JOURNAL OF REGIONAL SCIENCE, VOL. 48, NO. 5, 2008, pp. 941–967

SPATIAL ECONOMETRIC MODELING OF ORIGIN-DESTINATION FLOWS*

James P. LeSage
McCoy Endowed Chair of Urban and Regional Economics, Texas State University, San Marcos, Department of Finance and Economics, San Marcos, TX 78666.
E-mail: jlesage@spatial-econometrics.com

R. Kelley Pace
LREC Endowed Chair of Real Estate, Department of Finance, E.J. Ourso College of Business Administration, Louisiana State University, Baton Rouge, LA 70803-6308.
E-mail: kelley@pace.am

ABSTRACT. Standard spatial autoregressive models rely on spatial weight structures constructed to model dependence among n regions. Ways of parsimoniously modeling the connectivity among the sample of $N = n^2$ origin-destination (OD) pairs that arise in a closed system of interregional flows has remained a stumbling block. We overcome this problem by proposing spatial weight structures that model dependence among the N OD pairs in a fashion consistent with standard spatial autoregressive models. This results in a family of spatial OD models introduced here that represent an extension of the spatial regression models described in Anselin (1988).

1. INTRODUCTION

Spatial regression models have served as the workhorse in applied spatial econometric analysis, and the models introduced here can play an important role in modeling interregional flows. The focus of this study is to provide specifics regarding how spatial regression methods can be applied to spatial interaction models. Work by Porojon (2001) for the case of international trade flows and Lee and Pace (2004) for retail sales pointed out that residuals from conventional models were found to exhibit spatial dependence. Others such as Tiefelsdorf (2003) have noted that assuming independence of individual flows from origin i to destination j and from any pair of regions to other pairs of regions may be problematical. Nonetheless, empirical work still relies heavily on the assumption of independence among origin-destination (OD) flows.

*The authors would like to acknowledge Manfred M. Fischer, Randall W. Jackson, and Tony E. Smith for helpful discussions on this topic, as well as participants at the Fourth Spatial Econometrics and Statistics Workshop held in June, 2005 in Toulouse, and the Regional Science Association International Meetings in Las Vegas, NV held in November, 2005, and the March, 2006 Southern Regional Science Association Meetings in St. Augustine, FL, as well as an anonymous reviewer.

Received: April 2007; revised: December 2007; accepted: January 2008.

DOI: 10.1111/j.1467-9787.2008.00573.x

Gravity models have often been used to explain OD flows that arise in fields such as trade, transportation, and migration. However, the gravity model assumes independence among observations, and this assumption seems heroic for many fundamentally spatial problems. We extend the traditional gravity model using a combination of three spatial connectivity matrices for origin, destination, and origin-to-destination dependence as well as provide new technical results that greatly simplify maximum-likelihood estimation of the model.

This paper sets forth spatial econometric methods for modeling these general data structures that arise in a variety of economic, geography and regional science research contexts such as international trade flows, migration research, transportation, network, freight flow analysis, communications and information flow research, journey-to-work studies, as well as regional and interregional economic modeling. The term "spatial interaction models" has been used in the literature to label models that focus on flows between origins and destinations (Sen and Smith, 1995). A large literature on theoretical foundations for these models in the context of international trade models exist (see Anderson, 1979; Anderson and van Wincoop, 2004). These models rely on a function of the distance between an origin and destination as well as explanatory variables pertaining to characteristics of both origin and destination regions. Spatial interaction models assume that using distance as a variable will eradicate the spatial dependence among the sample OD pairs.

The notion that use of distance functions in conventional spatial interaction models effectively capture spatial dependence in interregional flows has long been challenged. Griffith (2007) provides an historical review of regional science literature on this topic in which he credits Curry (1972) as the first to conceptualize the problem of spatial dependence in flows. Griffith and Jones (1980, p. 190) in a study of Canadian journey-to-work flows noted that flows from an origin are "enhanced or diminished in accordance with the propensity of emissiveness of its neighboring origin locations." They also stated that flows associated with a destination are "enhanced or diminished in accordance with the propensity of attractiveness of its neighboring destination locations."

In contrast to typical spatial econometric models where the sample involves n regions, with each region being an observation, these models involve $n^2 = N$ OD pairs with each OD pair being an observation. This type of modeling seeks to explain variation in the level of flows across the sample of N OD pairs. There has been widespread recognition of the need for such models in disciplines such as population migration. Cushing and Poot (2003, p. 317) provide a survey of migration research in which they state that:

> As noted in the Introduction, no one has as yet seriously exploited the potential of spatial econometrics in the migration literature. This would seem to be a natural extension for migration research and one with potentially greater importance at greater levels of geographic disaggregation. A more complete consideration of the spatial dimension in migration research is one of the key contributions that regional science can make to this literature.

We introduce maximum-likelihood estimation procedures for these models. New technical results regarding calculation of log determinants that appear in the likelihood function greatly facilitate estimation. A family of spatial econometric model specifications is illustrated with a stylized migration example based on flows for the 48 contiguous U.S. states and the District of Columbia. The results indicate spatial dependence exists among the OD flows, contrary to the conventional assumption of independence.

One caveat for the maximum-likelihood method introduced here is that in cases where a large number of zero flows exist, these methods are not appropriate. Maximum-likelihood estimates require that the dependent variable vector follow a normal distribution or that it can be suitably transformed to achieve normality. There are numerous cases where these conditions are met, for example, the flow of airline passengers between airport nodes in the network where airport city-pair combinations are treated as OD pairs. Traffic networks provide another example, and we note that sparse flows measured at a finer spatial scale that do not meet the requirement of normality will become denser when measured over longer time periods, making application of these methods possible. When dealing with count data when the mean count takes on a large value, transformations to approximate normality are described in Sen and Smith (1995). For an example of a Poisson model that deals with the case of a large number of zero flows see LeSage, Fischer and Scherngell (2007), where spatially structured origin and destination effects parameters are used to model spatial dependence in this type of setting.

Section 2 introduces the notation, and develops a general gravity model with spatial dependence. Section 3 sets forth means for estimating the spatial gravity model that includes some new log-determinant results. The primary focus here is methodological, but Section 4 presents a stylized illustration using migration data, and Section 5 concludes with directions for future research.

2. INTERREGIONAL FLOWS IN A SPATIAL REGRESSION CONTEXT

We introduce the notation and conventions used in describing OD flows in Section 2.1, and discuss modeling these with conventional gravity models with independent observations in Section 2.2. In Section 2.3 we motivate spatial dependence in an OD setting. Section 2.4 describes various specifications for spatial dependence in flows, and spatial econometric model specifications based on these dependence structures are described in Section 2.5.

Origin-Destination Notation and Ordering

Let $\mathbf{Y}$ denote an n by n square matrix of interregional flows from each of the n origin regions to each of the n destination regions where the n columns represent different origins and the n rows represent different destinations as shown in (1). The flows considered here reflect a closed system.

$$
\begin{array}{c c}
 & \begin{array}{cccc} o_1 & o_2 & \cdots & o_n \end{array} \\
\begin{array}{c} d_1 \\ d_2 \\ \vdots \\ d_n \end{array} &
\begin{pmatrix}
o_1 \to d_1 & o_2 \to d_1 & \cdots & o_n \to d_1 \\
o_1 \to d_2 & o_2 \to d_2 & \cdots & o_n \to d_2 \\
\vdots & \vdots & \ddots & \vdots \\
o_1 \to d_n & o_2 \to d_n & \cdots & o_n \to d_n
\end{pmatrix}
\end{array}. \tag{1}
$$

Accordingly, $n^{-1}\mathbf{Y}\,\iota_n$ is an n by 1 vector representing an average of the flows from all of the n origins to each of the n destinations, where ι_n is an n by 1 vector of ones. Similarly, $n^{-1}\mathbf{Y}'\iota_n$ would produce an n by 1 vector that is an average of flows from all of the n destinations to each of the n origins.

We can produce an $N(=n^2)$ by 1 vector of these flows from the flow matrix in (1) in two ways, one reflecting an *origin-centric* ordering as in (2), and the other reflecting a *destination-centric* ordering as in (3).

$$
\begin{array}{ccc}
l^{(o)} & o^{(o)} & d^{(o)} \\
1 & 1 & 1 \\
\vdots & \vdots & \vdots \\
n & 1 & n \\
\vdots & \vdots & \vdots \\
N-n+1 & n & 1 \\
\vdots & \vdots & \vdots \\
N & n & n,
\end{array} \tag{2}
$$

$$
\begin{array}{ccc}
l^{(d)} & o^{(d)} & d^{(d)} \\
1 & 1 & 1 \\
\vdots & \vdots & \vdots \\
n & n & 1 \\
\vdots & \vdots & \vdots \\
N-n+1 & 1 & n \\
\vdots & \vdots & \vdots \\
N & n & n.
\end{array} \tag{3}
$$

The indices $l^{(o)}, l^{(d)}$ denote the overall index from $1, \ldots, N$ for the origin-centric and destination-centric orderings, while the origin, destination indices o, d go from $1, \ldots, n$. Beginning with a matrix $\mathbf{Y}$ whose columns reflect origins and rows destinations, we obtain the origin-centric ordering with $y = vec(\mathbf{Y})$, and the destination-centric ordering by setting $y^{(d)} = vec(\mathbf{Y}')$. These two orderings are related by the vec-permutation matrix so that $Py = y^{(d)}$, and by the properties of permutation matrices $y = P^{-1}y^{(d)} = P'y^{(d)}$. For most of the discussion, we will focus on the origin-centric ordering where the first n elements in the stacked vector y reflect flows from origin 1 to all n destinations. The last n elements of this vector represent flows from origin n to destinations 1 to n.

