IHS Economics Series
Working Paper 188
May 2006

Incorporating Transportation Network Structure in Spatial Econometric Models of Commodity Flows

James P. Lesage
Wolfgang Polasek

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

Impressum

Author(s):

James P. Lesage, Wolfgang Polasek

Title:

Incorporating Transportation Network Structure in Spatial Econometric Models of Commodity Flows

ISSN: Unspecified

2006 Institut für Höhere Studien - Institute for Advanced Studies (IHS)

Josefstädter Straße 39, A-1080 Wien

E-Mail: office@ihs.ac.at

Web: www.ihs.ac.at

188

Reihe Ökonomie

Economics Series

Incorporating Transportation Network Structure in Spatial Econometric Models of Commodity Flows

James P. LeSage, Wolfgang Polasek

188
Reihe Ökonomie
Economics Series

Incorporating Transportation Network Structure in Spatial Econometric Models of Commodity Flows

James P. LeSage, Wolfgang Polasek

May 2006

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

James P. LeSage
University of Toledo
Department of Economics
Toledo, OH 43606, USA
email: jlesage@spatial-econometrics.com

Wolfgang Polasek
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria
☎: +43/1/599 91-155
fax: +43/1/599 91-163
email: polasek@ihs.ac.at

Abstract

We introduce a regression-based gravity model for commodity flows between 35 regions in Austria. We incorporate information regarding the highway network into the spatial connectivity structure of the spatial autoregressive econometric model. We find that our approach produces improved model fit and higher likelihood values. The model accounts for spatial dependence in the origin-destination flows by introducing a spatial connectivity matrix that allows for three types of spatial dependence in the origins to destinations flows. We modify this origin-destination connectivity structure that was introduced by LeSage and Pace (2005) to include information regarding the presence or absence of a major highway/train corridor that passes through the regions. Empirical estimates indicate that the strongest spatial autoregressive effects arise when both origin and destination regions have neighboring regions located on the highway network. Our approach provides a formal spatial econometric methodology that can easily incorporate network connectivity information in spatial autoregressive models.

Keywords
Commodity flows, spatial autoregression, Bayesian, maximum likelihood, spatial connectivity of origin-destination flows

JEL Classification
R1, R41, L92, C21

Contents

1 Introduction

This paper extends the spatial econometric methods for modeling origin-destination matrices containing interregional flows introduced in LeSage and Pace (2005). These are general data structures used in a variety of economic, geography and regional science research contexts. Our focus is on interregional flows where a network structure exists to connect the regions. The network literature often makes a distinction between networks that are "open-access" versus "closed-access". Our approach would accommodate either type of network, focusing only on the presence or absence of a network route in the regions under study. Following LeSage and Pace (2005), our methodology allows for three types of spatial/network connectivity between origin and destination regions. We overlay information regarding the network structure and regions, providing an extension of the LeSage and Pace (2005) methodology.

We use truck and train commodity flows (measured in tons per kilometer) in our empirical example, and as a concrete example for discussion purposes. We note that numerous other flows such as telecommunication, airline passengers, train travel and shipping, and automobile and truck traffic are also heavily dependent on the transport network infrastructure used.

LeSage and Pace (2005) make the intuitively plausible argument that: 1) large commodity flows from region A (origin) to region Z (destination) might be accompanied by similarly large flows from neighbors to region A to region Z; 2) large commodity flows from region A to region Z might be accompanied by similarly large flows from region A to neighbors to region Z; and 3) large commodity flows from region A to region Z might be accompanied by large flows from neighbors to region A to neighbors of region Z.

Based on this, they devise formal spatial weight matrices that reflect these three types of spatial connectivity between origin and destination regions. These spatial weights can be used in the family of spatial econometric models popularized by Anselin (1988) to estimate the relative strength of these three types of spatial connectivity relations between origin regions such as A and destination regions Z. They label 1) above as origin-based dependence, 2) as destination-based dependence and 3) as origin-destination dependence.

In the context of our commodity flows, origin dependence of type 1) would be particularly convincing if the transportation network connecting the origin region A to the destination region Z included highway/railway routes from regions neighboring the origin A to the destination region Z. Similar arguments could be made regarding destination dependence of type 2) as well as origin-destination dependence of type 3) above. That is, highway/railway routes would seem an essential aspect of the argument in favor of spatial clustering of flow magnitudes (the dependent variable) that represent the hallmark of the spatial autoregressive/lag econometric models under consideration here.

The focus of this study is on a formal method for adjusting the spatial weights introduced by LeSage and Pace (2005) to reflect a general dependence structure between origin and destination regions that incorporates the nature of the transport network infrastructure. We are also interested in whether this type of adjustment will improve the estimates, inferences and predictions of the model.

2 The spatial econometric flow model

Models for origin-destination flows start by vectorizing the n by n square matrix of interregional flows from each of the n origin regions to each of the n destination regions,. This produces an n^2 by 1 vector of flows by stacking the columns of the flow matrix into a variable vector that we designate as y. For our model, the $n = 35$ columns reflect origin regions whereas the $n = 35$ rows represent destination regions. The objective of flow models is to explain variation in the magnitude of flows between each origin-destination pair. Since our focus is on interregional flows where spatial dependence is important, we set the diagonal elements of the flow matrix containing intraregional flows to zero.

Conventional least-squares regression gravity models use explanatory variables matrices containing characteristics of both the origin and destination regions in an attempt to explain variation in the vector y containing interregional flows. In addition, an intercept term and n^2 by 1 vector of distances between all origins and destinations are typically used as additional variables. This produces the model in (1). In (1), 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 matrix X_d is constructed using characteristics of the destination node for each of the origin-destination (O-D) pair observations, and the matrix X_o is similarly constructed from the origin node in the O-D pairs representing the sample of observations. The vector D denotes the n^2 by 1 origin-destination distances and γ a scalar parameter. Typically, these regression models assume $\varepsilon \sim N(0, \sigma^2 I_{n^2})$.

$$y = \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + \varepsilon \tag{1}$$

The term 'spatial interaction models' has been used in the literature to label models of the type in (1), Sen and Smith (1995). With a few exceptions, use of spatial lags typically found in spatial econometric methods have not been used in these 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 Porojan (2001) for the case of international trade flows, Lee and Pace (2004) for retail sales.

There has been widespread recognition of the need for such models in disciplines such as population migration, Cushing and Poot (2003, p. 317). There is considerably less recognition of issues related to spatial dependence in the transportation flow modeling literature. LeSage and Pace (2005) provide a parsimonious way to structure the connectivity of origin-destination regions in a fashion consistent with conventional spatial autoregressive models where each observation represents a region rather than an origin-destination pair. This seems to have been the stumbling block to extending conventional spatial econometric methods to origin-destination flow situations.

