Robin Flowerdew and
Andrew Lovett

Fitting Constrained Poisson Regression Models to Interurban Migration Flows

This paper demonstrates the effects of fitting singly and doubly constrained spatial interaction models using the Poisson regression approach. A large data set containing migration flows between labor market areas in Great Britain in 1970–71 is used. The results of fitting unconstrained, singly constrained, and doubly constrained models are compared with respect to goodness of fit and the interpretability of parameter estimates. The addition of other explanatory variables to the model is also explored.

1. INTRODUCTION

During the last thirty years there has been much research effort devoted to modeling interaction between spatially defined units, especially with respect to such topics as migration, commodity flows, journeys to work, and shopping trips. There have been two main approaches to spatial interaction modeling, one based on statistical estimation of flows as a function of various explanatory variables, usually including measures of distance and origin and destination zone size, and one based on a mathematical algorithm that produces the most likely set of flows given various constraints on origin and destination totals. The former is closely based on the Newtonian gravity model, and derives from the work of Stewart (1948) and Olsson (1965), while the latter is particularly associated with the work of Wilson (1970), who justified it in terms of entropy maximization. As usually employed, the former has the problem of giving rather poor fits to observed flows (see Senior 1979), while the latter has the problem of being relatively hard to adapt to incorporate additional explanatory variables.

Flowerdew and Aitkin (1982) argued that the problems of the regression approach to spatial interaction modeling stemmed from an incorrect specification of the model employed; if the dependent variable is a count—the number of trips, migrants, shipments, etc.—a form of regression should be employed based on a discrete distribution such as the Poisson rather than on the normal distribution as in ordinary-least-squares regression. Poisson regression can be carried out

Robin Flowerdew is senior research associate, Institute for Market and Social Analysis, Toronto, Canada, and is on leave from the department of geography, University of Lancaster, England, where Andrew Lovett is teaching fellow.

Geographical Analysis, Vol. 20, No. 4 (October 1988) © 1988 Ohio State University Press
Submitted 11/87. Revised version accepted 4/88.

conveniently within the generalized linear modeling framework using the widely available package GLIM, and Flowerdew and Aitkin were able to show the superiority of Poisson regression for modeling interurban migration over conventional regression models.

Baxter (1982) has shown the similarities between different approaches to spatial interaction modeling, in particular indicating the equivalence of entropy-maximizing models without origin and destination constraints to Poisson regression models using origin total, destination total, and distance as explanatory variables. The parameter estimates derived from the entropy-maximizing approach have maximum likelihood properties and are the same as those derived from a Poisson regression model with the same explanatory variables. Given this equivalence, Poisson regression has the advantages that the statistical properties of the model, including its goodness of fit, are made clear, and that different or additional explanatory variables can easily be fitted.

The equivalence between Poisson regression and entropy-maximizing models still applies where origin and/or destination constraints are employed. Baxter (1984) has indicated how constrained spatial interaction models can be fitted using a Poisson regression approach. In this paper we apply his approach to the interurban migration data studied by Flowerdew and Aitkin (1982), investigating the effectiveness of origin and destination constraints on goodness of fit and on the interpretability of the model. After a discussion of the method and an introduction to the data, the results of fitting unconstrained and constrained models are reported, together with some additional models. Finally the utility of the approach is assessed.

2. POISSON REGRESSION ANALYSIS

Several introductions to Poisson regression analysis are now available, including Flowerdew and Aitkin (1982), Lovett (1984), and Lovett, Whyte, and Whyte (1985), so only a brief discussion will be supplied here. The main characteristics of Poisson regression can be most easily described by referring to the concept of a generalized linear model (Nelder and Wedderburn 1972), where the observed values of the response variable y_i are regarded as values taken on by a random variable Y_i, whose mean μ_i is linked to a linear combination of explanatory variables. Ordinary-least-squares regression is a model in which the random variable Y_i is normally distributed and its mean μ_i is equal to a linear combination of the explanatory variables. In Poisson regression, Y_i is assumed to have a Poisson distribution whose parameter is logarithmically linked to a linear combination of the explanatory variables. Other models, including the binomial logit, can also be fitted.

