values of β are assumed to no longer be contaminated by spatial autocorrelation. Of course, more specifications of $\mathbf{W}$ might be tried and there may be additional tests of the model, but I believe this is a fair characterization of a common modeling strategy.

The first part of this approach is perfectly valid. If we are concerned that the base model does not adequately capture all spatial effects, there is no reason whatsoever not to use the spatial AR and spatial error models as tests of the underlying model specification. But the real question is whether we have obtained accurate estimates of the marginal effects of each explanatory variable on the dependent variable. Do these relationships vary over space, and are they sensitive to the underlying model specification? These questions are more readily answered with nonparametric methods.

Though fully nonparametric methods, such as the locally weighted regression procedure of Cleveland and Devlin (1988), are quite useful in models with few explanatory

[1] LeSage and Pace (2007) present a clever approach for avoiding the calculation of eigenvalues. Their approach makes the sample size a relatively insignificant issue.

variables, they are subject to a "curse of dimensionality" that produces high variances in larger models. Semiparametric approaches become quite useful in this context. The simplest form might supplement a base model specification with a function that controls for omitted spatial effects, i.e.,

$$\mathbf{Y} = \mathbf{X}\beta + f(\mathbf{z_1}, \mathbf{z_2}) + \mathbf{u}, \tag{2}$$

where, as before, $\mathbf{z}_1$ and $\mathbf{z}_2$ represent the geographic coordinates for each observation. The common approach of including fixed effects for census tracts or other districts is a special case of equation (2). A virtually universal finding is that including many fixed effects for location greatly reduces any evidence of spatial autocorrelation in the residuals. An alternative is to use Robinson's (1988) semiparametric approach to directly estimate f(•) using the geographic coordinates as explanatory variables. This approach implies a smooth surface for the f(•) and uses far fewer degrees of freedom than the fixed effects approach. The semiparametric approach also makes it easy to assess sensitivity of the results to the bandwidth, which is roughly analogous to employing fixed effects for various district sizes such as census tracts, blocks, block groups, etc. The guiding strategy is to test the model specification and to determine the robustness of the results rather than to impose a single model structure on the data *a priori*.

5. SPATIAL CAUSATION

So far I have argued that the primary role of a spatial AR or spatial lag model is as another tool for diagnostic and specification testing. There is one main exception to this relatively limited role for traditional spatial models—when the objective of the exercise is to directly test for the causal effect of neighboring values of the dependent variable on **y**.

Brueckner (2006) provides a general framework for a class of theoretical models of "spatial interaction" among local governments that lead directly to the spatial AR model for empirical implementation. In the first, the "spillover" model, each jurisdiction chooses the level of some variable *z*, which in turn generates a spillover that affects other jurisdictions. Changing Brueckner's notation slightly to avoid conflicts with symbols used here, the jurisdiction's objective function is $V(g_i, \mathbf{g}_{-i}; \mathbf{X}_i)$, where g_i is the decision variable, $\mathbf{g}_{-i}$ is the vector of g's for the other jurisdictions, and $\mathbf{X}_i$ is a vector of characteristics representing the jurisdiction's preferences over g. Under Nash behavior the jurisdiction takes g_{-i} as given when choosing g_i. The solution to the maximization problem implies a reaction function with the form $g_i = R(g_{-i}; \mathbf{X}_i)$. A linear version of this equation leads directly to the spatial AR model.

Brueckner (2006) also presents an alternative formation of the choice problem, the "resource-flow" model. In this model, a jurisdiction again chooses g_i, but it also cares about the amount of a "resource," s_i, whose level depends on the g choices of all jurisdictions. For example, the level of welfare benefits chosen by a state will depend on the number of poor households who live in the state, which in turn may depend on the level of welfare benefits provided by other states if low-income households are mobile. The objective function is written $V(g_i, s_i; \mathbf{X}_i)$, where in this case g_i represents welfare levels in state i, s_i is the number of poor households in the state, and $\mathbf{X}_i$ again is a vector of preferences. The number of poor households depends on the level of welfare payments in state i as well as the level in all other states, i.e., $s_i = H(g_i, g_{-i}; \mathbf{X}_i)$. Substituting this expression into the objective function implies that the state is once again choosing g_i to maximize an objective function with the form $V(g_i, g_{-i}; \mathbf{X}_i)$.

The goal of either the spatial interaction model or the resource-flow model is to estimate the reaction function, i.e., the effect of g_i on g_{-i}. Since each jurisdiction's choice of g_i depends on all $n-1$ other jurisdiction's choice of g, there clearly is no hope of identifying

TABLE 1: Spline and Spatial AR Population Density Estimates

	OLS Spline	Spatial AR (1)	Spatial AR (2)	Spatial AR (2)
Constant	8.5895	4.8648	4.9194	3.7074
	(0.1545)	(0.3041)	(0.3239)	(0.4173)
Distance to CBD	0.4687	−0.0384	0.2530	0.1814
(*DCBD*)	(0.0572)	(0.0041)	(0.0560)	(0.0578)
$DCBD^2$	−0.0567		−0.0308	−0.0222
	(0.0060)		(0.0060)	(0.0063)
$DCBD^3$	0.0016		0.0009	0.0006
	(0.0002)		(0.0002)	(0.0002)
$(DCBD - 15.119)^3$	−0.0023		−0.0013	−0.0009
$\times I(DCBD > 15.119)$	(0.0004)		(0.0004)	(0.0004)
Ln(House Age)				0.3334
				(0.0736)
ρ		0.5132	0.4333	0.4218
		(0.0305)	(0.0343)	(0.0344)
R^2, reduced form	0.3328	0.2657	0.3317	0.3477
R^2, structural		0.4129	0.4185	0.4263

Notes. The data set comprises 1,313 census tracts in Cook County, Illinois. The dependent variable is the natural log of population density per square mile. Population data are drawn from the 2000 U.S. Census. The average age of homes in a census tract is calculated from the Cook County Assessor's assessment file. Tracts with no residents or data on homes were excluded from the analysis.

the model without imposing structure. The spatial AR model is the logical choice for these classes of models. However, it is important to recognize that the models arise from a rigorous theoretical foundation, and that the primary objective is to estimate the effect of g_i on g_{-i}. It is reasonable to adopt a highly structured model when its form is predicted by theory, which is a point emphasized by Corrado and Fingleton (2011). However, it is not reasonable to impose restrictive model structure in more descriptive empirical exercises.[2] Or, as Pinkse and Slade (2010) state when also arguing that applications of spatial models should be guided by theory, "spatial econometric theory should be inspired by actual empirical applications as opposed to being directed by what appears to be the most obvious extension of what is currently available" (p. 103).

