Article

The Biggest Myth in Spatial Econometrics

James P. LeSage [1],* and R. Kelley Pace [2]

[1] Department of Finance and Economics, McCoy College of Business Administration,
   Texas State University, 601 University Drive, San Marcos, TX 78666, USA

[2] Department of Finance, E.J. Ourso College of Business Administration, Louisiana State University,
   Baton Rouge, LA 70803, USA; E-Mail: kelley@pace.am

* Author to whom correspondence should be addressed; E-Mail: james.lesage@txstate.edu;
   Tel.: +1-512-245-0256.

External Editor: Kerry Patterson

Abstract: There is near universal agreement that estimates and inferences from spatial regression models are sensitive to particular specifications used for the spatial weight structure in these models. We find little theoretical basis for this commonly held belief, if estimates and inferences are based on the true partial derivatives for a well-specified spatial regression model. We conclude that this myth may have arisen from past applied work that incorrectly interpreted the model coefficients *as if* they were partial derivatives, or from use of misspecified models.

Keywords: indirect effects; spatial regression estimates; sensitivity to spatial weights

JEL classifications: C18, C21

1. Introduction

It has become a near universal criticism of spatial regression models that estimates and inferences are sensitive to the spatial weight matrix used in the model. For concreteness in our discussion we let a spatial regression (SAR) model take the form:

$$
\begin{aligned}
y &= \alpha \iota_n + \rho W y + X\beta + \varepsilon \\
\varepsilon &\sim N(0, \sigma^2 I_n)
\end{aligned}
\tag{1}
$$

where y is an $n \times 1$ vector of observations on the dependent variable and X is an $n \times k$ matrix of observations on the explanatory variables. Each of these observations on the dependent and explanatory variables comes from regions or points in space. Also, β is a $k \times 1$ vector of parameters associated with the explanatory variables, ι_n is a vector of ones, α is the associated intercept parameter, and ρ is the scalar dependence parameter (commonly in the interval $[0, 1]$).[1] The n disturbances ε are distributed normally with constant variance σ^2 and zero covariance across observations. The matrix W is the spatial weight matrix that contains non-zero elements w_{ij} if observations j and i are *neighbors* and zero otherwise. Typically, the matrix W is row-stochastic, so the $n \times 1$ *spatial lag vector* Wy contains values constructed from an average of *neighboring observations*.[2] A more elaborate variant of the SAR model, labeled the spatial Durbin model (SDM) in the literature, replaces the matrix X with $\left(\begin{array}{cc} X & WX \end{array} \right)$.

In their introduction to a special issue of *Papers in Regional Science* entitled: "New spatial econometric techniques and applications in regional science," Arbia and Fingleton [2] summarize the criticism of spatial regression models as being sensitive to weight matrix specifications stating:

> This problem of what the spatial lag actually represents is bound up with the problem of definition of the spatial weights matrix, which is assumed to be a nonstochastic matrix capturing our hypothesis about the nature of the spatial interactions we are modelling. The problem is that, unlike the simple notion of a time series lag, the spatial lag is a very fluid and complex entity open to multiple definitions within a single study. Critics of spatial econometrics almost always in our experience home in on the arbitrary nature of the weights matrix, asking "how is it defined and why is it precisely like that when it could easily have been like this, what does it mean, and are not the results obtained conditional on somewhat arbitrary decisions taken about its structure?". Some future research on the robustness of outcomes to variations in assumptions about the weight matrix structure would be helpful in allaying such criticisms, although ideally carefully structured arguments coming from theory and leading precisely to the typical reduced form spatial econometric model, with a spatial lag and exogenous lags also, are the preferred option.

We view the notion that the explanatory variable effects and inferences are sensitive to use of a particular weight matrix as perhaps *the biggest myth about spatial regression models*. For the model in Equation (1) consider two model specifications based on weight matrices W_a and W_b, with the same sample data y, X. If $W_a y$ and $W_b y$ are highly correlated, it would seem difficult to reach materially different conclusions about the partial derivative impact of changes in the explanatory variables in the

---

[1] Theoretical bounds for this parameter are set forth in LeSage and Pace [1], and depend on minimum and maximum eigenvalues of the spatial weight matrix W.

[2] In linear algebra a row-stochastic matrix has non-negative entries and each row sums to 1 while a doubly stochastic matrix is non-negative and both the rows and columns sum to 1. Symmetric, doubly stochastic weight matrices, although they have not been used as much in applications, have a number of simple theoretical properties (all real eigenvalues, eigenvectors, and are constant preserving). Sometimes, as in Arbia and Fingleton [2], weight matrices are said to be nonstochastic which means the elements are not random variables. So, from a statistical viewpoint a matrix could be termed nonstochastic while from a linear algebra view the same matrix could be said to be row- or doubly-stochastic. This is merely a difference in terminology that uses the term stochastic in different contexts. We make the traditional assumption that the matrix W is fixed in repeated sampling and therefore exogenously determined.

matrix X on the dependent variable y (which LeSage and Pace [1] label *effects estimates*) from models based on $W_a y$ versus $W_b y$. (See Section 2.3 for the definition of effects estimates.)

For the case of row-stochastic nearest neighbor matrices, we 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: $\text{corr}(W_a u, W_b u) = (m_a/m_b)^{0.5}$, where $m_a \leq m_b$, and m_a, m_b represent the number of neighbors used in constructing W_a, W_b.[3] 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.[4]

A similar statement applies to more exotic spatial weight specifications that involve inverse distance and a decay parameter γ in conjunction with some number of nearest neighbors m used as a cut-off, beyond which zero weights are assigned. For example Anselin [5] discusses a weight matrix taking the form:

$$W(i,j) = 1/d(i,j)_m^\gamma \tag{2}$$

where $d(i,j)_m$ denotes the distance between the mth nearest neighboring observations j to observation i, and γ is a decay parameter, say in the interval $[0, 2]$. Kostov [6] considers this type of specification based on $m = 1, \ldots, 50$ (distance-based) nearest neighbors, and values of γ the decay parameter in the interval $[0.4, 4]$. Using increments of 0.1 for γ, this leads to a discrete set of 1850 alternative weight matrices. Here again, small changes in the decay parameters γ and m should lead to highly correlated $W_a y$ and $W_b y$.

To address the myth, we begin with Section 2 that provides a number of theoretical results that show why high correlations often exist between the predictions and marginal effects from different weight matrices. We also undertake a reexamination of past literature that may have contributed to formation of this myth in Section 3. Specifically, in Section 3.2 we consider work by Bell and Bockstell [7], and in Section 3.3 we empirically examine more recent work by Kostov [6]. Finally, in Section 4 a generated data example is used to illustrate some of the issues that arise regarding interaction between spatial regression model and weight matrix specifications. Specifically, we show how a model based on a flexible spatial spillover specification makes it possible to recover accurate estimates of the spatial effects estimates advocated by LeSage and Pace [1], even in a case where *both* the matrix W and spatial regression model are misspecified.

2. Measures of Similarity between Weight Matrices

Directly comparing two $n \times n$ weight matrices W_a and W_b containing n^2 elements seems unlikely to result in any clear cut measure of similarity. However, it is possible to derive scalar summaries of similarity between spatial lags that result from use of these two weight matrices. Specifically, in Section 2.1 we calculate the correlation between $W_a u$ and $W_b u$ for row-stochastic W based on nearest neighbors and an $n \times 1$ vector of independent identically distributed normal deviates u. We show that

---

[3] We assume for simplicity that $m_a < m_b$, without loss of generality.

[4] The correlation results we present also apply to spatial lag vectors involving nearest neighbor weight matrices that have been normalized using scalars, such as the spectral radius proposed by Barry and Pace [3], or the maximum of row or column sums proposed by Kelejian and Prucha [4].

the correlation takes the very simple form of $(m_a/m_b)^{1/2}$, where m_a, m_b represents the number of neighbors used in constructing W_a, W_b and $m_a < m_b$.

Section 2.2 describes alternative measures of similarity for predictions from autoregressive models based on weight matrices W_a and W_b that apply to more general spatial weight structures. This measure of similarity has a simple form involving traces of the powers of the weight matrices, and does not depend on the parameter vector β. The measure of predictive similarity allows us to explore sensitivity of the partial derivative *effects estimates* proposed by LeSage and Pace [1], a topic taken up in Section 2.3. Section 2.4 provides a numerical illustration of the alternative measures of similarity between models based on varying matrices W.

2.1. Correlation between $W_a u$ and $W_b u$ for Varying W

Let W_a and W_b represent row-stochastic weight matrices based on m_a, m_b nearest neighbors where each element in W_a, W_b has equal weight of m_a^{-1}, m_b^{-1}, and for simplicity we assume that $m_a \leq m_b$. Spatial regression models use spatial lag vectors $W_a u$, $W_b u$, where we assume u is an $n \times 1$ vector composed of standardized independent normal deviates. If $W_a u$, $W_b u$ act differently, this would support the myth that spatial regression model specifications are sensitive to the exact specification used for W. In contrast, if we can show conditions under which $W_a u$, $W_b u$ act similarly in the context of spatial regression models, this would support our contention that the exact composition of W is not critical.

Before beginning the analysis, we use extensively the relations,

$$u'Au = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} u_i u_j, \; u \sim N(0, I_n) \tag{3}$$

$$E(u'Au) = \sum_{i=1}^{n} E(u_{ii}^2) A_{ii} = \sum_{i=1}^{n} A_{ii} = \text{tr}(A) \tag{4}$$

$$\text{tr}(A'B) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij} = \iota_n'(A \odot B)\iota_n = E(A'B) \approx u'A'Bu \; \text{ for large } \; n \tag{5}$$

where A, B are n by n matrices, ι_n is a n by 1 vector composed of ones, and $\odot$ represents element-by-element multiplication. The quadratic form $u'Au$ in Equation (3) has expectation $\text{tr}(A)$ in Equation (4) because $E(u_i^2)$ equals 1 for unit normal random variable (χ^2 with one degree-of-freedom) and $E(u_i u_j) = 0$ when $i \neq j$ due to the independence assumption ($u \sim N(0, I_n)$). In addition, as emphasized by Equation (5), $\text{tr}(A'B)$ depends on the common elements in A and B.

