An Intervening Opportunities Model of U.S. Interstate Migration Flows

Siaw Akwawua
University of Regina*

James A. Pooler
University of Saskatchewan**

Unlike other constrained, spatial interaction models the intervening opportunities (IO) model is not often employed for migration analysis. In this paper, an attempt is made to perform an empirical examination of the IO model, using migration data for the 48 conterminous states of the U.S. including the District of Columbia for the period 1975–1980. Ruiter's (1967) numerical method for estimating the basic parameter of the model, the L value, is employed and a program is developed to effect the iterative calibration procedure. Empirical results demonstrate that the model provides a reasonable distribution of observed U.S. migrations and compares favorably with the production constrained gravity model. Further applications of the model and some areas of research are suggested.

Keywords: Intervening opportunities, migration flows, spatial interaction, gravity model, spatial choice, distance, decision-makers.

A frequent criticism of well-known, gravity-type, spatial interaction models is that they rely almost exclusively on the distance variable for the 'explanation' of observed patterns of spatial flows. The implication is that the models are weak behaviorally and theoretically as a result of their dependence on a single simple physical variable that is only related tentatively and indirectly to human behavior. It would be useful to have available a gravity-type model that is more amenable to interpretation as a behaviorally-based approach to spatial interaction.

It is difficult to believe that such a model of spatial interaction has existed for about sixty years now but has been virtually overlooked by practitioners in the field. Researchers have by and large ignored Stouffer's (1940) Theory of Intervening Opportunities even though it loans itself very readily to behavioral interpretations. Spatial interaction models are models of spatial choice and, in Stouffer's model, potential movers are seen to *choose explicitly* among the available opportunities. The model has a behavioral underpinning in that movement is motivated by the desire of an individual or household to satisfy certain basic needs (Jayet, 1990).

Although the Theory of Intervening Opportunities (Stouffer, 1940, 1960) is well known, and extensively discussed in introductory textbooks, empirical examinations of the intervening opportunities *(IO)* model, as practical tests of the theory, are almost non-existent. Some few scholars tested the Stouffer model with migra-

* Department of Geography, University of Regina, Regina, KL, SAS 0A2, Canada
** Department of Geography, University of Saskatchewan, Saskatoon, SK, S7N 5A5, Canada.

tion data in the 1940s (Bright and Thomas, 1941; Isbell, 1944). Stouffer's hypothesis (not the model) was also tested with migration data in the early 1970s (Miller, 1972; Haynes, et al., 1973). Apart from these instances, Slodczyk (1990) is the only empirical test of the model since the 1940s. What is even more surprising is the fact that although Stouffer developed his theory specifically in the context of a model of *population migration,* Slodczyk (1990) is apparently the only empirical application of the model to population migration in the sixty years since Stouffer's original work.

As used by Stouffer (1940) an opportunity refers to a place for the location of human activities or for the termination of trips; that is, the destination of travel. The concept of intervening opportunities refers to opportunities lying closer to the origins of decision-makers that can be considered for the destinations of trips. Some researchers (e.g., Wills, 1986; Fik and Mulligan, 1990) provide empirical evidence that closer, distance-ranked destinations serve to divert traffic. Roy (1993) similarly points out that theoretically, intervening opportunities between an origin and a destination can reduce the number of trips between that origin and the destination.

Schneider (1959) using notions of probability theory, reformulated Stouffer's original *IO* model. Schneider's version of the *IO* model (henceforth called the conventional *IO* model) was later derived from entropy maximization by Wilson (1967, 1970; Wilson and Bennett 1985). This conventional *IO* model was first applied to traffic flows in the Chicago Area Transportation Study (CATS, 1960; Heanue and Pyers, 1966; Jarema et al., 1967; Ruiter, 1967) and subsequently to traffic flows in the Pittsburgh Area Transportation Study (PATS, 1963). Only more recently has Slodczyk (1990) applied the model to population migration, analyzing the 1976–1980 migration flows among the nine provinces of western Poland.

In these previous practical applications of the conventional *IO* model, the empirical method of calibrating the basic parameter of the model, the *L* value, has been employed. Ruiter (1967) described the empirical method as difficult to apply (Pyers, 1965) and suggested, as an alternative, the use of iterative procedures employing the moment estimator (Rogerson, 1986). However, Ruiter's (1967) numerical method for calibrating the *L* parameter still remains a subject for empirical investigation.

The purpose of this paper is to present the results of an empirical examination of the conventional *IO* model in the context of 1975–80 state-to-state migration flow data for the 48 conterminous states of the United States including the District of Columbia (DC). The extent to which the model is supported empirically serves to validate the behavioral theory underlying the model. The results of this empirical application also contribute to understanding the behavior of the model as well as demonstrating its empirical utility in the context of population migration. For purposes of comparison, the production constrained gravity model is calibrated also with the same data set.

In the following section, the theories of destination competition are discussed and the point of view implicit in the *IO* model is elaborated. Following that, the models to be calibrated are described. A description of the interaction data used is presented next, and this is followed by a discussion of the calibration methodology.

The empirical results are then presented and finally, some conclusions and recommendations for further research are offered.

COMPARISON OF THE THEORIES OF DESTINATION COMPETITION

The theory of intervening opportunities is based on the assumption of a *sequential* decision making and information processing strategy. Two stages in the sequential search and choice process may be identified. First, decision-makers initially create a hierarchy of the choice alternatives by ranking them in terms of increasing spatial separation from their origins. Second, the decision-makers proceed step-by-step through the hierarchical ranking and make their selection on the basis of ease of access, starting with the closest one to the origin.