6. AN EMPIRICAL EXAMPLE

A simple population density function illustrates the role of a spatial AR model as a spatial smoother. The data set comprises 1,313 census tracts in Cook County, Illinois. The primary dependent variable is the natural log of gross population density, where density is defined as the number of people per square mile of total land area within a census tract. Population data are drawn from the 2000 U.S. Census. Distance is measured as straight-line miles from the census tract centroid to the traditional city center of Chicago at the intersection of State and Madison Streets, in the heart of the Chicago central business district (CBD). The spatial weight matrix is calculated by row-standardizing a first-order spatial contiguity matrix using a "queen" criterion.

Table 1 compares cubic spline estimates (Suits, Mason, and Chan, 1978; McMillen, 2006) to various spatial AR specifications. The spatial AR models are estimated by maximum likelihood estimates, and thus are referred to as "SARML" models. Although it is

[2] Examples of studies that estimated reaction functions include Brueckner (1998), Figlio et al. (1999), Frederiksson and Millimet (2002), Gerard et al. (2010), and Revelli (2003).

not obvious from glancing at the coefficient estimates, each of the models implies that gross population density declines with distance from the CBD, with the exception of an initial upward-sloping portion in the area near the CBD that is dominated by commercial land uses. A base linear specification for the spatial AR model produces a high estimate of 0.5132 for ρ. As expected given my claim that spatial AR models help to compensate for functional form misspecification, the estimated value of ρ falls significantly to 0.4333 when a cubic spline function is specified for $X\beta$. However, adding a potential missing variable—the average age of the housing stock for each census tract—has little effect on the estimated value of ρ despite being highly statistically significant.[3] In the remainder of this section, I will explore some of the implications and explanations for these findings.

First, note that there are several ways that we might evaluate the predictive power of the model (Cressie, 1993), and thus several ways to construct pseudo-R^2 measures. I define the pseudo-R^2s as the square of the correlation between the predicted and actual values of y. For the predicted values, I use (1) the reduced form measure, $\hat{\mathbf{y}} = (\mathbf{I} - \hat{\rho}\mathbf{W})^{-1}\mathbf{X}\hat{\beta}$; and (2) the underlying structural equation, $\hat{\mathbf{y}} = \hat{\rho}\mathbf{Wy} + \mathbf{X}\hat{\beta}$. These R^2 measures represent two extremes. The reduced form measure does not directly take advantage of any information on $\mathbf{Wy}$, while the structural equation measure treats $\mathbf{Wy}$ the same as any other explanatory variable for the purpose of calculating the goodness of fit measure. The R^2s show that the reduced form version of the model adds virtually no explanatory power relative to a simple spline function: $\mathbf{X}\hat{\beta}$ from the cubic spline and $(\mathbf{I} - \hat{\rho}\mathbf{W})^{-1}\mathbf{X}\hat{\beta}$ from the spatial AR model provide virtually the same explanatory power when $\mathbf{X}$ is the same. The additional explanatory power of the spatial AR model comes entirely from adding the spatial smoother, $\mathbf{Wy}$, as an explanatory variable.

This point is also seen clearly in Figure 1, which compares the predicted values from the spline function to the values of $(\mathbf{I} - \hat{\rho}\mathbf{W})^{-1}\mathbf{X}\hat{\beta}$ for the linear and cubic spline specifications of the spatial AR model. Since $(\mathbf{I} - \hat{\rho}\mathbf{W})^{-1}\mathbf{X}\hat{\beta} \approx (\mathbf{I} + \hat{\rho}\mathbf{W} + \hat{\rho}^2\mathbf{W}^2 + \hat{\rho}^3\mathbf{W}^3 + \ldots)\mathbf{X}\hat{\beta}$, it might be expected that allowing ρ to differ from zero would add significant explanatory power to the equation. In practice, the additional terms simply allow for some small local variation in the predictions around a base trend line that is estimated well by standard ordinary least squares (OLS) methods.

Although average housing age is clearly a statistically significant variable, it does little to alter the estimated values of ρ. This result might seem at odds with my claim that spatial AR models serve in part to compensate for the effects of missing explanatory variables that are correlated over space. In this application, ln(Age) follows nearly the same spatial trend as ln(density), so it does little to explain local variations around the overall trend in ln(density). Figure 2 shows that ln(Age) is explained well by distance from the CBD; apart from some rebuilding that has taken place near the city center, house ages clearly fall fairly smoothly with distance. Figure 2 helps to explain why it can be hard to separate the effects of individual variables in cross-sectional spatial models: key explanatory variables are nearly all functions of the underlying geographic coordinates.

Locally weighted regression (LWR) procedures are explicit spatial smoothers. The approaches commonly used for spatial modeling are generally derived from Cleveland and Devlin (1988). LWR provides estimates of $\mathbf{y}$ at a set of target points using a series of weighted least squares regressions, with more weight given to observations closer to the target points. The bandwidth or window size determines how rapidly the weights decline with distance. The spatial AR model is very similar to an LWR model with a very small window size. In the population density example, the average of the nonzero

[3] The age variable is added to the model because population densities are predicted to increase with the age of the housing stock in vintage urban growth models (e.g., Brueckner, 1981).