The GLIM package (Payne 1985) will fit a Poisson regression model to a set of data, generating a set of coefficients for the explanatory variables and a predicted value for each case in the data set. It uses an iteratively reweighted least-squares procedure which, as McCullagh and Nelder (1983) prove, converges to the maximum likelihood solution. The goodness of fit of generalized linear models is assessed on the basis of the log-likelihood ratio statistic, known as the deviance. In ordinary-least-squares regression, this quantity reduces to the residual sum of squares, but this is not true in the Poisson case. Because the Poisson distribution has only one parameter (its mean being equal to its variance), it is possible to test whether the value of the deviance is compatible with the model that has been fitted —in other words, whether the model could have generated the data. The deviance has a distribution approximating to chi-squared with degrees of freedom equal to the number of observations minus the number of parameters fitted (see Payne

1985, p. 111, for comments on the quality of the approximation). This property does not hold for two-parameter distributions, including the normal and the negative binomial, where a high deviance value could arise from misspecification or merely from a large standard error. For the Poisson, however, a high deviance can only arise from misspecification.

GLIM allows for any combination of interval and nominal variables to be used as explanatory variables. Nominal variables are known as factors and adding a factor to a model results in the estimation of a separate parameter for each value of the factor. A special case of Poisson regression is encountered if all explanatory variables are factors; in this situation Poisson regression is equivalent to log-linear modeling of contingency tables and Bowlby and Silk (1982) have demonstrated how GLIM can be used to carry out such analyses. Baxter (1984) shows that fitting an origin-constrained spatial interaction model is equivalent to fitting a factor in a Poisson regression model, where the factor has a different value for each origin zone. Adding this factor to the analysis will result in predicted flows satisfying the origin constraint and in origin-specific coefficients equivalent to the weights produced by an origin-constrained entropy-maximizing model. As Baxter indicates, this can be combined in a Poisson regression model with any other variables thought appropriate. A destination-constrained model can be fitted in an analogous manner, and a doubly constrained model by fitting two factors, one representing origins and one representing destinations.

3. THE DATA SET

The models described are fitted to census data indicating the number of interurban migrants in 1970–71, i.e., people whose 1971 address was in a different standard metropolitan labor area (SMLA) from their 1970 address. SMLAs were defined on the basis of commuter flows in an attempt to approximate as far as possible geographical labor market areas (Drewett et al. 1974), and they differ greatly in both area and population. In 1971 there were 126 SMLAs in Great Britain, and hence a set of 15,750 inter-SMLA flows was identified. Flowerdew and Salt (1979) discuss the data set in more detail. In addition to the number of migrants, population size, distance, and several spatial and socioeconomic variables were available as possible explanatory variables.

Besides the element of convenience, there are two main reasons for using this data set to fit the models discussed above. First, it is possible to compare the results of the analysis with those obtained in earlier studies (Flowerdew and Aitkin 1982; Flowerdew 1982). Second, with such a large data set it should be relatively easy to discriminate between the goodness of fit of different models.

4. RESULTS

In this section, the results of fitting spatial interaction models of various types to the data are described. The unconstrained model, being the simplest, is described first, with an explanation of the results and a discussion of the effects of including several additional variables, representing contiguity, unemployment, and the location of naval bases (which appear to have generated large numbers of migratory moves).

