Spatial Econometric Modeling of Origin-Destination flows

James P. LeSage[1]
University of Toledo
Department of Economics
Toledo, OH 43606
jlesage@spatial-econometrics.com
and
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
OFF: (225)-388-6256, FAX: (225)-334-1227
kelley@pace.am, www.spatial-statistics.com

October 18, 2005

[1]The authors would like to acknowledge Randall W. Jackson and Wolfgang Polasek for helpful discussions on this topic.

Abstract

The traditional gravity model used to provide econometric estimates of variables influencing origin-destination flows is extended to explicitly take into account spatial dependence in the flows. This is accomplished by introducing a spatial connectivity matrix that allows for three types of spatial dependence in the flows from origins to destinations. Introducing this type of connectivity or spatial weight structure for the flows allows conventional spatial econometric estimation procedures to be used in modeling variation in flows that arise when the origin-destination flow matrix is vectorized. A family of alternative spatial econometric model specifications is set forth along with an applied illustration based on state-level migration flows for the 48 contiguous US states and the District of Columbia.

KEYWORDS: migration flows, spatial autoregression, Bayesian, maximum likelihood, spatial connectivity of origin-destination flows.

1 Introduction

This paper sets forth spatial econometric methods for modeling origin-destination (OD) matrices containing interregional flows. These are general data structures used in a variety of economic, geography and regional science research contexts such as international trade flows, migration research, transportation, network, and freight flow analysis, communications and information flow research, journey-to-work studies, and regional and interregional economic modeling. In contrast to typical spatial econometric models where the sample involves n regions, with each region being an observation, these models involve n^2 OD pairs.

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). An objective of this type of modeling is to explain variation in the level of flows between the n^2 OD pairs. These models typically rely on a function of the distance between an origin and destination and explanatory variables pertaining to characteristics of both origin and destination regions. They typically assume that spatial dependence between the sample of N^2 OD pairs will be captured by the distance function. With a few exceptions, use of spatial lags typically found in spatial econometric methods have not been used in these models. 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."

Questions surrounding how to parsimoniously structure the connectivity of the sample of n^2 origin-destination pairs that arise in modeling interregional flows has remained a stumbling block. Conventional spatial autoregressive models rely on spatial weight structures constructed to reflect connectivity between n regions. One focus of the presentation here is a proposal for spatial weight structures that model dependence between the n^2 OD pairs in a fashion consistent with conventional spatial autoregressive models.

The notion that use of distance functions in conventional spatial interaction models effectively capture spatial dependence in the interregional flows being analyzed has been challenged in recent work by Porojon (2001) for the case of international trade flows, Lee and Pace (2004) for retail sales and unpublished work that utilizes both German and Canadian transportation network flows. The residuals from conventional models were found to exhibit spatial dependence, which could be exploited to improve the precision of inference as well as prediction accuracy.

A family of successive spatial filtering models is introduced here that represent an extension of the spatial regression models introduced in Anselin (1988). Spatial regression models have served as the workhorse in applied spatial econometric analysis, and the models introduced here should play an important role in modeling interregional flow matrices. Another focus of this study is maximum likelihood and Bayesian estimation of the models introduced here. We demonstrate how simple extensions of widely available software algorithms for implementing conventional spatial regression models can be employed to estimate the models set forth here.

2 Interregional flows in a spatial regression context

Let Y represent an n by n square matrix of interregional flows from each of the n origin regions to each of the n destination regions where each of the n columns of the flow matrix represents a different origin and the n rows reflect destinations. We can produce an n^2 by 1 vector of these flows in two ways, one reflecting an "origin-centric" ordering as in (1), and the other reflecting a "destination-centric" ordering as in (2). We obtain the origin-centric ordering via $y = vec(Y)$, and the destination-centric ordering via $y^{(d)} = vec(Y')$. 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.[1]

[1]Typically, the diagonal elements of a flow matrix containing flows within a region, e.g., from origin 1 to destination 1, origin 2 to destination 2, etc., will be large relative to the off-diagonal elements representing interregional flows. In fact, many of the interregional flows will take on zero values, indicating the absence of flows from some origins to particular destinations. In the following discussion, we ignore these issues with discussion

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

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

A conventional gravity model least-squares regression approach to explaining the variation in the vector of origin-destination flows would rely on an n by k explanatory variables matrix that we label x_d, containing k characteristics for each of the n destinations. Given the format of our vector y, where observations 1 to n reflect flows from origin 1 to all n destinations, this matrix would be repeated n times to produce an n^2 by k matrix of destination characteristics that we represent as X_d for use in the regression. Each vector j of X_d equals $\iota \otimes X_j$, where ι is a n by 1 vector of ones. A second matrix containing origin characteristics which we label X_o would be constructed for use in the gravity model. This matrix would repeat the characteristics of the first origin n times to form the first n rows of X_o, the characteristics of the second origin n times for the next n rows of X_o and so on, resulting in an n^2 by k matrix of origin characteristics. Similarly, each vector j of X_o is $X_j \otimes \iota$. Typically, the distance from each origin to destination is also included as an explanatory variable vector in the gravity model, and perhaps non-linear terms such as distance-squared. We let D represent an n^2 by 1 vector of these distances from each origin to each destination formed by stacking the columns of the origin-destination distance

and suggestions for dealing with these taken up later.

matrix into a variable vector.[2] This results in a regression model of the type shown in (3)

$$y = \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + \varepsilon \tag{3}$$

In (3), the explanatory variable matrices X_d, X_o represent n^2 by k matrices containing destination and origin characteristics respectively and the associated k by 1 parameter vectors are β_d and β_o. The vector D denotes the vectorized origin-destination distance matrix, and γ, α are scalar parameters. For now we assume $\varepsilon \sim N(0, \sigma^2 I_{n^2})$, but generalizations will be taken up later.

2.1 Spatial dependence in origin-destination flows

In contrast to the traditional regression-based gravity model, a spatial econometric model of the variation in origin-destination flows would be characterized by: 1) reliance on spatial lags of the dependent variable vector, which we refer to as a spatial autoregressive model (SAR); 2) spatial lags of the disturbance terms, which we label a spatial error model (SEM); or 3) perhaps spatial lags of both kinds, which we denote as the general spatial model (SAC). Spatial weight matrices represent a convenient and parsimonious way to define the spatial dependence or connectivity relations among observations.

In a typical cross-sectional model with n regions where each region represents an observation, the spatial weight matrix labelled W represents an n by n sparse matrix. This matrix captures dependency relations between the observations which represent regions. In this conventional case, the rows $i = 1, \ldots, n$ of the matrix W are specified using the set of neighboring observations to each observation i. If we designate neighboring observations using Υ_i, spatial dependence arises when an observation at one location, say y_i is dependent on "neighboring" observations $y_j, y_j \in \Upsilon_i$. If W_{ij} represents the individual elements of the matrix W, then $W_{ij} > 0$ when observation $y_j \in \Upsilon_i$, that is y_i depends upon y_j. By convention, $W_{ii} = 0$ to prevent an observation from being defined as a neighbor to itself, and the matrix W is typically row-standardized to have row sums of unity.