Stouffer (1940) originally makes the assumption that the decision-makers move in space to search for opportunities. This idea may be generalized to the notion that the decision-makers can do their spatial search without necessarily moving from their origins. Decision-makers can rely on various sources such as family and friends, the news media, or other sources of information to acquire their spatial information. The processing of the spatial information is done mentally at the origins. The hierarchical ranking of the spatial alternatives and the sequential elimination of undesirable alternatives from the choice set do not require movement from the origins as Stouffer originally assumes. The assumption of a sequential decision making and information processing is based on Stouffer's (1940) assertion that the decision-maker may not have detailed knowledge of distant opportunities.

The theory of competing destinations assumes that spatial decision making and information processing is a *hierarchical* process. According to Fotheringham (1988) this hierarchical or sequential choice process involves two stages. In the first stage, decision-makers choose a broad region or cluster of choice alternatives, such as destinations. In the second stage, the decision-makers choose specific sites from within the selected cluster. The idea of a hierarchical processing of information is based on the argument that decision-makers may not have information about all the alternatives and so they cannot evaluate all of them. Also the choice set of spatial alternatives may be too large and the decision-makers consider only a subset of the choice set.

An alternative viewpoint is the theory of spatial dominance (Pooler, 1992, 1998). This theory also postulates a *hierarchical* processing of spatial information. There are two stages in this hierarchical spatial decision making process. In the first stage, decision-makers reduce the size of the choice set by grouping the destinations into clusters based on the spatial dominance exerted on the origins. Spatial dominance is calculated as size divided by distance of destinations. In the second stage, the decision-makers choose one of the destinations in the reduced sets.

All of the theories discussed above adopt a two-stage approach to model spatial decision making. While the theories of competing destinations and spatial dominance assume a hierarchical process, the theory of *IO* assumes a sequential process. As already indicated, a sequential decision making process implies a hierarchical *ranking* of the choice alternatives by the decision-makers. The theories on destination competition are complementary and the spatial decision-making process involves first, a hierarchical process and second, a sequential consideration of alternatives. The theory of *IO* involves a ranking and sequential selection of alternatives. However, in both the theory of competing destinations and the theory of spatial dominance, there is no such specific selection rule.

In the theory of competing destinations, the alternatives in the chosen cluster are spatially contiguous. On the other hand, in the theory of spatial dominance, the alternatives in the reduced choice sets may or may not be in spatial proximity to one another. In both the theories of competing destinations and spatial dominance, the decision-makers do not consider all the choice alternatives; a process of simplification is assumed. This is based on the argument that the choice set of spatial alternatives may be very large and, therefore, decision-makers have to reduce the size of the choice set. Also, it is implicit in the theory of intervening opportunities that decision-makers do not consider all the choice alternatives since the search process is terminated as soon as a satisfactory opportunity is found. Stouffer (1940) claims that the spatial decision-maker may not have the same detailed knowledge of distant opportunities as he/she has of nearby opportunities. This assertion implies that the decision-maker does not evaluate all the choice alternatives. Rather, the uncertainty surrounding the decision making and choice process is reduced by a limitation of the size of the choice set to those alternatives which are clearly perceived.

THE INTERACTION MODELS

In this section the two spatial interaction models to be calibrated are described. These are the intervening opportunities *(IO)* model and the production constrained (PC) model. These two models are the most similar trip distribution models (Sheppard, 1986) and it is deemed instructive to assess the extent to which they compare when confronted with observed migration data.

The Intervening Opportunities Model

The original *IO* model developed by Stouffer (1940) is based on the hypothesis that '...the number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities...' (1940:846). This hypothesis postulates a linear relationship between the number of trips and the spatially cumulative opportunities. This theory shifts attention away from the traditional conception of spatial distance as an impediment to population movement. The emphasis instead is put on the fact that migrants move in space in order to satisfy certain needs. Migrants from a place A do not move to a more distant place C because there is an intervening opportunity at an intermediate place B, which can satisfy their needs. Individuals are willing to move only as far as they must in order to find an appropriate opportunity. The more intervening opportunities there are, the less the migration to the more distant place. As Stouffer (1940) argues, distance itself is important only as a measure of the number of intervening opportunities (Wilson, 1967; Jayet, 1990). The intervening opportunities between an origin and a destination could be used in place of an absolute measure of linear distance. In essence, Stouffer replaces the ratio scale of linear distance with a concept of ordinal or rank ordered distance.

Stouffer's (1940) original ideas were adopted and supplemented later by Schneider (1959) with two main hypotheses. These are: (i) total travel time from a point is minimized, subject to the condition that every destination point has a stated probability of being selected if it is considered; and (ii) Tthe probability of a destination being accepted, if it is considered, is a constant, independent of the order in which destinations are considered. The above hypotheses are based on the premise that a trip remains as short as possible, lengthening only as it fails to find an acceptable destination at a lesser distance (Heanue and Pyers, 1966). In deciding to make a trip, a household considers the nearest destination to its origin, and if that is unacceptable, considers the next nearest, and so on. The probability that trips from a given origin zone *i* will stop at a given destination zone *j*, depends therefore on the number of destinations in zone *j*, and on the number of destinations closer to the origin than zone *j* is to the origin.

The mathematical formulation of the conventional *IO* model derived from the above two hypotheses takes the following form:

$$M_{ij} = k_i O_i [\exp(-LV_{j-1}) - \exp(-LV_j)] \tag{1}$$

where M_{ij} is the predicted number of movers from origin *i* to the *jth* distance-ranked destination away from *i*, O_i is the known total outflow originating from origin *i*, V_{j-1} is the cumulative number of opportunities up to the immediately preceding zone, *j-1; L* is a constant probability that a random destination will satisfy the needs of a traveler; and where *k* is a constant or balancing factor which ensures

$$\sum_j M_{ij} = O_i \cdot \tag{2}$$

The value of the scaling constant *k*, which ensures that the matrix M_{ij} satisfies the origin constraint equation (2), is obtained from

$$\sum_j M_{ij} = k_i O_i [1 - \exp(-LV_n)] = O_i \tag{3}$$

and hence

$$k_i = \frac{1}{1 - \exp(-LV_n)} \tag{4}$$

where *n* is the total number of zones in the system.