Unconstrained Models

Fitting an unconstrained model is equivalent to a Poisson regression analysis of the number of migrants using origin and destination size and distance as explanatory variables. The results of this analysis are given in Flowerdew and Aitkin (1982), where they are fully discussed, but they will be recapitulated here to aid comparison with the constrained models discussed below. The regression equation estimates the parameter of the Poisson distribution ($\widehat{\lambda_{ij}}$), of which the observed migration flow (y_{ij}) is a realization. Using the logarithms of origin and destination populations ($\ln P_i$ and $\ln P_j$) and the logarithm of distance ($\ln d_{ij}$) as explanatory variables, the following equation was produced:

$$\widehat{\lambda_{ij}} = \exp(\underset{(0.049)}{-\ 14.94} + \underset{(0.002)}{0.954} \ln P_i + \underset{(0.002)}{0.804} \ln P_j - \underset{(0.004)}{1.134} \ln d_{ij}) \quad (1)$$

Standard errors are given in parentheses under the coefficients to which they refer. Values of 1.0 for the coefficients of $\ln P_i$ and $\ln P_j$ would indicate that migration is proportional to population size. All the coefficient values seem reasonably easy to interpret, and all are in accord with traditional gravity model theory. It may be noted that a model using distance rather than the logarithm of distance (equivalent to a negative exponential distance function) gave a considerably worse fit (deviance of 106,600) and this specification was not pursued further.

As stated above, the standard method for assessing the goodness of fit of a Poisson regression model is through the use of the deviance statistic; if the Poisson assumption is correct and the model correctly formulated, the deviance should have an approximate chi-squared distribution with 15,746 degrees of freedom (the number of observations minus four, the number of parameters fitted). This means that the calculated chi-squared value should be less than 16,075 for the model fitted to avoid rejection at the 0.05 significance level. In this case, the deviance value for model (1) was 77,190; clearly this model is a long way from being a satisfactory fit. On the other hand, the deviance of the null model (obtained by using the overall mean $\bar{y}$ as an estimate for y_{ij}) is 368,800, so model (1) can be regarded as accounting for 79 percent of the null deviance.

The standard errors reported for this model, and for subsequent models, are calculated by GLIM on the basis that the Poisson assumption is correct. The unsatisfactory fit may result from a failure to include all relevant explanatory variables or, as argued in section 5 below, from the inadequacy of the Poisson assumption. In the latter case, the standard errors are likely to be underestimated although, under certain conditions, the parameter estimates are valid. Davies and Guy (1987) discuss this issue and suggest how quasi-likelihood methods can be used to adjust the standard errors.

Given this high deviance, the obvious objective of the spatial interaction modeler is to find a regression model which does fit the data to an acceptable degree, or at least represents a significant step in this direction. The main aim of this paper is to assess the results of using a constrained form of the spatial interaction model. Before doing this, however, the effect of adding other possible explanatory variables to the unconstrained version of the model will be investigated. Appropriate variables can be identified by inspection of the residuals.

As shown in Table 2 of Flowerdew and Aitkin (1982), many of the largest residuals represent flows between contiguous SMLAs: these flows may have been underpredicted through the inclusion of relatively short-distance moves across a boundary, treated in this model as if they were between the main population centers of the SMLA. Again, a dummy variable can be introduced which represents the contiguity effect; adding this variable to model (1) produces the following

TABLE 1
Summary of Models Fitted

Model	Variables Included	Deviance
1 Unconstrained	$\ln P_i, \ln P_j, \ln d_{ij}$	77,190
2	$\ln P_i, \ln P_j, \ln d_{ij}, C_{ij}$	72,750
3	$\ln P_i, \ln P_j, \ln d_{ij}, C_{ij}, B_i, B_j, B_{ij}$	69,800
4	$\ln P_i, \ln P_j, \ln d_{ij}, C_{ij}, U_i, U_j, U_{ij}, G_i, G_j, G_{ij}$	71,110
5	$\ln P_i, \ln P_j, \ln d_{ij}, C_{ij}, B_i, B_j, B_{ij}, U_i, U_j, U_{ij}, G_i, G_j, G_{ij}$	65,250
6 Origin-constrained	$\ln P_j, \ln d_{ij}$	64,750
7	$\ln P_j, \ln d_{ij}, B_j, B_{ij}, U_j$	62,990
8	$\ln P_j, \ln d_{ij}, C_{ij}, B_j, B_{ij}, U_j$	60,440
9 Destination-constrained	$\ln P_i, \ln d_{ij}$	61,000
10	$\ln P_i, \ln d_{ij}, B_i, B_{ij}, U_i$	60,120
11	$\ln P_i, \ln d_{ij}, C_{ij}, B_i, B_{ij}, U_i$	56,690
12 Doubly Constrained	$\ln d_{ij}$	53,360
13	$\ln d_{ij}, C_{ij}, B_{ij}$	50,110