A key issue is how to construct a meaningful spatial weight matrix in the case where the n^2 by 1 vector of observations reflect flows from all origins to all destinations, rather than the typical case where each observation

[2] The diagonal elements of the distance matrix containing distances from origin 1 to destination 1, origin 2 to destination 2, etc., will be zero. We will have more to say about this later.

represents a region. We can create a typical n by n first-order contiguity or m nearest neighbors weight matrix W that reflects relations between the n destinations/origin regions. This can be repeated using $I_n \otimes W$ to create an n^2 by n^2 row-standardized spatial weight matrix that we label W_o, shown in (4), where $\underline{0}$ represents an n by n matrix of zeros.

$$W_o = \begin{pmatrix} W & \underline{0} & \dots & \underline{0} \\ \underline{0} & W & \underline{0} & \vdots \\ \vdots & \underline{0} & \ddots & \underline{0} \\ \underline{0} & \dots & \underline{0} & W \end{pmatrix} \quad (4)$$

Using this matrix to form a spatial lag of the dependent variable, $W_o y$, (where $W_o = I_n \otimes W$ with W row-standardized), we capture "origin-based" spatial dependence relations using an average of flows from neighbors to each origin region to each of the destinations. Intuitively, it seems plausible that 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. This is what the spatial lag $W_o y$ captures.

As an example, consider a single row i of the spatial lag vector $W_o y$ that represents flows from the origin state/region of Florida to the destination state/region of Washington. First-order contiguous neighbors to the origin Florida are Alabama and Georgia, and neighbors to the destination Washington are Oregon and Idaho. The spatial lag $W_o y$ would represent an average of the flows from Alabama and Georgia (neighbors to the origin) to the destination state Washington.

A similar interpretation applies to other rows of the spatial lag $W_o y$. For example when examining flows from the origin state of Alabama to the destination state of Washington, the spatial lag $W_o y$ would represent an average of the flows from Florida, Georgia, Mississippi and Tennessee (neighbors to the origin) to the destination state Washington.

A second type of spatial dependence that could arise in the gravity model would be "destination-based" dependence. Intuitively, it seems plausible that forces leading to flows from an origin state to a destination state may create similar flows to nearby or neighboring destinations.

A spatial weight matrix that we label W_d can be constructed to capture this type of dependence using $W \otimes I_n$, producing an n^2 by n^2 spatial weight matrix that captures connectivity relations between the flows from an origin state to neighbors of the destination state.

To provide an example of this we consider four regions located in a row as presented in Table 1.

Table 1: Location of 4 Regions in Space

Region #1	Region #2	Region #3	Region #4

The row-standardized first-order contiguity matrix associated with this regional configuration is shown in (5).

$$W = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{5}$$

For this example, $W_d = W \otimes I_n$ takes the form shown in (6), where $\underline{0}$ represents an n by n matrix of zeros, and $n = 4$ in this example.

$$W_d = \begin{pmatrix} \underline{0} & I_n & \underline{0} & \underline{0} \\ (1/2)I_n & \underline{0} & (1/2)I_n & \underline{0} \\ 0 & (1/2)I_n & \underline{0} & (1/2)I_n \\ \underline{0} & \underline{0} & I_n & \underline{0} \end{pmatrix} \tag{6}$$

Using our example of flows from the origin state of Florida to the destination state of Washington, the spatial lag vector $W_d y$ represent an average of flows from Florida to Idaho and Oregon, states that neighbor the destination state of Washington. In our other example, where the origin state was Alabama, the spatial lag vector $W_d y$ would represent an average of flows from Alabama to Idaho and Oregon, states that neighbor the destination state of Washington.

To provide a more formal development of destination based dependence, we employ the vec-permutation matrix P introduced previously. If we adopted the destination-centric ordering, specification of the destination weight matrix would be $I \otimes W$ by the same logic as introduced in the development of the origin weight matrix. Consequently, a destination weight matrix under the origin-centric ordering would be $P'(I \otimes W)P$. Some results on Kronecker products lead to a further simplification of $P'(I \otimes W)P$. Given that P is the vec-permutation matrix, by Corollary 4.3.10 in Horn and Johnson (1991, p. 290), $W \otimes I = P'(I \otimes W)P$, and thus $W_d = W \otimes I$.

A third type of dependence to consider is reflected in the product $W_w = W_o \cdot W_d = (I_n \otimes W) \cdot (W \otimes I_n) = W \otimes W$. This spatial weight matrix reflects an average of flows from neighbors to the origin state to neighbors of the destination state. One motivation for this matrix product might be a spatial filtering perspective. We might envision a spatial autoregressive model of the type shown in (7) based on successive filtering. We transform or filter the dependent variable successively by $(I_{n^2} - \rho_1 W_o)$, and $(I_{n^2} - \rho_2 W_d)$. The motivation is that we are removing destination dependence first and subsequently origin dependence, or *vice-versa*.

$$(I_{n^2} - \rho_1 W_o)(I_{n^2} - \rho_2 W_d)y = \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + \varepsilon \tag{7}$$

This leads to a model that includes the interaction term $W_w = W_o \cdot W_d$ in the sequence of spatial lags:

$$y = \rho_1 W_o y + \rho_2 W_d y - \rho_1\rho_2 W_w y + \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + \varepsilon \tag{8}$$

Using our example of flows from the origin state of Florida to the destination state of Washington, the spatial lag vector $W_w y$ represent an average of: flows from Alabama and Georgia (neighbors to the origin state) to Idaho (a neighbor to the destination state), and flows from Alabama and Georgia (neighbors to the origin state) to Oregon (a neighbor to the destination state). In the case of our other example based on flows from the origin state of Alabama to the destination state of Washington, the spatial lag vector $W_w y$ represent an average of: flows from Florida, Georgia, Mississippi and Tennessee (neighbors to the origin state) to Idaho (a neighbor to the destination state) and flows from Florida, Georgia, Mississippi and Tennessee (neighbors to the origin state) to Oregon (a neighbor to the destination state).

Note, the implementation of this does not require the actual formation of the n^2 by n^2 matrices W_o, W_d, or W_w. Given arbitrary, conformable matrices A, B, C, $(C' \otimes A)vec(B) = vec(ABC)$ (Horn and Johnson, 1991, p. 255, Lemma 4.3.1). Since $W_o y = (I \otimes W)\vec{(Y)}$, then $W_o y = vec(WY)$. Similarly, $W_d = vec(YW')$, and $W_w = vec(WYW')$. These expressions also aid in the interpretation of origin-destination dependence. The algebra of Kronecker products can be used to form moment matrices without dealing directly with n^2 by n^2 matrices. For example, $X'_{d_i} X_{o_j}$ equals $(\iota \otimes X_i)'(X_j \otimes \iota) = \sum X_i \cdot \sum X_j$. Also, the moment sub-matrices involving only origin variables or destination variables are very simple $(X'_{d_i} X_{d_j} = X'_{o_i} X_{o_j} = nX'_i X_j)$.