The value of the parameter L is determined from observed data to ensure that the model estimated mean trip length matches the observed mean trip length (alternatively called mean migration length/distance). The formula is written as:

$$\bar{r}_o = \frac{\sum_{j=1}^{n} d_{oj}\left[\exp(LV_{j-1})\exp(LV_j)\right]}{1-\exp(LV_n)} \tag{5}$$

(Ruiter, 1967), where $\bar{r}_o$ is the average trip length for zone o and d_{oj} is the distance from zone o to zone j. The L value or probability factor describes the rate at which migration flows from an origin decline as several destinations are encountered and as the length of migration increases from the origin. The parameter O_i is measured by the volume of outmigration from each origin; the total inflows to each destination is used as a surrogate measure for V_j.

The Production Constrained (PC) Model

The other model to be estimated in this study is the production constrained (PC) migration model given by the following equations

$$M_{ij} = A_i O_i D_j \exp(-\beta d_{ij}), \tag{6}$$

where

$$A_i = \frac{1}{\left[\sum_j D_j \exp\left(-\beta d_{ij}\right)\right]}, \tag{7}$$

and where M_{ij} is the predicted interaction from origin i to destination j, O_i is the known total number of movers originating at i, D_j represents the attractiveness of destination j (which is here measured by the known total inflow attracted to j), d_{ij} is the distance between i and j, and β is the distance decay parameter.

Unlike the *IO* model, the PC model is based on the gravity hypothesis, which postulates an inverse relationship between distance and migration flows (Wilson, 1974; Haynes and Fotheringham, 1984; Wilson and Bennett, 1985). In both models, the number of migrants leaving an origin is known and the task of the modeler is to predict to which destinations the migrants move. The major difference between the models is that in the PC model, absolute spatial distance is a variable that influences migration flows, whereas in the *IO* model migration flows are influenced by the number of destinations closer to an origin than any particular destination being considered.

THE DATA

The empirical evaluation of the models utilizes the state-to-state population migration data for the 48 conterminous states of the U.S. including the District of Columbia (DC) for the period 1975–1980. These data represent a 19 percent sample of the 1980 U.S. Census of Population and Housing and are published by the U.S. Bureau of the Census, 1983. The 1980 U.S. Census of Population and Housing included a question that asked for retrospective information on the state of residence of individuals in 1975. The Bureau of the Census infers migration from these data by comparing the state of residence in 1980 with the state of residence in 1975. The data therefore measure the migrations for the period 1975–1980 and refer to the mobility of persons aged 5 years and over. In total, 19.7 million persons migrated between the 48 conterminous states and DC during the period. Inasmuch as this paper is concerned solely with the empirical performance of the models the datedness of the data is not considered to be an important issue.

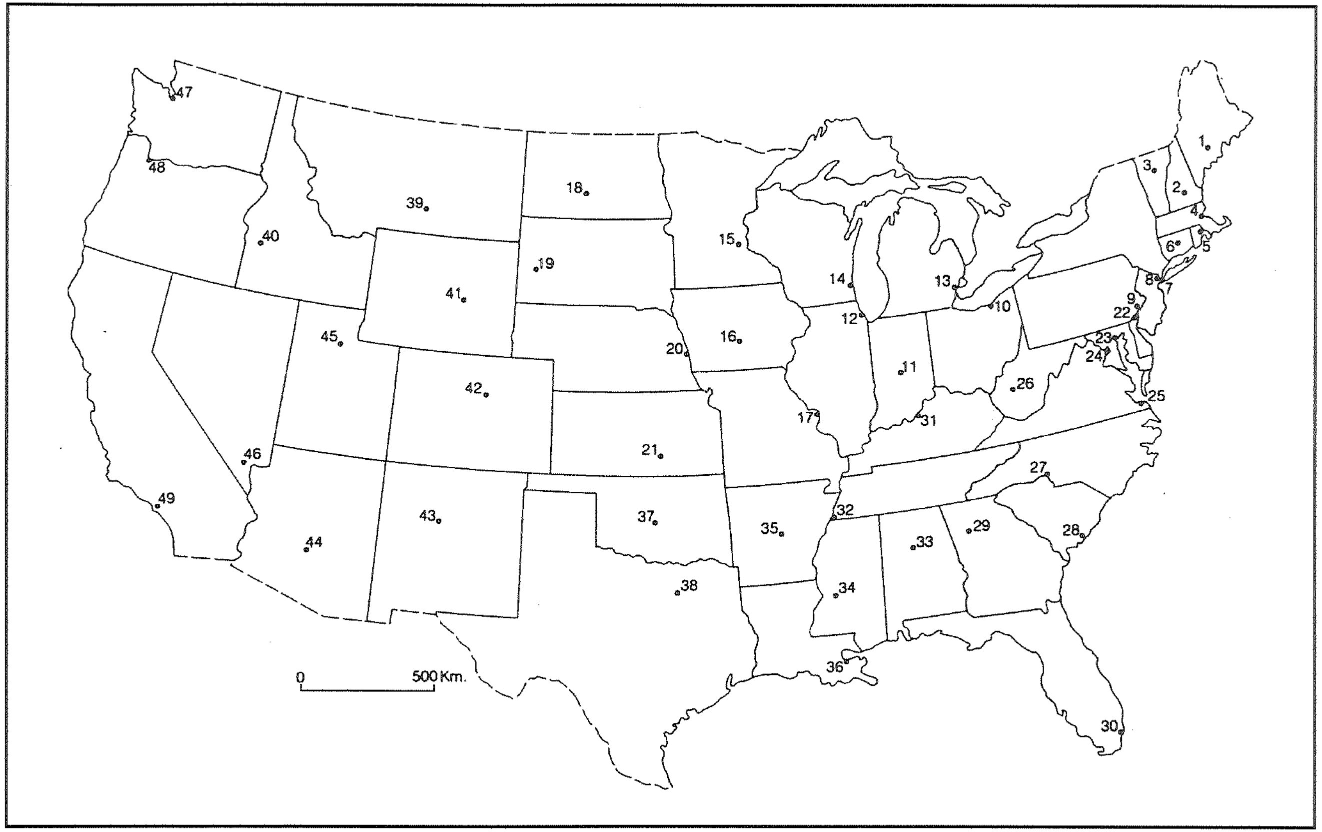
In the conventional *IO* model, the measure of spatial separation, defined with respect to distance below, is not used in absolute terms as a variable but instead, as an ordinal measure, as the basis for rank-ordering destinations in increasing order away from each origin. The data for the distance matrix were obtained from calculations using map grid references (Stillwell, 1991) for centers of population of the 48 conterminous states of the U.S. including DC. The distances were measured as inter-state (i.e., inter-area) distances between these centers of population and are in terms of map grid units, where 1 map grid unit ≑ 58.33 km. An outline map of the U.S. showing the 48 conterminous states and the locations of the centers of population are shown in Figure 1. The centers of population are listed in Appendix 1.