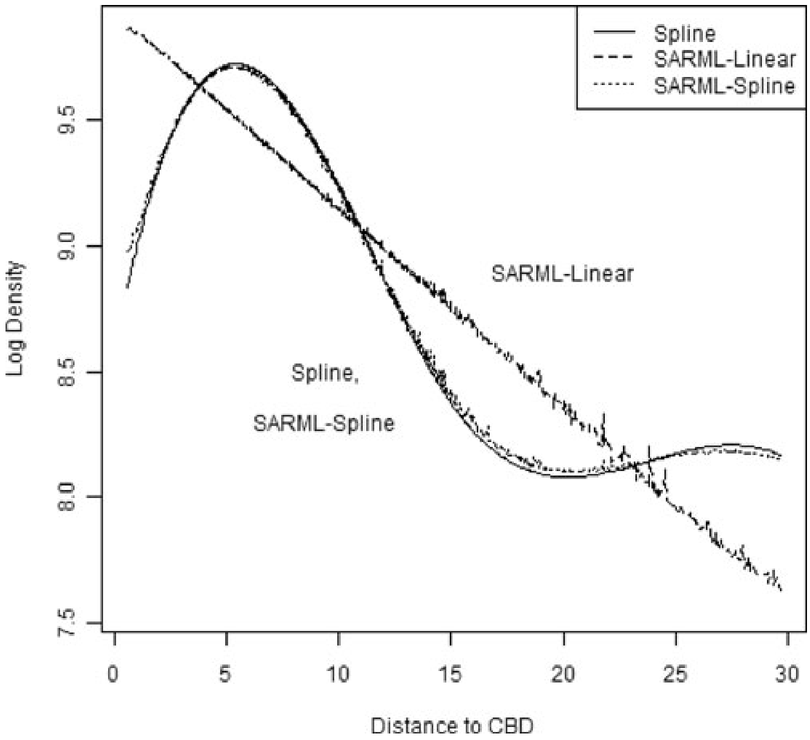

FIGURE 1: Log Population Density Estimates.

entries of **W** is 6.64, which means that the average tract is contiguous to fewer than seven other tracts, with a range of 2–14. With 1,313 observations, **Wy** is roughly equivalent to a nonparametric model with a uniform kernel of 6.64/1313 = 0.506 percent. In contrast, most applications of LWR models use windows in the 10–30 percent range.

To show the link between LWR smoothing and the spatial AR model, I estimate spatial AR models using the predicted values from conditionally parametric (CPAR) versions of the LWR model as the sole explanatory variable. The base CPAR version of the LWR model can be written ln(density) = Xβ(latitude, longitude) + u, where X includes a constant, distance from the CBD, and ln(Age). The model is estimated by a series of weighted least squares regressions of ln(density) on distance from the CBD and ln(Age), with weights defined using a tri-cube function based on the straight-line geographic distance between the target point and the other data points. I vary the window sizes from 5 percent to 30 percent; higher values lead to more smoothing of the dependent variable.

Table 2 shows the estimated values of ρ from spatial AR models with the LWR predictions as the sole explanatory variable. As expected, the estimated value of ρ declines as the window size falls. When the window size is 5 percent (meaning that about 65 of the observations are included in each LWR regression), the estimated value of ρ is effectively zero. No spatial autocorrelation is present in an LWR model with a very small window size. Either approach—nonparametric regression or the spatial AR model—accomplishes a similar objective of smoothing the data over space.

Table 2 also presents the generalized cross-validation (GCV) statistic. As discussed in Loader (1999) or McMillen and Redfearn (2010), the GCV statistic is a convenient way of choosing the appropriate window size for an LWR model. The 5 percent window

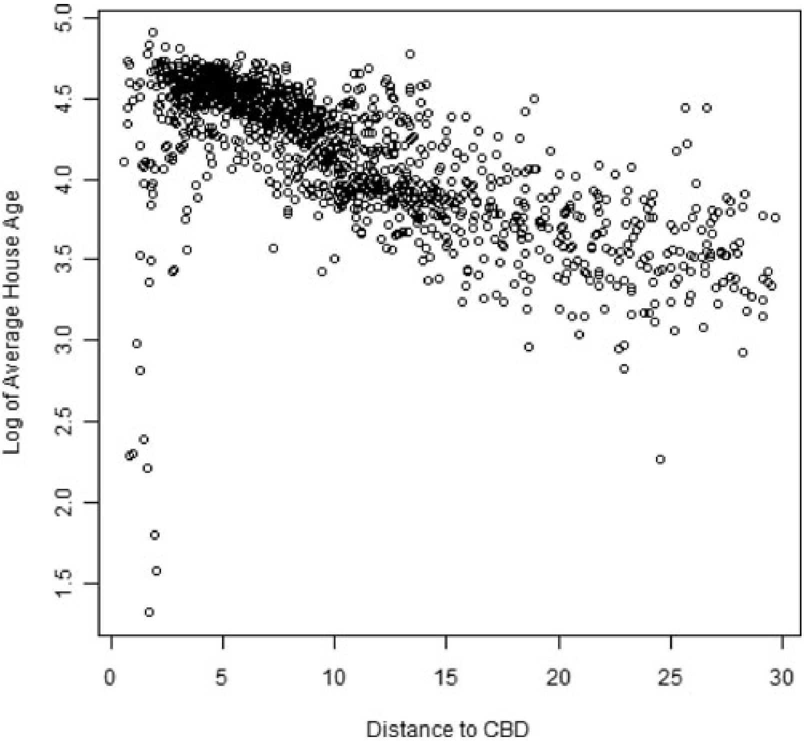

FIGURE 2: Log of Average House Age by Distance from the CBD.

TABLE 2: SARML Estimates of ρ Using CPAR Predictions as the Explanatory Variable

Window Size	GCV for CPAR Model	ρ	Z-Value for ρ
0.05	0.563	0.014	0.345
0.10	0.616	0.199	5.052
0.15	0.638	0.254	6.577
0.20	0.648	0.285	7.499
0.25	0.653	0.304	8.078
0.30	0.659	0.319	8.581

Notes. The CPAR model is ln(Population Density) = Xβ(latitude, longitude) + u, where X includes a constant, distance from the CBD, and ln(AGE). Predictions from the CPAR model are then used as the sole explanatory variable in the spatial AR model for population density.

provides the best tradeoff between bias and variance among the window sizes considered in Table 2.