In concluding we note that spatial lags involving the disturbance process could also be constructed using weight matrices W_o, W_d and the product W_w. This would allow for a model where spatial dependence arises in the error terms of the model. In the successive filtering case this would take the form in (9).

$$
\begin{aligned}
y &= \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + u \\
u &= (I_{n^2} - \rho_1 W_o)(I_{n^2} - \rho_2 W_d)u + \varepsilon
\end{aligned} \tag{9}
$$

2.2 Spatial model specifications for origin-destination flows

A family of nine different model specifications is proposed. We take as a starting point a slight generalization of the successive spatial filtering specification from (7) and (8), shown in (10) for the SAR model. The generalization in (10) stems from relaxing the restriction from (8) that $\rho_3 = -\rho_1\rho_2$.

$$
y = \rho_1 W_o y + \rho_2 W_d y + \rho_3 W_w y + \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + \varepsilon \tag{10}
$$

The generalized successive filtering model specification in (10) involves the origin, destination weight matrices and their cross-product, along with no restrictions on the parameters ρ_1, ρ_2 and ρ_3. We set forth nine different models that can be derived by placing various restrictions on the parameters $\rho_i, i = 1, \ldots, 3$. Since the statistical theory for testing parameter restrictions is well-developed, this seems desirable from an applied specification search viewpoint.

1. The restriction: $\rho_1 = \rho_2 = \rho_3 = 0$, produces the least-squares model where no spatial autoregressive dependence exists.

2. The restriction: $\rho_2 = \rho_3 = 0$, results in a model based on a single weight matrix W_o, reflecting origin autoregressive spatial dependence.

3. The restriction: $\rho_1 = \rho_3 = 0$, produces a sibling model based on a single weight matrix W_d for spatial dependence at the destinations.

4. The restriction: $\rho_1 = \rho_2 = 0$, creates another single weight matrix model containing only W_w, reflecting dependence based on interaction between origin and destination neighbors.

5. The restriction: $\rho_1 = \rho_2, \rho_3 = 0$, results in a model based on a single weight matrix constructed using $W_o + W_d$. This reflects a lack of separability between the impacts of origin and destination dependence relations in favor of a cumulative impact.

6. The restriction: $\rho_1 = \rho_2, \rho_3 = -\rho_1^2 = -\rho_2^2$, produces another single weight matrix model based on $W_o + W_d + W_w$. This reflects a lack of separability between the impacts of origin, destination and origin-destination interaction effects in favor of a cumulative impact.

7. The restriction: $\rho_3 = 0$, leads to a model with separable origin and destination autoregressive dependence embodied in the two weight matrices W_o and W_d, while ruling out dependence between neighbors of the origin and destination locations that would be captured by W_w.

8. The restriction: $\rho_3 = -\rho_1\rho_2$ results in a successive filtering or model involving both origin W_o, and destination W_d dependence as well as product separable interaction W_w, constrained to reflect the filter: $(I_{n^2} - \rho_1 W_o)(I_{n^2} - \rho_2 W_d) = (I_{n^2} + \rho_1 W_o + \rho_2 W_d - \rho_1\rho_2 W_w)$.

9. No restrictions produces the ninth member of the family of models based on an unrestricted variant of the filter: $(I_{n^2} - \rho_1 W_o)(I_{n^2} - \rho_2 W_d) = (I_{n^2} + \rho_1 W_o + \rho_2 W_d + \rho_3 W_w)$

Each of the single spatial weight matrix model specifications in 1) to 6) would obey the usual properties of row-normalized weight matrices, allowing use of existing algorithms for maximum likelihood (Pace and Barry, 1997), Bayesian (LeSage, 1997) or generalized method of moments estimation estimation (Kelejian and Prucha, 1999).

We note that specifications 1) to 6) based on single weight matrices are also amenable to variants of spatial regression models of the type shown in (11) to (13), which we label SAR, SEM and SAC models, respectively. In these equations, we use W_j to denote the single spatial weight matrix.[3]

$$y = \rho W_j y + \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + \varepsilon \quad (11)$$

$$\begin{aligned} y &= \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + u \quad (12) \\ u &= \rho W_j u + \varepsilon \end{aligned}$$

$$\begin{aligned} y &= \rho W_j y + \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + u \quad (13) \\ u &= \lambda W_j u + \varepsilon \end{aligned}$$

We note that use of conventional algorithms for maximum likelihood, Bayesian or generalized method of moments estimation of the spatial econometric origin-destination interregional flow models becomes difficult as the

[3]There are widely available algorithms for estimation of these alternative specifications, e.g., the spatial econometrics toolbox, www.spatial-econometrics.com and spatial statistics toolbox, www.spatial-statistics.com.

number of observations increases. For example, use of an origin-destination flow matrix for the sample of approximately 3,100 US counties would result in sparse spatial weight matrices of dimension n^2 by n^2 where $n^2 = 9,610,000$. Maximum likelihood and Bayesian estimation both require calculation of the logged determinant for the n^2 by n^2 matrix $(I_{n^2} - \rho W_j)$. While specialized approaches to calculating log-determinants of very large matrices have been proposed by Pace and LeSage (2004) and Smirnov and Anselin (2001), it turns out there are much more efficient approaches that can exploit the special structure of matrices like $W_d = I_n \otimes W$, $W_o = W \otimes I_n$ and $W_w = W_o \cdot W_d = W \otimes W$. We turn attention to this topic in the next section.

3 Estimation of spatial flow models

As already noted, SAR, SEM and SAC model specifications based on a spatial weight matrix $W_j = W_o, W_d, W_w$, as well as sums of these such as, $W_j = W_o + W_d, W_j = W_o + W_d + W_w$, can be implemented using standard algorithms. For a model based on $n = 50$ states, this results in $n^2 = 2,500$, a situation where the log-determinant calculation as well as all other calculations required for maximum likelihood or Bayesian estimation can be completed rapidly (LeSage and Pace, 2004). Despite this, the structure of the matrices W_o, W_d and W_w allow for computational improvements.

3.1 The case of a single weight matrix $W_k, k = o, d$

We note that the concentrated log-likelihood function for model specifications 1) to 6) based on a single spatial weight matrix, which we denote W_j, concentrated with respect to the parameters β and σ will take the form:

$$\text{Log}L(\rho) = C + log|I_{n^2} - \rho W_j| - (n^2/2)log(e'e(\rho)) \tag{14}$$

Where $e'e(\rho)$ represents the sum of squared errors expressed as a function of the scalar parameter ρ alone after concentrating out the parameters β, σ, and C denotes a constant not depending on ρ (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 a 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.