MODEL CALIBRATION

Calibration

Calibration of the *IO* model is not an easy task and this may explain why the model has been so seldom employed. One of the problems involved in calibrating the conventional *IO* model concerns the appropriate calibration methodology to adopt. Several researchers (for example Pyers, 1965; Ruiter, 1967; Jarema et al., 1967) point out that the two calibration methods developed for the Chicago Area Transportation Study (CATS, 1960) are lacking in ease of applicability. The first method involves manual derivation of the L value empirically, from observed data. According to Ruiter (1967:8) the empirical method is 'costly and error-prone'. Since previous researchers have found that manual calculation of the L value is not a useful method, it is not employed in this study. This study utilizes a 49 by 49 data matrix (of 19.7 million persons), and an attempt to derive the L value by manual calculations may only serve to confirm Ruiter's conclusion. Also, attempting to solve non-linear equations, such as the *IO* model equations, by manual calculations may be deemed obsolete in view of current technological innovations and the availability of computer programs.

Figure 1: Outline map of the United States locations of centers of population of the 48 conterminous states including DC.



The second method, which is also not employed in this study, uses multiple regression techniques on logarithmic transformations of the variables in model equation (1). Ruiter (1967:20) points out that this statistical calibration method 'is neither fast nor simple'. Openshaw (1976) points out that parameters estimated by regression techniques may not necessarily satisfy the constraints on the model. Wilson (1976) further discusses the inappropriateness of the use of regression method in calibrating spatial interaction models.

As an alternative to these two calibration methods, Ruiter (1967) proposes the use of iterative methods for solving non-linear equations. This iterative technique is based on the moment estimator and so uses the mean trip length, equation (5), as the calibration statistic. The application of this iterative method in the migration context has been rare. This study applies the iterative technique to migration data and a program is developed to calibrate the model for the purposes of this study.

The solution of equation (5) depends on finding the value of the L parameter that satisfies the origin constraint equation (2). Convergence of equation (2) occurs when an optimal value of the parameter is obtained and is used for estimating equation (1) to obtain the predicted trip matrix. An initial L value of 10^{-7}, calculated theoretically as 1/trips, is input into the calibration program to start the iteration. The program first rank-orders the destinations in terms of nearness to each origin; that is, the closest destination to the origin is given rank 1; the next nearest destination is given rank 2; and so on. The program then arranges the destination inflow data so that they are in ordinal distance rank order. The inflow data are then summed by destination rank order.

The calibration program is designed to make adjustments, at the end of every iteration, by either incrementing or decrementing the L value by very small amounts, say 10^{-8}, so as to converge on a solution. After a predetermined number of iterations, say 200, the difference between the observed and predicted mean trip lengths is examined. Ideally this difference must be zero, but time can be saved if the stop criterion is set to a certain minimum value. Therefore the iterative process is stopped when

$$\left| \bar{d}_{obs} - \bar{d}_{\text{mod}} \right| < 0.30 \; \textit{map grid units} \tag{8}$$

where $\bar{d}_{obs}$ is the observed mean trip length and $\bar{d}_{\text{mod}}$ is the mean trip length predicted by the model. In order to allow for flexibility the actual stop criterion is put between 0.10 map grid units and 0.50 map grid units. The PC model is calibrated using the method of maximum likelihood as developed by Baxter (1976). Like the iterative technique developed by Ruiter, the method of maximum likelihood uses the mean trip length as the calibration criterion (Evans, 1971; Batty and Mackie, 1972).

Goodness-of-Fit

A number of goodness-of-fit statistics are used to assess the performance level of the model. These are:

(i) the percentage misallocated which calculates the percentage of migrants allocated to the wrong cells in the matrix. The measure is defined as:

$$Percentage\ Misallocated = \frac{1}{2} \sum_{i-1}^{n} \sum_{j=1}^{n} \left| \hat{M}_{ij} - M_{ij} \right| \tag{9}$$

where M_{ij} is the observed number of migrants moving from i to j, $\hat{M}_{ij}$ is the number of migrants predicted by the model.

(ii) the mean absolute error statistic *(MABE)*, which describes the absolute number of migrants the model fails to capture.

$$MABE = \frac{\sum_{ij} \left| M_{ij} - \hat{M}_{ij} \right|}{n^2} \tag{10}$$

A discussion of these goodness-of-fit indicators can be found in Smith and Hutchinson (1981); Baxter (1983); and Knudsen and Fotheringham (1986).