The family of models introduced by LeSage and Pace (2005) rely on a spatial autoregression filtering shown in (2).

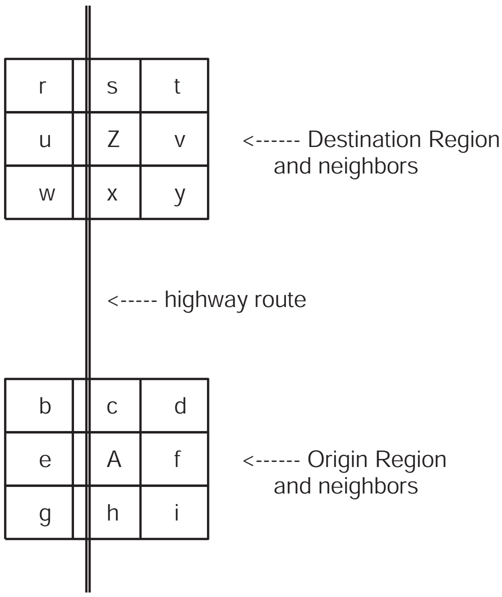
$$y = \rho_1 W_o y + \rho_2 W_d y + \rho_3 W_w y + \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + \varepsilon \tag{2}$$

In this model, $W_o = I_n \otimes W$, where W represents an n by n spatial weight matrix based on first-order contiguity, or some number say m of nearest neighbors. The matrix W_o captures origin-based spatial dependence of the type labelled 1) above. Similarly, $W_d = W \otimes I_n$ is used to capture type 2) dependence, or destination-based dependence relations. Finally, $W_w = W \otimes W$ reflects type 3) dependence that we referred to as origin-destination based dependence.

LeSage and Pace (2005) point out that the model in (2) can give rise to a family of other models by placing various restrictions on the parameters ρ_1, ρ_2 and ρ_3. For example, a restriction that: $\rho_1 = \rho_2 = \rho_3 = 0$ would produce the regression model from (1). Other restrictions would result in models that allow for only origin-based dependence ($\rho_2 = \rho_3 = 0$), only destination-based dependence ($\rho_1 = \rho_3 = 0$), and so on. Of course, estimates of the parameters ρ_1, ρ_2 and ρ_3 would provide an inference regarding the relative importance of the three different types of spatial dependence between the origin and destination regions.

Our example of flows from origin A to destination Z is depicted in Figure 1, where Queen-type contiguity has been used to define neighbors to the origin region A and destination region Z. These neighbors to origin region A are labelled b, c, d, e, f, g, h, i and neighbors to destination region Z are r, s, t, u, v, w, x, y. The spatial lag vector $W_o y$ would be constructed by averaging flows from neighbors to the origin region A, those labelled b, c, d, e, f, g, h, i in the figure. The parameter ρ_1 associated with this spatial lag would capture the magnitude of impact from this type of neighboring observation on the dependent variable vector y (averaged over all sample observations as is typical of regression models). Similarly, the spatial lag vector $W_d y$ would be constructed by averaging flows from neighbors to the destination region Z, those labelled r, s, t, u, v, w, x, y in the figure. The parameter ρ_2 for this spatial lag would measure the impact and significance on flows from all origins to all destinations from this type of neighboring observation. Finally, the third spatial lag in the model $W_w y$ is constructed using an average over all neighbors to both the origin and destination regions A and Z, that is: $b, c, d, e, f, g, h, i, r, s, t, u, v, w, x, y$. Here, the associated parameter ρ_3 represents the overall impact of this particular type of interaction effect.

Figure 1: Origin-Destination region contiguity relationships



Our approach is to consider regions through which the transportation routes pass and to use this information in modifying the spatial weight structure contained in the matrices W_o, W_d and W_w. As an example, consider a highway extending from region A to Z that passes through regions h, A, c on the way to and from the origin region A, and through regions x, Z, s as it passes through the destination region Z. If accessibility to this highway from other regions such as b, d, e, f, g, i or r, t, u, v, w, y is difficult or impossible, we should modify the matrices W_o, W_d and W_w to reflect this prior information.

For this example, the modification would construct $W_o y$ based on an average of regions h and c on the highway route neighboring the origin region A, and $W_d y$ would be an average of regions x and s also on the highway route neighboring the destination region Z, with $W_w y$ reflecting the interaction term consisting of an average over regions h, c, x, s. For this modification of the model of LeSage and Pace (2005), we might expect a large and significant magnitude of impact to arise from the spatial lag associated with the interaction term, $W_w y$. This is because the highway route passing through these regions would have the effect of raising the level of commodity flows to a more uniform level than in regions where the highway does not pass.

It is of interest to note that LeSage and Pace (2005) found the parameter estimate for ρ_3 to be insignificantly different from zero in their application involving state-level migration flows. That is, after taking into account the separate effects of neighbors to the origin and neighbors to the destination captured by the spatial lags $W_o y$ and $W_d y$, the interaction of neighbors to the origin and neighbors to the destination had no impact on variation in the state-to-state migration flows.

At this point, we are abstracting from issues related to the number of entry and exit points on the highway in each region, and we are assuming that access to the highway is limited to those regions through which it passes. We have assumed for simplicity that the matrix W is a binary Queen-type contiguity matrix that is row-normalized, where contiguous neighboring regions have a value of 1 and others 0 before normalization. The modified matrices we suggest represent a subset of the Queen-type contiguous regions, only those through which the highway passes. However, one could rely on more sophisticated approaches to forming an initial row-normalized matrix W that would take into account the number of entry and exit points on the highway in each region, the relative accessibility to the highway from each region that neighbors the origin and destination regions, etc. All of the modelling and estimation methods we set forth and illustrate here would work for these more informative weight structures, provided they were row-normalized. We provide specific illustrations and further discussion of extensions along these lines in Sections 3.3 and 3.4.

One issue that could be of great importance is that of accessibility. This could be quite different for rail versus road networks. For the case of commodity flows under examination here, an important factor would be the relative amounts of rail versus road transportation of commodities. In many parts of the United States where an extensive road network exists and commodities are primarily transported by road with few natural barriers such as mountains, rivers, or lakes, the unmodified approach to forming the spatial weight structure set forth in LeSage and Pace (2005) should work well. Our empirical illustration involves rail and truck commodity flows between 35 regions in Austria where mountains and other natural barriers as well as more limited road networks place limitations on access.

3 An empirical illustration

To illustrate the ideas discussed in Section 2 we produced estimates for the model in (2) using commodity flows transported by both road and rail between 35 regions in Austria during the years 1999, 2000 and 2001. The regions were based on the NUTS3 regions.[1] The flows that were used represent tons per kilometer, with the source of the data being Statistik Austria (with the permission of the Ministry of Transportation). Flows within regions were set to values of zero to emphasize interregional flows that exhibit spatial dependence of the type we are attempting to model. As is conventional, the interregional flow magnitudes were transformed using logs.