The log-determinant of the transformation 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 $I_{n^2} - \rho W_j$, shown in (15).

$$\ln|I_{n^2} - \rho W_j| = tr\left(\ln(I_{n^2} - \rho W_j)\right) = -\sum_{t=1}^{\infty} \frac{\rho^t tr(W_j^t)}{t} \qquad (15)$$

For the case of destination or origin weight matrices, $W_d = I_n \otimes W$ or $W_o = W \otimes I_n$, which we designate W_k, $k = o, d$,

$$tr(W_k^t) = tr(I_n^t \otimes W^t) = tr(I_n^t) \cdot tr(W^t) = n \cdot tr(W^t), \qquad (16)$$

and thus the trace of a square matrix of order n^2 is simplified to a scalar (n) times a trace of the n by n square matrix W.

$$\ln|I_{n^2} - \rho W_k| = -n \sum_{t=1}^{\infty} \frac{\rho^t tr(W_k^t)}{t} = n \ln|I_n - \rho W| \qquad (17)$$

Summarizing, for the case of a single spatial weight matrix W_k, $k = o, d$, it should be possible to rely on algorithms for computing the logged determinant of an n by n matrix $\ln|I_n - \rho W|$, when working with a vector of n^2 origin-destination flows. For the earlier example of $n = 3,100$ US counties and $n^2 = 9,610,000$, we should be able to 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).

3.2 Calculating logged determinants for the filtering models 7) to 9)

For the more general successive spatial filtering model specifications 7) to 9), we observe that the order in which one transforms the dependent variable makes no difference. The explanation for this can be seen by considering the cross products in (18). The mixed-product rule for Kronecker products indicates that the cross-product of $W \otimes I_n$ and $I_n \otimes W$ is $W \otimes W$, which is the same as the cross-product of $I_n \otimes W$ and $W \otimes I_n$.

$$
\begin{aligned}
(I_{n^2} - \rho_1 W_o)(I_{n^2} - \rho_2 W_d) &= (I_{n^2} - \rho_1 (I_n \otimes W))(I_{n^2} - \rho_2 (W \otimes I_n)) \\
&= I_{n^2} - \rho_1 (I_n \otimes W) - \rho_2 (W \otimes I_n) \qquad (18) \\
&\quad + \rho_1 \rho_2 (W \otimes W)
\end{aligned}
$$

Because the log-determinant of a product is the sum of the log-determinants, the overall log-determinant arising from the successive filtering approach is quite simple as shown in (19).

$$
\ln \left| (I_{n^2} - \rho_1 W_o)\,(I_{n^2} - \rho_2 W_d) \right| = n(\ln |I_n - \rho_1 W| + \ln |I_n - \rho_2 W|) \qquad (19)
$$

As in the case of single spatial weights and conventional model specifications from the previous section, the logged determinant required for maximum likelihood estimation of successive filtering models can be calculated using only n by n matrices rather than the large n^2 by n^2 matrices.

Drawing on the earlier discussion surrounding (17), the estimation challenge for the case of the most general spatial filtering model 9), with all three parameters ρ_1, ρ_2 and ρ_3 unrestricted is to easily compute $tr(W_f^t)$ for $t = 1 \cdots m$, where m is the largest moment computed, and W_f is defined in (20).

$$
W_f = \rho_1 (I_n \otimes W) + \rho_2 (W \otimes I_n) + \rho_3 (W \otimes W) \qquad (20)
$$

The case of $tr(W_f)$ where $t = 1$ is immediate, and equals zero since $tr(W) = 0$. The case of $tr(W_f^2)$ is slightly more challenging as shown in (21).

$$
\begin{aligned}
W_f^2 &= \rho_1^2 (I_n \otimes W^2) + \rho_2^2 (W^2 \otimes I_n) + \rho_3^2 (W^2 \otimes W^2) \qquad (21) \\
&+ 2\rho_1\rho_2 (W \otimes W) + 2\rho_1\rho_3 (W^2 \otimes W) + 2\rho_2\rho_3 (W \otimes W^2)
\end{aligned}
$$

For the quadratic, there are 9 possible terms and 6 of these are unique. Note, $tr(W_f^2)$ is the highest order term associated with W. Extrapolating, computations of $tr(W_f^t)$ only require computing $tr(W^t)$ based on the n by n weight matrix W, a much less demanding task. Individual terms have the form in (22).

$$
\rho_1^i \rho_2^j \rho_3^k tr\left(W_o^i W_d^j W_w^k\right) = \rho_1^i \rho_2^j \rho_3^k tr(W^{(i+k)}) tr(W^{(j+k)}) \qquad (22)
$$

Given a table of $tr(W^t)$ for $t = 1 \cdots m$, each term involves the multiplication of five scalars. However, there are 3^m terms, and this becomes difficult for large m. Other than computing $tr(W^t)$, none of these computations are dependent upon n, and so it takes just as long for a problems with many origins and destinations as for smaller problems. For small n, calculating exact $tr(W^t)$ requires little time. For large n, calculating $tr(W^t)$ can be approximate as in Barry and Pace (1999) who show how to do this with an $O(n)$ algorithm.

Given the m moments and the conditions on W, it becomes easy to compute an relatively short interval containing the log-determinant as shown in (23).

$$-\sum_{t=1}^{m} \frac{tr(W_f^t)}{t} \geq \ln |I_{n^2} - W_f| \geq -\left(\sum_{t=1}^{m} \frac{tr(W_f^t)}{t} + \sum_{t=m+1}^{\infty} \frac{tr(W_f^m)}{t} \right) \tag{23}$$

Pace and LeSage (2002) show how the moments $tr(W_f^t)$ must monotonically decline for $t > 1$, and this sets up the bounds. The interval is narrow provided $(\rho_1 + \rho_2 + \rho_3)^{m+1}/(m+1)$ is reasonably small. A requirement for stability is that $\rho_1 + \rho_2 + \rho_3 < 1$, making this a reasonable presumption.

Summarizing, we derived a family of nine model specifications that emphasize different spatial connectivity relations between origin and destination regions. Since members of the family of specifications reflect models based on parameter restrictions, these can be easily tested to draw inferences regarding the nature of spatial dependence in any applied problem. Potential computational problems that might plague estimation for models involving n^2 observations on origin-destination flows were eliminated by reducing the troublesome logged determinant calculation to one involving only traces of n by n matrices. As already noted, successive filtering of the type described here could also be applied to the disturbance process, producing a family of nine models of the type we have labelled SEM, or to both the dependent variable and disturbance vectors resulting in nine more models of the type we labelled SAC.