RESULTS OF THE CALIBRATION

Parameters and Overall Model Performance

Table 1 shows the observed and model estimated mean lengths of migration for each of the 49 origins under study. Examination of the entries in the table shows that, to a large extent, there is generally good agreement between the observed and model estimated mean lengths of migration. Whereas the *IO* model estimates an average migration length, which is within 0.5 percent of the observed value, the PC model's estimated average migration length is within 0.4 percent of the observed value. These results indicate that the two models are capable of satisfying the internal requirement of trip length constraint within a reasonable degree of accuracy.

The spatial variations in the distribution of the mean migration distances suggest that out-migrants from centralized origins such as DC and Delaware move, on the average, over short distances. On the other hand, out-migrants from peripheral origins such as California and Washington states move, on the average, over longer distances. This relationship between location and mean migration distance is obviously a function of accessibility. Centralized origins, at high population densities, are potentially more accessible than peripheral origins, which may be considered less accessible. Also the density of opportunities in the destination choice set at closer distances is larger for out-migrants from centralized origins than it is for out-migrants from peripheral origins. This explains why out-migrants from peripheral locations have to move, on the average, longer distances to satisfy their needs.

The parameter values obtained by calibrating the two models are described in Table 2. The final system-wide *L* value which predicts a migration matrix that satisfies the origin constraint and the observed mean migration length of 24.58 map grid units per person is 1.58×10^{-8}. The *L* value obtained here confirms previous findings

Table 1: Observed and estimated mean migration distances for calibrated spatial interaction models. (Distance in Map Grid Units. 1 Map Grid Unit = 58.33 km.)

State	*Observed*	*Estimated*	
		IO Model	*PC Model*
DC	12.15	12.05	12.03
West Virginia	17.88	17.77	17.78
Kentucky	18.37	18.26	18.27
Delaware	18.67	18.56	18.57
Arkansas	18.69	18.58	18.58
Mississippi	18.96	18.85	18.85
Tennessee	19.31	19.20	19.21
Alabama	19.31	19.20	19.21
Oklahoma	19.43	19.32	19.31
Kansas	19.84	19.73	19.74
Georgia	20.32	20.21	20.21
North Carolina	20.99	20.88	20.86
Iowa	21.15	21.04	21.15
Missouri	21.34	21.23	21.24
South Carolina	21.51	21.40	21.31
South Dakota	21.72	21.61	21.62
Indiana	21.89	21.78	21.72
Nebraska	21.96	21.86	21.86
Wyoming	22.04	21.94	21.94
Maryland	22.36	22.25	22.26
Idaho	22.99	22.88	22.86
Louisiana	23.25	23.14	23.15
Rhode Island	23.80	23.70	23.71
Pennsylvania	24.53	24.42	24.43
New Hampshire	24.79	24.68	24.69
Vermont	24.99	24.87	24.89
Utah	25.18	25.07	25.08
New Mexico	25.25	25.14	25.14
North Dakota	25.51	25.40	25.41
Wisconsin	25.56	25.06	25.45
New Jersey	25.57	25.47	25.46
Illinois	25.83	25.72	25.73
Montana	25.98	25.87	25.87
Connecticut	26.07	26.00	26.07
Nevada	26.27	26.16	26.16
Virginia	26.30	26.19	26.20
Minnesota	26.53	26.42	26.43
Colorado	27.16	27.06	27.07
Texas	27.48	27.38	27.37
Ohio	27.77	27.38	27.43
Massachusetts	27.93	27.84	27.83
Oregon	28.37	28.26	28.25
Michigan	29.26	29.05	29.16
New York	30.17	30.07	30.08
Arizona	30.77	30.66	30.66
Maine	31.47	31.36	31.38
Florida	36.81	36.72	36.71
Washington	37.66	37.56	37.55
California	43.32	43.22	43.32
Overall Average	24.58	24.46	24.47
Variance	29.70	29.66	29.78

based on calibration of the *IO* model, which show that the probability factor is small and positive. In this application, the interpretation of the L value is that for the U.S. state-to-state migrants during the 1975–80 period the probability of destination selection is low.

To facilitate comparison of the conventional *IO* model with the PC model, one global β is calibrated for the latter model. The calibrated value of β that predicts a trip matrix satisfying the PC model origin constraint, as well as the trip length constraint, is -0.03. This low negative β value suggests that during the period 1975–80 distance did not exert a large impediment to U.S. state-to-state migrants.

Table 2: Parameter values and goodness-of-fit for calibrated spatial interaction models.

Parameter	*IO Model*	*PC Model*
L value	1.58×10^8	–
β	–	-0.03
A_i	–	2.43
Percentage Misallocated	34.70	31.28
	(68.07)*	(40.03)*
Mean Absolute Error	5094	5073
	(17446801)**	(24152511)**

* Figures in brackets are variances.

** Variance large (but included here because of reviewer's comment).

In Table 3 the results of measuring the performance level of the models using the two goodness-of-fit indicators are presented. Inspection of the magnitude of the error levels suggests that the *IO* model provides a reasonable distribution of migration flows. The errors range from a low of about 21 percent in three origins, Michigan, Virginia, and Florida to a high of about 57 percent in DC. Only in two other origins is the percentage of migrants misallocated slightly greater than 50 percent; and these are Oregon and Idaho with 52 and 54 percent respectively. The overall average percentage misallocated is 34.70 with a variance of 68.07. This overall mean shows that the conventional *IO* model allocates about 65 percent of all migrants to the correct destinations.