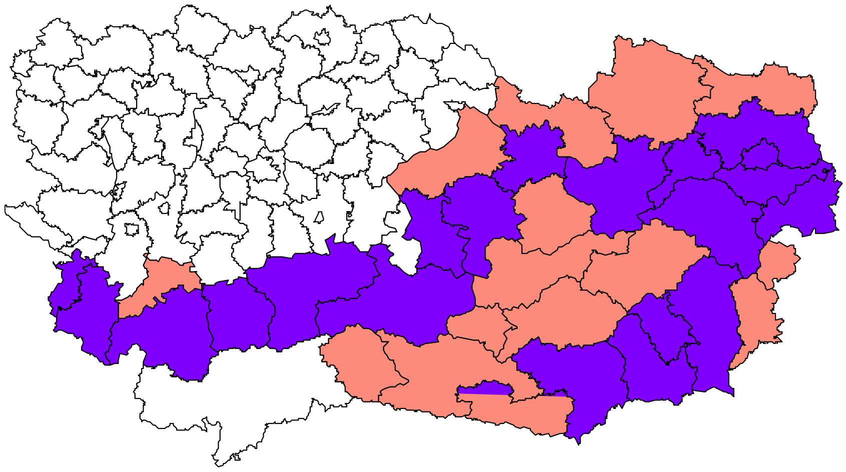
A map of the 35 regions is shown in Figure 2, where regions containing the main road/rail routes are blue and those not on these routes red. As already noted, this example illustrates a case where a clear differentiation can be made between regions that are located along the main transport routes and those that are not. This should provide a good test of whether explicitly incorporating such prior information into the spatial connectivity structure of the model results in substantial differences in the estimates and inferences.

The map in Figure 3 shows the total flows to all regions as destinations. (Darker blue colors reflect lower levels of flows while lighter blue and orange colors indicate higher flow levels.) Examining this map in conjunction with that of the road/rail network in Figure 2, it is clear that the level of flows to destination regions that are on the road/rail network is higher than for

[1] NUTS is the French acronym for Nomenclature of Territorial Units for Statistics used by Eurostat. In this nomenclature NUTS1 refers to European Community Regions and NUTS2 to Basic Administrative Units, with NUTS3 reflecting smaller spatial units most similar to counties in the US.

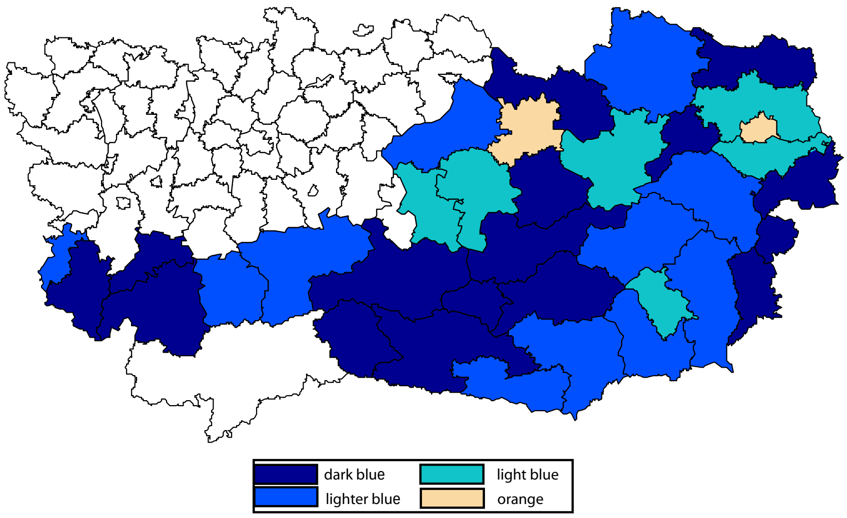
regions not on the network.

Figure 2: Austrian regions on the main road/rail network



The algorithms used to produce the estimates were those described in LeSage and Pace (2005), which involve maximizing the log-likelihood function concentrated with respect to the parameters β and σ in the model. This results in a three-parameter optimization problem involving the parameters ρ_1, ρ_2, ρ_3. Having found optimal values for the $\rho_i, i = 1, \ldots, 3$ parameters, estimates for β can be recovered using: $\hat{\beta} = (X'X)^{-1}X'(I_{n^2} - \rho_1 W_o - \rho_2 W_d - \rho_3 W_w)y$. Similarly, the estimate for $\hat{\sigma}^2$ is constructed using $(e'e)/(n^2 - k)$, where the vector e denotes the residuals from the model in (2). Estimates of

Figure 3: Total commodity flows to all destinations



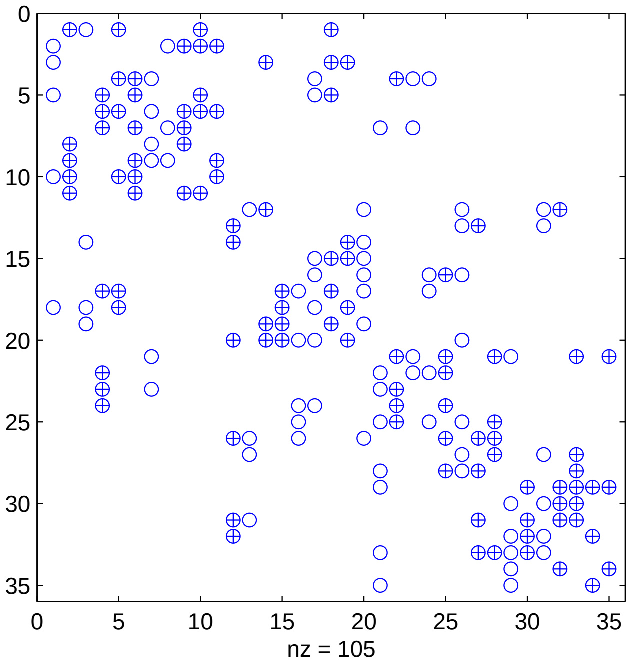
the variance-covariance and measures of dispersion for the parameters used to construct asymptotic $t-$statistics and associated marginal probabilities were based on a numerically constructed Hessian.

Two variants of the model were estimated, one based on the spatial weight structure proposed by LeSage and Pace (2005) and another reflecting the modification to reflect the road/rail transport routes discussed in Section 2. The first approach relied on a matrix W based on the first-order contiguous neighboring regions as the basis for constructing the weight matrices W_o, W_d, W_w used by the model. The mean number of first-order contiguous neighbors was 5.2, with a standard deviation of 1.23, so similar results would have been obtained by using the 5 nearest neighbors to each region. The modified weight matrix proposed here selected a subset of the contiguous neighboring regions that were located on the road/rail network.

