Environment and Planning A, 1983, volume 15, pages 15-36

A new set of spatial-interaction models: the theory of competing destinations†

A S Fotheringham
Department of Geography, Indiana University, Bloomington, IN 47405, USA
Received 17 August 1981, in revised form 6 January 1982

Abstract. Members of the family of spatial-interaction models commonly referred to as gravity models are shown to be misspecified. One result of this misspecification is the occurrence of an undesirable 'spatial-structure effect' in estimated distance-decay parameters and this effect is examined in detail. An alternative set of spatial-interaction models is formulated from which more accurate predictions of interactions and more accurate parameter estimates can be obtained. These new interaction models are termed competing destinations models, and estimated distance-decay parameters obtained in their calibration are shown to have a purely behavioural interpretation. The implications of gravity-model misspecification are discussed.

1 Distance-decay parameters and gravity models

A gravity model is defined here to be any aggregate spatial-interaction model in which interaction volume is a function of three variables: nodal propulsiveness, nodal attractiveness, and the cost of overcoming the spatial separation of nodes (frequently measured by distance). Discussion of these models deals primarily with estimates of distance-decay parameters obtained when the models are calibrated, since the spatial pattern of these estimates can be used to indicate misspecification within a particular model (Fotheringham, 1981).

When a distance-decay parameter is estimated in the calibration of a gravity model, the usual interpretation of the estimate is that, ceteris paribus, it describes the relationship between the observed interaction pattern and distance. An origin-specific distance-decay parameter has the same interpretation for an interaction pattern which is specific to one origin. However, there is fairly strong empirical evidence to suggest that this interpretation is a false one (Fotheringham, 1981). In particular, when a set of origin-specific distance-decay parameters is mapped, there is usually a marked spatial pattern inherent in the data: accessible origins have less-negative estimates and inaccessible origins have more-negative estimates(1). Examples of such a pattern occur frequently (for instance, Gould, 1975; Chisholm and O'Sullivan, 1973; Leinbach, 1973; Stilwell, 1978), which suggests that the estimates are a function of spatial structure, where spatial structure is defined as the configuration of origins and destinations within a spatial system. In fact, since the evidence for this 'spatial-structure effect' is derived from the spatial pattern of origin-specific estimates, it must be the configuration of destinations around an origin that determines, in part, the estimated distance-decay parameter. Although this spatial-structure effect has long been suspected (Porter, 1956) and several attempts have been made to identify it (Curry, 1972; Curry et al, 1975; Sheppard et al, 1976; Sheppard, 1979b; Johnston, 1973; 1975; 1976; Fotheringham and Webber, 1980), it has never been convincingly demonstrated (Fotheringham, 1981) nor, consequently, commonly accepted.

† The contents of this paper were presented in the Nystrom Award Session at the Seventy-seventh Annual Meeting of the Association of American Geographers, Los Angeles, April 1981 where the author was a joint winner of the Warren G Nystrom Award.

(1) It is assumed that in the precalibrated model formulation, the distance-decay parameter is written with a positive sign.

Its apparent presence, however, is disconcerting since it is a symptom of malaise within the gravity-modelling procedure.

The gravity models investigated here are the constrained members of the traditional 'family' of gravity models outlined by Wilson (1971). They are written below in the form which will be later calibrated, that is, in their origin-specific form[2]. Consider a spatial system consisting of m origins each denoted by i and n destinations each denoted by j. The following gravity models can then be described[3]:

(a) *An origin-specific production-constrained gravity model* is given by

$$I_{ij} = Z_i O_i m_j d_{ij}^{\beta_i} , \tag{1}$$

where

$$Z_i = \left(\sum_{j=1}^{n} m_j d_{ij}^{\beta_i} \right)^{-1} . \tag{2}$$

I_{ij} represents the interaction volume between i and j; m_j represents the attractiveness of destination j; d_{ij} represents the distance (or cost of the interaction) between i and j; O_i represents the known total outflow from origin i; and Z_i is a balancing factor which ensures that the constraint,

$$\sum_{j=1}^{n} \hat{I}_{ij} = \sum_{j=1}^{n} I_{ij} , \qquad \forall i , \tag{3}$$

is met. In words, equation (3) states that for each origin, the predicted total outflow is equal to the known total outflow. The symbol '^' is used throughout this paper to denote an estimated value.

(b) *An origin-specific doubly constrained gravity model* is given by

$$I_{ij} = Z_i O_i B_j D_j d_{ij}^{\beta_i} , \tag{4}$$

where

$$Z_i = \left(\sum_{j=1}^{n} B_j D_j d_{ij}^{\beta_i} \right)^{-1} , \tag{5}$$

and

$$B_j = \left(\sum_{i=1}^{m} Z_i O_i d_{ij}^{\beta_i} \right)^{-1} . \tag{6}$$

D_j is the known total inflow into destination j and B_j is a balancing factor which ensures that the constraint

$$\sum_{i=1}^{m} \hat{I}_{ij} = \sum_{i=1}^{m} I_{ij} , \qquad \forall j , \tag{7}$$

is met. Equation (7) states that for each destination the predicted total inflow is equal to the known total inflow. The balancing factor Z_i again ensures that the constraint in equation (3) is met.