The distribution of percentage misallocated by the PC model for the 49 origins are shown in the third column of Table 3. The lowest percentage of migrants misallocated by the model is 18 percent in Arizona. The District of Columbia has the highest value of 54 percent. This gives an overall average error of 31 percent with a variance of 40.03 (Table 3). The magnitude of the overall average percentage misallocated indicates that the PC model correctly allocates nearly 69 percent of all migrants in the trip matrix.

The results of the computation of the mean absolute error statistics are analyzed in the fourth and fifth columns of Table 3. The values of the mean absolute error for

Table 3: Goodness-of-Fit for calibrated spatial interaction models.

	Goodness-of-Fit			
	Percent Error		*MABE*	
State	*IO Model*	*PC Model*	*IO Model*	*PC Model*
DC	56.65	54.12	4012	3833
West Virginia	38.54	32.78	2407	2048
Kentucky	31.86	28.55	3672	3294
Delaware	34.35	26.64	1181	916
Arkansas	37.78	31.84	3219	2713
Mississippi	36.48	29.55	3246	2629
Tennessee	30.03	25.68	4320	3695
Alabama	31.26	25.43	3518	2861
Oklahoma	38.13	32.35	4173	3540
Kansas	37.25	30.00	4430	3567
Georgia	31.27	26.67	5818	4963
North Carolina	27.01	23.23	5018	4316
Iowa	37.44	36.29	4230	4100
Missouri	29.80	30.73	5571	5745
South Carolina	29.85	25.87	3253	2819
South Dakota	46.68	39.30	1729	1456
Indiana	26.11	27.80	5047	5373
Nebraska	34.54	31.97	2574	2382
Wyoming	43.21	34.20	1293	1023
Maryland	31.46	32.06	6058	6174
Idaho	53.79	39.30	2755	2013
Louisiana	37.05	35.75	4250	4101
Rhode Island	37.65	38.16	1448	1466
Pennsylvania	23.06	26.73	7999	9273
New Hampshire	41.85	42.50	1964	1995
Vermont	39.39	40.71	986	1020
Utah	44.34	26.92	2482	1507
New Mexico	38.98	25.39	2830	1844
North Dakota	40.05	34.45	1433	1233
Wisconsin	29.36	33.63	3967	4553
New Jersey	29.26	32.09	9267	10165
Illinois	24.04	29.13	10911	13221
Montana	47.95	35.42	2078	1533
Connecticut	32.14	34.30	4598	4907
Nevada	42.79	23.12	2141	1157
Virginia	21.49	23.33	5563	6038
Minnesota	35.22	36.73	4615	4811
Colorado	30.10	20.87	5211	3614
Texas	26.23	27.19	9270	9609
Ohio	23.26	30.11	8973	11614
Massachusetts	36.18	38.56	8108	8640
Oregon	51.56	32.61	5083	3215
Michigan	20.53	26.76	5826	7594
New York	31.27	33.68	22302	24026
Arizona	31.71	17.64	4585	2551
Maine	33.46	34.33	1465	1503
Florida	21.42	28.08	8638	11323
Washington	37.56	25.89	5276	3472
California	28.96	32.17	20818	23127
Overall Average	34.70	31.24	5094	5073
Variance	68.07	40.03	17446801**	24152511**

*Distance in Map Grid Units (1 Map Grid Unit = 58.33 km).

**Variance large (but included here because of reviewer's comment).

the *IO* model's predictions range from a minimum of 986 in Vermont to a maximum of 22,302 in New York. The overall average of the mean absolute error statistic for all the 49 origins is 5,094. The mean absolute errors for the PC model range from a low of 916 in Delaware to a high of 24,026 in New York; with an overall average of 5,073 for the 49 origins.

The foregoing results of the statistical indicators are tested for significant differences using Student's *t*-test. The two models are compared using two tailed test with both $\alpha = 0.5$ and 0.10 significant levels. The *t*-value for the percentage misallocated is 0.022 (not significant) and that for the mean absolute error is 0.9815 (also not significant). These results imply that each of the two models provides satisfactory predictions.

MODEL COMPARISONS

From the analysis of the aggregate characteristics of the two models, it seems fair to say that the results as described in Table 2 are similar in terms of goodness-of-fit and that both models are able to provide a relatively good fit to the U.S. state-to-state migration flows for the 1975–80 period. The differences between the overall averages of the two statistical indicators shown in Table 2 are not large and results of the Student's *t*-test reveal no significant differences between the predictions of the two models.

Examination of the models' percent error levels in Table 3 shows that the two models exhibit a similar pattern of spatial allocation of migrants. In order to explain the pattern of spatial allocation by the models, it is necessary to explore the relationship between the location of migrants and the models' predictions. A measure of location that is conveniently used is the mean migration distance. The Pearson product-moment correlation between mean migration distance and the *IO* model's error level (i.e., percentage misallocated) results in a low *r*-value of -0.30 (not significant at 0.01 two-tailed). The correlation between the mean migration distance and the PC model percent error is low and negative, -0.22, but is also not significant at 0.01 level (two-tailed).

The mean migration distance also is a function of the accessibility of destinations to the origins, as measured by an index of population potential (Stillwell, 1991). A low mean migration distance implies more accessible locations and a high mean migration distance is an indication of inaccessibility (i.e., peripheral locations). It follows that although the result of the statistical test is not significant, the negative coefficients imply nevertheless that, for the U.S. data analyzed in this study, there is a tendency for both the *IO* model and the PC model to produce the greatest errors where accessibility is high. In contrast, where migrants move long distances, that is, in less accessible and hence peripheral locations, both models' prediction errors tend to be moderate or small.

Comparison of Diversities Using Shannon's Entropy Measure

A further statistical test of the models' predicted interactions is performed using Shannon's (1948) entropy measure H. This is a measure of the diversity of the interactions (Haynes and Machunda, 1988). If the predicted interactions exhibit the same or similar degree of spatial diversity as the actual flow data set, as measured by the entropy index, then this gives an indication of the validity of the model's predictions. The statistical significance of the difference between the entropy of the observed and predicted interactions is tested with Student's t-test (Hutchenson, 1970).

