DIRECTIONAL EXPANSION IN MIGRATION MODELLING

Timothy C. Pitts
Department of Geography
University at Buffalo
Amherst, NY 14261

ABSTRACT Migration is typically considered to be directional, yet our models rarely take into account that directionality. Certainly migrants have preference for places in certain directions and show disfavor for those in other directions. The purpose of this paper is to extend the present knowledge of the effect of expansion on migration models. This paper uses an expanded spatial interaction model to elicit the directionality of migration. The results of this model lend insight into the stability of parameter estimates and the regional affiliations of states.

Migration research has recognized the directionality in migration for many years. Such concepts as beaten paths, regional shifts, and migration streams all include an implicit directionality. Migration models, on the other hand, have assumed that direction does not matter. The typical gravity (or spatial interaction) model uses distance and population (or attractiveness) as the prime determinates of the flow between two places while the friction of distance and the attractiveness of a place can vary quite significantly over direction.

Clark (1986), for example, discussed the regional flows of net migration. He considered the regional shift from the Northeast and Midwest to the South and West. However, when he modelled these shifts, the flows return to non-directional movements. The parameter estimates of a model that ignores direction will be unstable if the direction of migration is an important variable.

Casetti and Can (1986) have proposed an expansion that will take into account the direction of the migration. It uses a trigonometric expansion of the population and distance parameters. The purpose of this paper is to extend that analysis and lend insight into the stability of parameter estimates in the model.

SPATIAL INTERACTION MODELS

One of the commonly used models in migration research is the gravity or spatial interaction model. It posits that migrants will be attracted by larger places (where population is usually used as a proxy for other attractions) and discouraged by longer distances. The traditional origin-specific gravity model is:

$$I_i = \alpha \, POP_i^{\beta} \, DIST_i^{\Gamma} \tag{1}$$

where, I_i = the flow from the origin to destination i, POP_i = the population of destination i, and $DIST_i$ = the distance from the origin to destination i. One should expect the estimate of the parameter β to be positive (interaction should increase as attractiveness increases) and the estimate of the parameter Γ to be negative (interaction should decrease as distance increases). To estimate the parameters in Equation 1 using ordinary least squares estimation techniques, the equation must be log-log transformed to create a linear relationship:

$$\ln I_i = \ln \alpha + \beta \ln POP_i + \Gamma \ln DIST_i. \tag{2}$$

In estimation, a will be used to estimate α, b to estimate β, and c to estimate Γ.

TRIGONOMETRIC EXPANSION

Casetti (1972) suggested that parameters could be expanded to allow them to vary over some other variable. Foster (1991) supports this by saying that geographers are interested in how phenomena vary over space and time, but rarely consider that the very parameters they view as constant can also vary over the same scales. That is to say, as is commonly known, that parameter estimates calculated today might be significantly different from those calculated twenty years ago. Similarly, as is suggested in descriptive research, parameters may also vary over space or direction. For a variety of reasons, a migrant may view distances in one direction to be less than they actually are while they view those in another direction to be greater (see Figure 1).

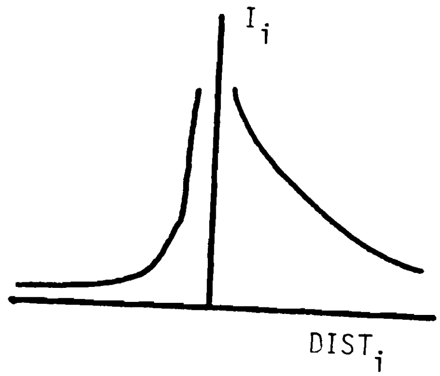

Figure 1. Interaction between the origin and locations at increasing distances for an origin with a directionally variant distance decay function.

While the traditional *x-y* expansion original proposed by Casetti is better known, Casetti and Can (1986) present a method of expansion that allows direction to be taken into account. The direction of each flow out of an origin-specific gravity model to each destination can be calculated, and each parameter can then be expanded as follows:

$$\alpha = \alpha_0 + \alpha_1 \cos (T - \Theta) \quad (3)$$
$$\beta = \beta_0 + \beta_1 \cos (T - \Theta) \quad (4)$$
$$\Gamma = \Gamma_0 + \Gamma_1 \cos (T - \Theta) \quad (5)$$

where T is the angle between the meridian of the origin and the great circle containing both the origin and the destination measured clockwise, and Θ_j is the directional parameter. Interaction with destinations towards angle Θ_α would have a parameter value of just α_0 (respectively, β_0, Γ_0), and interaction towards angle 180 - Θ_α would have a parameter value of $\alpha_0 + \alpha_1$ ($\beta_0 + \beta_1$, $\Gamma_0 + \Gamma_1$). Interaction towards intermediate angles would have a parameter value in the interval $[\alpha_0, \alpha_0 + \alpha_1]$ ($[\beta_0, \beta_0 + \beta_1]$, $[\Gamma_0, \Gamma_1]$). To easier calculate these using ordinary least squares, Equations 3, 4, and 5 can be transformed:

$$\alpha_0 + \alpha_0 \cos (T - \Theta) = \alpha_0 + \alpha_1 (\sin T \sin \Theta + \cos T \cos \Theta)$$
$$= \alpha_0 + \alpha_1 \sin T \sin \Theta + \alpha_1 \cos T \cos \Theta$$

Let $\alpha_{11} = \alpha_1 \sin \Theta$ and $\alpha_{12} = \alpha_1 \cos \Theta$,

$$= \alpha_0 + \alpha_{11} \sin T + \alpha_{12} \cos T. \quad (6)$$

Respective transformations can be performed on the other parameters. This formula is linear, then, and is the same as the cartesian expansion method standardized to the unit circle.