Gravity Models with Independent Observations

A conventional gravity model least-squares regression approach to explaining variation in the vector of OD flows relies on an n by k matrix of explanatory variables that we label $\mathbf{X}$, containing k characteristics for each of the n regions. Without loss of generality, let each column of $\mathbf{X}$ have a mean of 0 (mean-differences). Given the format of the vector y, where observations 1 to n reflect flows from origin 1 to all n destinations, the matrix $\mathbf{X}$ would be repeated n times to produce an N by k matrix representing destination characteristics that we label $\mathbf{X_d}$. We note that $\mathbf{X_d}$ equals $\iota_n \otimes \mathbf{X}$. A second matrix can be formed to represent origin characteristics that we label $\mathbf{X_o}$. This would repeat the characteristics of the first region n times to form the first n rows of $\mathbf{X_o}$, the characteristics of the second region n times for the next n rows of $\mathbf{X_o}$ and so on, resulting in an N by k matrix that we label $\mathbf{X_o} = \mathbf{X} \otimes \iota_\mathbf{n}$. The distance from each origin to each destination is also included as an explanatory variable vector in the gravity model. We let $\mathbf{G}$ represent the n by n matrix of distances between origins and destinations, and thus $g = vec(\mathbf{G})$ is an N by 1 vector of these distances from each origin to each destination formed by stacking the columns of the OD distance matrix into a variable vector (since $\mathbf{G}$ is symmetric, $g = vec(\mathbf{G}')$ would give the same result). Again, without loss of generality let g have a mean of 0, which can be achieved by transforming this vector to deviation from means form. This results in a regression model of the type shown in (4).[1]

$$y = \alpha\iota_N + \mathbf{X_d}\beta_\mathbf{d} + \mathbf{X_o}\beta_\mathbf{o} + \gamma\mathbf{g} + \varepsilon. \tag{4}$$

In (4), the explanatory variable matrices $\mathbf{X_d}, \mathbf{X_o}$ represent N by k matrices containing destination and origin characteristics, respectively, and the associated k by 1 parameter vectors are β_d and β_o. The scalar parameter γ reflects the effect of distance g, and α denotes the constant term parameter on ι_N. The N by 1 vector $\varepsilon = vec(\mathbf{E})$ represents disturbances and we assume $\varepsilon \sim N(0, \sigma^2\mathbf{I_N})$, but generalizations to the case of spatial dependence will be discussed in Section 2.5.

Some elementary manipulations of the moment matrix can illuminate the simplicity of gravity models based on the independence assumption for the case of a square matrix where each origin is also a destination. The algebra

[1]If one starts with the standard gravity model and applies a log-transformation, the resulting structural model takes the form of (4) (c.f., equation (6.4) in Sen and Smith, 1995).

of Kronecker products can be used to form moment matrices without dealing directly with N by N matrices. Given arbitrary, conformable matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, then $(\mathbf{C}' \otimes \mathbf{A})\ \mathbf{vec}(\mathbf{B}) = \mathbf{vec}(\mathbf{ABC})$ (Horn and Johnson, 1991, p. 255, Lemma 4.3.1). Using $\mathbf{Z} = (\iota_N \quad \mathbf{X_d}\ \ \mathbf{X_o}\ \ g)$ yields the moment matrix shown in (5).

$$\mathbf{Z}'\mathbf{Z} = \begin{pmatrix} N & 0 & 0 & 0 \\ 0 & n\mathbf{X}'\mathbf{X} & 0 & \mathbf{X}'\mathbf{G}\iota_n \\ 0 & 0 & n\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{G}\iota_n \\ 0 & \iota_n'\mathbf{G}'\mathbf{X} & \iota_n'\mathbf{G}'\mathbf{X} & tr(\mathbf{G}^2) \end{pmatrix}. \tag{5}$$

Due to mean-centering of $\mathbf{X}$ and g, many of the entries in (5) are 0. We note that $tr(\mathbf{G}^2)$ can be efficiently calculated using $\iota_n'(\mathbf{G}' \odot \mathbf{G})\iota_n$, where the operator $\odot$ refers to elementwise (Hadamard) multiplication. Similarly, $\mathbf{Z}'y$ can be expressed as

$$\mathbf{Z}'\mathbf{y} = \left(\iota_n'\mathbf{Y}\iota_\mathbf{n} \quad \mathbf{X}'\mathbf{Y}\iota_\mathbf{n} \quad \mathbf{X}'\mathbf{Y}'\iota_\mathbf{n} \quad tr(\mathbf{GY})\right), \tag{6}$$

where computing $\iota'_n(\mathbf{G}' \odot \mathbf{Y})\iota_n$ would be more efficient than $tr(\mathbf{GY})$.

The moment matrices can be rewritten to obtain some insights into the least-squares estimator, $\hat{\beta} = (\frac{\mathbf{Z}'\mathbf{Z}}{\mathbf{N}})^{-1}(\frac{\mathbf{Z}'y}{\mathbf{N}})$, for this model.

$$\frac{\mathbf{Z}'\mathbf{Z}}{\mathbf{N}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{\mathbf{X}'\mathbf{X}}{\mathbf{n}} & 0 & \frac{\mathbf{X}'}{\mathbf{n}}\frac{\mathbf{G}\iota_n}{\mathbf{n}} \\ 0 & 0 & \frac{\mathbf{X}'\mathbf{X}}{\mathbf{n}} & \frac{\mathbf{X}'}{\mathbf{n}}\frac{\mathbf{G}\iota_\mathbf{n}}{\mathbf{n}} \\ 0 & \frac{\iota_n'\mathbf{G}'}{\mathbf{n}}\frac{\mathbf{X}}{\mathbf{n}} & \frac{\iota_n'\mathbf{G}'}{\mathbf{n}}\frac{\mathbf{X}}{\mathbf{n}} & \frac{tr\left(\mathbf{G}^2\right)}{N} \end{pmatrix}$$

$$\frac{\mathbf{Z}'\mathbf{y}}{\mathbf{N}} = \left(\frac{\iota_\mathbf{n}'\mathbf{Y}\iota_\mathbf{n}}{\mathbf{N}} \quad \frac{\mathbf{X}'}{\mathbf{n}}\cdot\frac{\mathbf{Y}\iota_\mathbf{n}}{\mathbf{n}} \quad \frac{\mathbf{X}'}{\mathbf{n}}\cdot\frac{\mathbf{Y}'\iota_\mathbf{n}}{\mathbf{n}} \quad \frac{tr(\mathbf{GY})}{N}\right).$$

The quantities $\frac{\mathbf{X}'}{\mathbf{n}}\frac{\mathbf{G}\iota_\mathbf{n}}{\mathbf{n}}$ and $\frac{\iota_n'\mathbf{G}'}{\mathbf{n}}\frac{\mathbf{X}}{\mathbf{n}}$ measure the covariance between the explanatory variables and distance. In the case where every origin is also a destination, for any two regions i and j there are two OD pairs: an ij pair and ji pair. If the explanatory variable value is different for i and j, there will be two different explanatory variable values associated with the same distance for the ij and ji OD pairs, and this will also be the case for all $n(n-1)/2$ pairs where $i \neq j$. This should result in small covariances between the distance and explanatory variables. We note that sufficiently small values for $\frac{\mathbf{X}'}{\mathbf{n}}\frac{\mathbf{G}\iota_\mathbf{n}}{\mathbf{n}}$ and $\frac{\iota_n'\mathbf{G}'}{\mathbf{n}}\frac{\mathbf{X}}{\mathbf{n}}$ would result in a block diagonal $\mathbf{Z}'\mathbf{Z}$ where the inverse can be obtained block by block.

We can exploit the block diagonal structure that would arise if the origin, destination, and distance variables exhibit small covariances to produce an approximate estimate associated with the destination characteristics: $\hat{\beta}_d = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{n}^{-1}\mathbf{Y}\iota_\mathbf{n}]$, where $(n^{-1}\mathbf{Y}\ \iota_\mathbf{n})$ represents an average of the flows from all of the n origins to each of the n destinations. Similarly, an approximate estimate of the origin characteristics is: $\hat{\beta}_o = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{n}^{-1}\mathbf{Y}'\iota]$, where $n^{-1}\mathbf{Y}'\iota$ represents an average of flows from all of the n destinations to each of the n origins.

If the covariances between distance and explanatory variables are small, the approximate least squares estimate of the distance parameter γ is: $tr(\mathbf{G}Y)/tr(\mathbf{G}^2) = g'y/g'g$. This is the slope parameter estimate of the simple regression of y on g. Therefore, a low correlation of distance with the other explanatory variables implies that the addition of distance to this model would improve the fit (and thus possibly affect inference), but not affect estimates associated with the other explanatory variables.

In conclusion, the square ($n^2 = N$) gravity model under the assumption of independence between origin and destination flows has a very simple structure that relates average flows to the origin and destination characteristics. Moreover, distance is likely to have a material effect on the fit, but not on the estimates associated with the independent variables. The simplicity of the gravity model with independent observations may mean that it cannot account for the spatial richness of OD flows. To enhance this simple model, we augment it with richer forms of spatial dependence set forth in Section 2.5.

Spatial Dependence in Local Origin-Destination Flows

Although the gravity model makes an attempt at modeling interdependence among observations using distance, this may prove inadequate for many types of flow data where each region might affect its neighbors. For example, events such as new plant openings, retail outlets, or highway infrastructure, that occur in one region are thought by regional scientists to exert spatial spillover effects on neighboring regions. In the case of OD flows neighboring regions include neighbors to the origin, neighbors to the destination, and perhaps a link between neighbors of the origin and neighbors of the destination region.

Before addressing the general case with these forms of dependence, we develop the simpler case of a local OD model where we consider a single origin and n destinations or n origins and one destination. By themselves, local models have their uses. For example, a particular tourist destination might be interested in explaining the flows from the n origins or a university might wish to look at the placement of their graduates across n destinations. However, our main reason to discuss local OD models is to provide a simpler scenario for motivating spatial dependence. Our extension to the more general case with n^2 OD flows is set forth in the next section.

Both types of local OD models involve n observations, and thus provide situations more similar to traditional spatial econometric applications. In these applications, spatial weight matrices provide a convenient and parsimonious way to specify spatial dependence among observations. In a typical cross-sectional model with n regions with one observation per region, the spatial weight matrix labeled W represents an n by n nonnegative, sparse matrix. This matrix captures dependency relations among the observations (regions). For example,

$W_{ij} > 0$ if region i is contiguous to region j. Besides contiguity, various measures of proximity such as cardinal distance (e.g., kilometers), and ordinal distance (e.g., the 10 closest neighbors) have been used to specify **W**. By convention, $W_{ii} = 0$ to prevent an observation from being defined as a neighbor to itself, and the matrix **W** is typically standardized to have row sums of unity.

Given an origin-centric organization of the sample data, a local model for a single origin would represent an n by 1 vector of flows from a single origin i to all destinations, $j = 1, \ldots, n$. The conventional contiguity-based n by n matrix **W** would contain positive elements for neighbors to each of the regions treated as destinations. A specific example would be commuting flows from the central business district (CBD) region (the origin) to all other regions in the metropolitan area (the destinations), where our sample might reflect flows over a period of one month.