Figure 4 presents the non-zero elements from the two 35 by 35 spatial weight structures. The presentation in the figure is in terms of the $n = 35$ by $n = 35$ square matrix reflecting connectivity relations between the regions, with the contiguous neighbors labelled with the symbol 'o' and the subset of contiguous neighbors located on the road/rail network indicated by a 'plus' sign (+). For example if region 6 is a contiguous neighbor to region 1, then a symbol 'o' would appear in row 1, column 6. Similarly, if the region 6 neighbor also represents a region through which the road/rail routes pass, there would be a + symbol as well. In comparison to the average of 5.2 contiguous neighbors, the average number of neighbors with road/rail routes was 3, and the standard deviation was 1.11.

We note that use of some number say m of nearest neighbors in place of first-order contiguity can allow for more regions along a road/rail route to enter the subset of regions used to produce the averages that become the spatial lag variables. As an example, consider the simple case of regions organized along a line which also contains the road/rail route. Use of first-order contiguity would allow one neighbor to the left and another to the right to enter into creation of the spatial lag variable. In contrast, use of the six nearest neighbors relation would allow for the 3 nearest neighbors to the left and 3 nearest neighbors to the right to enter when creating the spatial lag variable. One way to view this is that one can construct spatially lagged variables that trace out longer segments along the transportation routes by increasing the number of neighbors used to produce the initial matrix W in

Figure 4: Comparison of two weight matrix structures



the model. In conjunction with the restriction that only neighbors on the road/rail routes will be included in formation of the spatial lags, this will result in a direct relationship between increased numbers of nearest neighbors and the length of the road/rail segments that enter into formation of the spatial lag variables. We will illustrate this aspect of model specification using our sample data for the 35 Austrian regions in the next section.

As explanatory variables used to form the matrices X_o and X_d we used: population density of the region; the log of area in each region; and the change in employment, population and GDP per capita over the previous year. Note that we produced parameter estimates for three years 1999, 2000 and 2001, so the 1999 estimation of the model parameters relied on the change in employment, population and per capita GDP for the years 1998 to 1999, the year 2000 estimates were based on changes from 1999 to 2000 and so on. A vector of (logged) distances between the centroids of each regions was also included as an explanatory variable along with an intercept vector. We would expect that changes in employment, population, and per capita GDP would exhibit positive signs, leading to higher levels of commodity flows at both the origin and destination regions. The coefficient estimate on distance should be negative indicating a decay of flows with distance, whereas the impact of population density when controlling for growth in employment, population and per capita GDP is less clear.

3.1 Estimation results

One focus of estimation is comparison of the model based on spatial weights constructed from simple contiguity relationships versus the model based on road/rail network considerations. Table 1 presents the log-likelihood function values for these two types of models from the 1999, 2000 and 2001 data samples along with the sum of squared errors (divided by n^2, the number of observations). From the table we see that the modified model that takes into account the road/rail network produces higher log-likelihoods and smaller errors for all three years.

A second question that arises regards the nature of the estimates and inferences from the two types of models. Table 2 presents the parameter estimates for the contiguity-based spatial weight model and Table 3 shows estimates from the modified model.

Comparing the estimates for the spatial dependence parameters ρ_1, ρ_2, ρ_3

Table 1: Comparison of the contiguity-based and road/rail modified spatial models 1999, 2000, 2001

Model/Year	Log Likelihood	$(e'e)/n^2$
Contiguity 1999	-2,354.3	5.3655
Road/Rail 1999	-2,319.6	5.0412
Contiguity 2000	-2,312.9	5.0271
Road/Rail 2000	-2,297.7	4.8885
Contiguity 2001	-2,255.0	4.6358
Road/Rail 2001	-2,212.8	4.2697

Table 2: Estimates from the contiguity-based spatial Model 1999, 2000, 2001

	1999		2000		2001	
Variable	$\hat{\beta}_o, \hat{\beta}_d$	t-statistic (p-level)	$\hat{\beta}_o, \hat{\beta}_d$	t-statistic (p-level)	$\hat{\beta}_o, \hat{\beta}_d$	t-statistic (p-level)
constant	-3.108	-1.74(0.080)	-0.229	-0.11(0.908)	-3.853	-2.31(0.0209)
popdensity_o	-1.385	-3.51(0.000)	-1.504	-3.38(0.000)	0.529	1.58(0.1125)
area_o	0.462	3.01(0.002)	0.289	1.89(0.058)	0.494	3.45(0.0006)
demp_o	0.122	1.54(0.123)	0.115	2.81(0.005)	0.089	1.10(0.2702)
dpop_o	0.092	0.77(0.440)	0.232	2.14(0.032)	0.280	2.98(0.0029)
dgdp_o	3.822	3.42(0.000)	2.009	3.67(0.000)	2.090	2.67(0.0076)
popdensity_d	-0.800	-2.02(0.042)	-1.016	-2.28(0.022)	0.608	1.81(0.0697)
area_d	0.586	3.74(0.000)	0.565	3.25(0.001)	0.742	4.95(0.0000)
demp_d	0.066	0.82(0.409)	0.120	2.94(0.003)	0.061	0.76(0.4465)
dpop_d	0.113	0.95(0.340)	0.159	1.46(0.143)	0.362	3.85(0.0001)
dgdp_d	3.058	2.75(0.006)	1.708	3.14(0.001)	2.155	2.75(0.0060)
distance	-0.075	-5.66(0.000)	-0.077	-5.87(0.000)	-0.066	-5.40(0.0000)
ρ_1	0.250	7.35(0.000)	0.214	6.28(0.000)	0.105	2.86(0.0043)
ρ_2	0.161	4.41(0.000)	0.155	4.36(0.000)	0.067	1.79(0.0733)
ρ_3	-0.179	-4.03(0.000)	-0.290	-7.03(0.000)	-0.043	-0.96(0.3337)

from the two models we see a distinctly different pattern of values over the three years. In the contiguity model, ρ_1 and ρ_2 are positive in all three years while ρ_3 is negative. The (positive) magnitude of ρ_1 is always larger than the (positive) ρ_2, pointing to more importance assigned to the spatial lag involving neighbors to the origin, relative to neighbors to the destination region. In fact, the parameter ρ_2 is not significantly different from zero at the 0.95 level for the year 2001 sample. The parameter ρ_3 that measures the influence of the interaction term reflecting connectivity between neighbors to the origin and neighbors to the destination is negative in all three years, but not significantly different from zero for the year 2001 sample. A negative sign for this parameter indicates negative spatial dependence between flows from an origin-destination pair and flows from neighbors to the origin and neighbors to the destination regions. LeSage and Pace (2005) provide a motivation for the model in (2) from a spatial filtering perspective as shown in (3).

$$(I_{n^2} - \rho_1 W_o)(I_{n^2} - \rho_2 W_d)y = \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + \varepsilon \qquad (3)$$