Figures 3–5 compare the predictions of the spatial AR model to the conditionally parametric LWR model with the 5 percent window. The spatial AR predictions in Figure 3 are based on reduced-form estimates, $\hat{\mathbf{y}} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\hat{\beta}$, while Figure 4 shows the structural estimates, $\hat{\mathbf{y}} = \hat{\rho}\mathbf{W}\mathbf{y} + \mathbf{X}\hat{\beta}$. The LWR estimates provide a much higher pseudo-R^2—0.5405—than either spatial AR model, even though the 5 percent window size is much higher (though small by LWR standards) than the implicit window for **Wy**. Figure 3 shows that the reduced form spatial AR predictions look much like the concentric rings of a standard urban economics textbook. Figure 4 shows that the linear smoother, **Wy**, detects areas with high densities north of the CBD, along the lakefront. The LWR

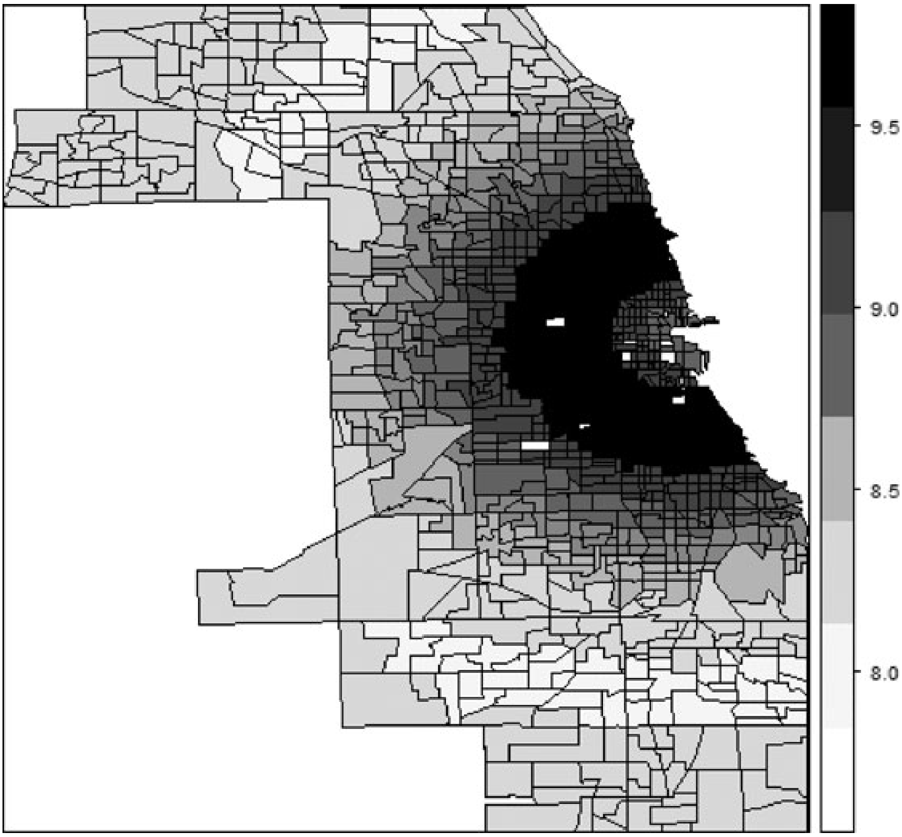

FIGURE 3: Reduced Form SARML Estimates.

estimates, shown in Figure 5, are similar to the structural spatial AR estimates, but the high-density area extends a bit farther inland. The north side of Chicago has much higher land values than the south side, and as urban spatial theory implies, higher land values are translated into higher population densities.

Either approach is a form of spatial smoothing; LWR is just better at it. The models show that the base cubic spline misses an important feature of Chicago's spatial structure; namely, the tendency for population density to be higher near Lake Michigan and on the Near North Side than in comparable areas on the South Side. This slight extension of the monocentric city model to recognize that the city is not completely symmetric is something that either smoothing approach is capable of revealing. The spatial AR model is not the *right* model; it is a way of revealing that a simple, symmetric model provides an incomplete picture of Chicago's spatial structure. The LWR model accomplishes the same purpose, but it provides more flexibility in the way that window sizes can be specified, and it is not derived from a maintained hypothesis of an correct functional form.

7. MONTE CARLO

The population density function is an example of a situation in which the spatial AR model provides a better fit than a simple functional form that does not take into account the spatial variation in the marginal effects of the explanatory variable. The empirical