4 An applied illustration using state-level population migration flows

To illustrate the family of spatial econometric models described here we use state-level population migration flows as the dependent variable. Specifically, the growth rates in migration flows for the population 5 years and over from the period covering 1985 to 1990 and flows for the period 1995 to 2000 were used.[4] The sample was restricted to the 48 contiguous states plus the District of Columbia resulting in $n = 49$ and $n^2 = 2,401$ observations. The growth rates for flows of population within each state were set to zero to emphasize flows between states which should exhibit the type of spatial dependence of interest here.

Another benefit of the growth rates transformation is alleviation of the problem noted earlier that arises with flows that are very large within regions relative to many zero values for interregional flows. Figure 1 shows a histogram of the annualized growth rates in the flows alongside a normal probably density plot. From the figure we see some evidence of fat tails reflecting more extremely large or small growth rates than one would expect in a normal distribution. We also see the impact of setting 49 within-region flows to zero values. An approach to dealing with the fat-tailed nature of the distribution of flows during estimation will be illustrated in Section 5.

Explanatory variables for the matrices X_o, X_d for each state were taken from the 1990 Census, with the exception of the unemployment rate variable, which was constructed as the ratio of state-level unemployment rates in 1995 to 1990. These variables are documented in Table 2.

The motivation for including the age variables 'near retirement', and 'retired' is that these should exert an impact on migration decisions. Apriori, we would expect that 'retired' would increase flows from the origin, whereas 'near retirement' should decrease flows from both destinations and origins. We note that an increase in flows is indicated by a positive coefficient estimate and a decrease by a negative estimate.

Population that lived in another state in 1985 might increase flows at both the origin and destination as this is an indicator of population mobility. The effect of foreign born population seems unknown apriori, depending on the mobility of this population relative to the average.

Population holding graduate and professional degrees should be the most mobile, leading to increased flows at both the origin and destination, whereas persons with less than ninth grade education should be less mobile. The impact of associate degrees, college degrees, and sales jobs is less clear.

Rents and unemployment rates should increase flows at the origin and decrease flows at the destination, whereas per capita income should decrease flows at the origin and increase flows at the destination. The 'area' variable

[4]Available on the internet State-to-State Migration Flows: 1995 to 2000 Census 2000 Special Reports. The data are based on a sample. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see http://www.census.gov/prod/cen2000/doc/sf3.pdf.

was included to control for the impact of variation in the size of the states on migration growth rates.

In addition to the variables included in the matrices X_o, X_d, the log of distance from each origin to each destination was included in the model, along with a constant term.

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 3, ordered from high to low, along with a likelihood ratio (LR) test of the restrictions imposed by each model versus the unrestricted model. It is clear from the table that the Models 7, 8 and 9 based on the filtering specification that contains separate spatial weight matrices for the origin and destination provide a significantly higher likelihood than Models 2 through 6 that use a single spatial weight matrix. There is also a noticeable drop in the likelihood values when going from models 5 and 6 to models models 2, 3 and 4. We note that models 5 and 6 are based on single weight matrices constructed by summing information from both origin and destination weight matrices, whereas to models 2, 3 and 4, are based on only and origin or only destination, or only the interaction weight matrices. This would seem to support the notion that both origin and destination dependence/connectivity information are important.

The LR tests indicate that the Model 7 restriction $\rho_3 = 0$ does not significantly reduce the likelihood function value. This restriction eliminates the weight matrix W_w reflecting connectivity between neighbors to the origin and neighbors to the destination. We might interpret this result as indicating these relations are relatively unimportant in explaining growth rates in the migration flows over our time period. The parameter estimates for the unrestricted model are presented in Table 4, where we see that the parameter ρ_3 is not significantly different from zero, consistent with the LR test results.

The LR test result for Model 8 based on the restriction that $\rho_3 = -\rho_1 \cdot \rho_2$, versus the unrestricted model rejects this restriction as consistent with the sample data, at the 95 percent level.

Finally, it is clear that least-squares which ignores spatial dependence in the growth rates of the migration flows and assumes these are independent produces a much lower likelihood function value.

Turning to the parameter estimates from least-squares and the unrestricted spatial Model 9, shown in Table 4, we see estimates for $\rho_1 = 0.313$ and $\rho_2 = 0.280$, indicating spatial dependence of equal importance between: neighbors to the origin and the destination, and neighbors to the destination and the origin. As indicated above, the estimate for $\rho_3 = -0.0072$ is not significantly different from zero, allowing us to infer that dependence between neighbors to the origin and neighbors to the destination specified by the weight matrix W_w is not important.

Turning attention to the parameter estimates, we see that distance is positive and significant in the least-squares model, but insignificant in the spatial model. (Distance was insignificant in all 8 spatial models.) Typically, regression-based gravity models produce a negative influence of distance on the flows, which seems intuitively appealing, whereas this is not the case here. We note that if the true data-generating process was in fact a model containing a spatial lag, then least-squares estimates are biased and inconsistent (see LeSage and Pace, 2004).

From Table 4, we see many cases where least-squares estimates (in absolute value terms) are larger than those from the spatial model, which is typical of least-squares, since it attributes variation assigned to the spatial lags of the dependent variable by the spatial model to explanatory variables. For example, spatial model estimates for: D_nearretirement, D_diffstate, D_rents, D_associate O_college, O_gradprof O_rents, and O_unemp, take on values around one-half those from least-squares, while D_unemployment is an exception.

Three of the variables exhibit a change from significantly different from zero to insignificant between the spatial and least-squares models: D_sales, O_sales, and distance are all insignificant in the spatial model. Three other variables change in the level of significance: D_college, D_unemp, have a higher level of significance in the spatial model and O_unemp has a lower level of significance.

The estimates from the spatial filtering model indicate that: higher rents, higher unemployment and more associate degrees lead to higher growth rates in origin flows, but lower destination flows, as might be expected. Per capita incomes also have the expected positive impact on destination flows, with an insignificant impact on origin flows.

Persons near retirement (aged 60 to 64) reduces flows at both the origin and destination, as do college graduates. Persons with graduate and professional degrees increase flows at both the origin and destination as do persons that lived in a different state in 1985 than in 1990. This suggest these groups reflect higher than average mobility, which seems plausible. In contrast, foreign-born population reduces flows at both the origin and destination, suggesting lower mobility. Retired persons (aged 65 to 74) increase flows only at the origin, having an insignificant impact on the destination.

Finally, area exerts a positive impact on flows at both the origin and destination suggesting that states with larger physical areas exhibit higher growth rates in migration flows when controlling for other factors.