We provide two econometric motivations for the use of spatial regression models that involve spatial lags of the dependent variable. The first motivation comes from viewing spatial dependence as a long-run equilibrium of an underlying spatiotemporal process and the second motivation shows that omitted variables that exhibit spatial dependence leads to a model with spatial lags of both the explanatory and dependent variables.

The first motivation for spatial dependence in cross-sectional flows measured at a point in time is based on a time-lag relationship describing a diffusion process over space, which we show in (7), where the matrix **W** reflects a conventional spatial dependence structure between observations in the n by 1 vectors y_t, y_{t-1} of cross-sectional observations measuring the dependent variable at times t and $t-1$. For concreteness, these might be commuting flows from the CBD region to all other metropolitan regions during months t and $t-1$. The exogenous matrix **W** specifies the spatial configuration of the regions.

$$y_t = \rho \mathbf{W} y_{t-1} + \mathbf{X}\beta + \varepsilon_t. \tag{7}$$

In (7) we omit the time subscript on the matrix of explanatory variables **X** to reflect a situation where the explanatory variables reflecting regional characteristics that describe regional variation in y change slowly over time, relative to the change in flows. Continuing with our example of commuting flows, we might consider a single explanatory variable representing the (residential) population density of each destination region in the metropolitan area. This would change slowly given our monthly time frequency. Other candidate explanatory variables that might explain commuting flows such as the age composition, employment status, occupational composition of population residing in each of the metropolitan regions could also be argued to change slowly over the monthly time frame. The disturbance vector ε_t obeys the usual constant variance zero covariance assumptions from regression theory, and without loss of generality can be assumed to follow a normal distribution centered on zero.

We can use the recursive relation: $y_{t-1} = \rho \mathbf{W} y_{t-2} + \mathbf{X}\beta + \varepsilon_{t-1}$ implied by the model in (7) to consider the state of our dynamic system after passage of q time periods, which is shown in (8).

$$
\begin{aligned}
y_t &= \rho^q \mathbf{W}^q y_{t-q} + (\mathbf{I} + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \cdots + \rho^q\mathbf{W}^q)\mathbf{X}\beta + u \\
u &= \rho^q \mathbf{W}^q \varepsilon_{t-q-1} + \rho^{q-1}\mathbf{W}^{q-1}\varepsilon_{t-q-1} + \cdots + \varepsilon_t.
\end{aligned} \tag{8}
$$

The steady-state equilibrium for the dynamic process in (8) can be found by letting $q \to \infty$, and taking the expectation. If $\rho\mathbf{W}$ is such that $\rho^q\mathbf{W}^q \to 0$, and $(\mathbf{I} - \rho\mathbf{W})^{-1} = (\mathbf{I} + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \cdots)$, using $E(\varepsilon_{t-r}) = 0, r = 0, \ldots, q-1$, so that $E(u) = 0$, leads to (9).[2]

$$
E(y_t) = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\beta. \tag{9}
$$

This is the expectation for a spatial autoregressive model that contains spatial lags of the dependent variable, whose model expression is shown in (10) and associated data-generating process in (11).

$$
y_t = \rho\mathbf{W}y_t + \mathbf{X}_t\beta + \varepsilon_t, \tag{10}
$$

$$
y_t = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}_t\beta + (\mathbf{I} - \rho\mathbf{W})^{-1}\varepsilon_t. \tag{11}
$$

Since cross-sectional spatial autoregressive models provide no explicit role for passage of time, we should interpret these models as reflecting an equilibrium outcome or steady state. That is our cross-sectional slice of commuting flow observations from the CBD to other metropolitan area regions at a single point in time can viewed as sampling a dynamic system in steady state. This also has implications for how we should interpret the impact of changes in the explanatory variables of these models on the flows. The model literally states that a change in $\mathbf{X}$ will lead to a simultaneous impact on commuting flows to all (destination) regions in the metropolitan area represented by the cross-section y. However, it seems more intuitive to view changes in $\mathbf{X}$ as setting in motion a series of changes that will lead to a new steady-state equilibrium at some future unknown point in time.[3] Using our commuting flow example, the model could be used to quantify how changes in the current population density of the regions would impact commuting flows from the CBD to all (destination) regions in the metropolitan area in the long-run.

The second motivation for the presence of spatial lags in flows is based on an omitted variables argument. It is difficult to find sample data that adequately reflects amenities, social milieus, entrepreneurial spirit, and a host of other influences that may be important for a particular flow modeling problem. For the case of our commuting flow example, an omitted variable might be school quality of each destination region.

---

[2]The properties ascribed to the term $\rho\mathbf{W}$ hold for conventional spatial dependence structures, as well as the spatial dependence structure we will introduce for the flow model.

[3]Since our model is based on cross-sectional observations, it is uninformative about the time dimension.

A nonspatial regression relationship is shown in (12) where we assume the existence of a single omitted variable vector z. For simplicity we also assume a single vector x, without loss of generality (see Pace and LeSage (2008) for a more general development). The variable x included in our model of commuting flows would be residential population density of each region in the metropolitan area.

$$y = x\beta + z, \tag{12}$$

$$z = \rho \mathbf{W} z + u, \tag{13}$$

$$u = x\gamma + \varepsilon. \tag{14}$$

We model the omitted variable z as exhibiting spatial dependence which we represent using a spatial autoregressive process consisting of the scalar spatial dependence parameter ρ and the n by n spatial weight matrix W as shown in (13). Expression (14) indicates that the omitted and included variables are correlated when the scalar parameter $\gamma \neq 0$, which is the typical assumption made in the omitted variables literature. In the case of our commuting flow example, unobservable school quality z would be correlated with residential population density x of the metropolitan regions. The spatial process in (13) assigned to govern the omitted variable (school quality) suggests that school quality in each region of our metropolitan area is related to that of neighboring regions. For example, suburban regions school quality would be similar to that of neighboring regions as would be school quality in central city regions.

The assumption of spatial dependence in the omitted explanatory variable z is consistent with the findings by Porojon (2001) as well as Lee and Pace (2004) that residuals from conventional flow models were found to exhibit spatial dependence. We will show that an omitted variable correlated with the included variable x in the presence of spatial dependence leads to a model that contains a spatial lag of the dependent variable. Under these circumstances, omitting the spatial lag of the dependent variable as is done in conventional gravity models will lead to bias in the coefficient estimates. This provides a purely econometric motivation for use of a spatial lag model as protection against bias arising from possible omitted variables.

Using the relations in (12)–(14) we arrive at the data-generating process in (15).

$$y = x\beta + (\mathbf{I} - \rho \mathbf{W})^{-1} x\gamma + (\mathbf{I} - \rho \mathbf{W})^{-1}\varepsilon. \tag{15}$$

Transforming the left- and right-hand sides of (15) by $(I - \rho\mathbf{W})$ yields a transformed DGP in (16) with independent, identically distributed (iid) disturbances.

$$y = \rho \mathbf{W} y + x(\beta + \gamma) + \mathbf{W} x(-\rho\beta) + \varepsilon. \tag{16}$$

Note, the transformed DGP in (16) contains a spatial lag of the dependent variable ($\mathbf{W}y$) as well as the explanatory variable ($\mathbf{W}x$). In the case of our commuting flow example, $\mathbf{W}y$ represents an average of flows from the CBD to neighboring destination regions, and $\mathbf{W}x$ is the average population density of regions neighboring the destinations.

Given the DGP (15) and iid transformation of the DGP in (16), estimating the model (17) will yield estimates for large samples so that $\beta_1 = \beta + \gamma$ and $\beta_2 = -\rho\beta$.

$$y = \rho \mathbf{W}y + x\beta_1 + \mathbf{W}x\beta_2 + \varepsilon. \tag{17}$$

Let $\hat{\beta}_1, \hat{\beta}_2$, and $\hat{\rho}$ represent consistent estimates of β_1, β_2 and ρ so that for sufficiently large samples the estimates converge to the underlying parameters. Given $\hat{\rho}$, $\hat{\beta}_1$, and $\hat{\beta}_2$, $\hat{\beta} = \hat{\beta}_2(\hat{\rho})^{-1}$, $\hat{\gamma} = \hat{\beta}_1 - \hat{\beta}$. This result suggests that if we are interested in the impact of population density on commuting flows from the CBD in the presence of an omitted variable such as school quality that is correlated with population density, we should rely on a model that includes a spatial lag of both the dependent and independent variable. This would allow us to produce consistent estimates of the model parameters and draw valid inferences about the impact of changes in population density on commuting flows from the CBD, in the long-run steady-state equilibrium sense described for our first example.

When the parameter $\gamma = 0$ so the included and excluded variables are not correlated, and the restriction $\hat{\beta}_2 = -\hat{\rho}\hat{\beta}_1$ holds, a spatial error model (SEM)emerges: $(y - \rho\mathbf{W}y) = (x - \rho\mathbf{W}x)\beta + \varepsilon$. This suggests an empirical test for the presence of omitted variables that are correlated with the included variables. A likelihood-ratio (LR) test based on log-likelihood function values from the error model and the spatial lag model tests the restriction $\hat{\beta}_2 = -\hat{\rho}\hat{\beta}_1$ for the coefficients on x and $\mathbf{W}x$. Of course, this restriction can only hold when the parameter $\gamma = 0$, indicating no omitted variables exist that are correlated with those included in the model.[4]

If the restriction ($\hat{\beta}_2 = -\hat{\rho}\hat{\beta}_1$) is consistent with the sample data, then least squares estimates of the regression parameters (which ignore the spatial lag term and the spatial lag of the explanatory variable) are unbiased (Anselin, 1988). In this situation, the omitted school quality variable z is uncorrelated with the included population density variable x, and the relation between population density x and the spatial lag of population density $\mathbf{W}x$ is such that spatial dependence arises only in the disturbance term in the model. However, least squares estimates would produce inconsistent inferential statistics unless $\rho = 0$. That is, measures of dispersion for the coefficient estimates would be wrong, even in large samples.

There is an asymmetric risk that arises with regard to incorrect model specification that also provides a motivation for use of a model that includes both spatial lags of the dependent and independent variables. To see this, consider

[4]We should note that a test of the restriction might show this to be inconsistent with the sample data even when $\gamma = 0$, so failing the test would not necessarily imply the presence of omitted variables. For example, failing the test could arise from spatial dependence in the dependent variable, requiring a spatial lag in the model.

the situation where $\rho = 0$. In this situation, including spatial lags of the explanatory and dependent variables will reduce estimation efficiency, but will not lead to bias. For sufficiently large samples, the problems arising from bias will dominate those arising from inefficiency. In contrast, when $\rho \neq 0$ omitting a spatial lag of the dependent variable will result in biased and inconsistent estimates. This suggests a strategy where we might rely on a model that includes spatial lags of the dependent and explanatory variables even if this seems counter to the underlying theory behind our model. This result is similar in spirit to the use of lagged dependent variables in time-series models to account for omitted variables. While this might not be the optimal solution to the problem, it provides a practical approach to the problem posed by omitted variables.