NOTE: See the text for definitions of variables and further explanation of the results.

regression equation:

$$\widehat{\lambda_{ij}} = \exp(\underset{(0.051)}{-15.54} + \underset{(0.002)}{0.925} \ln P_i + \underset{(0.002)}{0.784} \ln P_j - \underset{(0.005)}{0.894} \ln d_{ij} + \underset{(0.011)}{0.728}\, C_{ij}) \quad (2)$$

where C_{ij} has the value 1 if SMLAs i and j are contiguous, and 0 otherwise. This model has a deviance of 72,750 with 15,745 degrees of freedom (to aid comparisons, Table 1 shows the deviance obtained for this and all other models fitted). Clearly, the model is still far from satisfactory, but the deviance has been reduced by 4,440 for the loss of 1 degree of freedom (a very significant reduction). It may be noted that the population coefficients are fairly similar to those in model (1) but the distance coefficient has been reduced substantially. This is because of the relationship between distance and contiguity; including the contiguity variable increases the effect of flows between noncontiguous SMLAs on the estimation of the distance parameter.

Other large residuals include many flows involving the SMLAs of Plymouth, Portsmouth, and Dunfermline (which includes Rosyth). This suggests that a "naval-base" effect may be important. This effect can be modeled using dummy variables, which are treated as factors in GLIM; three dummy variables were used, one picking out flows that originated from a naval base, one for flows to a naval base, and one for flows between naval bases. Adding these three variables to model (2) produces a further regression equation:

$$\widehat{\lambda_{ij}} = \exp(\underset{(0.053)}{-15.62} + \underset{(0.002)}{0.933} \ln P_i + \underset{(0.002)}{0.787} \ln P_j - \underset{(0.008)}{0.918} \ln d_{ij} + \underset{(0.011)}{0.730}\, C_{ij} + \underset{(0.019)}{0.797}\, B_i + \underset{(0.018)}{0.807}\, B_j + \underset{(0.008)}{0.028}\, B_{ij}) \quad (3)$$

where B_i has the value 1 for flows from one of the three naval bases and 0 otherwise, B_j has the value 1 for flows to a naval base and 0 otherwise, and B_{ij} has the value 1 only for flows between naval bases. This shows the existence of significant increases in migration to and from naval bases. The deviance is reduced to 69,800, a reduction of 2,950 from model (2) for the loss of 3 degrees of freedom. Again the reduction is highly significant, but the deviance value remains well in excess of the critical value.

It is standard in the economic theory of migration to relate migrant flows to employment variables, although empirical studies of gross migration have often failed to find convincing evidence for such links (see Shaw 1975). Data were obtained from the Department of Employment on unemployment rates and unemployment change for local areas and adjusted to conform to SMLA units. Six variables were tried, all in logarithmic form: the origin unemployment rate in 1971 (U_i), the destination unemployment rate in 1971 (U_j), the difference between the two (U_{ij}), percentage unemployment growth 1970–71 at the origin (G_i), percentage growth at the destination (G_j), and the difference between the two (G_{ij}).