5 Extensions of the general spatial filtering model to accommodate special issues

Some issues that were raised earlier regarding modeling of vectorized origin-destination flow matrices were: 1) the fat-tailed nature of the distribution of the vectorized flow matrix, even after transformation to growth rates over time (as illustrated by Figure 1 for our state-level migration flows example); 2) the presence of numerous zeros for large flow matrices as in the case of migration flows between US counties, reflecting a lack of interaction between numerous regions in the sample; and 3) the presence of large flows on the diagonal of the flow matrix reflecting a large degree of intra-regional connectivity;

Bayesian estimation procedures for conventional spatial models of the type labelled SAR, SEM and SAC here have been set forth in LeSage (1997) and for probit and tobit variants of these models in LeSage (2000). One important aspect of these estimation procedures which rely on Markov Chain Monte Carlo methods (MCMC) (Gelfand and Smith, 1990) is that they can accommodate sample y vectors that exhibit fat-tails, or follow a Student $t-$distribution, rather than the conventional normal distribution. This suggests that these methods might be able to overcome problems associated with issue 1) above. Another feature of Bayesian MCMC estimation of these models is that tobit extensions to accommodate sample censoring in the y vector are relatively straightforward to implement, allowing a possible solution to issues raised by 2) above. Regarding issue 3) above, it may be the case that a two-regime, hierarchical or mixture model could be used to describe intraregional variation in flows (within regions) separately from interregional flows (between regions). There is a large literature on use of Bayesian mixture and hierarchical spatial models that might be applicable to issue 3) above (see Besag, York, and Mollie, 1991, Besag and Kooperberg, 1995, and Cressie 1995).

We illustrate one of these possible extensions by presenting estimates based on a robust variant of the SAR model presented in LeSage (1997) for our state-level migration flow growth rates. This involves Bayesian MCMC estimation of a model that takes the same form as the general spatial filtering model, but relaxes the constant variance assumption regarding the disturbances in the data generating process. In place of $\varepsilon \sim N(0, \sigma^2 I_n^2)$, we assume that:

$$\begin{aligned} \varepsilon &\sim N[0, \sigma^2 \text{diag}(V)] \\ V &= v_1, v_2, \ldots, v_{n^2} \end{aligned} \tag{24}$$

To produce estimates for the n^2 variance scalars in (24), we follow an approach introduce by Geweke (1993), that places a $\chi^2(r)$ prior on the variance scalars v_i with a mean of unity and a mode and variance that depend on the hyperparameter r of the prior. Small values of r around 5 result in a prior that allows for the individual v_i estimates to be centered on their prior mean of unity, but deviate greatly from the prior value of unity in cases where the model residuals are large. Large residuals are indicative of outliers or origin-destination combinations that are atypical or aberrant relative to the majority of the sample of origin-destination flows. Geweke (1993) points to the equivalence of this modeling approach and the assumption of disturbances that follow a Student $t-$distribution.

The MCMC estimation method samples sequentially from the complete set of conditional distributions for all parameters in the model. Sampling from the conditional distributions for the parameters β and σ, when uninformative priors are assigned to these, is relatively straightforward as they take known distributional forms. The conditional distribution for the β_o, β_d parameters take the form of a $k-$variate multivariate normals, and that for the parameter σ is a $\chi^2(n^2)$ distribution.[5] In this model we must sample the three parameters ρ_1, ρ_2, ρ_3 as well as the variance scalars $v_i, i = 1, \ldots, n^2$. The conditional distribution for the parameter ρ_1, conditional on the remaining parameters $\theta = (\beta, \sigma, V, \rho_2, \rho_3)$ take the form the form shown in (25), with the conditionals for ρ_2 and ρ_3 taking similar forms.

$$\begin{aligned} p(\rho_1|\theta) &= \log|I_{n^2} - \rho_1 W_o - \rho_2 W_d - \rho_3 W_w| \\ &- \frac{n^2-k}{2} log\{[e(\rho_1, \rho_2, \rho_3)' V^{-1} e(\rho_1, \rho_2, \rho_3)]/(n^2 - k)\} \\ e(\rho_1, \rho_2, \rho_3) &= y - \alpha\iota - \rho_1 W_o - \rho_2 W_d - \rho_3 W_w - X_o\beta_o - X_d\beta_d - D\gamma \end{aligned} \tag{25}$$

Where we note the presence of the logged determinant term as in the case of maximum likelihood estimation. We can rely on the same algorithms for rapidly evaluating this expression in the context of Bayesian MCMC estimation as in maximum likelihood. Sampling for the parameters $\rho_i, i =$

[5]See LeSage (2004), pp. 232-233 for the exact expressions needed here.

$1, 2, 3$ is accomplished using expressions similar to (25) in a Metropolis-Hastings algorithm based on a tuned normal random-walk proposal.

Table 5 presents the posterior means and highest posterior density (HPD) intervals based on 0.05 and 0.95 percentiles for the parameters of the most general spatial filtering model. Maximum likelihood estimates are also included in the table to facilitate comparison. We see 19 parameters whose posterior mean is different from zero based on the 0.95 HPD intervals, which contrasts with 18 such parameters from maximum likelihood estimation using the 95 percent level of significance. Differences arise for four destination characteristics: D_sales, D_foreignborn, D_grade9 and D_associate; and two origin characteristics: O_nearretirement, and O_sales.

There are also differences in the magnitudes of the parameter estimates, even when both Bayesian and maximum likelihood estimates are significantly different from zero. For example, the MCMC estimate of D_sales is twice as large as maximum likelihood, whereas maximum likelihood estimates for O_college and O_gradprof are about twice those from the robust Bayesian model. and O_rents is twice as large.

Finally, we see evidence of stronger spatial dependence in larger posterior mean estimates for the parameters ρ_1 and ρ_2. We note that the 0.05 and 0.95 HPD intervals for these two parameters do not include the maximum likelihood estimates, suggesting a substantial increase in spatial dependence when we account for non-constant variance.

Turning attention to the variance scalar estimates, these provide a diagnostic for observations (or OD pairs) that do not conform well to the model relationship. Large estimates for these variances suggest an outlier or aberrant observation. Of the 2,401 observations, 2,233 of the posterior mean values for the scalars v_i had values of 3 or less, and 2,354 values of 6 or less. There were 17 v_i estimates whose posterior means exceeded a value of 10, indicative of exceptionally high or low migration growth rates that could not be explained by the model's origin and destination variables/characteristics or the spatial autoregressive dependence structure.

Table 6 shows the origin-destination state pairs for the 17 cases where the posterior mean v_i values exceeded 10 along with the mean v_i estimate. These observations would be downweighted by the inverse of the v_i values during MCMC estimation of the robust model. This is in contrast to the maximum likelihood estimation procedure where all observations are assigned equal weight. This accounts for the differences between the Bayesian and maximum likelihood estimates. From the table, we see that origin-destination state pairs identified as outliers conform to intuition, reflecting the growth rates in migration flows from mostly small states to other small states. This is where we would expect to see large variances in the growth rates of migration flows over the 1985-90 and 1995-2000 periods, likely due to volatility that arises in growth rates calculated on the basis of small levels of flows.