The Shannon (1948) entropy measure takes the following form:

$$H = -\sum_{i=1}^{n} p_i \ln p_i, \tag{11}$$

where

$$p_i = x_i / \sum_{i=1}^{n} x_i , \tag{12}$$

and where

$$\sum_{i=1}^{n} p_i = 1. \tag{13}$$

In this formulation, x_i represents the observed or predicted migration flow for each of the n = 48 destinations with reference to one origin i.

The entropy values are calculated for both the observed and predicted interactions and are used to test the hypothesis of equal diversities in the observed and predicted data sets. The entropy values presented in Table 4 are in bits. In this application, the units of bits represent the minimum amount of information needed to represent the data. A higher value of this quantity indicates a slightly greater diversity in the predicted interactions than in the observed. The figures in Table 4 indicate that the PC model's predictions have slightly lower entropy than the *IO* model's predictions. Judging by this result the PC model slightly outperforms the *IO* model.

To test the hypothesis of equal diversities in the observed data set and the models' predictions, t-ratios are calculated and are shown in Table 4. The result of the significance test indicates that all t-values are not significant at $\alpha = 0.01$ level (two-tailed). This result suggests that the hypothesis of equal diversities cannot be rejected.

Table 4: Comparison of diversities in the observed and predicted interactions using Shannon's Entropy Measure.

	Entropy Value			
Model Type	*Observed*	*Predicted*	*Difference*	*t-ratio*
IO Model	3.1338	3.5595	0.4257	1.8327*
PC Model	3.1338	3.5133	0.3795	1.4573*

* Not significant p<0.01.

The implication is that there is good agreement between both models' predictions and the actual flow data since both data sets exhibit similar patterns of dispersion.

CONCLUSION AND RECOMMENDATIONS

This paper has presented the results of an empirical analysis of the conventional *IO* model in comparison with the gravity-based PC model, using the 1975–80 U.S. state-to-state migration flows for the 48 conterminous states. Although the PC model appears to perform slightly better than the *IO* model, the results of the various statistical tests show no significant differences between the two models' predicted interactions. Both models are thus capable of distributing migration flows with a reasonable degree of accuracy. Clearly, the results obtained in this study could be improved and as such a tentative conclusion is that the exercise has demonstrated the applicability of the *IO* model of trip distribution to migration analysis.

The theoretical basis of the *IO* model is that population movement is motivated by the satisfaction of individual/household needs. This basis provides the model with a behavioral underpinning that distinguishes it from the traditional family of gravity-type interaction models. It is clearly a model wherein the *spatial choice* of the movers is explicit. The validation of the model in this paper provides empirical support for the behavioral implications of the theory as a model of spatial choice.

Two recommendations are made following this calibration of the conventional *IO* model using U.S. state-to-state migration data. First, the use of a constant probability, *L* value, for all possible destinations certainly contributed to the large differences between model predictions and the observed data. This suggests that there is the need for a variable *L* value. The empirical results obtained here confirm theoretical suggestions advanced by Harris (1964) and Ruiter (1967) that the *L* value must be specified as a function. The problem with this notion is what to use as the specific type of function for the probability of destination selection. While Harris (1964) suggested a gamma function, Ruiter (1967) suggested a power function. Unfortunately their modified versions of the *IO* model based on these alternative functional forms have not been operationalized. Further research in this area is indeed warranted. Secondly, there is the need for further discussions of the spatial choice and behavioral implications of the model.

REFERENCES

Batty, M. and Mackie, S. (1972) The calibration of gravity, entropy, and related models of spatial interaction. *Environment and Planning A*, 4:205–203.

Baxter, M. (1976) *Computer and Statistical Techniques for Planners*. London: Methuen.

——. (1983) Model misspecification and spatial structure in spatial-interaction models. *Environmental and Planning A*, 15:319–327.

Bright, M.L. and Thomas, D.S. (1941) Interstate migration and intervening opportunities. *American Sociological Review,* 6:773–783.

CATS (Chicago Area Transportation Study) (1960) Final Report II.

Evans, A.W. (1971) The calibration of trip distribution models with exponential or similar cost functions. *Transportation Research,* 5:15–38.

Fik, T.J. and Mulligan, G.F. (1990) Spatial flows and competing central places: Towards a general theory of hierarchical interaction. *Environmental and Planning A,* 22:527–549.

Fotheringham, A.S. (1988) Consumer store choice and choice set definition. *Marketing Science,* 7:299–310.

Harris, B. (1964) A note on the probability of interaction at a distance. *Journal of Regional Science,* 5:31–35.

Hutchenson, K. (1970) A test for comparing diversities based on the Shannon formula. *Journal of Theoretical Biology,* 29:151–154.

Haynes, K.E. and Fotheringham, A.S. (1984) *Gravity and Spatial Interaction Models.* Beverly Hills: Sage.

Haynes, K.E., Machunda, Z.B. (1988) Decomposition of change in spatial employment concentration: An information-theoretic extension of shift-share analysis. *Papers of the Regional Science Association,* 65:101–113.

Haynes, K.E., Poston, D.L. and Schnirring, P. (1973) Intermetropolitan migration in high and low opportunity areas: Indirect tests of the distance and intervening opportunities hypothesis. *Economic Geography,* 49:68–73.

Heanue, K.E. and Pyers, C.E. (1966) A comparative evaluation of trip distribution procedures. *Highway Research Record,* 114:20–50.

Isbell, E.C. (1944) Internal migration in Sweden and intervening opportunities. *American Sociological Review,* 9:627–639.