Adding these variables to model (1) produces the following result:

$$\begin{aligned}\widehat{\lambda_{ij}} = \exp(&\underset{(0.060)}{-\ 14.72} + \underset{(0.002)}{0.933}\ \ln P_i + \underset{(0.002)}{0.785}\ \ln P_j - \underset{(0.004)}{1.159}\ \ln d_{ij}\\ &+ \underset{(0.009)}{0.010}\ U_i - \underset{(0.009)}{0.279}\ U_j + \underset{(0.002)}{0.020}\ U_{ij} - \underset{(0.021)}{0.790}\ G_i\\ &- \underset{(0.021)}{0.938}\ G_j - \underset{(0.003)}{0.008}\ G_{ij}).\end{aligned} \tag{4}$$

The deviance is 71,110, a reduction of 6,080 from model (1) for the loss of 6 degrees of freedom, a significant reduction, but perhaps a disappointingly small one considering the importance of employment considerations in migration theory. The effects of unemployment conditions at the destination are as might be expected: the coefficients are significant and negative. Origin conditions are less clear-cut, however. The coefficient of origin unemployment is positive (i.e., there are more out-migrants from high unemployment areas) but is not significantly different from zero, and the coefficient of origin unemployment growth is significant and negative (areas of high unemployment growth have fewer out-migrants). The latter finding is in conflict with classical economic theory but is in accord with earlier findings (e.g., Gleave and Cordey-Hayes 1977). Again, although the sign of U_{ij} is as expected, that of G_{ij} is negative (i.e., the greater the difference in unemployment growth between places i and j, the fewer migrants from i to j).

Adding the contiguity and naval-base variables to model (4) gives a further variation on the unconstrained Poisson regression model in which the deviance is reduced to 65,250 with 15,736 degrees of freedom—another significant reduction, but still a very long way from being an adequate fit. The equation is

$$\begin{aligned}\widehat{\lambda_{ij}} = \exp(&\underset{(0.064)}{-\ 15.34} + \underset{(0.002)}{0.914}\ \ln P_i + \underset{(0.002)}{0.767}\ \ln P_j - \underset{(0.005)}{0.949}\ \ln d_{ij}\\ &+ \underset{(0.011)}{0.681}\ C_{ij} + \underset{(0.020)}{0.662}\ B_i + \underset{(0.019)}{0.666}\ B_j + \underset{(0.008)}{0.011}\ B_{ij} + \underset{(0.009)}{0.015}\ U_i\\ &- \underset{(0.009)}{0.252}\ U_j + \underset{(0.002)}{0.019}\ U_{ij} - \underset{(0.021)}{0.663}\ G_i - \underset{(0.021)}{0.827}\ G_j\\ &- \underset{(0.003)}{0.020}\ G_{ij}).\end{aligned} \tag{5}$$

The coefficients of U_i and B_{ij} are not significantly different from zero in this model.

Origin-constrained Models

Using the methods outlined above, an origin-constrained model was fitted to the data. In its basic form, this involved fitting a factor representing the 126 origins and variables representing destination size and distance. An origin-constrained model assumes that the total number of out-migrants leaving each origin is fixed. The total arriving at each destination is not fixed, however, and is presumably responsive to the attractiveness of the destination. It seems more natural, therefore, to use $\ln P_j$, the logarithm of destination population, as an explanatory variable, rather than $\ln D_j$, the logarithm of the total number of in-migrants to place j. Fitting the origin factor, $\ln P_j$, and $\ln d_{ij}$ results in a model with a deviance of 64,750 with 15,622 degrees of freedom, a reduction of 12,400 from model (1); the origin factor is responsible for 126 degrees of freedom. In this model, 82 percent of the deviance of the null model is accounted for.

This model is equivalent to a standard origin-constrained spatial interaction model; a comparison of the deviance with that of the unconstrained model allows a quantitative assessment to be made of the effects on goodness of fit of introducing the constraint. As suggested earlier, it is easy within the Poisson regression framework to introduce new explanatory variables, and the next step is therefore to fit an origin-constrained model introducing some of the variables discussed above as additional explanatory variables.