6 Conclusions

We set forth a method for incorporating spatial autoregressive structures in conventional regression-based gravity models. This extension allows a family of conventional spatial regression models that explicitly model the spatial dependence structure between cross-sectional observations to be employed in modeling origin-destination flows.

The approach introduced here allows for application of conventional spatial regression algorithms for estimation and inference in the case of small samples. In addition, we provide a solution for more realistic cases where large samples involving flows between US counties numbering nearly 10 million observations can be estimated in a matter of seconds. This requires only slight modification to existing algorithms. Much of what we have learned about maximum likelihood and Bayesian estimation of spatial regression models can be immediately applied to origin-destination flow modeling.

As an extension to conventional spatial autoregressive and spatial error models, we introduce a family of models that subsume these conventional 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.

Three special issues that arise in origin-destination flow modeling were discussed: 1) the fat-tailed nature of the distribution of the vectorized flow matrix; 2) the presence of numerous zeros for large flow matrices reflecting a lack of interaction between numerous regions in the sample; and 3) the presence of large flows on the diagonal of the flow matrix reflecting a large degree of intra-regional connectivity. Solutions to some of these problems may be possible by drawing on past work from conventional spatial econometrics. As an illustration, robust Bayesian Markov Chain Monte Carlo estimates for the model introduced in this study were presented. These estimates accommodate the fat-tailed nature of the distribution of the vectorized flow matrix.

7 References

Besag, J. E., York, J.C. and Mollie, A. (1991). "Bayesian image restoration, with two applications in spatial statistics" (with discussion), *Annals of the Institute of Statistical Mathematics,* Vol. 43, pp. 1-59.

Besag, J. E. and Kooperberg, C.L. (1995). "On conditional and intrinsic autoregressions", *Biometrika*, Vol. 82, pp. 733-746.

Cressie, N. (1995). "Bayesian smoothing of rates in small geographic areas", *Journal of Regional Science*, Vol. 35, pp. 659-673.

Cushing, Brian, and Jacques Poot (2003), "Crossing Boundaries and Borders:: Regional Science Advances in Migration Modelling," *Papers in Regional Science*, Vol. 83, pp. 317-338.

Gelfand, Alan E., and A.F.M Smith. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association,* Vol. 85, pp. 398-409.

Geweke J. (1993), "Bayesian Treatment of the Independent Student t Linear Model," *Journal of Applied Econometrics*, Vol. 8, pp. 19-40.

Horn, R.A., and C.R. Johnson (1991). *Topics in Matrix Analysis*, Cambridge: Cambridge University Press.

LeSage, J.P. (1997) "Bayesian Estimation of Spatial Autoregressive Models," *International Regional Science Review*, Volume 20, number 1&2, pp. 113-129.

LeSage, J.P. (2000) "Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models," *Geographical Analysis*, Volume 32, number 1, pp. 19-35.

LeSage, J.P. (2004) "Spatial Regression Models," in *Numerical Issues in Statistical Computing for the Social Scientist*, John Wiley & Sons, Inc., Micah Altman, Jeff Gill and Michael McDonald (eds.), pp. 199-218.

LeSage, J.P. and R. Kelley Pace (2004) "Introduction" *Advances in Econometrics: Volume 18: Spatial and Spatiotemporal Econometrics*, (Oxford: Elsevier Ltd), pp. 1-32.

Lee,Ming-Long and R. Kelley Pace (2005) "Spatial Distribution of Retail Sales," *Journal of Real Estate Finance and Economics*, Volume 31, number 1, pp. 53-69.

Pace, R.K. and James P. LeSage (2002), "Semiparametric Maximum Likelihood Estimates of Spatial Dependence," *Geographical Analysis*, January 2002, Volume 34, Number 1, pp. 75-90.

Pace, R.K, and J.P. LeSage (2004) "Techniques for Improved Approximation of the Determinant Term in the Spatial Likelihood Function," *Computational Statistics and Data Analysis*, 2004, Volume 45, pp. 179-196.

Porojan, A. (2001) "Trade Flows and Spatial Effects: The Gravity Model Revisited," *Open Economic Review*, Volume 12, pp. 265-280.

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

Smirnov, O. and L. Anselin (2001). "Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: a Characteristic Polynomial Approach." *Computational Statistics and Data Analysis* 35, 301-319.

Table 2: Explanatory variables used in the model

Variable name	Description
young	log (population aged 22-29/population in 1990)
near retirement	log (population aged 60-64/ population in 1990)
retired	log (population aged 65-74/ population in 1990)
sales	log (proportion of work force in sales in 1990)
diffstate	log (proportion of population living in a different state in 1985)
foreign born	log (proportion of foreign born population in 1990)
grade9	log ($<$ 9th grade as highest degree in 1990/population $>$ age 25)
associate	log (associate degree in 1990/population $>$ age 25)
college	log (college as highest degree in 1990/population $>$ age 25)
grad prof	log (graduate or professional degree in 1990/population $>$ age 25)
rents	log (median rent in 1990)
unemp	ratio of 1995 unemployment rate to 1990 unemployment rate
pc income	log (per capita income in 1990)
area	log (1990 state area in square miles)

Table 3: Log Likelihoods for alternative models

Model	Log Likelihood	LR test versus Model 1	Critical Value ($\alpha = 0.05$)
Model 9: ρ_1, ρ_2, ρ_3 unrestricted	605.1585		
Model 7: $\rho_3 = 0$	605.1245	0.0680	$\chi^2(1) = 3.84$
Model 8: $\rho_3 = -\rho_1 \cdot \rho_2$	602.9628	4.3914	$\chi^2(1) = 3.84$
Model 5: $\rho_1 = \rho_2, \rho_3 = 0$	597.7079	14.9012	$\chi^2(2) = 5.99$
Model 6: $\rho_1 = \rho_2, \rho_3 = -\rho_1^2$	582.5549	45.2072	$\chi^2(2) = 5.99$
Model 2: $\rho_2 = \rho_3 = 0$	534.8818	140.5534	$\chi^2(2) = 5.99$
Model 3: $\rho_1 = 0, \rho_3 = 0$	516.1341	178.0488	$\chi^2(2) = 5.99$
Model 4: $\rho_1 = 0, \rho_2 = 0$	469.8518	270.6134	$\chi^2(2) = 5.99$
Model 1: $\rho_1 = 0, \rho_2 = 0, \rho_3 = 0$	388.5083	433.3004	$\chi^2(3) = 7.82$