Our approach augments theoretical models developed in Fingleton and López-Bazo (2006), and Ertur and Koch (2007, 2008) that directly include spatial dependence structures in underlying theoretical economic relationships, which also give rise to regression models that contain spatial lags of the dependent variable. In the case of Ertur and Koch (2007), the model deals with knowledge flows, and Behrens, Ertur and Koch (2008) provide a theoretical model that gives rise to spatial lags in the context of international trade flows. A notable contrast of our development and these is that we do not assume at the outset any dependence in the vector of flows y. Many regional scientists, particularly those involved in migration modeling, object to assuming spatial dependence in migration flows, as this seems contrary to current theoretical models motivated by utility considerations. Our starting point is consistent with theoretical models that posit a nonspatial theoretical relationship underlying migration flows.

Specifying Spatial Dependence in Flows

Having motivated situations in which spatial lags of the dependent variable might be useful when modeling OD flows, we turn attention to specification of the spatial structure of dependence for these models. We begin with a typical row-standardized n by n first-order contiguity or m nearest neighbors weight matrix $\mathbf{W}$ that reflects relations among the n regions. Referring to (2), $\mathbf{Y}_1$ contains the OD flows for the first origin and thus $\mathbf{W}Y_1$ provides the spatial average around each destination $i = 1, \ldots, n$ holding the origin constant. This can be repeated n times leading to a Kronecker product representation $\mathbf{I}_n \otimes \mathbf{W}$. We label this N by N row-standardized spatial weight matrix $\mathbf{W}_d$, shown in (18), where $\mathbf{0}_n$ represents an n by n matrix of zeros.

$$\mathbf{W}_d = \begin{pmatrix} \mathbf{W} & \mathbf{0}_n & \ldots & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{W} & \mathbf{0}_n & \vdots \\ \vdots & \mathbf{0}_n & \ddots & \mathbf{0}_n \\ \mathbf{0}_n & \ldots & \mathbf{0}_n & \mathbf{W} \end{pmatrix}. \tag{18}$$

The spatial weight matrix $\mathbf{W}_d$ will be used to capture *destination-based dependence* in the gravity model. This type of dependence reflects the intuition that forces leading to flows from an origin to a destination may create similar flows to nearby or neighboring destinations. Using the spatial weight matrix $\mathbf{W}_d = \mathbf{I}_n \otimes \mathbf{W}$ we can model this type of dependence. The N by N spatial weight matrix $\mathbf{W}_d$ reflects connectivity relations between an origin and neighbors of the destination. This matrix in conjunction with the vector y produces a spatial lag vector $\mathbf{W}_d y$ that formally captures the notion of Griffith and Jones (1980). They note that flows associated with a destination are "enhanced or diminished in accordance with the propensity of attractiveness of its neighboring destination locations."[5]

Reasoning similar to that used in developing the matrix $\mathbf{W}_d$, suggests that $\mathbf{W}\,(\mathbf{Y}')_1$ provides spatial averages around each origin for the first destination. Doing this for all destinations yields $\mathbf{WY}'$. Since $vec(\mathbf{WY}') = (\mathbf{W} \otimes \mathbf{I}_n)vec(\mathbf{Y})$, we can also create an N by N row-standardized spatial weight matrix that we label $\mathbf{W}_o = \mathbf{W} \otimes \mathbf{I}_n$. Using this matrix to form a spatial lag of the dependent variable, $\mathbf{W}_o y$ captures *origin-based spatial dependence* relations using an average of flows from neighbors to each origin region to each of the destinations. Intuitively, forces leading to flows from any origin to a particular destination region may create similar flows from neighbors to this origin to the same destination, and the spatial lag $\mathbf{W}_o y$ captures this effect. This formally captures the point of Griffith and Jones (1980) that flows from an origin are "enhanced or diminished in accordance with the propensity of emissiveness of its neighboring origin locations."[6]

We can use the spatial weight matrices $\mathbf{W}_d$, $\mathbf{W}_o$ in conjunction with scalar parameters ρ_d, ρ_o that indicate the strength of dependence to replace our previous $\rho\mathbf{W}$ specification for the case where we had only n observations containing local flow data. The new spatial structure $\rho_d\mathbf{W}_d$ reflects destination-based dependence and $\rho_o\mathbf{W}_o$ captures origin-based dependence. Since both types of dependence are likely to exist in the context of OD flows, this suggest a model shown in (19).[7]

[5]They also discuss formation of spatial lags for these models without using the Kronecker product relations set forth here.

[6]We note that the vec-permutation matrix $\mathbf{P}$ introduced previously in Section 2.1 can be used to translate between origin-centric and destination-centric ordering of the sample data. For example, if we adopted the destination-centric ordering (as opposed to the origin-centric ordering we use in the text), specification of the destination weight matrix would be $\mathbf{W}_d = \mathbf{W} \otimes \mathbf{I}_n$. This can be seen using the relation: $\mathbf{P}'\mathbf{W}_d\mathbf{P} = \mathbf{P}'\,(\mathbf{I}_n \otimes \mathbf{W})\mathbf{P}$, to translate the origin-centric destination weight matrix W_d to the destination-centric ordering scheme. Rules for multiplication using Kronecker products allow us to simplify the expression $\mathbf{P}'(\mathbf{I}_n \otimes \mathbf{W})\mathbf{P}$, using corollary 4.3.10 in Horn and Johnson (1991, p. 290), so that $\mathbf{P}'(\mathbf{I}_n \otimes \mathbf{W})\mathbf{P} = \mathbf{W} \otimes \mathbf{I}_n$, and thus $\mathbf{W}_d = \mathbf{W} \otimes \mathbf{I}_n$, under the destination-centric ordering of the sample data.

[7]The time lag motivation for the existence of spatial lags of the dependent variable could be developed by starting with: $y_t = \rho_d\mathbf{W}_d y_{t-1} + \rho_o\mathbf{W}_o y_{t-1} - \rho_d\rho_o\mathbf{W}_d\mathbf{W}_o y_{t-1} + \mathbf{X}\beta + \varepsilon_t$. Similarly, one could use an omitted variable development to motivate the use of spatial lags of flows.

$$(\mathbf{I}_N - \rho_d \mathbf{W}_d)(\mathbf{I}_N - \rho_o \mathbf{W}_o)y = \mathbf{X}\beta + \varepsilon. \tag{19}$$

Expanding the product $(\mathbf{I}_N - \rho_d \mathbf{W}_d)(\mathbf{I}_N - \rho_o \mathbf{W}_o) = \mathbf{I}_N - \rho_d \mathbf{W}_d - \rho_o \mathbf{W}_o + \rho_d \rho_o \mathbf{W}_d \cdot \mathbf{W}_o = \mathbf{I}_N - \rho_d \mathbf{W}_d - \rho_o \mathbf{W}_o - \rho_w \mathbf{W}_w$, leads us to consider a third type of dependence reflected in the product $\mathbf{W}_w = \mathbf{W}_o \cdot \mathbf{W}_d = (\mathbf{I}_n \otimes \mathbf{W}) \cdot (\mathbf{W} \otimes \mathbf{I}_n) = \mathbf{W} \otimes \mathbf{W}$.[8] This spatial weight matrix reflects an average of flows from neighbors to the origin to neighbors of the destination, which we label *origin-to-destination dependence* to distinguish it from *origin dependence* and *destination dependence*.

The model in (19) could be viewed as a successive spatial filter. Remarkably, one can filter successively by $(\mathbf{I}_N - \rho_d \mathbf{W}_d)$, and $(\mathbf{I}_N - \rho_o \mathbf{W}_o)$ or *vice-versa* and obtain the same results. The order does not matter because the cross-product of $\mathbf{W} \otimes \mathbf{I}_n$ and $\mathbf{I}_n \otimes \mathbf{W}$ is $\mathbf{W} \otimes \mathbf{W}$, which is the same as the cross-product of $\mathbf{I}_n \otimes \mathbf{W}$ and $\mathbf{W} \otimes \mathbf{I}_n$ via the mixed-product rule for Kronecker products. Therefore, in an OD context a successive spatial filter that removes destination dependence first and origin dependence second yields the same results as removing origin dependence first and destination dependence second.

Spatial Model Specifications for Origin-Destination Flows

We propose the following general spatial autoregressive model that takes into account origin, destination, and origin-to-destination dependence.

$$y = \rho_d \mathbf{W}_d y + \rho_o \mathbf{W}_o y + \rho_w \mathbf{W}_w y + \alpha \iota_N + \mathbf{X}_d \beta_d + \mathbf{X}_o \beta_o + \gamma g + \varepsilon. \tag{20}$$

This general model leads to a number of more specific models of interest. We set forth nine different models that result from various restrictions on the parameters ρ_i, $i = d, o, w$. Since the statistical theory for testing parameter restrictions in a maximum-likelihood setting using LR tests is well developed (Cressie 1993; Stein 1999), this seems desirable from an applied specification search viewpoint.

Model #1. The restriction: $\rho_d = \rho_o = \rho_w = 0$, produces the nonspatial model where no spatial autoregressive dependence exists.

Model #2. The restriction: $\rho_o = \rho_w = 0$, results in a model based on a single weight matrix W_d, reflecting destination autoregressive spatial dependence.

Model #3. The restriction: $\rho_d = \rho_w = 0$, produces a sibling model based on a single weight matrix W_o for spatial dependence at the origins.

Model #4. The restriction: $\rho_d = \rho_o = 0$, creates another single weight matrix model containing only W_w, reflecting dependence based on interaction between origin and destination neighbors.

[8]We note that this specification implies a restriction that $\rho_w = -\rho_o \rho_d$, but this restriction need not be enforced in applied work. Of course, restrictions on the values of the scalar dependence parameters ρ_d, ρ_o, ρ_w must be imposed to ensure stationarity in the case where ρ_w is free of the restriction.

Model #5. The restriction: $\rho_d = \rho_o$, and $\rho_w = 0$, results in a model based on a single weight matrix constructed using $0.5(W_d + W_o)$ with a parameter equal to $2\rho_d = 2\rho_o$. This reflects a lack of separability between the impacts of origin and destination dependence relations in favor of a cumulative impact.