Using the relations in Equations (3)–(5), simple expressions exist for covariance in Equation (6) and correlation between $W_a u$, $W_b u$ for the case of row-stochastic nearest neighbor matrices W as shown in Equation (7). The simplicity stems from the known common elements in W_a and W_b which equal m_a^{-1} and m_b^{-1}. Using Equation (5) this means that the elements of $W_a \odot W_b$ equal $m_a^{-1} m_b^{-1}$ and that each row of $W_a \odot W_b$ contains m_a such elements (otherwise zeros). Summing across all rows and averaging cancels out and so the $\text{cov}(W_a u, W_b u)$ equals m_b^{-1}. Following a similar method leads to the respective variances and thus the correlation in Equation (7).

$$\text{cov}(W_a u, W_b u) = n^{-1} u' W_a' W_b u = n^{-1} \text{tr}(W_a' W_b) = n^{-1} \sum_{i=1}^{n} m_a \left[\frac{1}{m_a} \frac{1}{m_b} \right] = m_b^{-1} \tag{6}$$

$$\text{var}(W_a u) = n^{-1} u' W_a' W_a u = n^{-1} \text{tr}(W_a' W_a) = n^{-1} \sum_{i=1}^{n} m_a \left[\frac{1}{m_a^2} \right] = m_a^{-1}$$

$$\text{var}(W_b u) = n^{-1} u' W_b' W_b u = n^{-1} \text{tr}(W_b' W_b) = n^{-1} \sum_{i=1}^{n} m_b \left[\frac{1}{m_b^2} \right] = m_b^{-1}$$

$$\text{corr}(W_a u, W_b u) = \frac{m_b^{-1}}{m_a^{-1/2} m_b^{-1/2}} = \left(\frac{m_a}{m_b} \right)^{1/2} \tag{7}$$

The result in Equation (7) indicates that when $m_a = 10$ and $m_b = 20$, the correlation between $W_a u$ and $W_b u$ is 0.707. For $m_a = 10$ and $m_b = 12$, the correlation is 0.9129. This suggests that for the case of nearest neighbor weight matrices, spatial lag vectors used in spatial regression models may behave similarly in the face of different specifications for W. This result also highlights a source of the similarity between spatial lags, namely the number of non-zero elements (neighbors) that W_a and W_b have in common.

This result generalizes beyond nearest neighbor matrices, since other weight matrices almost always contain at least the contiguous or nearby neighbors. Given our result, this should lead to $W_a u$ and $W_b u$ having at least a moderately high level of correlation. This result also suggests that it would be possible for very different scalings or sparsity levels (e.g., a sparse contiguity matrix W_a versus a dense inverse distance matrix W_b) to produce lower correlations.

We show later that estimates and inferences regarding the true partial derivatives (or effects) are robust in circumstances where W_a and alternative specifications W_b exhibit the same type of scaling, and therefore result in moderate to high levels of correlation between $W_a u$ and $W_b u$. This means that spatial regression models produce inferential results that frequently avoid the criticism of sensitivity to specification of W.

A practical implication of this correlation is that a specification based on say an r nearest neighbor matrix W_a, is likely to produce estimates and inferences (for the effects) that are robust with respect to alternative weight matrix specifications W_b based on s nearest neighbors. Similarly, a weight matrix specification based on inverse distance with decay of influence based on a cut-off (zero weight) beyond some distance or number of nearest neighbors, is likely to produce effects estimates and inferences that are robust with respect to weight matrices based on alternative choices of decay or cut-off.

Note, weight matrices with different non-scalar scalings may show a lower correlation with each other. However, as will be shown later, the use of row-stochastic scalings follows from the long-run equilibrium of a spatiotemporal process which underlies the motivation behind the SAR cross-sectional model. Therefore, we will largely focus on the consequences of using a different W for estimation than used in the DGP within the class of row-stochastic or doubly stochastic scalings.

Spatial regression models also involve use of vectors involving higher-order spatial lags such as $W_a^j u$, $W_b^j u$ for $j > 1$. The correlation between $W_a^j u$, $W_b^j u$ for large j for *doubly stochastic* W_a, W_b has an even simpler form, shown in Equation (8), where ι_n is a $n \times 1$ vector of ones.

$$
\begin{aligned}
\lim_{j \to \infty} W^j &= n^{-1} \iota_n \iota_n' \\
\lim_{j \to \infty} \mathrm{corr}(W_a^j u, W_b^j u) &= 1
\end{aligned}
\tag{8}
$$

Therefore, all doubly stochastic symmetric weight matrices, whether based on nearest neighbors, inverse distance, contiguity, common border lengths, or any other method of construction result in identical limiting values of $W_a^j u$, $W_b^j u$.

Other types of weight matrices may partly share this trait. As the order of the neighboring relations (non-zero elements) rise in W^j, various forms of weight matrices will place weight on neighbors of neighbors of neighbors, and so on. To the extent that two matrices W_a^j and W_b^l place higher weights on first-order neighbors and lower weights on higher-order neighbors and that they share common elements, a positive correlation between $W_a^j u$ and $W_b^l u$ will exist.[5] Therefore, we would expect that the exact form of W would become less important for many of the higher order neighboring relations that play an important role in determining effects estimates for spatial regression models.

2.2. Measures of Correlation between Predictions from Varying W

We begin the analysis of the relation between W_a and W_b with separate examples of the SAR data generating processes (DGP) in Equations (9) and (10).

$$
y_a = (I_n - \rho_a W_a)^{-1} X \beta_a + (I_n - \rho_a W_a)^{-1} \varepsilon_a
\tag{9}
$$
$$
y_b = (I_n - \rho_b W_b)^{-1} X \beta_b + (I_n - \rho_b W_b)^{-1} \varepsilon_b
\tag{10}
$$

These DGPs have the following expectations for the dependent variable which are predictions for the SAR model. Note, the expectation operation in this case removes the randomness from the error term. However, the resulting $E(y_a)$ and $E(y_b)$ still exhibit variation.

$$
E(y_a) = (I_n - \rho_a W_a)^{-1} X \beta_a
\tag{11}
$$
$$
E(y_b) = (I_n - \rho_b W_b)^{-1} X \beta_b
\tag{12}
$$

For simplicity, assume that $E(X_j) = 0$ for $j = 1 \ldots k$ so $E(y_a)$, $E(y_b)$ equal 0. In this case, the covariance (cross-products) between the predictions (expected values of the dependent variables) have the form in Equation (13).

$$
E(y_a)' E(y_b) = \beta_a' X' (I_n - \rho_a W_a')^{-1} (I_n - \rho_b W_b)^{-1} X \beta_b
\tag{13}
$$

Also, the own products in Equations (14) and (15) have a similar form.

$$
E(y_a)' E(y_a) = \beta_a' X' (I_n - \rho_a W_a')^{-1} (I_n - \rho_a W_a)^{-1} X \beta_a
\tag{14}
$$
$$
E(y_b)' E(y_b) = \beta_b' X' (I_n - \rho_b W_b')^{-1} (I_n - \rho_b W_b)^{-1} X \beta_b
\tag{15}
$$

---

[5]   If W_a is much sparser than W_b, the maximum positive correlation between $W_a^j u$ and $W_b^l u$ might occur for powers $j > l$.

Given the cross product $E(y_a)'E(y_b)$ and own products $E(y_a)'E(y_a)$, $E(y_b)'E(y_b)$, we can form the correlation between the expected values of the dependent variables from the two models a and b (denoted by $\text{corr}(E(y_a), E(y_b))$). Note, $E(y_a)$ and $E(y_b)$ still exhibit variation and the correlation measures their association.

$$\text{corr}(E(y_a), E(y_b)) \quad = \quad \frac{E(y_a)'E(y_b)}{[E(y_a)'E(y_a)]^{0.5}[E(y_b)'E(y_b)]^{0.5}} \tag{16}$$

If β_a, β_b are scalars, these would cancel out of $\text{corr}(E(y_a), E(y_b))$ and result in further simplicity.

We also note that given Equations (14) and (15), spatial dependence in the explanatory variables X could potentially increase $\text{corr}(E(y_a), E(y_b))$. Pace, LeSage and Zhu [8] provide evidence that explanatory variables used in spatial regression models often exhibit very high levels of spatial dependence. Specifically, they show that county, census tract and block group variables measuring age, race, income, employment, educational attainment, etc. exhibit spatial dependence such that the first-order spatial autoregressive dependence parameter ϕ from the model: $x = \alpha\iota_n + \phi W x + u$, exceeds 0.9.

For a univariate explanatory variable $x = (I_n - \phi W)^{-1}u$ where $u \sim N(0, I_n)$, we can arrive at a simple expression for the covariance between predictions by substituting this relation into Equation (13) which yields Equation (17).

$$E(y_a)'E(y_b) \quad = \quad \text{tr}[(I_n - \phi W')^{-1}(I_n - \rho_a W_a')^{-1}(I_n - \rho_b W_b)^{-1}(I_n - \phi W)^{-1}] \tag{17}$$

which shows that the covariance between y_a and y_b may be influenced even more by higher order spatial relations than Equation (13). Since we have already shown that higher order spatial relations depend less on the specific form of the matrix W, this implies that predictions are less sensitive to alternative specifications for W.

2.3. Measures of Correlation for Effects Estimates Based on Varying W

As shown in the previous section, we can examine the correlation between spatial regression model predictions. However, the dependence of this correlation on spatial dependence in the explanatory variables X might lead to very high correlations in the face of high levels of dependence in X. High levels of correlation would mean that this measure would not provide a useful basis for discriminating between alternative weight matrices.

One way around this is to consider the cumulative scalar summary measures proposed by LeSage and Pace [1] that measure the impacts or partial derivative responses of the dependent variable to changes in the explanatory variables in spatial lag regression models. These *effects estimates* are not a function of X for the SAR or SDM models. For the case of the SAR model DGPs in Equations (9) and (10), the partial derivatives take the form of $n \times n$ matrices shown in Equations (18) and (19) for $k = 2, \ldots, K$ (assuming the first explanatory variable is a constant ι_n).

$$D_{ak} \quad = \quad (I_n - \rho_a W_a)^{-1}\beta_{ak} \tag{18}$$
$$D_{bk} \quad = \quad (I_n - \rho_b W_b)^{-1}\beta_{bk} \tag{19}$$

A direct comparison of the n^2 elements in D_{ak} and D_{bk} for similarities would of course be difficult. We can reduce the dimension of the problem to that of comparing two $n \times 1$ vectors using the

total differential for each observation. The total differential shown in Equation (20) is a weighted combination of individual partial derivatives, where the weights are changes or perturbations in the underlying variables.

$$df \;=\; \sum_{i=1}^{n} \frac{\partial f}{\partial z_i} dz_i \tag{20}$$

In a spatial context, this equates to perturbing each observation or region. In other words, we examine the total change in the expectation of the dependent variable for each observation or region that arises from simultaneously changing the independent variable in all regions. Similarly, we can form a total differential for each observation for a given perturbation vector u, which we assume is composed of independent standard normal deviates.

$$t_{ak} \;=\; D(a)u = (I_n - \rho_a W_a)^{-1}\beta_{ak}u \tag{21}$$

$$t_{bk} \;=\; D(b)u = (I_n - \rho_b W_b)^{-1}\beta_{bk}u \tag{22}$$

Each row of t_{ak}, t_{bk} is the total differential for observation i given the perturbation vector u. Since u has an expectation of 0, the total differential vectors also have an expectation of zero ($E(t_{ak}) = E(t_{bk}) = 0$). Consider the expectation of the cross-product of the total derivatives:

$$E(t'_{ak}t_{bk}) \;=\; \beta_{ak}\beta_{bk}E(u'(I_n - \rho_a W'_a)^{-1}(I_n - \rho_b W_b)^{-1}u) \tag{23}$$

$$E(t'_{ak}t_{bk}) \;=\; \beta_{ak}\beta_{bk}tr((I_n - \rho_a W'_a)^{-1}(I_n - \rho_b W_b)^{-1}) \tag{24}$$

and we note that the own products take a similar form:

$$E(t'_{ak}t_{ak}) \;=\; \beta_a^2 tr([(I_n - \rho_a W_a)'(I_n - \rho_a W_a)]^{-1}) \tag{25}$$

$$E(t'_{bk}t_{bk}) \;=\; \beta_b^2 tr([(I_n - \rho_b W_b)'(I_n - \rho_b W_b)]^{-1}) \tag{26}$$

Given the cross product $E(t'_{ak}t_{bk})$, the own products $E(t'_{ak}t_{ak})$ and $E(t'_{bk}t_{bk})$, we can form the correlation between the total differentials from the two models a and b, denoted by $\mathrm{corr}(t_{ak}, t_{bk})$ in Equation (27).

$$\mathrm{corr}(t_{ak}, t_{bk}) \;=\; \frac{tr[(I_n - \rho_a W'_a)^{-1}(I_n - \rho_b W_b)^{-1}]}{tr([(I_n - \rho_a W_a)'(I_n - \rho_a W_a)]^{-1})]^{0.5}tr[(I_n - \rho_b W_b)'(I_n - \rho_b W_b)]^{-1}]^{0.5}} \tag{27}$$

Note, for the autoregressive model the regression model parameters cancel and so the result does not depend on their magnitude. Also, if $\rho = 0$, so there is no spatial dependence or weight matrix in our model, then $\mathrm{corr}(t_{ak}, t_{bk}) = 1$. This implies that models where spatial dependence is weak should exhibit less sensitivity to the weight matrix specification, an intuitively plausible result.