Table 4: Estimates from least-squares and the unrestricted spatial Model 9

	Spatial model		Least-squares	
Variable	Coefficient	t-statistic(plevel)	Coefficient	t-statistic(plevel)
constant	-7.2281	-11.230 (0.0000)	-11.4183	-5.503 (0.0000)
D_nearretirement	-0.5407	-2.465 (0.0138)	-0.8669	-3.236 (0.0012)
D_retired	0.1420	1.001 (0.3166)	0.1848	1.096 (0.2729)
D_sales	0.1245	1.152 (0.2491)	0.2438	2.022 (0.0432)
D_diffstate	0.3454	2.498 (0.0125)	0.7137	4.724 (0.0000)
D_foreignborn	-0.3464	-1.580 (0.1140)	-0.1772	-0.733 (0.4633)
D_grade9	-0.0817	-3.443 (0.0006)	-0.1013	-3.886 (0.0001)
D_associate	-0.1329	-3.620 (0.0003)	-0.1995	-4.918 (0.0000)
D_college	-0.1371	-1.976 (0.0482)	-0.1289	-1.659 (0.0971)
D_gradprof	0.0289	0.508 (0.6110)	-0.0265	-0.416 (0.6768)
D_rents	-0.5884	-4.360 (0.0000)	-0.9086	-6.205 (0.0000)
D_area	0.0126	1.549 (0.1213)	0.0258	2.867 (0.0041)
D_unemp	-0.7892	-3.444 (0.0006)	-0.6102	-2.465 (0.0137)
D_pcincome	0.4928	3.314 (0.0009)	0.6564	3.880 (0.0001)
O_nearretirement	-0.4955	-2.695 (0.0071)	-0.8755	-3.268 (0.0010)
O_retired	0.4520	3.377 (0.0007)	0.7546	4.476 (0.0000)
O_sales	-0.1505	-1.373 (0.1697)	-0.2365	-1.961 (0.0498)
O_diffstate	0.8783	6.160 (0.0000)	1.6001	10.593 (0.0000)
O_foreignborn	-1.3203	-5.908 (0.0000)	-1.7875	-7.400 (0.0000)
O_grade9	0.0163	0.697 (0.4856)	0.0375	1.437 (0.1506)
O_associate	0.0718	1.913 (0.0558)	0.1969	4.856 (0.0000)
O_college	-0.2834	-4.232 (0.0000)	-0.5823	-7.497 (0.0000)
O_gradprof	0.1971	3.502 (0.0005)	0.3167	4.976 (0.0000)
O_rents	0.6938	5.446 (0.0000)	1.2287	8.391 (0.0000)
O_area	0.0456	5.486 (0.0000)	0.0571	6.342 (0.0000)
O_unemp	0.4999	2.308 (0.0211)	1.2287	4.964 (0.0000)
O_pcincome	-0.1371	-1.256 (0.2091)	-0.3004	-1.775 (0.0758)
log(distance)	0.0018	0.492 (0.6226)	0.0084	2.172 (0.0299)
ρ_1	0.3135	13.105 (0.0000)		
ρ_2	0.2800	11.198 (0.0000)		
ρ_3	-0.0072	-0.172 (0.8627)		
σ^2	0.0675		0.0819	

Table 5: Estimates from maximum likelihood and robust Bayesian versions of spatial Model 9

	Bayesian MCMC			Maximum Likelihood	
Variable	Posterior mean	Lower 0.05 HPD	Upper 0.95 HPD	Coefficient	t-statistic (plevel)
constant	-5.1139	-7.6374	-2.7329	-7.2281	-11.230 (0.0000)
D_nearretirement	-0.4835	-0.8174	-0.1512	-0.5407	-2.465 (0.0138)
D_retired	0.1346	-0.0674	0.3472	0.1420	1.001 (0.3166)
D_sales	0.2328	0.0954	0.3757	0.1245	1.152 (0.2491)
D_diffstate	0.2122	0.0204	0.3861	0.3454	2.498 (0.0125)
D_foreignborn	-0.4218	-0.7004	-0.1506	-0.3464	-1.580 (0.1140)
D_grade9	-0.0309	-0.0619	0.0001	-0.0817	-3.443 (0.0006)
D_associate	-0.0455	-0.0957	0.0015	-0.1329	-3.620 (0.0003)
D_college	-0.1222	-0.2137	-0.0329	-0.1371	-1.976 (0.0482)
D_gradprof	0.0589	-0.0174	0.1334	0.0289	0.508 (0.6110)
D_rents	-0.5811	-0.7516	-0.4058	-0.5884	-4.360 (0.0000)
D_area	-0.0003	-0.0115	0.0109	0.0126	1.549 (0.1213)
D_unemp	-0.7896	-1.0800	-0.4954	-0.7892	-3.444 (0.0006)
D_pcincome	0.4612	0.2689	0.6606	0.4928	3.314 (0.0009)
O_nearretirement	-0.2168	-0.5271	0.0966	-0.4955	-2.695 (0.0071)
O_retired	0.2413	0.0446	0.4425	0.4520	3.377 (0.0007)
O_sales	-0.1673	-0.3067	-0.0153	-0.1505	-1.373 (0.1697)
O_diffstate	0.7839	0.5884	0.9744	0.8783	6.160 (0.0000)
O_foreignborn	-1.0458	-1.3278	-0.7674	-1.3203	-5.908 (0.0000)
O_grade9	0.0143	-0.0162	0.0456	0.0163	0.697 (0.4856)
O_associate	0.0701	0.0191	0.1212	0.0718	1.913 (0.0558)
O_college	-0.1808	-0.2746	-0.0883	-0.2834	-4.232 (0.0000)
O_gradprof	0.1013	0.0266	0.1735	0.1971	3.502 (0.0005)
O_rents	0.5575	0.3791	0.7335	0.6938	5.446 (0.0000)
O_area	0.0365	0.0245	0.0480	0.0456	5.486 (0.0000)
O_unemp	0.5656	0.2792	0.8501	0.4999	2.308 (0.0211)
O_pcincome	-0.0936	-0.2870	0.1064	-0.1371	-1.256 (0.2091)
log(distance)	0.0045	-0.0010	0.0091	0.0018	0.492 (0.6226)
ρ_1	0.3821	0.3424	0.4262	0.3135	13.105 (0.0000)
ρ_2	0.3408	0.2952	0.3907	0.2800	11.198 (0.0000)
ρ_3	-0.0512	-0.1240	0.0197	-0.0072	-0.172 (0.8627)

Table 6: Origin-Destination pairs for variance scalar estimates greater than 10

Origin State	Destination State	Posterior mean v_i
RI	ND	144.32
DE	DC	58.38
VT	SD	44.65
VT	NE	38.81
ND	RI	32.65
DE	ND	28.47
VT	WY	24.57
MT	DE	18.54
RI	WV	18.53
WY	RI	18.23
DE	SD	18.11
ND	CT	14.63
SD	CT	14.54
NH	SD	13.06
VT	MN	12.40
ID	DE	11.33
IA	VT	10.52

Figure 1: Distribution of Annualized Migration Growth Rates