This leads to a model that includes the interaction term $W_w = W_o \cdot W_d$ in the sequence of spatial lags with a coefficient equal to $-\rho_1\rho_2$, as shown in (4).

$$y = \rho_1 W_o y + \rho_2 W_d y - \rho_1\rho_2 W_w y + \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + \varepsilon \qquad (4)$$

This might provide a partial motivation for the negative sign on the coefficient ρ_3 from the contiguity-based spatial model. It appears clear that the estimated parameters ρ_3 do not obey the implied restriction that $\rho_3 = -\rho_1\rho_2$. However, another motivation is that when one is attempting to model flows in the presence of a network structure, the relationship between neighbors to origin and neighbors to destination regions is simply not important. For the case of household migration decisions, it might be intuitively plausible that the costs and benefits of moving from region A to Z are similar to the costs and benefits of moving from regions that neighbor A to regions that neighbor Z. The presence of an interaction effect such as this is likely to be enhanced if the variables that come into play in determining household costs and benefits are positively spatially correlated. For example, employment and income opportunities in neighboring counties or states may be similar because of regional economic conditions. In contrast, for the situation where network routes come into play, there is far less motivation for the importance of neighbors to the origin and neighbors to the destination regions if they do not have access to the network.

3.2 The corridor neighborhood model

Turning attention to the corridor (or road/rail modified) model we see a pattern of estimates for ρ_1, ρ_2, ρ_3 where all three parameters are positive. This should not be surprising since the spatial lags for the origin and destination (associated with parameters ρ_1 and ρ_2) average over neighboring regions on the corridor network which should be positively associated with the level of commodity flows. In addition, the spatial lag for the interaction term averages over neighbors to the origin and neighbors to the destination that are also on the corridor, suggesting that flows between an origin and destination region that are both on the network corridor should be greater than flows between regions where only one of the two is located on the corridor. Given this type of interpretation for the three parameters ρ_1, ρ_2, ρ_3 in this model, it should come as no surprise that the magnitude of ρ_3 is the largest, reflecting the positive impact on levels of commodity flows that arise from both origin and destination regions being located on the network corridor. There is also a consistent pattern of larger values for the parameter ρ_1 than ρ_2 in all three years, suggesting that neighbors to the origin region on the corridor represent the second most important determinant of high levels of commodity flows between O-D pairs.

The estimates and inferences for the explanatory variables in the two models suggest that distance is negative and significantly related to the level of flows for all three years in both models, as we would expect. The area of the origin and destination regions is positively related to the level of flows in all three years for both models, but differing somewhat in terms of the level of significance. With one exception, changes in employment, population and GDP per capita over the previous year are positively related to the level of flows at both the origin and destination regions for both models and all three years. The exception being population change in 1999 for the corridor model which is negative, but not significantly different from zero. Although the signs of these coefficients are positive, the levels of significance vary across the two models and the time periods. Finally, population density is negative and significant at the 0.95 level or above for both origin and destination regions in both models for the years 1999 and 2000. For the year 2001 sample, we find positive but weakly significant estimates for origin and destination regions in both models.

Table 3: Estimates from the road/rail corridor spatial Model 1999, 2000, 2001

Variable	1999 $\hat{\beta}_o, \hat{\beta}_d$	1999 t-statistic (p-level)	2000 $\hat{\beta}_o, \hat{\beta}_d$	2000 t-statistic (p-level)	2001 $\hat{\beta}_o, \hat{\beta}_d$	2001 t-statistic (p-level)
constant	-6.684	-3.24(0.001)	-4.050	-2.42(0.015)	-6.725	-3.54(0.000)
popdensity_o	-1.893	-4.94(0.000)	-1.283	-2.94(0.003)	0.450	1.43(0.151)
area_o	0.393	2.34(0.019)	0.247	1.68(0.093)	0.383	2.39(0.016)
demp_o	0.137	1.80(0.071)	0.135	3.39(0.000)	0.101	1.37(0.168)
dpop_o	-0.037	-0.31(0.749)	0.120	1.12(0.260)	0.198	2.19(0.028)
dgdp_o	5.295	4.85(0.000)	1.950	3.63(0.000)	2.224	3.00(0.002)
popdensity_d	-1.636	-4.27(0.000)	-0.882	-2.04(0.041)	0.645	2.09(0.036)
area_d	0.553	3.51(0.000)	0.495	3.43(0.000)	0.675	4.79(0.000)
demp_d	0.131	1.72(0.085)	0.141	3.54(0.000)	0.112	1.54(0.122)
dpop_d	0.002	0.01(0.984)	0.134	1.26(0.207)	0.249	2.75(0.006)
dgdp_d	4.707	4.35(0.000)	1.545	2.89(0.003)	1.977	2.68(0.007)
distance	-0.055	-4.38(0.000)	-0.043	-3.58(0.000)	-0.049	-4.21(0.000)
ρ_1	0.183	5.79(0.000)	0.186	5.66(0.000)	0.101	2.96(0.003)
ρ_2	0.062	1.80(0.070)	0.091	2.62(0.008)	0.055	1.60(0.108)
ρ_3	0.437	8.88(0.000)	0.343	6.77(0.000)	0.462	9.09(0.000)

3.3 Alternative specifications based on varying numbers of nearest neighbors

We consider the impact of using a nearest neighbors scheme to define the initial spatial weight matrix W used to form the spatial lags W_oy, W_dy and W_wy in the model. Varying this aspect of model specification might allow practitioners to produce better model fit and more accurate predictions.

Before turning to the empirical results from this experiment we make some observations on the nature of a model based on our road/rail network corridor modification scheme in the context of nearest neighbor weight matrices. As already indicated, an increase in the number of nearest neighbors used to form the initial weight matrix W will result in spatial lags that place relatively more emphasis on regions located along the road/rail routes. We can interpret the extent to which increasing the number of nearest neighbors extends the spatial lags along the road/rail corridor by calculating the number of first-order contiguous neighbors, number of second order contiguous neighbors (these are neighbors to the first-order contiguous neighbors), and so on for higher order contiguity relationships.

As an example of the interpretative value, consider our case where there are around 5 first-order contiguous neighbors on average across all 35 regions in the sample. Use of 5 nearest neighbors should result in spatial lags W_oy, W_dy for the origin and destination regions that extend one neighbor in both the entry and exit directions of the road/rail corridor in this case. The spatial lag based on W_wy should represent an average over these four regions. If there were, say 15, second-order contiguous neighbors (again, on average across all 35 regions in the sample), then a weight matrix based on 15 nearest neighbors should on average result in spatial lags W_oy, W_dy constructed from two neighboring regions to the origin and destination regions. That is, our spatial lags now extend out to the two neighboring regions that lie in the direction of the entry and two regions that lie in the direction of exit along the road/rail corridor through the regions in the sample. Similarly, the spatial lag based on the interaction term $W_w y$ will reflect an average over these 8 regions.