The origin constraint fits one unique term for each origin; if it is included in a regression model, it will incorporate the effects of all origin-specific influences on migration. There is therefore no point in including variables describing the origin places. In view of the results of fitting the unconstrained models, the variables B_j, B_{ij}, and U_j were included in an origin-constrained model. This results in a reduction of deviance by a further 1,850 to 62,990 with 15,619 degrees of freedom. When the contiguity variable C_{ij} is also included, deviance is reduced to 60,440. Excluding the origin factor, the coefficient values fitted are as follows:

Variable	Parameter	Standard Error
$\ln P_j$	0.780	0.002
$\ln d_{ij}$	− 1.024	0.005
C_{ij}	0.572	0.011
B_j	0.739	0.019
B_{ij}	0.610	0.063
U_j	− 0.176	0.008

In addition to these coefficients, 126 values are generated representing the effects of each origin: these incorporate the effects of origin size, and range from positive values for the largest places (3.449 for London; 1.687 for Birmingham) to negative values for those places with the fewest out-migrants (− 2.183 for Rhondda; − 1.931 for Ellesmere Port).

These origin parameters are related, but not identical, to the balancing factors produced by spatial interaction models of the type developed by Wilson (1970). It is standard to fit origin-constrained models using a formula such as

$$\widehat{I_{ij}} = A_i O_i D_j d_{ij}^{\beta} \tag{6}$$

where $\widehat{I_{ij}}$ is the estimated migration between i and j, A_i is an origin-specific balancing factor, and O_i is the total number of out-migrants from i. The equivalent formulation in Poisson regression is

$$\widehat{I_{ij}} = \exp(\beta_i + \beta_1 \ln D_j + \beta_2 \ln d_{ij}) \tag{7}$$

where β_i is the origin parameter and β_1 is constrained to be 1. It can be seen, therefore, that β_i is equal to the natural logarithm of $A_i O_i$. It is not surprising, therefore, that the values of the origin parameters appear to be related to origin size. They presumably also reflect other variables affecting the number of out-migrants from 1 and its accessibility to the other places in the data set, but they cannot be identified precisely with any set of exogenously measurable variables. The destination parameters discussed later have an analogous relationship to the balancing factors B_j that arise in a standard destination-constrained spatial interaction model; the natural logarithm of $B_j D_j$ is equal to the destination parameter for place j derived from the Poisson model.

Evaluation of the relative worth of the origin-constrained model must take account of both the improvement in goodness of fit and the extra parameters that must be estimated. It is clear from these results that the improvement in fit is significant, despite the loss in degrees of freedom. Problems do arise, however, in interpreting the origin parameters, as in any constrained spatial interaction model. The main criterion in comparing the models should be the appropriateness or otherwise of the model: for some applications of spatial interaction modeling, such as journey to work, an origin-constrained model is clearly appropriate; for migration, cases can be made either way.

Destination-constrained Models

In a similar fashion, a series of destination-constrained models was fitted to the data, using a factor to represent destination-specific effects for all 126 SMLAs. The initial model also included $\ln P_i$ and $\ln d_{ij}$ as explanatory variables. It produced a deviance value of 61,000, 16,190 less than that of the unconstrained model (1), and also considerably less than that of the corresponding origin-constrained model. It is perhaps not surprising that the destination-constrained version should perform better than the origin-constrained version on a model calibrated with population size data: certain places, such as New Towns and suburban areas, grew much faster than others, and so destination population is not as closely related to in-migration as origin population is to out-migration.