In the next section, we provide a numerical application of the correlation measure in Equation (27) as well as the other measures developed in Sections 2.1 and 2.2.

2.4. Applied Illustrations of Similarity Measures between Varying W

To provide an illustration of these measures of correlation between varying W matrix specifications, we use a 1000 observation sample of generated data. We examined the correlation between $W_a u$ and $W_b u$ using an $n \times 1$ vector u containing independent, identically distributed, standard normal deviates. W_a is a (row-stochastic) contiguity-based weight matrix, and W_b is a symmetricized (row-stochastic) nearest neighbor weight matrix, where the number of neighbors (m) varied between 5 and 30. The second column of Table 1 shows correlations between $W_a u$ and $W_b u$. These correlations range between 0.7157 in the case of 5 neighbors and decline to 0.3735 for the case of 30 nearest neighbors. Since a typical contiguity weight matrix will have an average of approximately six neighbors for each observation for spatially random data on a plane, the decline in correlation with a larger number of neighbors should not be surprising.

Table 1. Correlations across nearest neighbor W relative to contiguity W.

Neighbors m	1st Order corr($W_a u, W_b u$)	10th Order corr($W_a^{10} u, W_b^{10} u$)	Effects Estimates	Predictions
5	0.7157	0.7642	0.9331	0.9817
6	0.6980	0.7930	0.9392	0.9829
7	0.6860	0.8287	0.9478	0.9836
8	0.6748	0.8491	0.9533	0.9848
9	0.6596	0.8676	0.9568	0.9848
10	0.6469	0.8787	0.9590	0.9851
11	0.6267	0.8825	0.9598	0.9851
12	0.6133	0.8801	0.9586	0.9839
13	0.5862	0.8689	0.9563	0.9827
14	0.5702	0.8653	0.9542	0.9816
15	0.5566	0.8593	0.9523	0.9798
16	0.5415	0.8486	0.9498	0.9786
17	0.5172	0.8377	0.9469	0.9775
18	0.4993	0.8246	0.9443	0.9763
19	0.4810	0.8114	0.9419	0.9751
20	0.4688	0.8012	0.9399	0.9738
21	0.4623	0.7854	0.9375	0.9721
22	0.4460	0.7772	0.9350	0.9709
23	0.4405	0.7706	0.9337	0.9703
24	0.4343	0.7643	0.9320	0.9696
25	0.4208	0.7551	0.9300	0.9687
26	0.4066	0.7464	0.9281	0.9681
27	0.3913	0.7368	0.9263	0.9671
28	0.3854	0.7255	0.9244	0.9663
29	0.3760	0.7135	0.9224	0.9651
30	0.3735	0.7056	0.9209	0.9644

The third column of the table show that correlations increased for higher order relations in all weight matrices. Specifically, the third column shows correlations between W_a^{10} and W_b^{10}. For the case of 10th order weight matrices, the correlations ranged from 0.7056 for 5 neighbors to 0.8825 for 11 nearest neighbors.

The correlations between the *effects estimates* shown in the fourth column of the table exhibit a high level of correlation. These high correlations exist across the entire range of varying W matrices. For values of $\rho_a, \rho_b = 0.8$ in both DGPs, the correlations range from 0.9209 for 30 nearest neighbors to 0.9598 for 11 nearest neighbors. We will have more to say about this result in the next section where we discuss possible origins of the myth.

The last column of the table shows correlations between predictions based on the assumption that the univariate x used to generate the sample data followed a first-order spatial autoregressive process, with a dependence parameter equal to 0.95. The predictions exhibit even higher levels of correlation than the effects estimates, ranging from 0.9644 for 30 nearest neighbors to 0.9851 for 10 and 11 nearest neighbors.

3. Origins of the Myth

The theoretical developments and numerical illustrations from Section 2 suggests that practitioners should not encounter situations where estimates and inferences exhibit sensitivity to particular weight matrix specifications. An interesting question is—how did the myth regarding sensitivity to weight matrices arise? Of course, it is trivially obvious from Equation (1) that the matrix W plays a role in spatial regression models. Dramatically different choices for W could lead to very different estimates and inferences. But the myth we focus on is something quite different. The myth perpetrates the idea of the need to fine-tune weight matrix specifications, because estimates and inferences are sensitive to small changes in these specifications. If this were true, then spatial regression models could be considered ill-conditioned and would represent an unreasonable method for analyzing relationships involving spatially dependent data.

3.1. Past Literature

Many authors cite Anselin [5], perhaps because there is a great deal of discussion regarding alternative approaches that can be used to construct spatial weight structures. However, Anselin [5] makes no explicit statement that estimates and inferences (from a well-specified model) can be very sensitive to the choice of weight matrix. Rather his emphasis is on the flexibility these models provide in alternative choices of weight matrices, and the need to appropriately specify this aspect of the model in any particular application. For example, when modeling cross-border shopping by smokers to avoid high levels of taxes on cigarettes, an intuitively appealing weight matrix might be constructed based on miles of border in common between each state and its neighbors.

Even if misspecification of the weight matrix leads to incorrect estimates of β and ρ in spatial regression models, the impact of this on inferences regarding the partial derivative response of the dependent variable to changes in the explanatory variables is unclear. Both ρ and β appear in the

(non-linear) partial derivative expressions, meaning that upward bias in ρ could be offset by downward bias in β in such a way as to produce effects estimates that remain quite stable.

There are numerous cases where applied spatial econometric work includes statements that sensitivity of the estimates for β and ρ were checked with respect to choice of the weight matrix.[6] However, this is not the appropriate basis for inference about sensitivity of estimates and inferences from these models, since these should rely on the partial derivatives (or effects estimates).

In Section 3.2 we examine an article that is widely cited as a justification for sensitivity of spatial regression estimates to small changes in the weight matrix specification. Section 3.3 turns attention to a more recent work by Kostov [6] who sets forth a methodology for fine-tuning the weight matrix specification. In Section 3.4 we provide an illustration of how one might diagnose sensitivity of estimates and inferences in applied practice, using a publicly available county-level data set, and public domain software algorithms.

3.2. A Re-Examination of Bell and Bockstael (2000)

An article often cited to justify the myth is Bell and Bockstell [7], who explicitly argue that estimates and inference are sensitive to small changes in the weight matrix. They use this line of argument to further contend that the generalized moments method of estimation from Kelejian and Prucha [9] allow flexible weight structures to be more easily implemented than maximum likelihood estimation methods. In an application involving land parcels, they explore three (row-normalized) contiguity-based weight matrices that assign values of 0 or 1 to neighboring observations that are within 200, 400 and 600 m distances of each observation. These three matrices are compared to a fourth (row-normalized) matrix where inverse distance-based weights were assigned to each neighboring observation within 600 m. They rely on a spatial error model (SEM) shown in equation Equation (28) and compare estimates from least-squares, maximum likelihood and generalized moments, with the latter two sets of estimates constructed using the four different types of spatial weight matrices.

$$ y = X\beta + u \tag{28} $$
$$ u = \rho W u + \varepsilon \tag{29} $$

They conclude:

> What emerges from the example is that our results are more sensitive to the specification of the spatial weight matrix than to estimation technique. Compared to the variation across estimation methods, the results across spatial weight matrices are much less stable.

> Where qualitative results change, they are almost universally associated with changes in the spatial weight matrix and not with changes in the estimation method. For three of the estimated coefficients, one spatial weight matrix produces results qualitatively different from the others, and, for three more of the estimated coefficients, two spatial weight matrices produce results

---

[6] It is not surprising there are few reports of cases where estimates were found to be sensitive to the choice used by the practitioner.

qualitatively different from the other two. There is no particular pattern to these reversals, nor is there a pattern when comparing the spatial correlation-corrected results to the OLS results.

To put the work of Bell and Bockstell [7] into perspective, consider that for a *correctly specified* SEM model, the only difference between least-squares (OLS) and SEM model estimates should be in the measures of dispersion, not the coefficient estimates for β. This is because coefficient estimates for β from OLS are unbiased under the null hypothesis of an SEM specification (Anselin [10], p. 59). This follows because spatial dependence in the disturbances leads to an efficiency problem for OLS, but no bias in the estimates for β. Changes in the weight matrix should lead to changes in the $t-$statistics that we observe from OLS versus SEM model estimates, but not in the coefficient estimates for β. Pace and LeSage [11] use this idea to develop a Hausman specification test for significant differences between OLS and SEM estimates for β. Intuitively, significant differences in OLS and SEM estimates for β point to model misspecification that should lead us to reject the SEM model as an appropriate choice.

In Bell and Bockstell [7] the sample size used was 1000 observations, so we would expect no small sample issues that would lead to differences between OLS and SEM estimates for the parameters β. As noted above, this should be true irrespective of the spatial weight matrix employed. Changes in the spatial weight specification could lead to changes in measures of dispersion (e.g., $t-$statistics), but not significant differences in the coefficients β. The discussion (quoted above) by Bell and Bockstell [7] appropriately focused on differences in significance or inference that arise in response to the four alternative weight matrices used to estimate their model. However, they neglect to note that five of the ten coefficients β from OLS estimation versus maximum likelihood estimation of the SEM model differ by more than 1.67 standard deviations, suggesting model misspecification.[7] Table 2 presents their OLS and maximum likelihood SEM estimates, along with standard errors and a $t-$test for significant differences between these.[8]

There is one coefficient where the SEM estimate is 2.8 standard deviations away from the OLS, two cases where the two estimates are 1.99 standard deviations apart and two more that are different using the 90% level of significance. Of the ten coefficients five are likely to be significantly different, suggesting the SEM model represents a misspecification.