A point to note is that for reasonably small samples as we increase the number of nearest neighbors, the spatial lag based on the interaction term may become a source of redundant information. As the spatial lags based on $W_o y$ and $W_d y$ are extended to include all regions on the road/rail corridor, there is less need to incorporate an average of these two sets of regions. To see this, consider that as we extend out along the transport corridor there will come a point at which the spatial lag $W_o y$ and the spatial lag $W_d y$ are constructed based on many of the same regions. As these two variable vectors begin to look more similar due to the overlap of regions used to construct the spatial lags at the origin and destinations, there will be less of a role for the spatial lag based on the interaction term $W_w y$, which is constructed using observations from neighbors based on both origins and destinations. In fact, the spatial lags $W_o y$ and $W_d y$ will come to look more and more like the spatial lag $W_w y$ based on the interaction term.

These ideas are important for interpreting estimates and inferences from model specifications based on a weight matrix W constructed using an increasing number of nearest neighbors. One implication is that we should change our interpretation of the parameters ρ_1, ρ_2, ρ_3 as we increase the number of neighbors used in the model specification. At some point, the redundancy of information in the spatial lag vectors will produce a classic collinear relationship between these three variable vectors. As in the collinear variables situation, we might expect to see all of the importance placed on a single spatial lag variable (a large and significant coefficient) with the other two variables becoming small and insignificantly different from zero. As a limit to the process of increasing the number of nearest neighbors used to produce W, we will have a single spatial weight matrix that produces a spatial lag vector that reflects an average of flows from all regions on the road/rail corridor. In this situation it should be clear that there is only a role for a single spatial lag vector.

Another implication of these ideas is that simple optimization of the likelihood function over models specified based on varying numbers of nearest neighbors may not produce a solution to the model comparison problem that exhibits desirable statistical operating characteristics. It may be the case that models based on more neighbors result in a single weight matrix that represents a more parsimonious model structure capable of producing a better fit. This remains an area for future research, with Bayesian model comparison methods representing an approach that may hold an advantage in this type of situation. Bayesian model comparison requires calculation of the log-marginal likelihood for the model. For the case of reasonable model comparison priors on the parameters of this model, it is possible to analytically integrate the parameters β and σ out of the log-marginal likelihood function, leaving an integration problem involving only the parameters ρ_1, ρ_2, ρ_3, a 3-dimensional numerical integration problem. Simple grid-based numerical integration procedures of the type used by LeSage and Parent (2005) are not computationally efficient because of the relatively high cost of calculating the determinant of an n by n matrix that appears in the log-marginal likelihood after analytical integration of the parameters β and σ. This determinant of the potentially large n by n matrix would need to be calculated repeatedly for a large number of values for the parameters ρ_1, ρ_2, ρ_3 making up the region of support.

We stress that these ideas are probably not important for large data samples with a relatively sparse set of regions through which the transport routes pass. In these situations, overlap in the regions used to produce the spatially lagged variable vectors is less likely to occur. This means that interpretation of the role played by the spatial lags and their associated parameters ρ_1, ρ_2, ρ_3 is relatively constant as we vary the number of nearest neighbors used to construct the initial weight matrix W. We conjecture here that the estimated values for the parameters ρ_1, ρ_2 and ρ_3 in these situations would mirror those presented in the previous section. That is, the spatial lag vector $W_w y$ constructed based on flows from neighbors to both the origin and destination regions would exert the greatest impact on the level of flows between origin-destination pairs. This seems intuitively plausible since it reflects the fact that O-D flows between two regions both located on the road/rail network should be greater than those associated with other types of O-D region pairs. Further, an average of the magnitudes of flows from neighboring regions at both the origin and destination that lie on the road/rail network in this case would best be capable of explaining the high level of flows between these types of O-D region pairs.

An empirical investigation of these issues was carried out for our sample of 35 regions using an origin matrix W constructed using nearest neighbors that varied from 5 to 30. As already noted, the average number of first-order contiguous neighbors for our sample is 5.2, the average number of second-order neighbors 13.5, with a standard deviation of 3.45, and the average number of third-order contiguous neighbors is 24.1 with a standard deviation of 3.86. This suggests that use of 30 nearest neighbors would allow the spatial lags to extend outward beyond the three nearest regions on the road/rail routes. We note that with a sample of 35 regions, use of the 30 nearest neighbors results in spatial lags constructed on the basis of nearly the entire sample of 35 observations. Adding our modifying restriction that only neighbors lying on the road/rail corridor are included, this should result in a spatial lag that is constructed from almost all regions on the road/rail corridor.

Table 4 presents results based on the 1999 sample information in the form of a log-likelihood function value, the standardized sum of squared errors and the three estimates for the parameters ρ_1, ρ_2, ρ_3 for models based on the varying number of nearest neighbors. Table 5 presents results for the year 2000 sample in an identical format. Results for the year 2001 sample were similar to these two sets of results and were omitted to save space.

We first note that when we use 5 nearest neighbors, the log-likelihood function values, squared errors and estimates for the parameters ρ_1, ρ_2, ρ_3 are similar to those reported in the previous section where a first-order contiguity matrix was used. This seems intuitively correct since the average number of first-order contiguous neighbors in our sample was 5.2, close the the 5 nearest neighbors.

The results (from both years) suggest a monotonically increasing relationship between the log-likelihood function and the number of neighbors up to the very large number of 28 neighbors for both years 1999 and 2000. This suggests that use of more than 3 neighboring regions on the entry and exit of the road/rail corridor maximizes the log-likelihood function. However, there are perhaps reasons to be cautious about these conclusions regarding a well-defined maximum in the likelihood function. For example, in the year 2001 sample, no maximum was found, with 30 neighbors exhibiting the

Table 4: Estimates from the road/rail modified spatial Model for 1999 based on varying numbers of nearest neighbors