Because the destination factor subsumes the effects of all destination-specific variables, there is no point in adding other variables of this type to the model; however, origin-specific variables may be added. Introducing the variables B_i, B_{ij} and U_i produces a reduction in deviance to 60,120—in this model, the coefficient of B_{ij} is not significantly different from zero. The contiguity variable reduces the deviance to 56,690. In this model, the coefficients of the explanatory variables are as follows:

Variable	Parameter	Standard Error
$\ln P_i$	0.937	0.002
$\ln d_{ij}$	− 1.014	0.006
C_{ij}	0.682	0.012
B_i	0.663	0.020
B_{ij}	0.004	0.008
U_i	0.029	0.008

Again, the coefficients of the destination factor reflect size, ranging from 2.602 (London) and 1.142 (Birmingham) to − 2.496 (Rhondda) and − 2.190 (Hartlepool).

Doubly Constrained Models

A doubly constrained model involves fitting two factors—one origin-related and one destination-related—together with variables concerned with the interaction of origin and destination. The basic form of the model involves fitting the two factors and a distance variable; using $\ln d_{ij}$ produces a model with a deviance of 53,360 with 15,498 degrees of freedom. Again, there is a highly significant reduction from the unconstrained model (23,830) and indeed from each of the singly constrained models.

No further variables can be fitted referring to either the origin or the destination alone; however, the contiguity variable C_{ij} and the naval-base-interaction variable B_{ij} can be included. Doing so produces a model with a deviance of 50,110 on 15,496 degrees of freedom. The coefficients of the explanatory variables are

Variable	Parameter	Standard Error
$\ln d_{ij}$	− 1.071	0.006
C_{ij}	0.687	0.012
B_{ij}	0.609	0.064

The coefficients of the origin factor range from 3.234 (London) and 1.624 (Manchester) to − 2.155 (Rhondda) and − 1.938 (St. Albans), and the coefficients of the destination factor, covering a similar range of values, run from 2.693 (London) and 1.213 (Birmingham) to − 2.458 (Rhondda) and − 2.196 (Hartlepool).

5. CONCLUSIONS ABOUT THE DATA

As stated above, the use of Poisson regression, based as it is on a probability distribution with only one parameter, allows reasonably unambiguous conclusions to be drawn concerning the fit of model to data. It is clear that none of the models fitted provides anything like an adequate fit to the data. However, all the models described do account for very substantial proportions of the deviance of a null model, and there are substantial improvements introduced by the imposition of constraints and by the use of additional explanatory variables; size and distance do allow reasonably good estimates of migration to be made, and the other variables studied do help to improve these estimates (though the unemployment variables contribute relatively little considering their traditional theoretical importance).

The failure of the models to fit may be due to the inapplicability of the simple Poisson distribution. A Poisson distribution is derived from the assumption that each individual has an equal and independent probability of moving from i to j, but in practice migrants move as households, not as completely independent individuals. A generalized Poisson distribution may therefore be more appropriate, in which the movement of households follows a Poisson distribution, but the size of each household follows a known household size distribution. The coefficients and flows estimated in this model are identical to those for the Poisson, but the method of calculating deviance is different and thus different conclusions are to be drawn about the model's goodness of fit. Fitting a generalized Poisson distribution of this type, unconstrained and using only population and distance as explanatory variables, produces a much lower deviance (21,746), only some 40 percent in excess of the critical value at the 0.05 significance level.

There may be many processes, like this one, where a generalized or compound Poisson distribution is more theoretically attractive than an ordinary Poisson model. As Davies and Guy (1987) show, the parameter estimates derived from Poisson regression are consistent estimates of the parameters for these processes. The standard errors of these parameters, however, will be underestimated. Even if the nature of the probability process is not known, quasi-likelihood or pseudo-likelihood methods (Davies and Guy 1987, pp. 306–10) can be used to calculate standard errors. In any case, the Poisson parameter estimates can still be used even if the process is not strictly Poisson.

Returning to the Poisson models discussed above, it is clear that the constrained regression models produce better fits than the unconstrained models. They raise problems of interpretation, however; the fact that each place has a different and measurable effect on migration, whether as origin or destination, may not be especially helpful unless the differences in these effects can themselves be understood and modeled. It is also clear that adding additional explanatory variables to the constrained models does improve goodness of fit.