Though there is significant interpretational quality to the estimated values, it is of great interest to retransform these back into the directional form to find the directional parameter estimate. The coefficient parameter estimate, a_1 (b_1,c_2), should be adjusted so it is positive for easy interpretation. This can easily be done by subtracting 180 from the estimate of Θ_j if the sign of the coefficient is negative.

DATA AND STUDY AREA

The United States was chosen as the study area with the data collected at the state level. Because of the availability of the data, the study time selected was 1950 and the migration data were drawn from Lee, et al. (1957). This source lists the state of birth for migrants living in each state. That is, for example, "state of birth of Arkansas residents." There is one limitation to the data, they do not separate repeat migrants. Thus, a person that was born in Michigan, moved to Texas, then moved to Pennsylvania is considered a migrant from Michigan to Pennsylvania, rather than as a repeat mover who first migrated south, then northeast from there.

The distances between states and the angles between the great circles were calculated from the geographical centroids of the states (Douglas, 1932), using simple geodetic formulae. The populations of the states were found in the Bureau of the Census (1976).

Five states were selected on which to carry the analysis out: Illinois, Ohio, New Mexico, North Dakota, and Tennessee. Both large and small states were included so the effect of size on directionality could be evaluated. Similarly, both states in the center of the country and those on the periphery were selected to evaluate whether the location of the origin relative to the destinations was important.

RESULTS AND ANALYSIS

For each of the five states, stepwise regression was employed in expanding the base origin-specific gravity model. One should not consider the individual components (that is, the sine or cosine of the angle) of the expansion to be significant or insignificant by themselves if the direction is of interest. Rather, the two components should be accepted or rejected simultaneously. The proper test to use to evaluate the significance of the expanded variables in this case is the extra F-test (Clark and Hosking, 1986, p. 360) instead of the t-test. Regardless of whether the extra F-test is significant or not, no expanded model which has a lower overall F-value than the unexpanded model should be accepted.

Table 1 shows the values for the unexpanded and the expanded models for each of the five states. Expansion proved significant for three of the five states, in only New Mexico and Ohio was no expansion significant (they met the extra-F requirement, but not the increased overall significance requirement).

The expansions each have their own explanation. If the expansion of the intercept term is significant (which was not the case with any of the states in this analysis), it suggests that there is a directional bias in the absolute number of migrants. If the expansion of the population term is significant, migrants out that state are drawn more by population centers in the direction of bias than by those in the opposite direction. If the expansion of the distance parameter is significant, migrants in that state are less deterred by distance in the direction of bias than by distance in the other direction.

The level of improvement in the explanatory power of the model with the addition of the expanded parameters is high when expansion proved significant. The two states for which expansion was insignificant already had relatively high R^2 values, hence expansion did little for them beyond what would normally be expectated when increasing the number of explanatory variables. The three states for which expansion was significant, Illinois, North Dakota, and Tennessee, showed an increase in R^2 from, respectively, 0.50, 0.37, and 0.45 to 0.80, 0.60, and 0.65. The Illinois case seems unique in that it already had considerable explanation from the population and distance variables before expansion, and direction helped to increase the explanatory power of the model considerably. Ohio and New Mexico had similarly high R^2 values, but did not seem to benefit much from the addition of the expansions.

In addition to the R^2 improvements, the inclusion of the directional expansion helps stabilize the population parameter estimates. In the unexpanded models, the estimate ranges from 0.51 (North Dakota) to 1.07 (Ohio). With expansion, the estimates are less variant, with New Mexico the lowest at 0.73 (insignificant) and the rest clustered around the more intuitive value of unity. Thus, the directionality of the migration might be biasing the population parameter estimate downwards. On the other hand, the distance decay parameter estimate does not seem to be affected by the expansion. None of the estimates (expanded or unexpanded) seem to approach the traditional value of -2 or the more modern -1. Instead, several states cluster around a value of -2.5, while others are calculated to be above -1. It is possible that the addition of a variable that measures the accessibility of destinations to each other (Fotheringham, 1983) would help stabilize the distance-decay estimates.

Figures 2 and 3 illustrate the direction of bias for those states in which expansion was significant for the *DIST* and *POP* parameters, respectively. One notices several things about the directions. First, in the three states east of the Mississippi, the distance decay expansion was more significant, while in the other two, the population parameter expansion was more so. Second, the directionality for each state is different, indicating that preferences are different. This stands in contrast to the idea that everyone has similar perceptions of distance and attractions to populous places. Indeed, given the directions indicated, one might think that these show regional affiliations.