The sensitivity of estimates and inferences to changes in the weight matrix noted in Bell and Bockstell [7] was likely due to misspecification in their SEM model. For example, following the argument of LeSage and Pace [1], if their SEM model omitted variables that were correlated with included variables, this would lead to biased and inconsistent estimates for the parameters β. This meshes with their observation that "There is no particular pattern to these reversals, nor is there a pattern when comparing the spatial correlation-corrected results to the OLS results." In general, sensitivity to changes in the weight matrix may be indicative of model misspecification.

---

[7] We focus our discussion on their maximum likelihood estimates constructed using an inverse distance weight matrix based on a 600 m cut-off, but similar statements apply to maximum likelihood and GMM estimates based on the other three weight matrices.

[8] The standard errors from maximum likelihood SEM estimates were used to construct the $t-$test, since these are not adversely impacted by spatial dependence in the disturbances. However, this is not a multivariate test for the difference between the two vectors, but is a way of describing the differences between estimates.

Table 2. Bell and Bockstell(2000) OLS and maximum likelihood SEM estimates.

	OLS $\hat{\beta}_o\,(\hat{\sigma}_{\beta_o})$	**ML** $\hat{\beta}_{ml}\,(\hat{\sigma}_{\beta_{ml}})$	*t*–Statistic (*t*–Probability) $H_o : \hat{\beta}_o = \hat{\beta}_{ml}$
Intercept	4.7332 (0.2047)	5.1725 (0.2204)	1.9932 (0.0465)
LIV	0.6926 (0.0124)	0.6537 (0.0135)	2.8815 (0.0040)
LLT	0.0079 (0.0052)	0.0002 (0.0052)	1.4808 (0.1390)
LDC	−0.1494 (0.0195)	−0.1774 (0.0245)	1.1429 (0.2534)
LBA	−0.0453 (0.0114)	−0.0169 (0.0156)	1.8205 (0.0690)
POPN	−0.0493 (0.0408)	−0.0149 (0.0414)	0.8309 (0.4062)
PNAT	0.0799 (0.0177)	0.0586 (0.0212)	1.0047 (0.3153)
PDEV	0.0677 (0.0180)	0.0253 (0.0253)	1.6759 (0.0941)
PLOW	−0.0166 (0.0194)	−0.0374 (0.0224)	0.9286 (0.3533)
PSEW	−0.1187 (0.0173)	−0.0828 (0.0180)	1.9944 (0.0464)

3.3. A Re-Examination of Kostov (2010)

Another factor contributing to the myth arises in the case of models that include spatial lags of the dependent variable. There are different types of spatial regression specifications that include spatial lags of the dependent variable, but LeSage and Pace [1] argue that one specification, the spatial Durbin model (SDM) stands out as superior in a wide number of applied situations. The SDM shown in Equation (30) includes a spatial lag of the dependent variable (Wy) as well as explanatory variables (WX):

$$\begin{aligned} y &= \alpha\iota_n + \rho Wy + X\beta + WX\theta + \varepsilon \\ \varepsilon &\sim N(0,\sigma^2) \end{aligned} \tag{30}$$

In the case of OLS where observations are independent, changes in X_i can only influence y_i, so we use the coefficient estimates for the rth explanatory variable (β_r) to summarize the average (across the sample) impact of changing the rth explanatory variable on the dependent variable vector y. LeSage and Pace [1] point out that this is not the case when the dependent variable observation y_i exhibits dependence on other observations. They rewrite the model in Equation (30) as in Equation (31), which is useful for examining the partial derivatives of y with respect to a change in the rth variable x_r from X, which is shown in Equation (32).

$$\begin{aligned} y &= (I_n - \rho W)^{-1}[\alpha\iota_n + X\beta + WX\theta + \varepsilon] \tag{31} \\ \partial y/\partial x_r' &= (I_n - \rho W)^{-1}(I_n\beta_r + W\theta_r) \tag{32} \end{aligned}$$

The partial derivatives are an $n \times n$ matrix rather than the typical scalar expression β_r from OLS. The matrix arises because a change in a single observation x_{ir} can influence all observations of the vector $y_j, j = 1, \ldots, n$. Considering changes in each of the $x_{ir}, i = 1, \ldots, n$ observations and the associated $n \times 1$ vectors of y−responses gives rise to the $n \times n$ matrix of partial derivatives. The own-region or *direct* response is captured by the own-partial derivative $\partial y_i/\partial x_{ir}$ which are elements on the diagonal of the matrix in Equation (32). The cross-partial derivatives $\partial y_j/\partial x_{ir}, j \neq i$ reflect *indirect* or spillover responses, and these are located on the off-diagonal elements of the matrix in Equation (32).

Changes in coefficient estimates for β and ρ observed by practitioners who try alternative weight matrix specifications may have contributed to formulation of the myth. This is because practitioners who incorrectly believe that coefficient estimates β measure partial derivative responses in the dependent variable to changes in the independent variables would infer sensitivity of estimates and inferences arising from changes in W. As noted earlier, this focus on how changes in the matrix W impact coefficient estimates for ρ and β, is misplaced, since the focus should be on changes in the true partial derivatives, the direct and indirect effects estimates described above. In fact, changes in estimates for ρ and β may arise systematically as a response to changes in the matrix W, because these are required to maintain relatively stable partial derivatives in the face of changing W. Past misinterpretation of estimates from spatial regression models containing lags of the dependent variable may have contributed to the myth that estimates and inferences are sensitive to the choice of W. Ironically, these changes might be occurring in an effort to ensure a well-conditioned model where the true partial derivative responses remain relatively constant in the face of changing W.

We re-examine the model from Kostov [6], who used the data set from Harrison and Rubinfeld [12]. This data set containing 506 Boston census tract observations was augmented to have latitude-longitude coordinates from Gilley and Pace [13]. The 10 explanatory variables used in the model are shown in Table 3.

Table 3. Variables and definitions.

Variable	Description
CRIME	per capita crime rate by town
CHARLES	Charles River dummy variable (=1 if tract bounds river; 0 otherwise)
NOX	nitric oxides concentration (parts per 10 million)
ROOMS	average number of rooms per dwelling
DISTANCE	weighted distances to five Boston employment centers
RADIAL	index of accessibility to radial highways
TAX	full-value property-tax rate per $10,000
PTRATIO	pupil-teacher ratio by town
B	$1000(Bk - 0.63)^2$ where Bk is the proportion of blacks by town
LSTATUS	% lower status of the population

The parameterized weight structure used by Kostov [6] takes the form:

$$W(i,j) = 1/d(i,j)_m^\gamma \tag{33}$$

where $d(i,j)_m$ denotes the distance between the m nearest neighboring observations j to observation i and $\gamma > 0$ is a decay parameter. Other values for $W(i,j)$ for neighbors $m+1, m+2, \ldots$ are set to zero. The model used is the SAR: $y = \rho W y + X\beta + \varepsilon$.

Rather than estimate the parameter γ, Kostov [6] considers a "boosting" type of model search/comparison procedure that is applied to a discrete set of models based on a 0.1 grid of values for γ in the interval $[0.4, 4]$ and a range of $m = 1, \ldots, 50$. His approach identifies models based on γ values in the range 0.4 to 1, and $m = 6$ as representing the "best" weight structure for the SAR model and sample data.

Bayesian model comparison methods can be used to compare this discrete set of models. Specifically Hepple [14] shows that the log-marginal likelihood for the SAR model takes the form in Equation (34):[9]

$$
\begin{aligned}
p(y|M_i) &= \frac{1}{D}\Gamma\left(\frac{n-k}{2}\right) \cdot (2\pi)^{-\frac{n-k}{2}} \cdot |X'X|^{-1/2} \int |I_n - \rho W_i|(e'e)^{-\frac{n-k}{2}} d\rho \\
e &= y - \rho W y - X\beta
\end{aligned}
\tag{34}
$$

where we use M_i to denote model i based on spatial weight matrix W_i, and D to denote the interval defined by the minimum and maximum eigenvalues of the matrix W_i.[10]

Table 4 presents posterior model probabilities for a discrete set of models based on a grid of values for the parameter γ in the interval $[0, 2]$ based on increments of 0.1, and three values $m = 5, 6, 7$ of nearest neighbors that we considered. Of course, we chose these values based on the results from Kostov [6].

Table 4. Posterior model probabilities .

γ	$m = 5$	$m = 6$	$m = 7$
0	0.0001	0.0095	0.0007
0.1	0.0004	0.0288	0.0025
0.2	0.0013	0.0726	0.0083
0.3	0.0028	0.1381	0.0207
0.4	0.0041	0.1835	0.0368
0.5	0.0041	0.1672	0.0445
0.6	0.0029	0.1080	0.0376
0.7	0.0015	0.0525	0.0234
0.8	0.0007	0.0203	0.0114
0.9	0.0002	0.0065	0.0045
1	0.0001	0.0018	0.0015
1.1	0.0000	0.0004	0.0004
1.2	0.0000	0.0001	0.0001
1.3	0.0000	0.0000	0.0000
1.4	0.0000	0.0000	0.0000
1.5	0.0000	0.0000	0.0000
1.6	0.0000	0.0000	0.0000
1.7	0.0000	0.0000	0.0000
1.8	0.0000	0.0000	0.0000
1.9	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000

---

[9] This result is based on assigning no prior distributions for the parameters β, σ^2, and a uniform prior for ρ, over the interval D defined by the minimum and maximum eigenvalues of the matrix W_i (Hepple 2004, p. 111).

[10] LeSage and Pace (2009, p. 175–178) provide theoretical details regarding Bayesian comparison of models based on alternative weight matrices, as well as applied examples. One can also use J−tests proposed by Kelejian and Piras [15] as illustrated in Gerkman and Ahlgren [16].

From the table we see the highest posterior model probability associated with $\gamma = 0.4, m = 6$, a result consistent with those reported by Kostov [6] based on his alternative "boosting" type of model search/comparison procedure.

An interesting question is—do these fine-tuning adjustments of the spatial weight matrix make a difference in terms of the estimates and inferences?

To explore this issue we produce estimates for models based on values of γ ranging from 0.2 to 1, in 0.2 increments and for $m = 5$ and $m = 6$.[11] We note that Table 4 indicates there is virtually no posterior probability support for models based on $m = 5$, so one might expect estimates based on $m = 5$ to differ greatly from those based on $m = 6$. The lack of posterior probability support is also evident in the table for models with weight matrices based on values of γ equal to 0, 0.9 and 1.0, when $m = 6$. One would typically not want to use weight structures having such low support from the sample data, but we use these here to make the point that estimates and inferences will not differ greatly even for these weight matrices.

Tables 5 and 6 show (posterior median) direct effects estimates for the ten variables constructed using a set of 2000 retained draws from Bayesian Markov Chain Monte Carlo estimation of the model (see LeSage and Pace [1], Chapter 6).[12] These are equivalent to median direct effects values constructed using simulated draws based on the maximum likelihood estimates and variance-covariance matrix. Estimates for both $m = 5$ and $m = 6$ are presented along with lower and upper 95% confidence intervals based on plus or minus two standard deviations. (The standard deviations are from the $m = 6$ model.)