Model #6. The restriction: $\rho_d = \rho_o = \rho_w$, produces another single weight matrix model based on $(1/3)(\mathbf{W}_d + \mathbf{W}_o + \mathbf{W}_w)$ with $\rho = (3\rho_o = 3\rho_d = 3\rho_w)$. This reflects a lack of separability between the impacts of origin, destination and origin-to-destination interaction effects in favor of a cumulative impact.

Model #7. The restriction: $\rho_w = 0$, leads to a model with separable origin and destination autoregressive dependence embodied in the two weight matrices $\mathbf{W}_d$ and $\mathbf{W}_o$, while ruling out dependence between neighbors of the origin and destination locations that would be captured by $\mathbf{W}_w$.

Model #8. The restriction: $\rho_w = -\rho_d\rho_o$ results in a successive filtering or model involving both origin $\mathbf{W}_d$, and destination $\mathbf{W}_o$ dependence as well as product separable interaction $\mathbf{W}_w$, constrained to reflect the filter $(\mathbf{I}_N - \rho_d\mathbf{W}_d)(\mathbf{I}_N - \rho_o\mathbf{W}_o) = (\mathbf{I}_N - \rho_o\mathbf{W}_o)(\mathbf{I}_N - \rho_d\mathbf{W}_d) = (\mathbf{I}_N - \rho_d\mathbf{W}_d - \rho_o\mathbf{W}_o + \rho_d\rho_o\mathbf{W}_w)$.

Model #9. The unrestricted model shown in (20) involves three matrices $\mathbf{W}_d$, $\mathbf{W}_o$, and $\mathbf{W}_w$, which yields the ninth member of the family of models. Appropriate restrictions on ρ_d, ρ_o, and ρ_w can thus produce the other more specialized models.

An addition to the family of models arises if we define $\tilde{\mathbf{X}} = (\mathbf{X}\ \ \mathbf{WX})$. In this case we have a model that Anselin (1988) labeled the spatial Durbin model (SDM), which includes a spatial lag of the explanatory variables matrix. As noted in Section 2.3, this model nests the SEM as a special case. A LR test of the error model against the SDM provides a test for the presence of omitted variables that are correlated with the included variables. The estimation procedures we set forth in the next section use the matrix $\mathbf{X}$ for simplicity, but replacing $\mathbf{X}$ by $\tilde{\mathbf{X}}$ would allow one to apply the same procedures to the case of the SDM.

3. ESTIMATION OF SPATIAL FLOW MODELS

The likelihood provides the starting point for both maximum likelihood and Bayesian estimation. We note that the log-likelihood function for the model specifications concentrated with respect to the parameters β and σ will take the form:

$$\text{Ln}L(\rho_d, \rho_o, \rho_w) = C + \ln|\mathbf{I}_N - \rho_d\mathbf{W}_d - \rho_o\mathbf{W}_o - \rho_w\mathbf{W}_w| - \frac{N}{2}\ln(S(\rho_d, \rho_o, \rho_w)), \tag{21}$$

where $S(\rho_d, \rho_o, \rho_w)$ represents the sum of squared errors expressed as a function of the scalar parameters ρ_i, $i = d, o, w$ alone after concentrating out the parameters β, σ, and C denotes a constant not depending on ρ_i (see LeSage and Pace, 2004).

The log-determinant of a matrix plays an important role in both maximum likelihood and Bayesian estimation of transformed random variables. Specifically, the log-determinant ensures that the transformed continuous random variable has a proper density. Otherwise, multiplication of the dependent variable by a transformation such as ϵI, where ϵ is a small positive number, would reduce the magnitude of the estimation residuals to a negligible level. The log-determinant term serves as a penalty to prevent such pathological transformations from obtaining an advantage in estimation. Consequently, the likelihood is invariant to such scalings.

Standard algorithms for maximum likelihood, Bayesian or generalized method of moments estimation of the spatial econometric OD interregional flow models become difficult as the number of observations increases. For example, use of an OD flow matrix for the sample of approximately 3100 US counties can yield sparse spatial weight matrices of dimension N by N where $N = 9{,}610{,}000$. Maximum-likelihood and Bayesian estimation both require calculation of the log-determinant for the N by N matrix $(\mathbf{I}_N - \rho_d\mathbf{W}_d - \rho_o\mathbf{W}_o - \rho_w\mathbf{W}_w)$. Specialized approaches to calculating log-determinants of very large matrices have been proposed by Pace and LeSage (2004), Barry and Pace (1999), Smirnov and Anselin (2001), Griffith (1992, 2004), and Griffith and Sone (1995). However, more efficient approaches exist that can exploit the special structure of matrices like $\mathbf{W}_d = \mathbf{I}_n \otimes \mathbf{W}$, $\mathbf{W}_o = \mathbf{W} \otimes \mathbf{I}_n$ and $\mathbf{W}_w = \mathbf{W}_o \cdot \mathbf{W}_d = \mathbf{W} \otimes \mathbf{W}$. The following sections examine feasible means of computing estimates, sum-of-squared errors, and log-determinants used in calculating the log-likelihood.

Estimates and Moments

As the number of origins and destinations rises, difficulties in implementation of OD models increase. In particular, creating and storing the N by k dense matrices $\mathbf{X}_d$ and $\mathbf{X}_o$ create computer memory problems for large n, and the same problem arises for the N by N sparse matrices $\mathbf{W}_d$, $\mathbf{W}_o$, and $\mathbf{W}_w$. We demonstrate how the unique structure of the OD model can be exploited to allow estimation and inference for large data sets. In many cases straightforward calculations of quantities such as $\mathbf{X}'_d\mathbf{X}_d$ that would require $O(Nk^2)$ can be reduced to $O(Nk^2)$ by exploiting the unique structure of the problem. This means that the benefits from exploiting the special structure rise linearly with n. In addition, an efficient formulation of the estimation problem allows many of the calculations to be performed once, with subsequent updates to important quantities such as the sum-of-squared errors requiring trivial computational time.

First, implementation of OD models does not require formation of the N by N matrices $\mathbf{W}_d$, $\mathbf{W}_o$, or $\mathbf{W}_w$. Since $\mathbf{W}_d y = (\mathbf{I} \otimes \mathbf{W})\, vec(\mathbf{Y})$, then $\mathbf{W}_d y = vec(\mathbf{WY})$ using the relation, $(\mathbf{C}' \otimes \mathbf{A})vec(\mathbf{B}) = vec(\mathbf{ABC})$. Similarly, $\mathbf{W}_o y = vec(\mathbf{YW}')$, and $\mathbf{W}_w y = vec(\mathbf{WYW}')$. We use these forms to rewrite the unrestricted model from (20).

$$(22)\quad vec(\mathbf{Y}) - \rho_d\, vec(\mathbf{WY}) - \rho_o\, vec(\mathbf{YW}') - \rho_w\, vec(\mathbf{WYW}') = \mathbf{Z}\beta + vec(\mathbf{E}).$$

The overall left-hand side of (22) is a linear combination of four component dependent variables. Therefore, the overall parameter estimate is a linear combination of four separate terms that we label $\hat{\beta}^{(t)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' vec(\mathbf{F}^{(t)}(\mathbf{Y}))$, where $\mathbf{F}^{(t)}(\mathbf{Y})$ equals $\mathbf{Y}, \mathbf{WY}, \mathbf{YW}'$, or $\mathbf{WYW}'$ when $t = 1, \ldots, 4$. The overall parameter estimate becomes $\hat{\beta}(\tau) = (\hat{\beta}^{(1)}\ \ \hat{\beta}^{(2)}\ \ \hat{\beta}^{(3)}\ \ \hat{\beta}^{(4)})\tau(\rho)$ where $\tau(\rho) = (1\ \ -\rho_d\ \ -\rho_o\ \ -\rho_w)'$.

The independent variable moment matrix $\mathbf{Z}'\mathbf{Z}$ has the simple form in (5) from the discussion of the least-squares model. The cross-moments associated with the independent and dependent variables are shown in (23).

$$(23)\quad \mathbf{Z}'\, vec(\mathbf{F}^{(t)}(\mathbf{Y})) = \left(\iota_n'\mathbf{F}^{(t)}(\mathbf{Y})\iota_n \quad \mathbf{X}'\mathbf{F}^{(t)}(\mathbf{Y})\iota_n \quad \mathbf{X}'\mathbf{F}^{(t)}(\mathbf{Y})'\iota_n \quad \iota_n'(\mathbf{G}' \odot \mathbf{F}^{(t)}(\mathbf{Y}))\iota_n\right).$$

When $t = 1$, the expression in (23) reduces to that in (6), and the estimates reduce to those from the independent gravity model #1. Given $\hat{\beta}^{(t)}$ for $t = 1, \ldots, 4$, one can form the matrices of residuals $\hat{\mathbf{E}}^{(t)}$ by substituting the estimated parameters into Equation (24), where the overall residual matrix is $\hat{\mathbf{E}} = \hat{\mathbf{E}}^{(1)} - \rho_d\hat{\mathbf{E}}^{(2)} - \rho_o\hat{\mathbf{E}}^{(3)} - \rho_w\hat{\mathbf{E}}^{(4)}$.

$$(24)\qquad \hat{\mathbf{E}}^{(t)} = \mathbf{F}^{(t)}(\mathbf{Y}) - \hat{\alpha}^{(t)}\iota_n\iota_n' - \mathbf{X}\hat{\beta}_d^{(t)}\iota_n' - \iota_n(\hat{\beta}_o^{(t)})'\mathbf{X}' - \hat{\gamma}^{(t)}\mathbf{G}.$$

We wish to introduce the cross-product matrix of the various component residuals, $\mathbf{Q}$, where $\mathbf{Q}_{ij} = tr(\hat{\mathbf{E}}^{(i)'}\hat{\mathbf{E}}^{(j)}), i = 1, \ldots, 4, j = 1, \ldots, 4$. Alternatively, $\mathbf{Q}_{ij} = \iota_n'(\hat{\mathbf{E}}^{(i)} \odot \hat{\mathbf{E}}^{(j)})\iota_n$, and the sum-of-squared residuals for the OD model become $S(\rho) = \tau(\rho)'\mathbf{Q}\tau(\rho)$. Consequently, recomputing $S(\rho)$ for any given vector of ρ values, $\tau(\rho)$, requires a small number of operations that do not depend on n or k. This permits rapid optimization of the likelihood function and acceleration of Bayesian Markov Chain Monte Carlo (MCMC) estimation techniques.