6. CONCLUSIONS ABOUT THE METHODS

This example illustrates that the approach described can contribute to both the approaches to spatial interaction modeling outlined in the introduction. The statistical approach can benefit from the introduction of constrained models, where appropriate, which may greatly improve the goodness of fit, and the entropy-maximizing approach can benefit from the information the deviance statistic provides concerning goodness of fit and from the ability to experiment with the addition of extra explanatory variables. As shown above, the addition of other variables besides size and distance improved the goodness of fit whether the model was constrained or not.

In summary, this paper has demonstrated how the Poisson regression approach can be extended to fit singly or doubly constrained spatial interaction models in the context of a large data set on British interurban migration. The approach allows the evaluation of goodness of fit for these constrained models and the addition of new variables to the basic size and distance framework.

LITERATURE CITED

Baxter, M. (1982). "Similarities in Methods of Estimating Spatial Interaction Models." *Geographical Analysis* 14, 267–72.

______ (1984). "A Note on the Estimation of a Nonlinear Migration Model Using GLIM." *Geographical Analysis* 16, 282–86.

Bowlby, S., and J. Silk (1982). "Analysis of Qualitative Data Using GLIM: Two Examples Based on Shopping Survey Data." *Professional Geographer* 34, 80–90.

Davies, R. B., and C. M. Guy (1987). "The Statistical Modeling of Flow Data When the Poisson Assumption Is Violated." *Geographical Analysis* 19, 300–14.

Drewett, R., J. Goddard, N. Spence, C. Connock, and R. Pinkham (1974). "Urban Change in Britain: 1961–71." *Working Reports* No. 1 and 2, Department of Geography, London School of Economics.

Flowerdew, R. (1982). "Fitting the Lognormal Gravity Model to Heteroscedastic Data." *Geographical Analysis* 14, 263–67.

Flowerdew, R., and M. Aitkin (1982). "A Method of Fitting the Gravity Model Based on the Poisson Distribution." *Journal of Regional Science* 22, 191–202.

Flowerdew, R., and J. Salt (1979). "Migration between Labour Market Areas in Great Britain, 1970–1971." *Regional Studies* 13, 211–31.

Gleave, D., and M. Cordey-Hayes (1977). "Migration Dynamics and Labour Market Turnover." *Progress in Planning* 8, 1–95.

Lovett, A. A. (1984). "Poisson Regression Using the GLIM Package." *Computer Package Guide* No. 5, Department of Geography, University of Lancaster.

Lovett, A. A., I. D. Whyte, and K. A. Whyte (1985). "Poisson Regression Analysis and Migration Fields: The Example of the Apprenticeship Records of Edinburgh in the Seventeenth and Eighteenth Centuries." *Transactions, Institute of British Geographers* 10, 317–32.

McCullagh, P., and J. A. Nelder (1983). *Generalized Linear Models*. London: Chapman and Hall.

Nelder, J. A., and R. W. M. Wedderburn (1972). "Generalised Linear Models." *Journal of the Royal Statistical Society A* 135, 370–84.

Olsson, G. (1965). "Distance and Human Interaction: a Migration Study." *Geografiska Annaler B* 47, 3–43.

Payne, C. D., ed. (1985). *The GLIM System Release 3.77: Manual*. Oxford: Numerical Algorithms Group.

Senior, M. L. (1979). "From Gravity Modelling to Entropy Maximizing: a Pedagogic Guide." *Progress in Human Geography* 3, 175–210.

Shaw, R. P. (1975). *Migration Theory and Fact: A Review and Bibliography of Current Literature*. Bibliography Series No. 5, Regional Science Research Institute, Philadelphia.

Stewart, J. Q. (1948). "Demographic Gravitation: Evidence and Applications." *Sociometry* 11, 31–58.

Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. London: Pion.