From the tables, we see very little change in the (median) direct effects as values of the decay parameter vary from 0 to 1, despite the fact that there is little posterior probability support for models based on values of $\gamma = 0$ and $\gamma = 1$ (see Table 4). The direct effects estimates for models based on $m = 5$ and $m = 6$ are also remarkably similar, given there is virtually no support for models based on $m = 5$ (see Table 4). The estimates are within the lower and upper 95% confidence intervals for the model based on $\gamma = 0.4, m = 6$, which has the highest posterior model probability. This suggests no substantive changes in inference would arise from use of any of these weight matrix structures.

In the case of the Charles River dummy variable which is not significantly different from zero, we see a relatively dramatic change in the direct effects as we change weight matrices from $\gamma = 0$ to $\gamma = 1$. However, none of these effects magnitudes are different from zero given the lower and upper limits reported in the table.

---

[11] Decay values of $\gamma = 0$ were also calculated, but to save space and make the tables smaller this value was excluded from the tables.

[12] Pace, LeSage and Zhu [8] suggest using median effects estimates since the total effects can have a non-symmetric distribution. In our case with 506 observations the means and medians produced nearly identical results.

Table 5. Direct effects estimates, varying γ and m.

Direct Effects CRIME				
Decay	**-2σ**	**$m = 6$**	**$m = 5$**	**$+2\sigma$**
0.2	−0.0100	−0.0079	−0.0079	−0.0058
0.4	−0.0101	−0.0079	−0.0080	−0.0058
0.6	−0.0102	−0.0080	−0.0080	−0.0059
0.8	−0.0102	−0.0081	−0.0082	−0.0060
1	−0.0104	−0.0082	−0.0083	−0.0061

Direct Effects CHARLES				
Decay	**-2σ**	**$m = 6$**	**$m = 5$**	**$+2\sigma$**
0.2	−0.0374	0.0189	0.0263	0.0753
0.4	−0.0358	0.0198	0.0279	0.0754
0.6	−0.0336	0.0228	0.0286	0.0792
0.8	−0.0308	0.0263	0.0308	0.0835
1	−0.0271	0.0299	0.0342	0.0869

Direct Effects NOX^2				
Decay	**-2σ**	**$m = 6$**	**$m = 5$**	**$+2\sigma$**
0.2	−0.4767	−0.2889	−0.2879	−0.1010
0.4	−0.4908	−0.2966	−0.2927	−0.1023
0.6	−0.4951	−0.3085	−0.3147	−0.1219
0.8	−0.5136	−0.3190	−0.3252	−0.1245
1	−0.5184	−0.3267	−0.3365	−0.1350

Direct Effects ROOMS^2				
Decay	**-2σ**	**$m = 6$**	**$m = 5$**	**$+2\sigma$**
0.2	0.0051	0.0073	0.0074	0.0095
0.4	0.0052	0.0073	0.0074	0.0093
0.6	0.0052	0.0073	0.0074	0.0094
0.8	0.0052	0.0073	0.0074	0.0095
1	0.0053	0.0074	0.0074	0.0095

Direct Effects DISTANCE				
Decay	**-2σ**	**$m = 6$**	**$m = 5$**	**$+2\sigma$**
0.2	−0.1976	−0.1521	−0.1481	−0.1066
0.4	−0.1949	−0.1506	−0.1468	−0.1062
0.6	−0.1922	−0.1486	−0.1461	−0.1050
0.8	−0.1920	−0.1468	−0.1448	−0.1016
1	−0.1924	−0.1466	−0.1443	−0.1007

Table 6. Direct effects estimates, varying γ and m (continued).

		Direct Effects RAD		
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	0.0467	0.0766	0.0760	0.1065
0.4	0.0460	0.0769	0.0769	0.1078
0.6	0.0468	0.0776	0.0773	0.1083
0.8	0.0478	0.0783	0.0781	0.1088
1	0.0473	0.0791	0.0777	0.1109

		Direct Effects TAX		
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	−0.0005	−0.0003	−0.0003	−0.0001
0.4	−0.0005	−0.0003	−0.0003	−0.0001
0.6	−0.0005	−0.0003	−0.0003	−0.0001
0.8	−0.0005	−0.0003	−0.0003	−0.0001
1	−0.0005	−0.0003	−0.0003	−0.0002

		Direct Effects PTRATIO		
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	−0.0203	−0.0120	−0.0117	−0.0038
0.4	−0.0209	−0.0125	−0.0123	−0.0040
0.6	−0.0215	−0.0133	−0.0130	−0.0050
0.8	−0.0223	−0.0137	−0.0139	−0.0052
1	−0.0229	−0.0145	−0.0148	−0.0061

		Direct Effects B		
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	0.0001	0.0003	0.0003	0.0005
0.4	0.0001	0.0003	0.0003	0.0005
0.6	0.0001	0.0003	0.0003	0.0005
0.8	0.0001	0.0003	0.0003	0.0005
1	0.0001	0.0003	0.0003	0.0005

		Direct Effects LSTATUS		
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	−0.3059	−0.2641	−0.2619	−0.2223
0.4	−0.3019	−0.2603	−0.2592	−0.2186
0.6	−0.3015	−0.2594	−0.2578	−0.2173
0.8	−0.3000	−0.2581	−0.2571	−0.2161
1	−0.3002	−0.2574	−0.2584	−0.2146

Indirect effects are presented in Tables 7 and 8 in the same format used for the direct effects. We might expect to see indirect effects estimates that are slightly smaller (in absolute value terms) for models based on $m = 5$ neighbors weight matrices. This is because the scalar summary measures of the indirect effects reflect an average of spatial spillovers cumulated over all neighbors. Since $m = 5$ results in fewer

neighbors, this should have some impact on the cumulative indirect effects estimates. From the table, we see this is the case, but the differences are quite small.

Table 7. Indirect effects estimates, varying γ and m.

Indirect Effects CRIME				
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	−0.0094	−0.0073	−0.0068	−0.0052
0.4	−0.0094	−0.0073	−0.0067	−0.0052
0.6	−0.0091	−0.0069	−0.0065	−0.0048
0.8	−0.0088	−0.0067	−0.0063	−0.0046
1	−0.0086	−0.0064	−0.0060	−0.0042

Indirect Effects CHARLES				
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	−0.0393	0.0171	0.0226	0.0734
0.4	−0.0375	0.0181	0.0239	0.0737
0.6	−0.0369	0.0195	0.0230	0.0759
0.8	−0.0359	0.0212	0.0233	0.0784
1	−0.0341	0.0229	0.0245	0.0799

Indirect Effects NOX^2				
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	−0.4553	−0.2675	−0.2481	−0.0796
0.4	−0.4652	−0.2710	−0.2481	−0.0768
0.6	−0.4537	−0.2672	−0.2532	−0.0806
0.8	−0.4575	−0.2630	−0.2456	−0.0684
1	−0.4462	−0.2545	−0.2428	−0.0629

Indirect Effects ROOMS^2				
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	0.0046	0.0067	0.0064	0.0087
0.4	0.0046	0.0067	0.0063	0.0088
0.6	0.0042	0.0063	0.0060	0.0085
0.8	0.0040	0.0061	0.0057	0.0082
1	0.0036	0.0057	0.0054	0.0079

Indirect Effects DISTANCE				
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	−0.1855	−0.1406	−0.1275	−0.0957
0.4	−0.1823	−0.1366	−0.1249	−0.0909
0.6	−0.1745	−0.1290	−0.1182	−0.0834
0.8	−0.1664	−0.1209	−0.1111	−0.0754
1	−0.1582	−0.1130	−0.1036	−0.0679

Table 8. Indirect effects estimates, varying γ and m (continued).

Indirect Effects RAD				
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	0.0401	0.0708	0.0651	0.1016
0.4	0.0395	0.0701	0.0650	0.1007
0.6	0.0366	0.0674	0.0621	0.0983
0.8	0.0318	0.0636	0.0597	0.0955
1	0.0298	0.0612	0.0567	0.0926

Indirect Effects TAX				
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	−0.0005	−0.0003	−0.0003	−0.0001
0.4	−0.0005	−0.0003	−0.0003	−0.0001
0.6	−0.0005	−0.0003	−0.0003	−0.0001
0.8	−0.0004	−0.0003	−0.0003	−0.0001
1	−0.0004	−0.0003	−0.0002	−0.0001

Indirect Effects PTRATIO				
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	−0.0193	−0.0111	−0.0100	−0.0029
0.4	−0.0198	−0.0113	−0.0103	−0.0029
0.6	−0.0197	−0.0115	−0.0106	−0.0033
0.8	−0.0198	−0.0112	−0.0106	−0.0027
1	−0.0197	−0.0113	−0.0107	−0.0029

Indirect Effects B				
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	0.0001	0.0003	0.0003	0.0004
0.4	0.0001	0.0003	0.0003	0.0004
0.6	0.0001	0.0002	0.0002	0.0004
0.8	0.0001	0.0002	0.0002	0.0004
1	0.0000	0.0002	0.0002	0.0004

Indirect Effects LSTATUS				
Decay	-2σ	$m = 6$	$m = 5$	$+2\sigma$
0.2	−0.2876	−0.2458	−0.2283	−0.2039
0.4	−0.2804	−0.2388	−0.2215	−0.1971
0.6	−0.2668	−0.2247	−0.2097	−0.1827
0.8	−0.2545	−0.2126	−0.1976	−0.1706
1	−0.2416	−0.1988	−0.1870	−0.1560

As noted, the parameters β and ρ change in response to changes in the spatial weight matrix in an effort to produce consistent effects estimates (partial derivatives). To illustrate this point, we present a plot of the posterior median values for ρ for the 22 models based on m equal to 5 and 6 and γ ranging from 0 to 1 in Figure 1, where one should note the scale of the vertical axis ranges between 0.4 and 0.6 making the (small) changes appear large.

Figure 1. Variation in estimates for ρ over γ and m values.
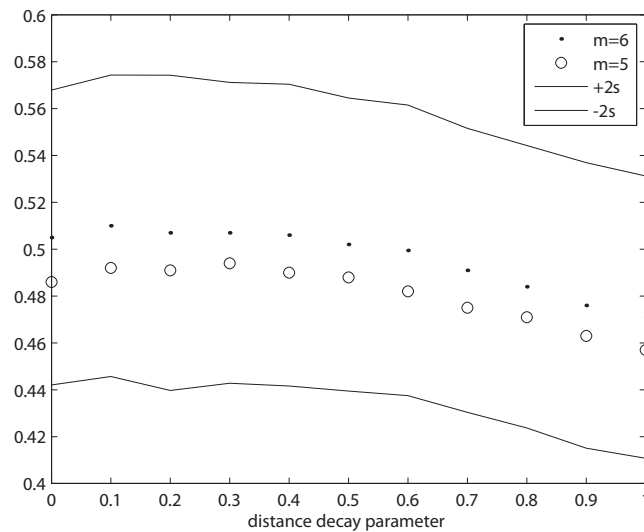
From the figure, we see fairly large variation in values for the spatial dependence parameter ρ in response to changes in the decay parameter γ and in the number of neighbors m used. The effects estimates remain relatively more stable as a result of changes in the coefficients β that offset the changes shown for the parameter ρ in the figure. This of course has lead practitioners to perceive sensitivity of estimates and inferences to the choice of weight matrix. However, as already noted this reflects an incorrect interpretation of the model estimates. For purposes of inference regarding response of the dependent variable to changes in the independent variables, the effects estimates are what is relevant, not the coefficients β and ρ.