Log-Determinants for a Single Weight Matrix

Models #2 through #6 from the family of nine models use a single weight matrix, denoted here by $\mathbf{W}_s$. The concentrated log-likelihood (21) for these models would contain the term $\log|\mathbf{I}_N - \rho\mathbf{W}_s|$, which would require $O(N^3) = O(n^6)$ calculations using conventional approaches. However, the structure of the OD problem allows efficient techniques that require $O(n)$ calculations for evaluating the log-determinant. The log-determinant of the transformation $\mathbf{I}_N - \rho\mathbf{W}_s$ is the trace of the matrix logarithm of the transformation, and the Taylor series expansion of this has a simple form for the positive definite matrix transformation $\mathbf{I}_N - \rho\mathbf{W}_s$, shown in (25).

$$(25)\qquad \ln\left|\mathbf{I}_N - \rho\mathbf{W}_s\right| = tr\left(\ln(\mathbf{I}_N - \rho\mathbf{W}_s)\right) = -\sum_{t=1}^{\infty}\frac{\rho^t tr\left(\mathbf{W}_s^t\right)}{t}.$$

For the case of single destination or origin weight matrices, $\mathbf{W}_d = \mathbf{I}_n \otimes \mathbf{W}$ or $\mathbf{W}_o = \mathbf{W} \otimes \mathbf{I}_n$, which arises in models #2 and #3. Let $\mathbf{W}_s$, $s = d, o$, and note:

$$tr\left(\mathbf{W}_s^t\right) = tr\left(\mathbf{I}_n^t \otimes \mathbf{W}^t\right) = tr\left(\mathbf{I}_n^t\right) \cdot tr(\mathbf{W}^t) = n \cdot tr(\mathbf{W}^t), \tag{26}$$

and thus the trace of a square matrix of order N is simplified to a scalar (n) times a trace involving the n by n square matrix $\mathbf{W}$.

$$\ln\left|\mathbf{I}_N - \rho\mathbf{W}_s\right| = -n\sum_{t=1}^{\infty}\frac{\rho^t tr\left(\mathbf{W}_s^t\right)}{t} = n\ln\left|\mathbf{I}_n - \rho\mathbf{W}\right|. \tag{27}$$

Summarizing, for the case of a single spatial weight matrix $\mathbf{W}_s$, $s = d, o$, users can employ algorithms for computing the log-determinant of an n by n matrix $\ln|\mathbf{I}_n - \rho\mathbf{W}|$, when working with a vector of N OD flows. For the earlier example of $n = 3{,}100$ US counties and $N = 9{,}610{,}000$, we can solve these estimation problems in a matter of seconds on desktop computers when using computationally efficient sparse algorithms for the n by n log-determinant portion of the problem (see LeSage and Pace, 2004).

Log-determinants for the Successive Filtering Model

For the successive spatial filtering model specification in model #8 the order of transformation of the dependent variable does not matter since:

$$\left(\mathbf{I}_N - \rho_d\mathbf{W}_d\right)\left(\mathbf{I}_N - \rho_o\mathbf{W}_o\right) = \left(\mathbf{I}_N - \rho_o\mathbf{W}_o\right)\left(\mathbf{I}_N - \rho_d\mathbf{W}_d\right). \tag{28}$$

The log-determinant term appearing in the concentrated likelihood (21) takes the simple expression in (29), because the log-determinant of a product is the sum of the log-determinants.

$$\ln\left|\left(\mathbf{I}_N - \rho_d\mathbf{W}_d\right)\left(\mathbf{I}_N - \rho_o\mathbf{W}_o\right)\right| = n\ln\left|\mathbf{I}_n - \rho_d\mathbf{W}\right| + n\ln\left|\mathbf{I}_n - \rho_o\mathbf{W}\right|. \tag{29}$$

Log-determinants for Combinations of Weight Matrices

As in the case of single spatial weights from Section 3.2, the log-determinant required for maximum-likelihood estimation of models with more general combinations of weight matrices can be calculated using only n by n matrices rather than the large N by N matrices.

Drawing on the earlier discussion surrounding (27), the estimation challenge for the case of the most general OD model #9 (shown in (20)), with all three parameters ρ_d, ρ_o and ρ_w unrestricted is to efficiently compute $tr(\mathbf{W}_f^t)$ for $t = 1, \ldots, m$, where m is the largest moment computed, and $\mathbf{W}_f$ is defined in (30).

$$\mathbf{W}_f = \rho_d\left(\mathbf{I}_n \otimes \mathbf{W}\right) + \rho_o\left(\mathbf{W} \otimes \mathbf{I}_n\right) + \rho_w(\mathbf{W} \otimes \mathbf{W}). \tag{30}$$

The case of $tr(\mathbf{W}_f)$ where $t = 1$ is immediate, and equals zero since $tr(\mathbf{W}) = 0$. The case of $tr(\mathbf{W}_f^2)$ is slightly more challenging as shown in (31).

$$\mathbf{W}_f^2 = \rho_d^2(\mathbf{I}_n \otimes \mathbf{W}^2) + \rho_o^2(\mathbf{W}^2 \otimes \mathbf{I}_n) + \rho_w^2(\mathbf{W}^2 \otimes \mathbf{W}^2) + 2\rho_d\rho_o(\mathbf{W} \otimes \mathbf{W}) + 2\rho_o\rho_w(\mathbf{W}^2 \otimes \mathbf{W}) + 2\rho_d\rho_w(\mathbf{W} \otimes \mathbf{W}^2). \tag{31}$$

$$tr(\mathbf{W}_f^2) = \rho_d^2 n \cdot tr(\mathbf{W}^2) + \rho_o^2 n \cdot tr(\mathbf{W}^2) + \rho_w^2 tr(\mathbf{W}^2)^2 + 2\rho_d\rho_o tr(\mathbf{W})^2 + 2\rho_o\rho_w tr(\mathbf{W}^2)tr(\mathbf{W}) + 2\rho_d\rho_w tr(\mathbf{W})tr(\mathbf{W}^2). \tag{32}$$

For the quadratic, there are nine possible terms and six of these are unique. Note, $tr(\mathbf{W}^2)$ is the highest-order term involved in calculating the expression $tr(\mathbf{W}_f^2)$. Extrapolating, computations of $tr(\mathbf{W}_f^t)$ only require computing $tr(\mathbf{W}^t)$ based on the n by n weight matrix $\mathbf{W}$, a much less demanding task.

Other than computing $tr(\mathbf{W}^t)$, none of these computations are dependent on n, so the time required does not depend on the number of origins or destinations.[9] For small n or t calculating the exact $tr(\mathbf{W}^t)$ requires little time. For large n, $tr(\mathbf{W}^t)$ can be approximated as in Barry and Pace (1999) to any desired accuracy using an $O(n)$ algorithm. However, the number of terms in the expansion of $tr(\mathbf{W}_f^t)$ does rise quickly with t, and this poses the main computational challenge. However, we have done this for $m = 16$ without too much difficulty.

Given the m moments and the conditions on $\mathbf{W}$, it is simple to calculate a relatively short interval containing the log-determinant as shown in (33).

$$-\sum_{t=1}^{m} \frac{tr(\mathbf{W}_f^t)}{t} \geq \ln|\mathbf{I}_N - \mathbf{W}_f| \geq -\left(\sum_{t=1}^{m} \frac{tr(\mathbf{W}_f^t)}{t} + \sum_{t=m+1}^{\infty} \frac{tr(\mathbf{W}_f^m)}{t}\right). \tag{33}$$

Given previous assumptions on $\mathbf{W}$, the principal eigenvalue of $\mathbf{W}_f$ is less than 1 in magnitude. Using similar reasoning, Pace and LeSage (2002) show that the moments $tr(\mathbf{W}_f^t)$ must monotonically decline when $t > 1$ for matrices $\mathbf{W}$ that are similar to symmetric matrices, and this sets up the bounds. The interval is narrow provided the last term in parentheses in (33) is reasonably small, and this will be revealed during the actual computations.

Summarizing, potential computational problems that might plague estimation for models involving N observations on OD flows have been eliminated by reducing the troublesome log-determinant calculation to one involving only traces of n by n matrices.

[9]We have focused on estimating or computing $tr(\mathbf{W}^t)$ as opposed to other methods for the calculation of the log-determinant term since this can be done in $O(n)$ time. However, one could employ eigenvalues to compute the log-determinant using some of the useful Kronecker product properties of eigenvalues and determinants. Assuming the spectral decomposition of $\mathbf{W}$ exists so that $\mathbf{W} = \mathbf{VDV}'$, $\lambda_i = \mathbf{D}_{ii}$, and $w = \iota_N - \rho_d\iota_n \otimes \lambda - \rho_o\lambda \otimes \iota_n + \rho_w\lambda \otimes \lambda$, $\ln|\mathbf{I} - \mathbf{W}_f|$ equals $\sum_{i=1}^{N} \ln(w_i)$. Note, computing eigenvalues requires $O(n^3)$ computations and the summation requires $O(n^2)$ computations. However, this is still small relative to calculating the eigenvalues of $\mathbf{W}_f$ which would require $O(n^6)$ calculations.

4. AN APPLIED ILLUSTRATION USING STATE-LEVEL POPULATION MIGRATION FLOWS

We use state-level population migration flows to provide a stylized illustration of the family of OD spatial econometric models. Specifically, we model state-to-state flows (for the population 5 years and over) over the period from 1995 to 2000, using 1990 Census characteristics of the states. These flows were all nonzero, and a log transformation produced a dependent variable vector that was nearly normal, with slightly fatter tails. In addition we use two explanatory variables that were available for 1995, the state population in 1995 and the state unemployment rate in 1995. These data are available from the Census 2000 Special Reports.[10] The flow matrix was transformed using log(vec($\mathbf{Y}$)) which produces a single cross-sectional vector representing the (logged) state-to-state migration flows over the five-year period from 1995 to 2000.

The sample was restricted to the 48 contiguous states plus the District of Columbia resulting in $n = 49$ and $N = 2{,}401$ observations. The flows of population within each state (those on the main diagonal of the flow matrix) were not set to zero as is often done when attempting to model interregional flows (see Tiefelsdorf, 2003 and Fischer et al., 2006). Variables such as $\mathbf{W}_d y$ represent local averages of the dependent variable that can aid in fitting the data. Setting some of the elements of y to zero defeats the purpose of using local averages.