There are several reasons why the above models are written in terms of a power function of distance, rather than, for example, an exponential or Tanner function or even a general distance function. One reason is that an empirical analysis of these

[2] Throughout this paper, the discussion is in terms of origin-specific models merely because the effects of spatial structure are most clearly seen in the parameter estimates from such models. The conclusions reached, however, are applicable to the more usual form of the models, that is, the models where one 'average' parameter is derived for a system consisting of many origins and many destinations.

[3] An origin-specific attraction-constrained gravity model is a trivial model and is not discussed here. For a given origin, there would be only one flow terminating at each destination and such a model would merely constrain each predicted interaction to be equal to the known interaction.

models is undertaken with the use of interurban interaction data, and it is assumed that the power function is a more accurate description of the perception of distances at this scale than is the exponential function. Support for this assumption is given by Frost (1969), Hyman (1969), Wilson (1970a), Gordon (1976), and Stillwell (1977). Although the Tanner function may also be appropriate, the interpretation of its two parameters is not always obvious. A second reason for the use of the power function is that references will be made to earlier empirical findings regarding distance-decay parameters and, as Black and Larsen (1972) and Taylor (1975) point out, the predominant distance function used in such studies has been the power function. Third, in the subsequent empirical analysis much emphasis will be placed on the spatial variation of the estimated distance-decay parameters. This variation is more visible when the estimates are derived from a power function rather than from an exponential function simply because of the nature of the two functions. Thus, for consistency, a power function is utilised throughout this study, but the conclusions reached regarding the distance-decay parameter apply to any distance function.

2 Interaction prediction from gravity models

To consider any possible misinterpretation of the parameter estimates obtained in the calibration of the models described in equations (1) and (4), it is useful to consider the behaviour of the models in the two different spatial systems shown in figure 1. In this figure, $O_1 = O_2$ and m_j is a constant for all j. It is assumed that the perception of distance as a deterrent to interaction is equal for the residents of both origins, that is, $\beta_1 = \beta_2$. From a production-constrained gravity model, the ratio of predicted interactions from origin i to two destinations, j and k, is given by

$$\frac{\hat{I}_{ij}}{\hat{I}_{ik}} = \left(\frac{d_{ij}}{d_{ik}}\right)^{\hat{\beta}_i} . \tag{8}$$

Whence, if $d_{ij} < d_{ik}$ and $\hat{\beta}_i < 0$, $\hat{I}_{ij} > \hat{I}_{ik}$. The result given in equation (8) is independent of the distribution of destinations if $\hat{\beta}_i$ is independent of spatial structure. Thus, in figures 1(a) and (b), where the spatial structure of destinations is very different, the ratios $\hat{I}_{1j}/\hat{I}_{1k}$ and $\hat{I}_{2j}/\hat{I}_{2k}$ are identical. For example, the ratio of the predicted volume of interaction terminating at D_1 to the amount terminating at D_2 is $(\frac{1}{2})^{\hat{\beta}_i}$ and the ratio of the predicted volume of interaction terminating at D_2 to the amount terminating at D_3 is $(\frac{2}{3})^{\hat{\beta}_i}$. Thus, the production-constrained gravity model can only predict variation in any $\hat{I}_{ij}/\hat{I}_{ik}$ ratio if $\hat{\beta}_i$ is allowed to vary, wherein lies the paradox of this model. Intuitively, because of the different pattern of opportunities in figures 1(a) and (b), it is very probable that actual interaction ratios, such as

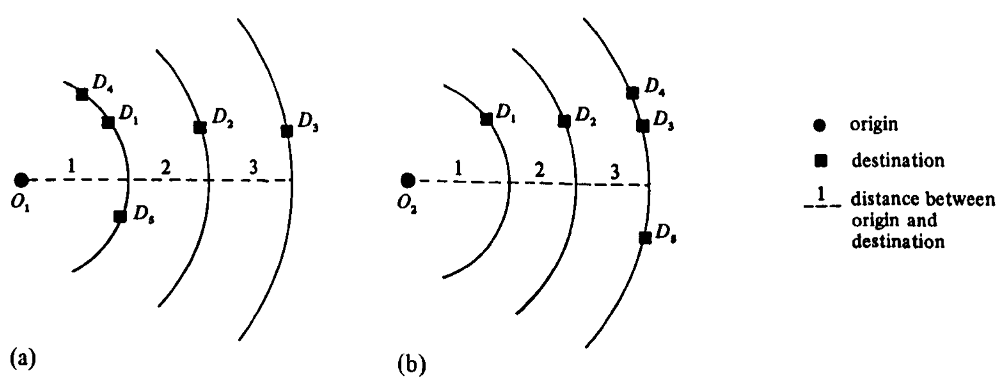


Figure 1. Scenarios of spatial structure: (a) an accessible origin, and (b) an inaccessible origin.