In conclusion, our answer to the question—do the fine-tuning adjustments of the spatial weight matrix advocated by Kostov [6] make a difference in terms of the estimates and inferences?—is no. In the context of the research question addressed by Harrison and Rubinfeld [12] who constructed the data set, a much better question to ask is whether the spatial spillovers from NOX^2 pollution represent a situation that is more appropriately modeled using a *global* or *local* spillover specification. LeSage and Pace [17] motivate that distinguishing between these two types of situations that arise in applied modeling situations has a great deal of influence on how one interprets estimates and inferences.

In the case of NOX^2 pollution, it seems likely that the appropriate specification is a *local* spillover model which implies *contextual effects* rather than *endogenous interaction* between economic agents. The SAR specification represents a global spillover scenario and is likely inappropriate here.

3.4. A Diagnostic Example

For an illustration of how one might proceed in applied practice to explore the issue of sensitivity of estimates and inferences to the weight matrix specification, we use the voter turnout data set from Gilley and Pace [13]. The data set contains observations on 3107 US counties for the 1980 presidential election, where the dependent variable is votes cast as a proportion of population over age 19 eligible to vote. The explanatory variables are: *educ*, population with college degrees as proportion of population over age 19 eligible to vote, *homeownership*, homeownership as proportion of population over age 19

eligible to vote, and *income*, income per capita of population over age 19 eligible to vote. All variables were transformed using logs, so the effects estimates can be interpreted as approximate elasticities.

Table 9 shows posterior model probabilities for a number of alternative spatial weight matrices, including a contiguity-based matrix and (equally weighted) nearest neighbors ranging from 3 to 20.[13] Both SAR and SDM models were examined and they both produced probabilities that provide very strong evidence in favor of models based on 15 nearest neighbors.

Table 9. Posterior model probabilities

Model	SAR Model Posterior Probability	SDM Model Posterior Probability
W-contiguity	0.0000	0.0000
neighbors = 3	0.0000	0.0000
neighbors = 4	0.0000	0.0000
neighbors = 5	0.0000	0.0000
neighbors = 6	0.0000	0.0000
neighbors = 7	0.0000	0.0000
neighbors = 8	0.0000	0.0000
neighbors = 9	0.0000	0.0000
neighbors = 10	0.0000	0.0000
neighbors = 11	0.0000	0.0000
neighbors = 12	0.0000	0.0000
neighbors = 13	0.0000	0.0000
neighbors = 14	0.0093	0.0030
neighbors = 15	0.9905	0.9940
neighbors = 16	0.0003	0.0030
neighbors = 17	0.0000	0.0000
neighbors = 18	0.0000	0.0000
neighbors = 19	0.0000	0.0000
neighbors = 20	0.0000	0.0000

One might suppose that given such strong data evidence in favor of a 15 nearest neighbor weight matrix, estimates and inferences would be sensitive to an incorrect choice of neighbors. Figure 2 shows a plot of the direct effects estimates for both SAR and SDM models for models based on 10 to 20 nearest neighbors. In addition to the direct effects for the three explanatory variables *educ, homeowners, income*, a lower and upper three standard deviation confidence interval is shown in the figure. Figure 3 shows a similarly formatted plot for the indirect effects estimates.

---

[13] These can be produced using the *lmarginal_cross_section* function from the Spatial Econometrics Toolbox available at: www.spatial-econometrics.com.

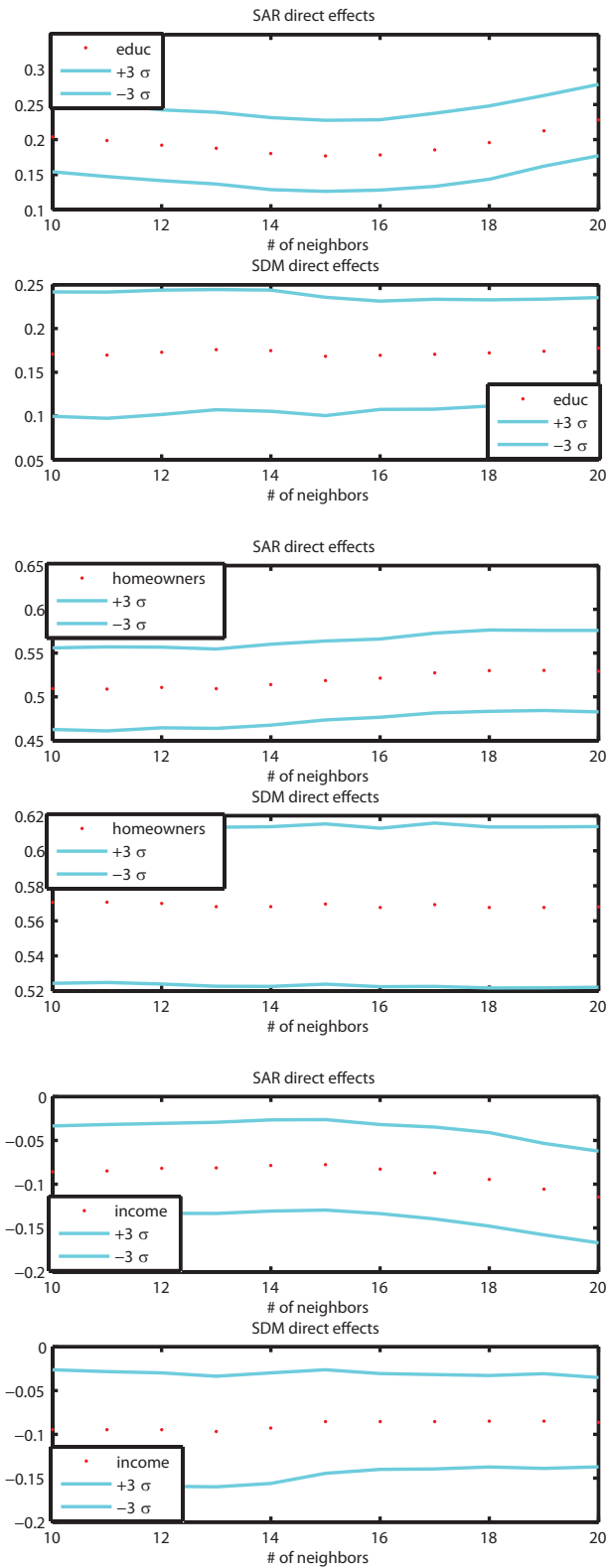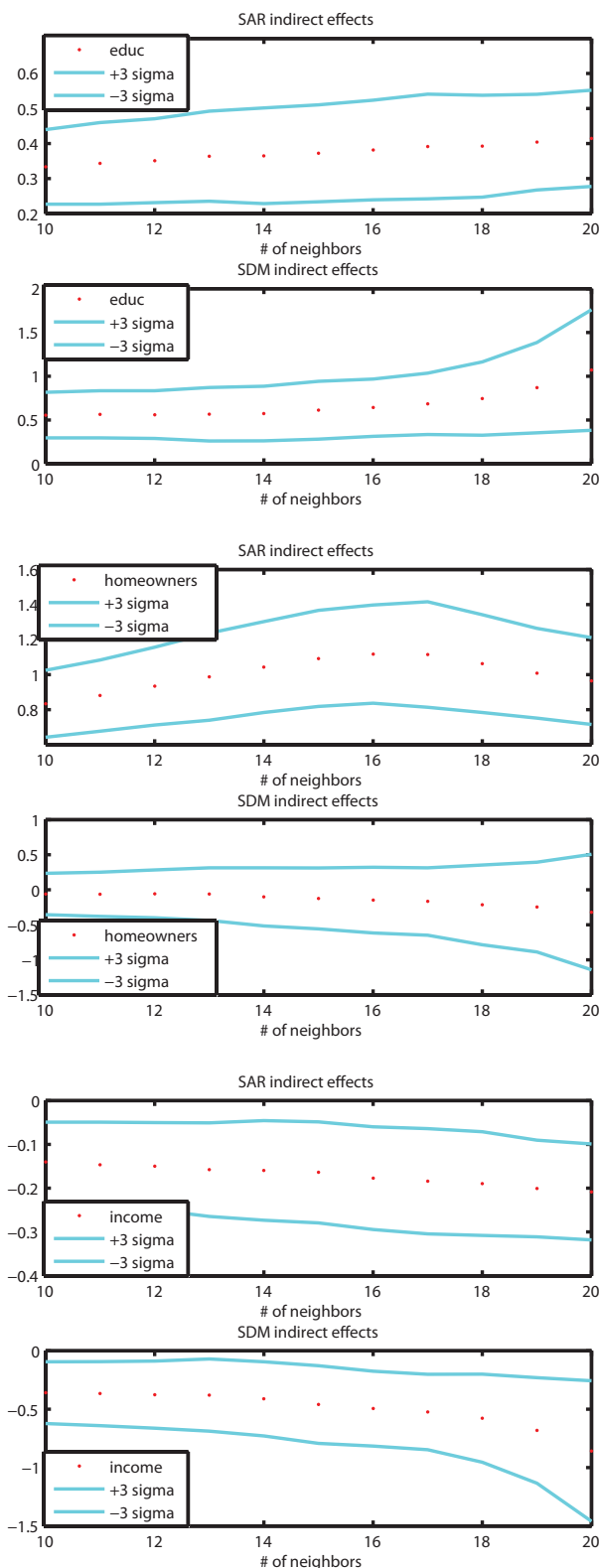**Figure 2.** Direct effects.

Figure 3. Indirect effects.

From the plots, we see that despite the very strong preference of the sample data for models based on 15 nearest neighbors, the direct effects estimates are relatively constant across the different models. In addition, the direct effects magnitudes for both the SAR and SDM models are similar. It appears that the SDM model consistently produces effects that are more stable as the number of neighbors used to

construct W vary. LeSage and Pace [1] provide an extensive development documenting the robustness and desirable statistical properties of the SDM model in applied modeling situations.

The indirect effects estimates in Figure 3 are also relatively constant across models with varying numbers of nearest neighbors, but we see some increase in the indirect effects for models based on 19 and 20 neighbors. Elhorst [18] provides a detailed discussion and simple illustration of important differences in the relative flexibility and sophistication of indirect effects for the SAR versus SDM models.

An interesting point is that differences between indirect effects arising from varying the number of neighbors is much smaller than the difference between the SAR and SDM indirect effects magnitudes. For example, the elasticity of voter turnout with respect to *educ* is around twice as large for the SDM as the SAR model. The *homeowner* variable has a large positive indirect effect (near unity) in the SAR model, but is not different from zero in the SDM model (based on the three standard deviation intervals). Finally, the (negative) SDM indirect effects for the *income* variable are around three times those of the SAR model.