We created a separate model for flows from the main diagonal of the flow matrix representing intrastate migration. This was done by setting all elements of the covariate matrices $\mathbf{X}_d$, $\mathbf{X}_o$ corresponding to the main diagonal of the flow matrix (flows within each state) to zero. This prevents these variables from entering the interregional migration flow model, forcing the nonzero observations in the explanatory variable matrices $\mathbf{X}_d$, $\mathbf{X}_o$ to explain variation in the between state or interregional flows. The intraregional model consisted of additional explanatory variables: state population in 1995, state area, the unemployment rate in 1995, and the proportion of state employment in farming. Only 49 observations for these four variables were used, those corresponding to the diagonal of the flow matrix, with the remaining elements of these variable vectors set to zero. These four N by 1 vectors were added as columns to the explanatory variables matrix. They will be used to explain intrastate migration flows. Intuitively, we would expect more intrastate migration for states with larger population and area, and we would expect less intrastate migration for states with more farming employment. The unemployment rate in 1995 might also lead to intrastate migration. Of course, one could devise a richer model for within-state migration flows, but typically these models focus on factors that explain between state migration flows.

[10]The data represent sample information. For confidentiality protection, sampling error, nonsampling error, and definitions, see census.gov/prod/cen2000/doc/sf3.pdf.

TABLE 1: Variables Used in the Model

Variable name	Description
y	Log (interstate and intrastate migration flows 1995–2000)
Population 1995	log (state population in 1995)
Area	Log (state area)
College	Proportion of population > age 25, college degree as highest
Born in state	Proportion of population > age 5 born in the state in 1990
Unemployment rate 1995	Unemployment rate in 1995
Mortgage	Log (median mortgage costs in 1990)
Executive	Proportion of employment in executive and managerial occupations
Sales	Proportion of employment in sales occupations
Local Govt.	Proportion of employment in local government
State Govt.	Proportion of employment in state government
Federal Govt.	Proportion of employment in federal government
Farming	Proportion of employment in farming
Distance	Log (distance between origin and destination state centroids)

Use of the separate model for intraregional migration flows should downweight the impact of the large values on the main diagonal of the flow matrix, preventing them from exerting undue impact on the resulting estimates for β_d, β_o, which are typically the focus of interest in these models.

Explanatory variables for the matrices $\mathbf{X}_d$, $\mathbf{X}_o$ for each state were taken from the 1990 Census, with the exception of the 1995 population and unemployment rate variables.[11] These variables are defined in Table 1. If one is interested in the partial derivative impact on flows arising from a change in say the rth destination variable $\mathbf{X}_{dr}$, this takes the form: $\partial y/\partial \mathbf{X}_{dr} = (\mathbf{I}_N - \hat{\rho}_d\mathbf{W}_d - \hat{\rho}_o\mathbf{W}_o - \hat{\rho}_w\mathbf{W}_w)^{-1}(\mathbf{I}_N\hat{\beta}_{dr})$, where β_{dr} is the coefficient estimate associated with the rth destination variable and $\hat{\rho}_i, i = d, o, w$ are estimates for the dependence parameters. See Pace and LeSage (2008) for a detailed discussion of this issue as well as scalar summary measures based on an average of the elements in the N by N matrix of partial derivative impacts. In the discussion that follows, it is important to note that the average elements of the matrix $(\mathbf{I}_N - \hat{\rho}_d\mathbf{W}_d - \hat{\rho}_o\mathbf{W}_o - \hat{\rho}_w\mathbf{W}_w)^{-1}\mathbf{I}_N$ for our model were positive, so the direction of impact on flows from changing the explanatory variables is determined by the sign of the coefficient estimates β_d, β_o. Therefore, interpretation of positive coefficient estimates β_d in this model are that an increase in the associated variable X_d would correspond to increased destination migration flows. Similarly, for the case of positive β_o estimates, increasing values of the associated variables in X_o would lead to increased flows at the origins.

[11]Use of 1995 population and unemployment rate variables might introduce endogeneity, but our application is more illustrative than substantive in nature.

TABLE 2: Log Likelihoods for Alternative Models

Model	Log Likelihood	LR test versus Model 9	Critical Value ($\alpha = 0.05$)
Model 9: ρ_d, ρ_o, ρ_w unrestricted	−17.82		
Model 7: $\rho_w = 0$	−242.13	448.63	$\chi^2(1) = 3.84$
Model 8: $\rho_w = -\rho_d \cdot \rho_o$	−245.10	454.57	$\chi^2(1) = 3.84$
Model 5: $\rho_d = \rho_o$, $\rho_w = 0$	−247.87	460.11	$\chi^2(2) = 5.99$
Model 2: $\rho_o = \rho_w = 0$	−434.13	832.63	$\chi^2(2) = 5.99$
Model 6: $\rho_d = \rho_o = \rho_w$	−480.02	924.39	$\chi^2(2) = 5.99$
Model 3: $\rho_d = 0$, $\rho_w = 0$	−546.86	1,058.08	$\chi^2(2) = 5.99$
Model 4: $\rho_d = 0$, $\rho_o = 0$	−791.36	1,547.09	$\chi^2(2) = 5.99$
Model 1: $\rho_d = 0$, $\rho_o = 0$, $\rho_w = 0$	−1,698.13	3,360.63	$\chi^2(3) = 7.82$

For example, higher mortgage costs at the origin should increase migration flows away from origin states and decrease flows to destination states suggesting $\hat{\beta}_o > 0$ and $\hat{\beta}_d < 0$. We might expect a higher proportion of population born in the state to reduce migration flows at both the origin and destination. This is because states with a higher proportion of population born in the state demonstrate a long-term propensity to remain rather than migrate, and it may also signal past low attractiveness to in-migrants. In contrast, a higher proportion of population with college degrees might increase flows at both origins and destinations, since college graduates might be more mobile at origins and act as a attractive force at destinations. In addition to the variables included in the matrices $\mathbf{X}_d$, $\mathbf{X}_o$, the log of distance from each origin to each destination was included in the model, along with a constant term and the variables included to model intrastate migration flows.[12]

The family of nine model specifications described in Section 3 were estimated using maximum-likelihood methods with a numerical Hessian approach used to compute estimates of dispersion and t-statistics. The log-likelihood function values for the family of nine models are shown in Table 2, ordered from high to low, along with a LR test of the restrictions imposed by each model versus the unrestricted model. It is clear from the table that Model #9 containing separate spatial weight matrices for the origin and destination and no restrictions on the parameters ρ_d, ρ_o, ρ_w dominates all other models. The second-best model (Model #7) also contains spatial lags involving both $\mathbf{W}_d$ and $\mathbf{W}_o$ matrices, with ρ_w restricted to be zero, and the third-best model (Model #8) also contains separate origin and destination spatial weight matrices. The next best model (Model #5) based on the sum of the destination and origin spatial weight matrices ($\mathbf{W}_d + \mathbf{W}_o$), with ρ_w restricted to zero, has a likelihood function value

[12]An anonymous reviewer suggested there is some debate in the literature concerning whether distance should be transformed using logs. We leave this as an issue for further research regarding interstate migration modeling.

that is significantly different from Model #7, based on separate $\mathbf{W}_d$, $\mathbf{W}_o$, with ρ_w restricted to zero.[13]

From this we conclude that models based on a single weight matrix, either $\mathbf{W}_d$, $\mathbf{W}_o$, $\mathbf{W}_w$, exhibit significantly lower likelihoods than models based on separate weight structures involving $\mathbf{W}_d$ and $\mathbf{W}_o$. This would seem to support the notion that both origin and destination dependence/connectivity information is important. The LR test clearly rejects least-squares that ignores spatial dependence in the migration flows.

We estimated the SDM and SEM for the case of the most general unrestricted models that contain $\mathbf{W}_d$, $\mathbf{W}_o$, $\mathbf{W}_w$. The results indicated a log-likelihood function value for the SDM that was 129.51 higher than the SEM. Twice this value equals 259.02, so a LR test of the parameter restrictions allows us to reject the SEM model in favor of the SDM model. The 99 percent critical value for the chi-squared statistic with 13 degrees of freedom (which equals the number of parameter restrictions) is 27.7. We also note that a test of the SDM model versus the restrictions implied by the spatial lag (SAR) model rejected the SAR model in favor of the SDM. Despite this, we rely on the SAR model in our applied illustration for brevity.

Table 3 presents estimates from least-squares and the unrestricted spatial Model #9 for comparison. The estimates for $\rho_d = 0.4581$ and $\rho_o = 0.5175$, indicate spatial dependence of almost equal importance between neighbors to the origin and neighbors to the destination. As indicated above, the estimate for $\rho_w = -0.3863$ is not consistent with the restriction $\rho_w = -\rho_d \cdot \rho_o$, and differs significantly from zero, suggesting that origin-to-destination dependence specified by the weight matrix $\mathbf{W}_w$ is also important.

In Table 3, variables included to model intrastate migration flows are labeled OD_pop95, OD_area, OD_urate, OD_farming, with variables from the matrix $\mathbf{X}_d$ labeled with $D\_$ preceding the variable name and those from $\mathbf{X}_o$ indicated by the symbol $O\_$.

In terms of the parameter estimates, distance is negative and significant in both least-squares and spatial models, but the spatial model that includes spatial lags of the dependent variable shows a decrease to nearly one-fourth in the magnitude of this coefficient estimate. However, a direct comparison of the magnitudes of the coefficients from least-squares and the spatial model is not valid. For least-squares the coefficients for the rth explanatory variable x_r represent $\partial y/\partial_{x_r}$, whereas those from the spatial lag model do not. For the spatial lag model partial derivative impacts arising from a change in the rth explanatory variable involve an N by N matrix inverse as noted earlier.

In Table 3, the least-squares estimates are often larger in magnitude than those from the spatial model. For example, spatial model estimates for D_pop95, and D_area are around one-half of those from least-squares. As already noted,

[13]Using a LR test, $-2[\mathrm{LR}(\#5) - \mathrm{LR}(\#7)] = 11.4800$, which exceeds the critical value for $\chi^2(1) = 3.84$. Model #5 has one additional restriction relative to Model #7, so the degrees of freedom for the LR test is 1.