# neighbors	Log Likelihood	$(e'e)/n$	ρ_1	ρ_2	ρ_3	$\sum_{i=1}^{3} \rho_i$
5	-2309.5151	4.9561	0.1830	0.0927	0.4999	0.7756
6	-2304.6175	4.9187	0.2411	0.1123	0.5351	0.8885
7	-2310.6778	4.9624	0.2630	0.1252	0.4898	0.8780
8	-2308.9902	4.9494	0.2955	0.1329	0.4858	0.9142
9	-2307.4583	4.9387	0.3294	0.1600	0.4420	0.9314
10	-2300.1608	4.8813	0.3394	0.1820	0.4404	0.9618
11	-2292.1025	4.8243	0.3583	0.1806	0.4362	0.9751
12	-2291.1543	4.8167	0.3772	0.2107	0.3898	0.9777
13	-2288.6949	4.7979	0.4166	0.2612	0.3052	0.9830
14	-2290.9517	4.8034	0.4488	0.3032	0.2200	0.9720
15	-2288.3789	4.7773	0.4669	0.3254	0.1754	0.9677
16	-2287.8616	4.7656	0.4927	0.3399	0.1292	0.9618
17	-2281.0262	4.7100	0.5100	0.3591	0.1004	0.9695
18	-2273.1579	4.6543	0.5292	0.3682	0.0804	0.9778
19	-2269.7722	4.6263	0.5384	0.3824	0.0567	0.9775
20	-2266.3698	4.6001	0.5687	0.4121	-0.0041	0.9767
21	-2265.2754	4.5894	0.5857	0.4353	-0.0482	0.9728
22	-2264.0015	4.5824	0.5847	0.4497	-0.0614	0.9730
23	-2263.6736	4.5833	0.5915	0.4509	-0.0682	0.9742
24	-2261.8623	4.5683	0.6003	0.4630	-0.0922	0.9711
25	-2261.5919	4.5618	0.6175	0.4786	-0.1373	0.9588
26	-2257.5913	4.5301	0.6187	0.4925	-0.1493	0.9619
27	-2257.3327	4.5289	0.6213	0.4979	-0.1584	0.9608
28	-2255.0180	4.5072	0.6292	0.5088	-0.1809	0.9571
29	-2255.8023	4.5070	0.6400	0.5245	-0.2186	0.9459
30	-2255.9766	4.5011	0.6602	0.5387	-0.2731	0.9258

Table 5: Estimates from the road/rail modified spatial Model for 2000 based on varying numbers of nearest neighbors

# neighbors	Log Likelihood	$(e'e)/n$	ρ_1	ρ_2	ρ_3	$\sum_{i=1}^{3} \rho_i$
5	-2287.2849	4.7998	0.1966	0.1045	0.4006	0.7017
6	-2285.0632	4.7804	0.2632	0.1457	0.3704	0.7793
7	-2292.5783	4.8390	0.2813	0.1618	0.2989	0.7420
8	-2286.3070	4.7771	0.3223	0.1777	0.3179	0.8179
9	-2282.1868	4.7405	0.3522	0.2018	0.3040	0.8580
10	-2274.4123	4.6748	0.3616	0.2243	0.3338	0.9197
11	-2265.4944	4.6082	0.3782	0.2361	0.3402	0.9545
12	-2262.0356	4.5777	0.4045	0.2658	0.2893	0.9596
13	-2257.9857	4.5486	0.4382	0.3117	0.2193	0.9692
14	-2256.5404	4.5297	0.4625	0.3435	0.1511	0.9571
15	-2251.4005	4.4861	0.4814	0.3596	0.1159	0.9569
16	-2250.5259	4.4728	0.5017	0.3794	0.0680	0.9491
17	-2240.6377	4.3984	0.5168	0.3976	0.0511	0.9655
18	-2233.2543	4.3455	0.5428	0.4196	0.0090	0.9714
19	-2229.5526	4.3167	0.5532	0.4322	-0.0138	0.9716
20	-2226.6024	4.2981	0.5789	0.4569	-0.0631	0.9727
21	-2224.4080	4.2798	0.5949	0.4821	-0.1069	0.9701
22	-2220.3753	4.2498	0.6035	0.5001	-0.1338	0.9698
23	-2219.8217	4.2469	0.6166	0.5049	-0.1526	0.9689
24	-2216.8742	4.2259	0.6235	0.5144	-0.1695	0.9684
25	-2217.1603	4.2240	0.6389	0.5293	-0.2138	0.9544
26	-2215.3852	4.2098	0.6345	0.5358	-0.2107	0.9596
27	-2214.8906	4.2151	0.6333	0.5427	-0.2213	0.9547
28	-2213.5894	4.1995	0.6405	0.5485	-0.2375	0.9515
29	-2214.0158	4.1971	0.6515	0.5602	-0.2718	0.9399
30	-2215.7165	4.2039	0.6679	0.5708	-0.3179	0.9208

largest log-likelihood function value.

Another cause for concern is the monotonically increasing relationship between the number of neighbors and estimates for the parameters ρ_1, ρ_2, ρ_3. We see that an increase in the number of neighbors places more weight on the parameters ρ_1 and ρ_2 and less on ρ_3. We note that after 18 neighbors for the 1999 year sample the estimate for ρ_3 became insignificantly different from zero, and this occurred for the year 2000 sample at 14 neighbors. It should also be noted that a stability restriction requires that the sum: $\rho_1 + \rho_2 + \rho_3 < 1$. At 21 neighbors for the 1999 sample (and 20 neighbors in the 2000 sample) a negative magnitude for ρ_3 is required to meet this restriction. For numbers of neighbors larger than 21 (20), the values of the parameters ρ_1 and ρ_2 continue to increase with their sum exceeding unity, while the parameter ρ_3 becomes increasingly negative as is required by the stability restriction.

We interpret these results as possible pathological behavior arising from the penalty function imposed during optimization to enforce the stability restriction. It may also arise from an identification problem that occurs when the spatial lags $W_o y, W_d y$ and $W_w y$ contain a large number of overlapping regions, and therefore reflect high correlation. This type of situation may interfere with the ability to identify or properly decompose variation in the O-D flow vector y that should be attributed to each of the three types of spatial connectivity relationships. It is of interest in this connection to consider the sum of the three parameters $\rho_1 + \rho_2 + \rho_3$ which are shown in the last columns of Tables 4 and 5.

We note that despite the possible pathological variation of the parameters $\rho_i, i = 1, 2, 3$ with respect to changing the number of neighbors, the parameters β were relatively consistent with those reported in the previous section. For example, in the year 2001 sample: distance was negative and significant for all neighbors between 5 and 30; population density, area and change in population, employment and per capita GDP were all positive for all neighbors between 5 and 30, identical in sign to the results reported for the year 2001 sample in the previous section.

One way to further investigate this potential problem would be to impose a restriction that $\rho_3 = 0$, or that $\rho_3 = -\rho_1\rho_2$. This might be useful for problems involving use of a large number of neighbors in a small sample environment. In conjunction with the restriction that $\rho_1 + \rho_2$ must be less than one for stability, this might provide enough prior information to overcome any weak data problems. Other solutions that generally work well in the face of weak sample data problems are Bayesian priors placed on the parameters involved.