It does not seem to be the case that states on the periphery of the country are any more or less likely to have expansion prove significant than those in the center. Expansion was insignificant for New Mexico but significant for North Dakota. Similarly, expansion was insignificant for Ohio, but significant for Illinois. It might be the case that the distance expansion proves significant for states in the center of the country (Illinois and Tennessee) where distances in different directions might be perceived differently. On the other hand, the population expansion might be more significant for peripheral states.

Whichever parameter estimate is significant, the level of significance of the directional expansion seems to be affected by the location of the state in relation to the region that it identifies with (e.g., Illinois with the Upper Plains, Tennessee with the Southeast).

If a state identifies with several regions or is in the center of the region it identifies with, expansion is insignificant (e.g., New Mexico with Texas and the West, Ohio in the center of the Midwest).

At the same time, there seems to be no correlation between expansion and the size of a state. Both the large states (Illinois and Ohio) and the small states (New Mexico and North Dakota) had one with a significant expansion and one without a significant one.

CONCLUSIONS

The results of this study agree with Casetti and Can (1986) in saying that it does not seem reasonable to ignore direction in obviously directional phenomena. The parameter estimate for the population variable in the gravity model seems to be biased downward when direction is not taken into account. That is, the draw of larger places is underestimated if one does not consider that directional bias exists.

There are several prospects for future research into the utility of directional expansion in gravity models. The effect of expansion on the population parameter estimate needs to be further studied. The present study only offers a single dataset and five examples to support these conclusions, but they still suggest a pattern. By expanding the dataset and using other sources, the stability of the population parameter estimate can be better understood.

Including other variables in the model may help to elucidate the effects of direction on the parameter estimates. The accessibility of destinations has been found to be an important variable in gravity models. Other variables such as temperature, unemployment, and house prices might help the explanatory power of the model.

Figure 2. States for which the direction parameter estimate for distance was most significant. The expansion for Ohio was significant in the XF-test, but the expanded model's F-value did not exceed the original model's value.

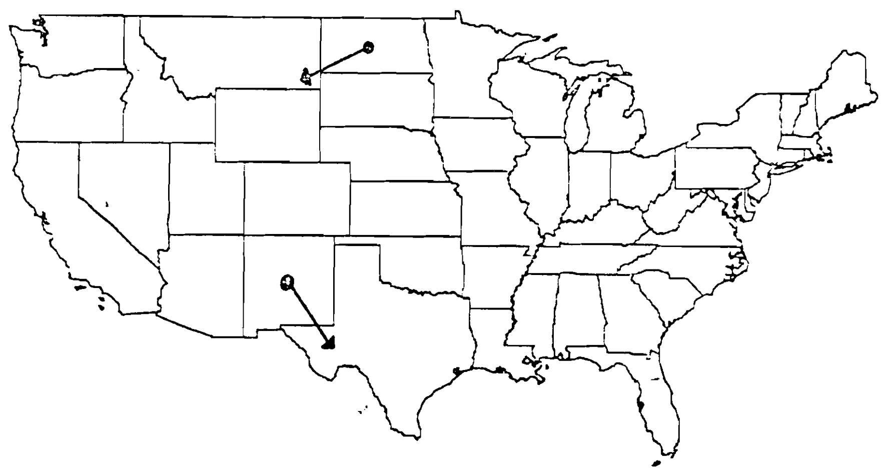

Figure 3. States for which the population parameter estimate was most significant. The expansion for New Mexico was significant in the XF-test, but the F-value for the expanded model did not exceed the value for the original model.

REFERENCES

Bureau of the Census. 1976. *Historical Statistics of the United States, Colonial Times to 1970.* Washington, DC: Department of Commerce.

Casetti, E. 1972. "Generating Models by the Expansion Method: Applications to Geographical Research." *Geographical Analysis* 4:81-91.

Casetti, E., and A. Can. 1986. "Trigonometric Expansions: An Application to the Study of Migration Fields." *Modeling and Simulation* 17:23-28.

Clark, W.A.V. 1986. *Human Migration.* Beverly Hills, CA: Sage Publications.

Clark, W.A.V., and P.L. Hosking. 1986. *Statistical Methods for Geographers.* New York: John Wiley and Sons.

Douglas, E.M. 1932. *Boundaries, Areas, Geographic Centers and Altitudes of the United States and Several States,* United States Geological Survey, Bulletin 817.

Foster, S.A. 1991. "The Expansion Method: Implications for Geographic Research." *Professional Geographer* 43(2):131-142.

Fotheringham, A.S. 1983. "A New Set of Spatial Interaction Models: The Theory of Competing Destinations." *Environment and Planning A* 15:15-36.

Lee, E.S., A.R. Smith, C.P. Brainerd, and R.A. Easterlin. 1957. *Population Redistribution and Economic Growth, United States 1870-1950.* Philadelphia: The American Philosophical Society.