4. Interaction between Specification of the Model and W Matrix

There are some issues pertaining to trade-offs between model versus weight matrix specifications. For example, model specifications such as the SARor popular SARAR (labeled SAC by LeSage and Pace [1]) that are not very flexible in accommodating spatial spillover effects might require more fine-tuning of the weight matrix specification. In contrast, a model specification that accounts for spatial spillovers in a flexible way might reduce the need to construct a weight matrix that closely replicates that from the data generating process.[14] In this section we provide a generated data example that illustrates these issues.

We begin with an example where we simulated 1000 random planar points and formed W_a based on the inverse distance between the points, with an absolute cutoff magnitude of 0.0025. By construction, W_a is symmetric. On average there were 49 neighbors per row (column) with a minimum of 16 neighbors per row (column) and a maximum of 82 neighbors per row (column). The resulting matrix was scaled to be doubly stochastic.

We used the SAR model DGP shown in Equation (35):

$$
\begin{aligned}
y &= (I_n - 0.5W_a)^{-1}X\beta + \varepsilon \\
X &\sim N(0, \sigma_x^2 I_n), \ \sigma_x = 1 \\
\varepsilon &\sim N(0, \sigma_\varepsilon^2 I_n), \ \sigma_\varepsilon = 0.1
\end{aligned}
\tag{35}
$$

The small value for the disturbance variance ($\sigma_\varepsilon^2 = 0.01$) was used to allow a focus on spatial spillovers arising in the model rather than noise in the disturbances. Given the autoregressive parameter value of 0.5, the total impacts equal $(1 - 0.5)^{-1} = 2$.

We produced model estimates using the sample data vectors y, X in conjunction with a misspecified extended variant of the SLX model in Equation (36):

$$
y = \alpha \iota_n + X\beta_1 + W_b X\beta_2 + W_b^2 X\beta_3 + W_b^3 X\beta_4 + \varepsilon
\tag{36}
$$

---

14 See Pace and Zhu [19] for a detailed discussion of this type of model specification.

where W_b is based on nearest neighbors and is symmetricized and scaled to be doubly stochastic. In this simple variant of the SLX model the total impacts correspond to the sum of the coefficients β, $(\sum_{i=1}^{4} \beta_i)$.

This example represents a case where *both* the weight matrix and spatial regression model are misspecified. Despite this two-fold misspecification, the model produced estimated total impacts shown in Table 10 that reasonably reflect the true values. The highest likelihood function value was associated with a model based on 18 nearest neighbors, which produced an estimated total impact magnitude of 1.94, very close to the true total impact of 2. Any choice of nearest neighbors in the range from 15 to 30 would have resulted in a total impact estimate between 1.9 and the true value 2.0.

Table 10. Estimation of Total Impacts Across W with Differing Number of Neighbors.

Neighbors	Total Impact	Likelihood
4	1.547	-1480.282
5	1.629	-1436.294
6	1.685	-1418.920
7	1.728	-1404.318
8	1.754	-1401.690
9	1.781	-1393.235
10	1.807	-1379.055
11	1.831	-1369.690
12	1.853	-1359.639
13	1.872	-1352.425
14	1.890	-1352.786
15	1.903	-1356.933
16	1.916	-1359.539
17	1.928	-1358.922
18	1.940	-1354.247
19	1.941	-1360.217
20	1.951	-1364.012
21	1.947	-1366.057
22	1.946	-1374.467
23	1.945	-1376.323
24	1.946	-1376.508
25	1.940	-1379.946
26	1.940	-1381.173
27	1.948	-1381.586
28	1.955	-1379.777
29	1.958	-1383.759
30	1.961	-1382.689

This simple example points to some more general issues regarding interaction between model and weight matrix specifications. To explore these, consider a situation where the data is generated by some unknown DGP as in Equation (37), but fitted with the model such as that in Equation (38) that

we label DFM (data fitted model). In Equations (37) and (38), $F(\cdot)$, $G(\cdot)$, $R(\cdot)$, and $S(\cdot)$ are distinct matrix functions (that could take the forms involving varying specifications of W such as $(I_n - \rho W)^{-1}$ (autoregressive), $e^{\alpha W}$ (matrix exponential), or $(I_n + \gamma W)(I_n - \rho W)^{-1}$ (ARMA(1,1)) where α, λ and ρ are scalar real parameters.

$$y = F(W_a)X\beta + R(W_a)\varepsilon \qquad (37)$$
$$y = G(W_b)X\theta + S(W_b)\varepsilon \qquad (38)$$

If the specification of the matrix function $G(W_b)$ is flexible enough, there may be some estimated parameter values where $F(W_a)X\beta$ equals $G(W_b)X\theta$ on average, as shown in Equation (39).

$$E\left[F(W_a)X\beta - G(W_b)X\theta\right] = 0_{n\times 1} \qquad (39)$$

We can think of Equation (39) as the need to have a high correlation between the expected values based on W_a in the context of the DGP and expected values based on W_b in the context of the fitted model. In addition, the fitted model could show similar scalar summary effects (estimates of average or median partial derivative measures of the impacts arising from changes in the K explanatory variables indexed by $k = 1, \ldots, K$) when using the correct specification, if the following were true:

$$E\left[F(W_a)\beta_t - G(W_b)\theta_t\right]\iota_n = 0_{n\times 1} \qquad (40)$$
$$\mathrm{tr}[E\left(F(W_a)\beta_t - G(W_b)\theta_t\right)] = 0_{n\times 1} \qquad (41)$$

Condition Equation (40) indicates that there is a need for high levels of correlation between the average total impact in the DGP using W_a and the DFM using W_b. In addition, condition Equation (41) indicates that there is a need for high levels of correlation between the average direct impact in the DGP using W_a and the DFM using W_b.

The ability of a model to match the magnitude of the levels and the effects estimates that form the basis for inference in spatial regression models in the face of misspecification of W would be enhanced by a more flexible matrix function of the estimated spillovers $G(W_b)$. Specifically, our applied model specification might rely on a more flexible matrix function for the spatial spillover component of the model than that which actually generated the data. In other words, if a spatial autoregressive process $(I_n - \rho W_a)^{-1}X\beta$ is thought to have generated the data it might be wise to use a spatial ARMA specification $(I_n + \gamma_1 W_b)(I_n - \gamma_2 W_b)^{-1}X\theta$ for the spillovers. Inflexible specifications for the spillovers in these models allow little leeway for the models to adjust to misspecification in W. Failure to appropriately model the spatial spillover component of the model can result in bias (violation of Equations (39)–(41)) in the estimates and inferences from these models.

This should not be construed as indicating that more flexible modeling of the disturbances would lead to improvements in these situations. Flexible modeling of the disturbances may improve the model fit, but does not necessarily help in estimating the spillover part of the model. In fact, since spillovers appear to play the most important role in producing estimates and inferences that are robust to both model and weight matrix misspecification, our conclusion is that one should endeavor for more flexibility in this aspect of the model rather than the disturbance part of the model. In other words, suppose the researcher must choose between fitting DFM A represented by $y = (I_n - \gamma_1 W_b)^{-1}X\theta + (I_n - \gamma_2 W_b)(I_n - \gamma_3 W_b)^{-1}\varepsilon$ which fits AR(1) spillovers and ARMA(1,1) disturbances or DFM B represented

by $y = (I_n - \gamma_1 W_b)^{-1}(I_n - \gamma_2 W_b)X\theta + (I_n - \gamma_3 W_b)^{-1}\varepsilon$ which fits ARMA(1,1) spillovers and AR(1) disturbances. Both DFM A and DFM B involve fitting three spatial parameters γ_1, γ_2, and γ_3 and so have more or less equal complexity. However, the DFM A devotes its efforts (two parameters) to fitting the disturbances and only one parameter to fitting the spillovers. In contrast, the DFM B devotes its efforts (two parameters) to fitting the spillovers and only one parameter to fitting the disturbances. Because incorrectly fitting the spillovers leads to such large subject matter consequences in terms of misintrepreting the spillovers, we claim that *ex-ante* it would be better to employ DFM B than DFM A, even if DFM A showed a better fit (which includes some consideration given to both spillovers and disturbances). Put in another way, suppose DFM A fits better than DFM B. A statistician would prefer DFM A on the basis of goodness-of-fit which, especially in the case with low signal-to-noise gives more weight to fitting the disturbances. However, an economist should prefer DFM B because it potentially provides a more valid estimate of the spillovers, even if it does not fit as well as DFM A. Essentially, statisticians and economists have different loss functions. These choices become clearer as n becomes large because the problems of inefficiency resulting from a poor fit of the disturbances become less important than the problems of bias resulting from a poor fit of the signal.

Although flexible models may allow recovery of the impacts, it is possible to construct counterexamples. For example, if the scaling of W_a is quite different from that of W_b, the approximation in Equations (39)–(41) may break down. Any matrix function composed of row or doubly stochastic matrices has fixed row sums. If the row sums for some other scaling of W_a differ dramatically by rows, most likely Equations (39)–(41) will break down.

A number of points can be made regarding scaling choices used in applied practice. First, there is an economic argument in favor of scalings where Wy reflects an average of neighboring observations. LeSage and Pace [1] provide an economic motivation for the existence of observed cross-sectional spatial dependence based on diffusion over time and space, beginning with Equation (42).

$$y_t = X\beta + \rho W y_{t-1} + \varepsilon_t \tag{42}$$

In this example, one may value a house based on explanatory variables such as size, age, or quality (captured by $X\beta$), as well some function of recent past prices in the neighborhood (captured by Wy_{t-1}). The parameter ρ can be viewed as governing the relative importance of previous price information versus explanatory variables.

The spatiotemporal specification in Equation (42) leads to an expected long-term steady state equilibrium for the cross-sectional specification shown in Equation (43) to Equation (45). Specifically, Equation (43) represents a first stage substitution of Equation (42), Equation (44) shows the expected value of the dependent variable based on an evolution of the dependent variable over q periods, and Equation (45) shows the expected value of the long-term cross-section of the dependent variable in period t.

$$y_t = X\beta + \rho W(X\beta + \rho W y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \tag{43}$$

$$E(y_t) = (I_n + \rho W + \rho^2 W^2 + \ldots + \rho^q W^q)X\beta + \rho^q W^q E(y_{t-q}) \tag{44}$$

$$\lim_{q \to \infty} E(y_t) = (I_n - \rho W)^{-1} X\beta \tag{45}$$

The expected long-run equilibrium expressed in Equation (45) is the same as the DGP of the SAR model and provides an economic motivation for the SAR model.

Given such an economic motivation for the SAR model, can the original spatiotemporal relation in Equation (42) help in specification of W for the SAR model? The economic motivation in our example reflects the fact that an individual estimates the price of their house using local prices, so the question becomes—what is the best specification for local prices? The most straightforward approach would be to take an average of local prices. If individuals use the average of neighboring house prices, this would dictate use of a row-stochastic or doubly-stochastic matrix W. Therefore, stochastic scaling of W does not arise *ex nihilo*, but comes from an assumption regarding the economic behavior of individuals.