Jayet, R. (1990) Spatial search processes and spatial interaction: 1. Sequential search, intervening opportunities, and spatial search equilibrium. *Environmental and Planning A,* 22:583–599.

Jarema, F.E., Pyers, C.E., and Reed J.A. (1967) Evaluation of trip distribution and calibration procedures. *Highway Research Record,* 191:106–129.

Knudsen, D.C. and Fotheringham, A.S. (1986) Matrix comparison, goodness-of-fit and spatial interaction modeling, *International Regional Science Review,* 10:127–147.

Miller, E. (1972) A note on the role of distance in migration: Costs of mobility versus intervening opportunities. *Journal of Regional Science,* 12:475–478.

Openshaw, S. (1976) An empirical study of spatial interaction models. *Environment and Planning A,* 8:23–41.

PATS (Pittsburgh Area Transportation Study) (1963) Final Report II.

Pooler, J.A. (1992) Spatial uncertainty and spatial dominance in interaction modeling: A theoretical perspective on spatial competition. *Environment and Planning A,* 24:995–1008.

——. (1998) Competition among destinations in spatial interaction models: A new point of view. *Chinese Geographical Science,* 8:2112–224.

Pyers, C.E. (1965) Evaluation of intervening opportunities trip distribution model. *Highway Research Record,* 114:71–98.

Rogerson, P.A. (1986) Parameter estimation in the intervening opportunities model. *Geographical Analysis,* 18:357–360.

Roy, J.R. (1993) Comments on 'The development of a new gravity-opportunity model for trip distribution'. *Environment and Planning A,* 25:1689–1691.

Ruiter, E.R. (1967) Improvements in understanding, calibrating, and applying the opportunity model. *Highway Research Record,* 165:1–21.

Schneider, M. (1959) Panel discussion on inter-area travel formulas. *Highway Research Record,* 253: 128–138.

Shannon, C.E. (1948) A mathematical theory of communication. *Bell Systems Technical Journal,* 27:379–423 and 623–656.

Sheppard, E. (1986) Modeling and Predicting aggregate flows. In Hanson, S. (ed.) *The Geography of Urban Transportation.* New York: The Guilford Press, pp. 91–118.

Slodczyk, J. (1990) Spatial structure of migrations from the standpoint of the model of intervening opportunities. In Zipseer, T. (ed.) *The Model of Intervening Opportunities in Theory and Practice of Territorial Arrangement.* Scientific papers of the Institute of History of Architecture, Arts and Technology. Wroclaw: Technical University of Wroclaw, Wydawnictwo Politechniki Wroclawskiej, 21:143–153.

Smith, D.P., and Hutchinson, B.C. (1981) Goodness-of-fit statistics for trip distribution models. *Transportation Research,* 15A:295–303.

Stillwell, J. (1991) Spatial interaction models and the propensity to migrate over distance. In Stillwell, J. and Congdon, P. (eds.) *Migration Models: Macro and Micro Approaches,* London: Ballhaven Press, pp. 34–56.

Stouffer, S. A. (1940) Intervening opportunities: A theory relating mobility and distance. *American Sociological Review,* 5:845–867.

——. (1960) Intervening opportunities and competing migrants. *Journal of Regional Science,* 2:1–26.

U.S. Bureau of the Census (1983) *Census of Population Supplementary Report* PC80–51–9. State of Residence in 1975 by State of Residence in 1980. Washington, D.C.: Government Printing Office.

Wills, M.J. (1986) A flexible gravity-opportunity model for trip distribution. *Transportation Research,* 29B:89–111.

Wilson, A.G. (1967) A statistical theory of spatial distribution models. *Transportation Research,* 1:253–269.

——. (1970) *Entropy in Urban and Regional Modeling.* London: Pion

——. (1974) *Urban and Regional Models in Geography and Planning.* New York: John Wiley.

——. (1976) Statistical notes on the evaluation of calibrated gravity models. *Transportation Research,* 10: 343–345.

Wilson, A.G. and Bennett, R.J. (1985) *Mathematical Methods in Human Geography and Planning.* Chichester: John Wiley and Sons.

Appendix 1. Centers of population for the 48 conterminous states of the U.S. including the District of Columbia.

Number	*State*	*Center*
1	Maine	Augusta
2	New Hampshire	Concorde
3	Vermont	Montpellier
4	Massachusetts	Boston
5	Rhode Island	Providence
6	Connecticut	Hartford
7	New York	New York
8	New Jersey	Jersey City
9	Pennsylvania	Philadelphia
10	Ohio	Cleveland
11	Indiana	Indianapolis
12	Illinois	Chicago
13	Michigan	Detroit
14	Wisconsin	Milwaukee
15	Minnesota	Minneapolis
16	Iowa	Des Moines
17	Missouri	St. Louis
18	North Dakota	Bismarck
19	South Dakota	Rapid City
20	Nebraska	Omaha
21	Kansas	Wichita
22	Delaware	Wilmington
23	Maryland	Baltimore
24	District of Columbia	District of Columbia
25	Virginia	Norfolk
26	West Virginia	Charleston
27	North Carolina	Charlotte
28	South Carolina	Charleston
29	Georgia	Atlanta
30	Florida	Miami
31	Kentucky	Louisville
32	Tennessee	Memphis
33	Alabama	Birmingham
34	Mississippi	Jackson
35	Arkansas	Little Rock
36	Louisiana	New Orleans
37	Oklahoma	Oklahoma City
38	Texas	Dallas
39	Montana	Billings
40	Idaho	Boise
41	Wyoming	Casper
42	Colorado	Denver
43	New Mexico	Albuquerque
44	Arizona	Phoenix
45	Utah	Salt Lake City
46	Nevada	Las Vegas
47	Washington	Seattle
48	Oregon	Portland
49	California	Los Angeles