TABLE 3: Estimates from Least-squares and the Unrestricted Spatial Model 9

	Least-Squares		Spatial Model	
Variable	Coefficient	t-statistic(plevel)	Coefficient	t-statistic(plevel)
Constant	−12.5281	−8.21 (0.0000)	−5.8195	−15.27 (0.0000)
D_pop95	0.9281	38.85 (0.0000)	0.4296	22.88 (0.0000)
D_area	0.2273	12.66 (0.0000)	0.0868	7.14 (0.0000)
D_college	8.5537	7.89 (0.0000)	4.5691	6.20 (0.0000)
D_borninstate	−2.9131	−22.10 (0.0000)	−1.0983	−11.65 (0.0000)
D_urate95	−0.5264	−10.55 (0.0000)	−0.1883	−5.67 (0.0000)
D_mortgage	−0.5958	−4.66 (0.0000)	−0.4149	−5.12 (0.0000)
D_exec	−0.4993	−1.76 (0.0781)	−0.0978	−0.53 (0.5941)
D_sales	0.4464	1.86 (0.0617)	0.3939	2.59 (0.0096)
D_local_govt	−2.5698	−3.83 (0.0001)	−0.5446	−1.23 (0.2156)
D_state_govt	−1.3228	−2.07 (0.0385)	−1.2894	−2.97 (0.0030)
D_fed_govt	0.8228	1.63 (0.1026)	0.6782	2.04 (0.0410)
O_pop95	0.8108	33.94 (0.0000)	0.4185	22.71 (0.0000)
O_area	0.2680	14.92 (0.0000)	0.1172	9.52 (0.0000)
O_college	6.0058	5.54 (0.0000)	4.5046	6.07 (0.0000)
O_borninstate	−1.5466	−11.73 (0.0000)	−0.4946	−5.50 (0.0000)
O_urate95	−0.4208	−8.43 (0.0000)	−0.1367	−4.19 (0.0000)
O_mortgage	0.0628	0.49 (0.6225)	−0.0616	−0.88 (0.3759)
O_exec	0.3886	1.37 (0.1702)	0.1198	0.68 (0.4945)
O_sales	−0.0039	−0.01 (0.9868)	0.2819	1.83 (0.0660)
O_local_govt	1.7072	2.54 (0.0108)	1.5082	3.47 (0.0005)
O_state_govt	−0.3572	−0.55 (0.5760)	−1.0193	−2.37 (0.0176)
O_fed_govt	−0.6902	−1.36 (0.1709)	0.4101	1.23 (0.2175)
OD_pop95	1.0150	8.68 (0.0000)	0.4468	5.79 (0.0000)
OD_area	0.2260	2.29 (0.0218)	0.1159	1.84 (0.0653)
OD_urate95	0.6557	2.61 (0.0089)	0.5319	3.54 (0.0004)
OD_farming	−0.6185	−3.34 (0.0008)	−0.3431	−3.22 (0.0013)
Log(distance)	−0.5759	−75.22 (0.0000)	−0.1671	−16.96 (0.0000)
ρ_d			0.4581	33.36 (0.0000)
ρ_o			0.5175	39.28 (0.0000)
ρ_w			−0.3863	−23.56 (0.0000)
σ^2	0.2438		0.1042	

if the true data-generating process was a model containing a spatial lag, then least-squares estimates are biased and inconsistent.

Summarizing, we found significant spatial dependence in the OD flows modeled here. Least-squares estimates and inferences that ignore this dependence are different from those produced using a spatial autoregressive model specification. We found evidence for three different types of spatial dependence that we have labeled, *origin*, *destination* and *origin-to-destination* based dependence. A test for the SDM that subsumes the SEM and spatial lag model (SAR) as special cases lead to a rejection of the SEM and SAR in favor of the SDM. This suggests that our model suffers from the presence of omitted variables that exhibit spatial dependence.

5. CONCLUSIONS

We argue that use of a traditional least-squares regression to estimate gravity or spatial interaction models ignores possible spatial dependence in the sample data vector y of OD flows. Beginning with a model that posits no spatial dependence in the vector y, we show that spatial dependence in y can arise from omitted variables correlated with included variables when omitted variables exhibit spatial dependence. Strategies that can be used to test for the problem of omitted variables were discussed. It is well-known that use of least-squares in the face of spatial dependence in the dependent variable vector leads to biased and inconsistent estimates (see LeSage and Pace, 2004).

To address this issue, we show ways of incorporating spatial autoregressive dependence in regression-based gravity models. An illustrative application to migration flows among the lower 48 U.S. states plus District of Columbia was used to provide a stylized illustration of our methods.

Our approach leads to a spatial OD filter specification that we apply to the vector of OD flows to capture three types of possible spatial dependence that may arise between OD flows: origin-based, destination-based, and origin-to-destination based dependence.

We provide new results to make maximum-likelihood estimation computationally feasible for OD flow data. These exploit the structured nature of the family of spatial regression gravity models introduced here to allow estimates and inferences for the case of OD flow data, which can be large since the number of observations N will equal the number of regions n in the sample squared, e.g., $N = n^2$. Specifically, we discuss $O(n)$ methods that can be used to compute the log-determinant of the spatial transformation, and for computing the relevant moment matrices. The use of the spatial OD filter makes the current corpus of work on maximum likelihood and Bayesian estimation of spatial regression models immediately relevant to OD flow modeling. The family of models introduced here subsumes least-squares gravity models as well as single weight matrix spatial regression models as a special case. Simple tests of parameter restrictions that produce varying specifications for spatial dependence can be carried out, resolving contentious model specification issues that often arise.

Maximum- likelihood estimation of the models introduced here is subject to the caveat that the OD flows follow a normal distribution, or can be suitably transformed to achieve normality. Additional work is needed to deal with OD flows where a large proportion of the flows take values of zero, since this situation may not be amenable to transformations that result in a normal distribution (see LeSage, Fischer and Scherngell (2007) for a discussion of these issues).

REFERENCES

Anderson, James E. 1979. “A Theoretical Foundation for the Gravity Model,” *American Economic Review*, 69(1), 106–116.

Anderson, James E. and Eric van, Wincoop. 2004. “Trade Costs,” *Journal of Economic Literature*, 42(3), 691–751.

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic Publishers.

Barry, Ronald and R. Kelley Pace. 1999. “A Monte Carlo Estimator of the Log Determinant of Large Sparse Matrices,” *Linear Algebra and its Applications*, 289, 41–54.

Behrens, Kristian, Cem Ertur, and Wilfried Koch. 2007. “Dual Gravity: Using Spatial Econometrics to Control for Multilateral Resistance,” *CORE Discussion Paper* No. 2007/59.

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

Curry, Les. 1972. “A Spatial Analysis of Gravity Flows,” *Regional Studies*, 6, 137–147.

Cushing, Brian and Jacques Poot. 2003. “Crossing Boundaries and Borders:: Regional Science Advances in Migration Modelling,” *Papers in Regional Science*, 83, 317–338.

Ertur, Cem and Wilfried Koch. 2007. Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, *Journal of Applied Econometrics*, 22, 1033–1062.

———. 2008. “International R&D Spillovers in the Multi-country Schumpeterian Growth Model.” Paper presented at the Southern Regional Science Association Meetings, April, 2008.

Fingleton Bernard and Enrique López-Bazo. 2006. “Empirical Growth Models with Spatial Effects,” *Papers in Regional Science*, 85, 177–198.

Fischer, Manfred M., Thomas Scherngell, and Eva Jansenberger. 2006. “The Geography of Knowledge Spillovers between High-Technology Firms in Europe Evidence from a Spatial Interaction Modelling Perspective,” *Geographical Analysis*, 38(3), 288–309.

Griffith, D. A. 1992. “Simplifying the Normalizing Factor in Spatial Autoregressions for Irregular Lattices,” *Papers in Regional Science*, 71(1), 71–86.

Griffith, Daniel A. 2004. “Faster Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: An Extension of the Smirnov-Anselin result,” *Journal of Statistical Computation and Simulation*, 74, 855–866.

———. 2007. “Spatial Structure and Spatial Interaction: 25 Years Later,” *The Review of Regional Studies*, 37(1), 28–38.

Griffith, Daniel A. and Akio Sone. 1995. “Trade-offs Associated with Normalizing Constant Computational Simplifications for Estimating Spatial Statistical Models,” *Journal of Statistical Computation and Simulation*, 51, 165–183.

Griffith, Daniel A. and Kelvyn G. Jones. 1980. “Explorations into the Relationship Between Spatial Structure and Spatial Interaction,” *Environment and Planning A*, 12, 187–201.

Horn, Ronald A. and Charles. R. Johnson. 1991. *Topics in Matrix Analysis*. Cambridge: Cambridge University Press.

Lee, Ming and R. Kelley Pace. 2005. “Spatial Distribution of Retail Sales,” *Journal of Real Estate Finance and Economics*, 31(1), 53–69.

LeSage, James P., Manfred M. Fischer, and Thomas Scherngell. 2007. “Knowledge Spillovers across Europe, Evidence from a Poisson Spatial Interaction Model with Spatial Effects,” *Papers in Regional Science*, 86(3), 393–421.

LeSage, James P. and R. Kelley Pace. 2004. “Introduction,” *Advances in Econometrics: Vol. 18: Spatial and Spatiotemporal Econometrics*. Oxford: Elsevier Ltd, pp. 1–32.

Pace, R. Kelley and James P. LeSage. 2002. “Semiparametric Maximum Likelihood Estimates of Spatial Dependence,” *Geographical Analysis*, 34(1), 75–90.

———. 2004. “Techniques for Improved Approximation of the Determinant Term in the Spatial Likelihood Function,” *Computational Statistics and Data Analysis*, 45, 179–196.

———. 2007. “Interpreting Spatial Econometric Models,” Paper presented at the North American Meetings of the Regional Science Association International, Toronto, CA, November 2006, Institute for Advanced Studies, Vienna, Austria, December, 2006, the First World Conference of the Spatial Econometrics Association, Cambridge, UK, July 2007, and the Allied Social Science Association meetings in New Orleans, January 2008.

———. 2008. "Biases of OLS and Spatial Lag Models in the Presence of an Omitted Variable and Spatially Dependent Variables," SSRN Working paper available at: http://ssrn.com/abstract=1133438.

Porojon, Anca. 2001. "Trade Flows and Spatial Effects: The Gravity Model Revisited," *Open Economies Review*, 12, 265–280.

Sen, Ashish and Tony E. Smith. 1995. *Gravity Models of Spatial Interaction Behavior*, Heidelberg: Springer-Verlag.

Smirnov, Oleg and Luc Anselin. 2001. "Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: A Characteristic Polynomial Approach," *Computational Statistics and Data Analysis*, 35, 301–319.

Stein, Michael L. 1999. *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer.

Tiefelsdorf, Michael. 2003. "Misspecifications in Interaction Model Distance Decay Relations: A Spatial Structure Effect," *Journal of Geographical Systems*, 5, 25–50.