We tested imposition of a zero restriction on the parameter ρ_3 and found that a unique maximum likelihood function value existed for all three sample years. For the year 2001 sample where no maximum of the likelihood was found for neighbors ranging up to 30 in the unrestricted model involving all three parameters, the restricted model produced parameter estimates for ρ_1, ρ_2 that increased monotonically until around 20 neighbors, where they took on values around: $\rho_1 = 0.50$ and $\rho_2 = 0.40$. As the neighbors increased from 20 to 30, the value of ρ_2 remained around 0.40, while ρ_1 increased to 0.5677, with the maximum likelihood estimate at 0.5503 and 28 neighbors. A similar result was found for the year 2000 sample, with a maximum likelihood at 29 neighbors, $\rho_1 = 0.5564$ and $\rho_2 = 0.4177$. For the year 1999, the maximum occurred at 28 neighbors with $\rho_1 = 0.5794$ and $\rho_2 = 0.4076$.

We note that these results are roughly consistent with the large number of neighbors identified by the unrestricted model for the years 1999 and 2000. In addition, the sum of the parameters $\rho_1 + \rho_2$ at the maximum of the likelihood function was around 0.95 in all three cases, which is consistent with the values reported for the sum of the three parameters in the unrestricted models reported in Tables 4 and 5 at 28 neighbors. These confirmatory results lend support for our previous conclusion that spatial lags should extend to include slightly more than 3 neighbors on the entry and exit of the road/rail route to each region.

A possible conclusion from these exercises is that one can adequately model spatial dependence in the flows using two approaches. One approach reflects that of the previous section, where a small number of neighbors based on first-order contiguity relations are used to construct the initial weight matrix W. This approach incorporates a more sophisticated spatial filtering model for the spatial autoregressive lags in the model, where the filtering specification uses three spatial lags to reflect the three possible types of dependence motivated by LeSage and Pace (2005). A second approach is based on increasing the number of neighbors used to produce the initial weight matrix W, but uses a simpler structure for the spatial autoregressive lags in the model. Simplification is achieved by eliminating one of the three types of dependence suggested by LeSage and Pace (2005). Both types of models produced similar estimates and inferences regarding the parameters β. Subject to the caveats noted regarding interpretation and identification of the parameters ρ_1, ρ_2 and ρ_3 in these two types of models, similar conclusions about the role of spatial dependence between origin-destination flows based on regions located along the road/rail corridor can also be inferred.

3.4 Other extensions and areas for further research

Extensions to the modification procedure suggested here could include weight matrices based on distances from the centroids of each regions to the road/rail network, where the weight decays with distance. This would allow nearby regions that are not on the road/rail corridor to enter into determination of the spatial lags, with the weight assigned decaying inversely with distance from the network. An additional parameter could be introduced to determine the rate of decay with distance.

Other characteristics of the regions and transport network could be used when modifying the matrix W that forms the basis for the model of LeSage and Pace (2005). For example, the number of entry and exit access points along the road or rail network, or the length of the route contained in each region could be used to produce a more tailored set of weights. Similarly, a combination of regional characteristics could be used to create an accessibility index or variable that might be the basis for assigning spatial weights.

In this vein, anisotropic neighbors could be used to construct the weight matrices, allowing for directionality in the model. For example, separate weight matrices reflecting neighbors to the north, south, east and west could be constructed. In this case, an increase in the number of neighbors could be used to move along the transportation corridor in separate directions, or different spatial dependence parameters could be introduced to capture directional aspects of dependence.

For customized weight structures of the type mentioned above, the problems found here in distinguishing between alternative weight structures based on likelihood function values may be aggravated. LeSage and Pace (2004) provide an illustration of these types of problems in a Bayesian model comparison setting. They find that: 1) the strength of spatial dependence exerts an influence on the quality of model comparison inferences; and 2) the sample size plays an important role. Intuitively, in the face of weak spatial dependence it will be difficult to distinguish between alternative weight structures because the role they play in explaining variation in the dependent variable is small. It is also intuitively plausible that, making fine distinctions between alternative weight structures will require a large sample with many regions that exhibit a potentially rich connectivity structure.

LeSage and Pace (2005) point out that the conventional assumption of a normal distribution for the disturbances in the data generating process (and the implied normal distribution of the origin-destination flow magnitudes) may not be a valid one. They suggest and illustrate implementation of a Bayesian Markov Chain Monte Carlo (MCMC) estimation procedure that allows for a fat-tailed error distribution (Gelfand and Smith, 1990, Geweke, 1993). This robust estimation approach should be useful for the case of commodity flows of the type considered here. Outliers or aberrant observations are downweighted during estimation to preclude these observations from exerting an undue influence on the resulting estimates and inferences.

They also discuss tobit variants of the model that can be estimated with the same MCMC procedures, which would be useful for data samples containing missing values for some of the origin-destination pairs. In fact, the sample data used in this study contained some missing values which were set to zero values.

A final point is that models based on spatial dependence in the error structure or models exhibiting both dependence in the dependent variable and the error structure can be treated in a similar fashion to those illustrated here. For example, LeSage and Pace (2005) point to models of the type shown in (5) and (6).

$$\begin{aligned} y &= \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + u \\ u &= (I_{n^2} - \rho_1 W_o)(I_{n^2} - \rho_2 W_d)u + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{5}$$

$$\begin{aligned} y &= \lambda_1 W_o y + \lambda_2 W_d y + \lambda_3 W_w y + \alpha\iota + X_d\beta_d + X_o\beta_o + D\gamma + u \\ u &= \rho_1 W_o u + \rho_2 W_d u + \rho_3 W_w u + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \tag{6}$$

Without loss of generality, the same modification scheme suggested here could be used to form the matrices W_o, W_d and W_w in these models.

4 Conclusions

Drawing upon work by LeSage and Pace (2005) for spatial autoregressive modeling of interregional flows, we propose an extension that seems suitable for a number of applications where a transport network exists between the regions. This would be the case for commodities flowing over rail and road networks, commuters travelling to work along major roads and highways, as well as international trade flows that must pass through specific ports of entry and exit. We provide a simple method for incorporating prior information regarding the path of the network into the spatial connectivity structure proposed by LeSage and Pace (2005) for modeling origin-destination flows.

Our modification involves forming spatial lags for the spatial autoregressive structure used in the model based only on neighboring regions that are located on the network. In addition to the intuitive appeal of this type of modification, we show that an improvement in the likelihood function value and fit of the model arises from the modification.

More sophisticated extensions of our approach to modification were discussed and illustrated, as well as unresolved issues that should be considered in future research.

References

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.

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.

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

LeSage, J.P. and R. Kelley Pace (2005) "Spatial Econometric Modeling of Origin-Destination flows", working paper available at: www.spatial-econometrics.com

LeSage, J.P. and Olivier Parent (2004) "Bayesian model averaging for spatial econometric models," working paper available at: www.spatial-econometrics.com

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.

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.

Authors: James P. LeSage, Wolfgang Polasek

Title: Incorporating Transportation Network Structure in Spatial Econometric Models of Commodity Flows

Reihe Ökonomie / Economics Series 188

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996

Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • http://www.ihs.ac.at

ISSN: 1605-7996