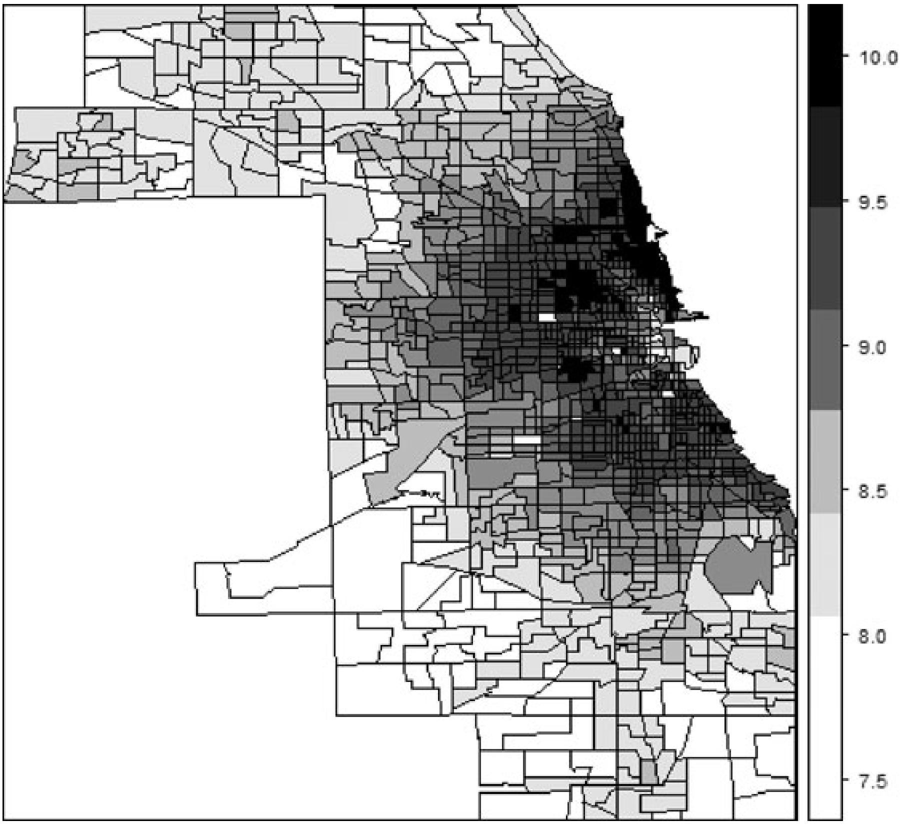

FIGURE 4: Structural SARML Estimates.

example demonstrates that a nonparametric approach is likely to be a better way of accounting for spatial variation in the functional form. In this section, I present a series of Monte Carlo experiments that illustrate the advantages of nonparametric modeling in a spatial setting.

The population density example serves as the basis for the Monte Carlo experiments. The base model is $y_i = 20 - DCBD_i + u_i$, where $DCBD_i$ is the straight-line distance between Chicago' CBD and the centroid of census tract i. I limit the sample to 861 tracts within the City of Chicago. As the base model can be estimated accurately by a simple regression of the dependent variable on distance from the CBD, the estimated value of ρ in a spatial AR model should be close to zero, and CPAR estimates should produce coefficients for *DCBD* that are close to −1 at all locations.

The coefficients for *DCBD* vary by location in the alternative model. To place the spatial AR and CPAR models on equal footing, the alternative model is $y_i = 20 - (1 - p) * DCBD_i + p * THETA_\mathrm{i} + u_i$, where p is a parameter and $THETA_i$ is the angle between observation i and the CBD. In this formulation of the model, which follows Cameron (2006), the CBD gradient is constant along each ray from the CBD. Although this model directly implies a CPAR model in which $y_i = \beta_0(THETA_i) + \beta_1(THETA_i)DCBD_i + u_i$, I estimate the model with longitude and latitude in place of the polar coordinates. Thus, both the CPAR and the spatial AR model are partially misspecified when $p > 0$. In the Monte Carlo experiments, p varies from 0 to 1 in increments of 0.25. The base OLS model will fit

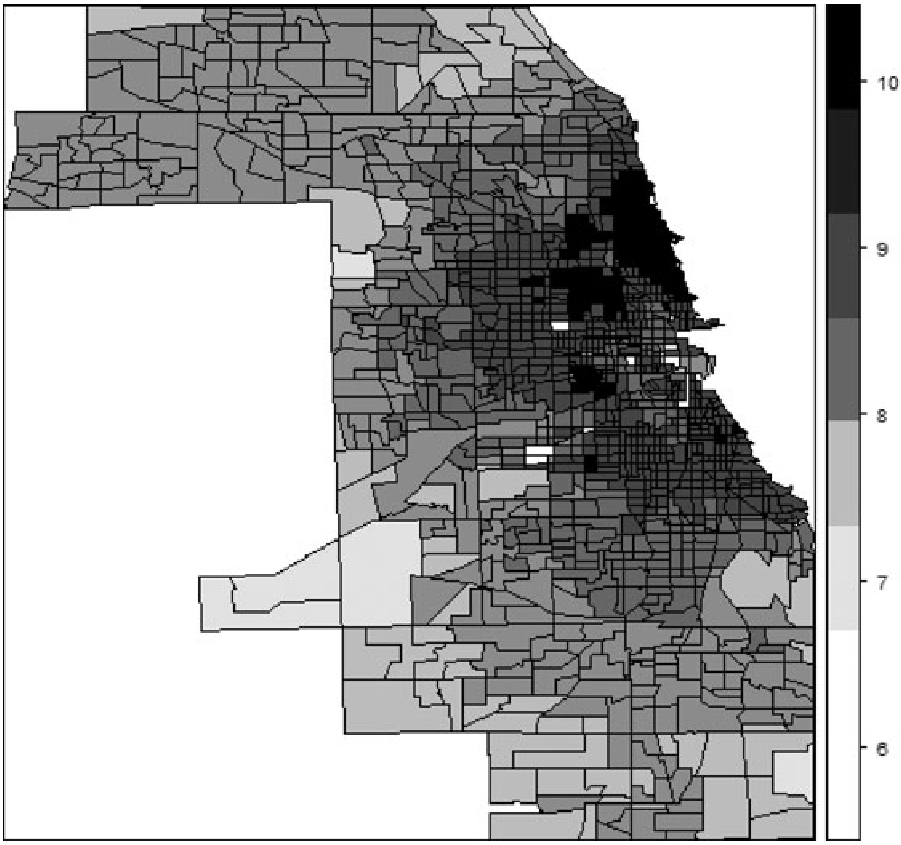

FIGURE 5: Conditionally Parametric LWR Estimates.

the data progressively worse as p increases and the spatial variation in the coefficients becomes more pronounced. I set the variance of the error term to $\sigma^2 = \mathrm{var}(DCBD)/3$, which implies an R^2 of approximately 0.75 for the OLS model. For each value of p, I conduct 1,000 experiments by drawing from a $N(0, \sigma^2)$ distribution for the vector of errors.