This development is different than the statistical motivation used by Kelejian and Prucha [4], who begin by examining the variance-covariance or inverse variance-covariance matrices. From this perspective, many scalings produce row and column sums of the variance-covariance (VC) matrix of the disturbances (and y) that are uniformly bounded in absolute value. They note that uniform bounds on row and column sums has the virtue of limiting the degree of correlation between elements of the disturbance vector and y, which ensures decay of influence with higher order neighbors analogous to fading memory in time series. Of course, Kelejian and Prucha [4] note that this correlation exists in small samples, but is limited in large sample analysis. They argue against use of row-normalization unless "theoretical issues suggest a row-normalized weights matrix." The drawback of row-normalization in their view is that use of a different factor for the elements of each row means "there exists no corresponding re-scaling factor for the autoregressive parameter that would lead to a specification that is equivalent to that corresponding to the un-normalized weights matrix." Our view is that this purely statistical approach neglects behavioral assumptions used in the economic justification (based on behavioral assumptions) of the SAR DGP set forth in LeSage and Pace [1]. This view seems consistent with the point made by Arbia and Fingleton [2] in the introduction that: "ideally carefully structured arguments coming from theory and leading precisely to the typical reduced form spatial econometric model, with a spatial lag and exogenous lags also, are the preferred option." It is comforting that row-stochastic scaling emerges as the most frequently used approach in applied practice, since stochastic scaling has support from both theoretical and econometric standpoints.

A related point is that many specifications rely on a spatial parameter that appears in both the spillover and disturbance terms. Consequently, any misspecification in one term contaminates the other term. LeSage and Pace [1] in Chapter 7 provide examples of this. Pace and Zhu [19] provide a more detailed discussion of this point, and argue that in the presence of misspecification it seems more prudent to separately fit spillovers and disturbances.

In one sense the critics pointing to sensitivity of estimates and inferences to varying specifications for W are correct. If the desire is to produce estimates for each of the $n^2 k$ partial derivatives of $E(y)$ with respect to X (observation-level effects estimates) that are correct, spatial regression models could be viewed as sensitive to specification of individual elements of W. It does not seem realistic that one could produce estimates and inferences at the observation-level that are robust with respect to changes in

W matrix specification.[15] LeSage and Pace [1] argue against an observation-level approach to drawing inferences from spatial regression models. They argue that scalar summary measures of the impacts that average over all observations are more consistent with typical regression model uses and interpretation. If a goal of using spatial regression models is to have approximately correct scalar summary measures of the average direct, indirect and total effects on the dependent variable that arise from changes in the explanatory variables, this seems quite realizable. In fact, we argue there are a number of reasons to believe this will be the case in applied practice.

5. Conclusions

We conclude that a myth may have arisen in the spatial econometrics discipline. Sources of the myth are likely twofold. First, even in cases where W does not theoretically make a difference (such as in correctly specified spatial error models with large n), practitioners have observed sensitivity of the results to specification of W. In this setting, practitioners have incorrectly flagged W as the culprit rather than misspecification of the model. Second, practitioners have frequently misinterpreted spatial regression estimates in the case of models containing spatial lags of the dependent variable. In these models, the parameters β are often interpreted *as if* they are partial derivatives reflecting the *ceteris paribus* impact of changes in explanatory variables on the dependent variable. As discussed here and in LeSage and Pace [1], this is not the case. In fact, two models differing only in W that have similar *ceteris paribus* impacts from changing explanatory variables (partial derivatives), must have different values of ρ and β.

Another strand of the literature recognizes the role W plays in estimated partial derivatives from spatial regression models, but sets an impossible goal. If the requirement is that estimating $k + 1$ parameters will yield correct estimates for each of the n^2k partial derivatives of $E(y)$ with respect to X (observation-level effects estimates), spatial regression models are sensitive to specification of individual elements of W. It does not seem realistic that one could produce estimates and inferences at the observation-level that are robust with respect to changes in W matrix specification. To address this issue LeSage and Pace [1] argue against an observation-level approach to drawing inferences from spatial regression models. They argue that scalar summary measures of the impacts that average over all observations are more consistent with typical regression model uses and interpretation. If a goal of using spatial regression models is to have approximately correct scalar summary measures of the average direct, indirect and total effects on the dependent variable that arise from changes in the explanatory variables, this seems quite feasible. Using the empirical examples and the theoretical developments in this manuscript, we argue there are a number of reasons to believe this will be the case in applied practice.

Yet another strand of the literature would like to see more economic motivation for W. For example, Arbia and Fingleton [2] suggest that " . . . ideally carefully structured arguments coming from theory and leading precisely to the typical reduced form spatial econometric model, with a spatial lag and exogenous lags also, are the preferred option." However, there are two problems with this desired linkage between

---

[15] See Kelejian and Mukerji [20], Kelejian, Tavlas and Hondronyiannis [21] for examples of this type of observation-level inference.

W and economic theory. First, most variants of W would likely share common elements and, as shown here, this often makes the results from the various W more similar than different. If the estimates and inferences are not all that sensitive to the specific weight matrix used, it is difficult to see how current economic theory can shed light on a specific "ideal" weight matrix. Second, basing W on economic variables may lead to some forms of interaction between W and X that are difficult to detect. Moreover, interpretation of such W could prove difficult as elements of W may change with the explanatory variables. Weight matrices based on location have the great advantage of exogeneity. In addition, we reiterate an economic motivation from LeSage and Pace [1] for row-normalized weight specifications that produce averages of neighboring observations. LeSage and Pace [17] and LeSage [22] argue that theory can provide guidance regarding whether a *local* or *global* spatial regression specification is appropriate in many applied modeling situations. This type of distinction has frequently been ignored in applied work.

We do not mean to imply that any weight matrix will perform equally. Bayesian model comparison methods can easily be used to produce posterior model probabilities for any discrete set of weight matrices. As illustrated here, these are easy to interpret. Additionally, unlike AIC and other model fit criterion such as model $R-$squared statistics, posterior model probabilities are unconditional on the model parameters β and ρ which have been integrated out of the log-marginal likelihood. We do mean to imply that far too much effort has gone into "fine-tuning" spatial weight matrices that depend on highly parameterized functions of distance, lengths of common borders, and so forth. However, due to the number of common elements in these weight matrices and selection of parameters that give the best fit for each W, good fitting models using these different forms of W are not likely to produce estimates and inferences that materially differ.

Arbia and Fingleton [2] point to the need for "research on the robustness of outcomes to variations in assumptions about the weight matrix structure," indicating that this would overcome some of the criticisms of spatial regression models. The considerations in this manuscript provide a step in this direction. However, more needs to be done to characterize the notion of equivalence among W.

Acknowledgments

Author Contributions

Both authors contributed equally.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. LeSage, J.P.; Pace, R.K. *Introduction to Spatial Econometrics*; CRC Press: Boca Raton, FL, USA; Taylor & Francis: Boca Raton, FL, USA, 2009.

2. Arbia, G.; Fingleton, B. New spatial econometric techniques and applications in Regional Science. *Pap. Reg. Sci.* **2008**, *87*, 311–317.

3. Barry, R.; Pace, R.K. Monte Carlo estimates of the log determinant of large sparse matrices. *Linear Algebr. Appl.* **1999**, *289*, 41–54.

4. Kelejian, H.H.; Prucha, I.R. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *J. Econom.* **2010**, *157*, 53–67.

5. Anselin, L. Under the hood: Issues in the specification and interpretation of spatial regression models. *Agric. Econ.* **2002**, *27*, 247–267.

6. Kostov, P. Model boosting for spatial weighting matrix selection in spatial lag models. *Environ. Plann. B* **2010** *37*, 533–549.

7. Bell, K.P.; Bockstael, N.E. Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data. *Rev. Econ. Statist.* **2000**, *87*, 72–82.

8. Pace, R.K.; LeSage, J.P.; Zhu, S. Spatial Dependence in Regressors. In *Advances in Econometrics*; Fomby, T.B., Hill, R.C., Jeliazkov, I., Escanciano, J.C., Hillebrand, E., Series Eds.; Terrell, D., Millimet, D., Volume Eds.; Elsevier Science: Amsterdam, The Netherlands, 2012; Volume 30, pp. 257–295.

9. Kelejian, H.H.; Prucha, I.R. A Generalized Spatial Two-State Least Squares Procedure for estimating a Spatial Autogressive Model. *J. Real Estate Financ. Econ.* **1999**, *17*, 99–121.

10. Anselin, L. *Spatial Econometrics: Methods and Models*; Kluwer Academic: Dordrecht, The Netherlands, 1988.

11. Pace, R.K.; LeSage, J.P. A Spatial Hausman Test. *Econ. Lett.* **2008**, *101*, 282–284.

12. Harrison, D.; Rubinfeld, D.L. Hedonic prices and the demand for clean air. *J. Environ. Econ. Manag.* **1978**, *5*, 81–102.

13. Gilley, O.W.; Pace, R.K. The Harrison and Rubinfeld Data Revisited. *J. Environ. Econ. Manag.* **1996**, *31*, 403–405.

14. Hepple, L.W. Bayesian model choice in spatial econometrics. In *Advances in Econometrics*; LeSage, J.P., Pace, R.K., Eds.; Elsevier Ltd.: Oxford, UK, 2004; Volume 18, pp. 101–126.

15. Kelejian H.H.; Piras, G. An extension of Kelejian's J-test for non-nested spatial models. *Reg. Sci. Urban Econ.* **2011**, *41*, 281–292.

16. Gerkman, L.M.; Ahlgren, N. Practical Proposals for Specifying k-Nearest Neighbours Weights Matrices. *Spat. Econ. Anal.* **2014**, *9*, 260–283.

17. LeSage, J.P.; Pace, R.K. Interpreting Spatial Econometric Models. In *Handbook of Regional Science*; Fischer, M.M., Nijkamp, P., Eds.; Springer: Berlin, Germany, 2014; pp. 1535–1552.

18. Elhorst, J.P. Applied Spatial Econometrics: Raising the Bar. *Spat. Econ. Anal.* **2010**, *5*, 9–28.

19. Pace, R.K.; Zhu, S. Separable spatial modeling of spillovers and disturbances. *J. Geogr. Syst.* **2012**, *14*, 75–90.

20. Kelejian, H.H.; Mukerji, P. Important dynamic indices in spatial models. *Pap. Reg. Sci.* **2011**, *90*, 693–702.

21. Kelejian, H.H.; Tavlas, G.S.; Hondronyiannis, G. A Spatial Modeling Approach to Contagion Among Emerging Economies. *Open Econ. Rev.* **2006**, *17*, 423–442.

22. LeSage, J.P. Spatial econometric panel data model specification: A Bayesian approach. *Spat. Stat.* **2014**, *9*, 122–145.

23. Lee, S.-Y. Three essays on spatial econometrics and empirical industrial organization. Ph.D. Dissertation, The Ohio State University, Columbus, OH, USA, 2008; p. 131. ProQuest Dissertation Publication 3321366.

24. Pace, R.K.; Barry, R.P. Quick computation of spatial autoregressive estimators. *Geogr. Anal.* **1997**, *29*, 232–246.

25. Pace, R.K.; LeSage, J.P. Omitted variables biases of OLS and spatial lag models. In *Progress in Spatial Analysis: Theory and Computation, and Thematic Applications*; Páez, A., Le Gallo, J., Buliung, R., Dall'Erba, S., Eds.; Springer: Berlin, Germany, 2010; pp. 17–28.