Following the series of Monte Carlo Experiments, I calculate root mean squared errors for the predicted values of y and the marginal effect of $DCBD$. The two measures are:

$$RMSE(\hat{y}) = \sqrt{\frac{1}{1000}\sum_{j=1}^{1000}\left\{\frac{1}{n}(\hat{y}_i - 20 + (1-p) * DCBD_i - p * THETA_i)^2\right\}}$$

and

$$RMSE\left(\frac{\partial \hat{y}}{\partial DCBD}\right) = \sqrt{\frac{1}{1000}\sum_{j=1}^{1000}\left\{\frac{1}{n}\left(-\frac{\partial \hat{y}_i}{\partial DCBD_i} - (1-p) * DCBD_i - p * THETA_i\right)^2\right\}}$$

Figures 6 and 7 show the implied values of y for models without errors in the two extreme cases of $p = 0$ and $p = 1$. The only difference between the two cases is that y tails off more rapidly to the west of the peak area at the CBD than to the north or south when $p > 0$.

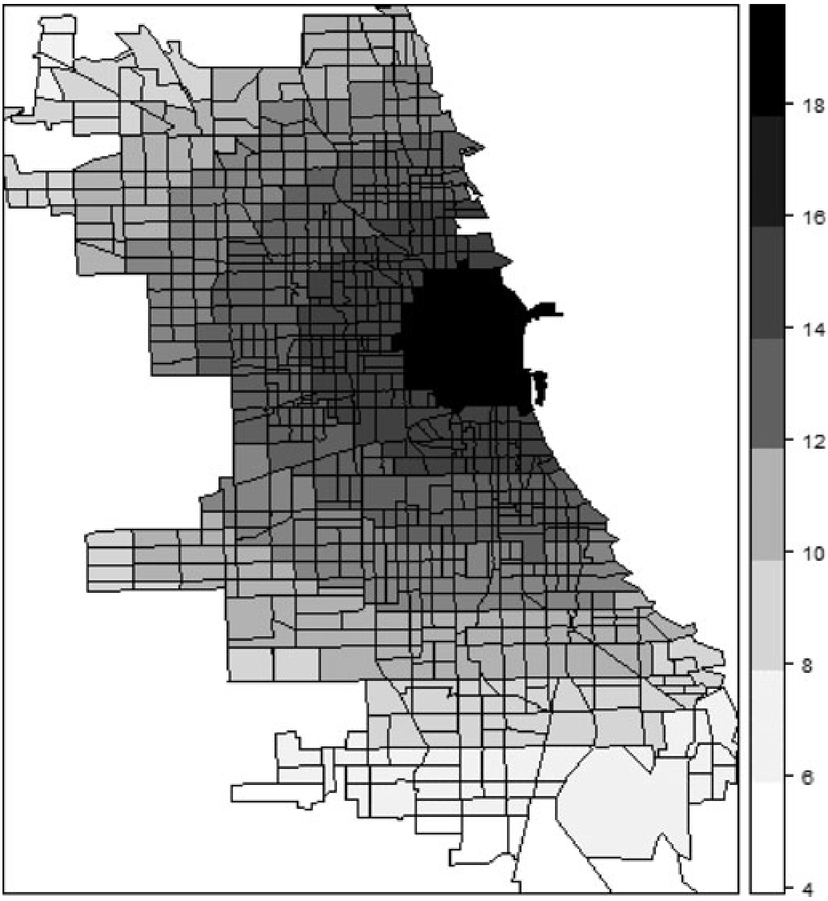

FIGURE 6: Base Monte Carlo Model ($p = 0$).

Thus, the Monte Carlo experiments represent a realistic comparison of a symmetric city to a spatial arrangement in which the distance gradient is lower along the Lakefront.

Tables 3 and 4 present the *RMSE*'s for the predicted values of *y* and the marginal effect of *DCBD*. A correctly specified OLS model would have *DCBD* and *THETA* as explanatory variables. Naturally, this specification has the lowest *RMSE* in every case, with values hovering around 0.115 for the predictions and around 0.026 for the marginal effects. Except when $p = 0$, the *RMSE*'s are much larger for a misspecified OLS model in which *DCBD* is the sole explanatory variable.

The *RMSE* for the spatial AR model's dependent variable predictions are comparable to OLS when $p = 0$, although the *RMSE* for the marginal effects are much higher than the OLS estimates for all values of *p*. The spatial AR model accounts for the spatial variation in the *DCBD* gradient through progressively higher values of $\hat{\rho}$ as *p* increases. The average value of $\hat{\rho}$ across the Monte Carlo experiments rises from approximately zero when $p = 0$ to 0.681 when $p = 1$. The *RMSE*'s rise also, and the *RMSE* for the marginal effects are more than twice the size of even the misspecified version of the OLS model when $p = 1$. It is not clear that much is gained by using a misspecified spatial AR model to account for omitted spatial effects, even when compared with a misspecified OLS model.

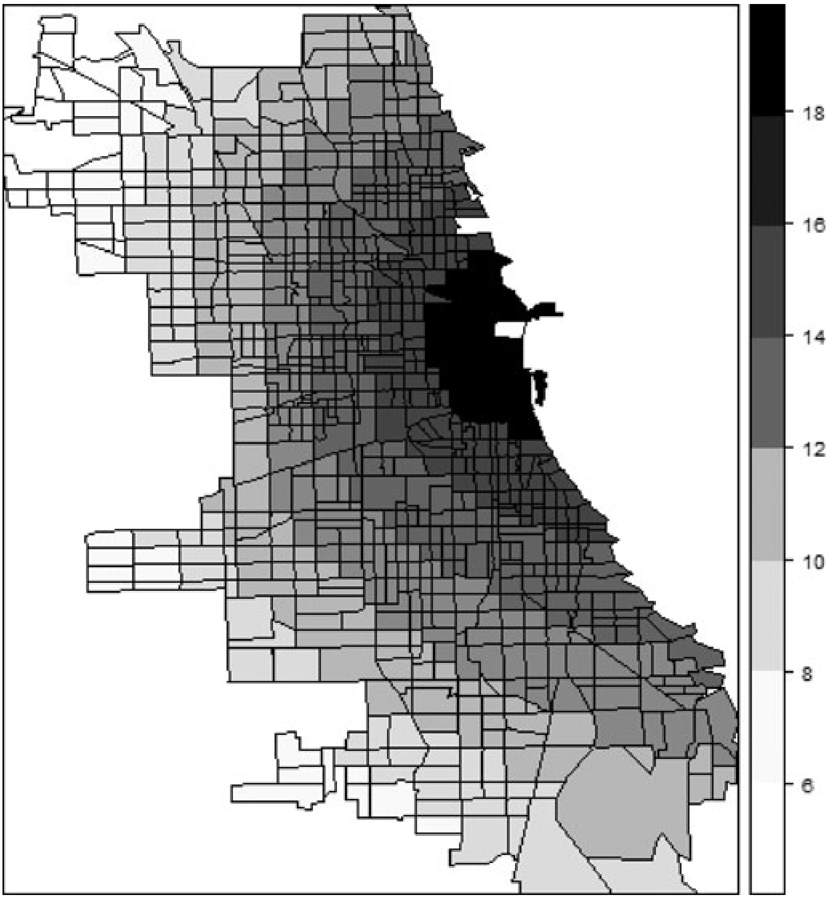

FIGURE 7: Monte Carlo Model with Spatial Variation in *DCBD* Coefficients ($p = 1$).

The CPAR model is also partially misspecified when the CBD gradient is a function of latitude and longitude rather than *THETA*. Nonetheless, the *RMSE* for both the predictions and the marginal effects are much lower than the spatial AR estimates when $p > 0.25$. The *RMSE*'s vary with the window size. For the dependent variable predictions, a window size of 25 percent tends to produce the lowest *RMSE*. Optimal window sizes are typically larger when the objective is to estimate marginal effects: in this series of Monte Carlo experiments, a window size of 50 percent tends to produce the lowest *RMSE* for the estimated marginal effects. When $p = 1$, the *RMSE* for the misspecified CPAR model's marginal effects is 0.123 when the window size is 50 percent, which compares with 0.846 and 0.816 for the two versions of the spatial AR estimates and 0.374 for the misspecified OLS model. Naturally, the *RMSE* for both the predications and the marginal effects are lower still when the CPAR is correctly specified so that the estimated coefficients vary with *THETA* rather than latitude and longitude.

Although the *RMSE* for the CPAR models vary by window size, they are consistently lower than the *RMSE* for the spatial AR model except when $p = 0$. Choosing the window size for a CPAR model is analogous to specifying the spatial weight matrix in a spatial AR model. An important difference, though, is that the literature on bandwidth and window size selection is enormous, whereas spatial weight matrices are much more likely to be

TABLE 3: Monte Carlo Root Mean Squared Errors for $\hat{y}$

p	0	0.25	0.50	0.75	1
OLS					
OLS, *DCBD,* and *THETA*	0.115	0.114	0.115	0.116	0.113
OLS, *DCBD* Only	0.076	0.445	0.877	1.312	1.747
Spatial AR Model with $X = (1, DCBD)$					
Mean $\hat{\rho}$ and	−0.014	0.140	0.387	0.566	0.681
standard deviation	(0.061)	(0.058)	(0.041)	(0.029)	(0.020)
$\hat{\rho} Wy + X\hat{\beta}$	0.104	0.407	0.639	0.687	0.761
$(1 - \hat{\rho} W)^{-1} X\hat{\beta}$	0.091	0.448	0.884	1.333	1.774
Conditionally Parametric Model: $y = X\beta(lo, la) + u$					
$h = 0.10$	0.289	0.292	0.306	0.327	0.353
$h = 0.25$	0.192	0.214	0.270	0.345	0.428
$h = 0.50$	0.136	0.218	0.363	0.523	0.686
$h = 0.75$	0.108	0.283	0.530	0.785	1.042
Conditionally Parametric Model: $y = X\beta(THETA) + u$					
$h = 0.10$	0.314	0.316	0.317	0.317	0.317
$h = 0.25$	0.195	0.201	0.213	0.235	0.260
$h = 0.50$	0.133	0.172	0.252	0.348	0.448
$h = 0.75$	0.108	0.201	0.353	0.514	0.679

Notes. The true model is $y_i = 20 - (1 - p) * DCBD_i + p * DCBD_i * THETA_i + u_i$, where *DCBD* represents distance from the CBD and $THETA_i$ is the angle between location i and the CBD. *DCBD* is the sole explanatory variable for the spatial AR model. All entries in the table represent root mean squared errors with the exception of the row labeled "*," which presents the means and standard deviations for $\hat{\rho}$ across 1,000 Monte Carlo estimates of the spatial AR model.

TABLE 4: Monte Carlo Root Mean Squared Errors for $\partial\hat{y}/\partial DCBD$

p	0	0.25	0.50	0.75	1
OLS					
OLS, *DCBD,* and *THETA*	0.026	0.026	0.027	0.027	0.026
OLS, *DCBD* Only	0.016	0.094	0.188	0.280	0.374
Spatial AR Model with $X = (1, DCBD)$					
$\hat{\beta}_{DCBD}$	0.050	0.206	0.492	0.701	0.846
$(1 - \hat{\rho} W)^{-1} \hat{\beta}_{DCBD}$	0.050	0.204	0.477	0.675	0.816
Conditionally Parametric Model: $y = X\beta(lo, la) + u$					
$h = 0.10$	0.219	0.218	0.223	0.226	0.232
$h = 0.25$	0.092	0.095	0.102	0.114	0.127
$h = 0.50$	0.046	0.055	0.074	0.098	0.123
$h = 0.75$	0.028	0.061	0.113	0.167	0.220
Conditionally Parametric Model: $y = X\beta(THETA) + u$					
$h = 0.10$	0.083	0.084	0.084	0.085	0.085
$h = 0.25$	0.049	0.051	0.053	0.059	0.065
$h = 0.50$	0.031	0.040	0.061	0.084	0.108
$h = 0.75$	0.025	0.041	0.073	0.107	0.140

Notes. The true model is $y_i = 20 - (1 - p) * DCBD_i + p * DCBD_i * THETA_i + u_i$, where *DCBD* represents distance from the CBD and $THETA_i$ is the angle between location i and the CBD. *DCBD* is the sole explanatory variable for the spatial AR model.

held as a maintained hypothesis. The commonly used first-order contiguity matrix used here is roughly equivalent to a very narrow window size. Nevertheless, the CPAR model outperforms the spatial AR model in the realistic case in which both models are partially misspecified.

8. CONCLUSION

Standard spatial econometric models have become over-used. Though they play a useful role in detecting various forms of model misspecification, they are apt to be viewed as the *correct* parametric form for a model when, in fact, they are simply a convenient way to control for unknown sources of spatial clustering among model residuals or the dependent variable. Using a spatial weight matrix to form weighted averages of a variable is nothing more than a form of linear smoothing with (typically) an unusually narrow bandwidth. Though clever procedures have been proposed for manipulating or simply avoiding working with the large matrices required by standard spatial models, researchers seldom truly believe in the model structure that made them necessary in the first place. If the true model structure is unknown and the primary objective is not to estimate a causal effect of $\mathbf{Wy}$ on $\mathbf{y}$, there is no compelling reason for choosing spatial AR or error models over other forms of spatial smoothing; indeed, the choice is likely to cause as much harm as good.

My emphasis in this paper has been on descriptive data analysis and models in which the primary objective is to estimate the marginal effect of an exogenous explanatory variable. Corrado and Fingleton (2011) argue that a spatial weight matrix must be assumed whether implicitly or explicitly when estimating spatial interaction models, and suggest that theory can be used to guide the specification of weight matrix. In practice, it is common to simply impose the spatial weight matrix as a maintained hypothesis, and to use simple specifications whose appeal lies more in the frequency of their use rather than their correspondence with economic theory. Since results can vary substantially depending on the spatial weight matrix, some humility is in order: adopt several weight matrices and employ alternative estimation methods to determine whether the results are robust. I have advocated nonparametric methods as an alternative because they encourage the researcher to vary window sizes and assess the accuracy of a base model structure. Spatial econometric models as well as nonparametric methods should be viewed as sets of tools for diagnostic and specification testing rather than as a literal description of a true underlying model.

REFERENCES

Brueckner, Jan K. 1981. "Testing a Vintage Model of Urban Growth," *Journal of Regional Science*, 21, 23–35.

———. 1998. "Testing for Strategic Interaction among Local Governments: The Case of Growth Controls," *Journal of Urban Economics*, 44, 438–467.

———. 2006. "Strategic Interactions among Governments," in Richard J. Arnott and Daniel P. McMillen (eds.), *A Companion to Urban Economics*. Malden, MA: Blackwell, 332–347.

Cameron, Trudy A. 2006. "Directional Heterogeneity in Distance Profiles in Hedonic Property Value Models," *Journal of Environmental Economics and Management*, 51, 26–45.

Cleveland, William S. and Susan J. Devlin. 1988. "Locally Weighted Regression Analysis by Local Fitting," *Journal of the American Statistical Association*, 83, 596–610.

Corrado, Luisa and Bernard Fingleton. 2011. "Where is the Economics in Spatial Econometrics?," *Journal of Regional Science*. DOI: 10.1111/j.1467-9787.2011.00726.x

Cressie, Noel A. 1993. *Statistics for Spatial Data*, New York: Wiley.

Figlio, David N., Van W. Kolpin, and William E. Reid. 1999. "Do States Play Welfare Games?," *Journal of Urban Economics*, 46, 437–454.

Fredriksson, Per G. and Daniel L. Millimet. 2002. "Strategic Interaction and the Determination of Environmental Policy across U.S. States," *Journal of Urban Economics*, 51, 101–122.

Gérard, Marcel, Hubert Jayet, and Sonia Paty. 2010. “Tax Interactions among Belgian Municipalities: Do Interregional Differences Matter?,” *Regional Science and Urban Economics*, 40, 336–342.

Gibbons, Steve and Henry Overman. Forthcoming. “Mostly Pointless Econometrics?,” *Journal of Regional Science.*

Kelejian, Harry H. and Dennis P. Robinson. 1993. “A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a County Expenditure Model,” *Papers in Regional Science*, 72, 297–312.

Kelejian, Harry H. and Ingmar Prucha. 1999. “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model,” *International Economic Review*, 40, 509–533.

LeSage, James P. and R. Kelley Pace. 2007. “A Matrix Exponential Spatial Specification,” *Journal of Econometrics*, 140, 190–214.

Loader, Clive. 1999. *Local Regression and Likelihood*. New York: Springer.

McMillen, Daniel P. 2006. “Testing for Monocentricty,” in Richard J. Arnott and Daniel P. McMillen (eds.), *A Companion to Urban Economics*. Malden, MA: Blackwell, pp. 128–140.

McMillen, Daniel P. 2010. “Issues in Spatial Data Analysis,” *Journal of Regional Science*, 50, 119–141.

McMillen, Daniel P. and Christian Redfearn. 2010. “Estimation and Hypothesis Testing for Nonparametric Hedonic House Price Functions,” *Journal of Regional Science*, 50, 712–733.

Ord, J. Keith. 1975. “Estimation Methods for Models of Spatial Interaction,” *Journal of the American Statistical Association*, 70, 120–126.

Pinkse, Joris and Margaret E. Slade. 2010. “The Future of Spatial Econometrics,” *Journal of Regional Science*, 50, 103–117.

Revelli, Federico. 2003. “Reaction or Interaction? Spatial Process Identification in Multi-Tiered Government Structures,” *Journal of Urban Economics*, 53, 29–53.

Robinson, Peter M. 1988. “Root-N-Consistent Semiparametric Regression,” *Econometrica*, 56, 931–954.

Suits, Daniel B., Andrew Mason, and Louis Chan. 1978. “Spline Functions Fitted by Standard Regression Methods,” *Review of Economics and Statistics*, 60, 132–139.

Tobler, Waldo. 1970. “A Computer Movie Simulating Urban Growth in the Detroit Region,” *Economic Geography*, 46, 234–240.