I_{12}/I_{13} and I_{22}/I_{23}, for example, vary[4]. Yet in order to model such variation, $\hat{\beta}_1$ cannot be equal to $\hat{\beta}_2$, although it has been stated that $\beta_1 = \beta_2$. Thus, the production-constrained gravity model is only capable of producing purely behavioural estimates of distance-decay parameters at the expense of accuracy in the replication of known interaction patterns. Since accurate replication is an important component of model calibration and the true distance-decay parameters are unknown, it is the former desirable property that is sacrificed.

The above results do not apply in the same way to the doubly constrained gravity model. It can be shown that from this latter model, the ratio of predicted interactions to two destinations, j and k, whose inflows are equal, is given by[5]

$$\frac{\hat{I}_{ij}}{\hat{I}_{ik}} = \frac{B_j}{B_k}\left(\frac{d_{ij}}{d_{ik}}\right)^{\hat{\beta}_i} . \tag{9}$$

The balancing factor, B_j, can be interpreted as a measure of the inaccessibility of destination j to all origins in the system (Thomas, 1977) and hence B_j and B_k will vary when the configuration of centres varies. The nature of this variation can be described with reference to figure 1 where the accessibility of D_1 with respect to the accessibility of D_2 is greater in figure 1(a) than in figure 1(b). If B_j measures the inaccessibility of a destination to all origins, and if every centre is both an origin and a destination, the ratio B_1/B_2 will be lower for the spatial system in figure 1(a) than for the spatial system in figure 1(b). Consequently, if $\hat{\beta}_i$ is constant for all i, the predicted interaction ratios will vary between origins in different spatial systems and this variation will be entirely due to variations in spatial structure. For example, given the above interpretation of B_j, it is clear from equation (9) that $\hat{I}_{11}/\hat{I}_{12} < \hat{I}_{21}/\hat{I}_{22}$ and $\hat{I}_{12}/\hat{I}_{13} < \hat{I}_{22}/\hat{I}_{23}$. The result from the production-constrained gravity model states that these ratios are equal if $\hat{\beta}_1 = \hat{\beta}_2$ and variation in the ratios is only possible if $\hat{\beta}_1 \neq \hat{\beta}_2$. Hence, in production-constrained gravity models, variations in spatial structure impose an artificial variation in $\{\hat{\beta}_i\}$, the set of origin-specific estimated distance-decay parameters. In the doubly constrained model, variations in spatial structure can, under certain conditions, be accounted for in the balancing factor B_j rather than in the distance-decay parameter. Similar results to those obtained for the doubly constrained gravity model would be obtained for an attraction-constrained gravity model applied to a spatial system with more than one origin. The reason for the similarity is the inclusion in both models of a balancing factor B_j which measures the accessibility of a destination to all origins in the interaction system under investigation.

The above discussion is based on a certain amount of speculation. It is necessary to assume that the ratios of known interactions will behave in the manner described. For instance, it is necessary to assume in figure 1 that $I_{22} < I_{21}$ and $I_{13} > I_{23}$. From the present understanding of interaction decisionmaking behaviour (as summarised in the first sentence of section 1), there is no reason to suspect that these relationships exist. However, a theory of interaction decisionmaking is presented below which, if valid, verifies these relationships and invalidates the gravity-modelling procedure.

3 The theory of competing destinations

Consider individuals in an origin who have already made a decision to interact with other centres, but whose destinations are unknown. Present interaction models imply that the destinations chosen by individuals result from a single decisionmaking process.

[4] The notation of I_{ij} is slightly unusual here. I_{22}, for example, refers to the interaction from origin 2 to destination 2 in figure 1(b). It does not refer to interactions beginning and terminating within the same centre.

[5] The assumptions are made that there is more than one origin interacting with each destination and that D_j is a constant for all j.

The expected net utility derived from interacting with any particular destination (in terms of the attractiveness of the destination and its distance from the origin) is compared to the expected utilities for all other destinations and interactions are predicted on the basis of this comparison. For example, if a destination is relatively unattractive and peripheral to the origin, relatively few individuals will be predicted to interact with that destination.

However, many types of interaction can be considered to be a result of a two-stage decisionmaking process. The first stage is that individuals choose a broad region with which to interact. The second stage is that individuals then choose a specific destination from the set of destinations contained within the broad region. Consider the decision to migrate in search of employment. An individual in a region of high unemployment will be aware that there are other regions which hold better prospects of employment. Once the individual has decided to move, his first locational decision is to choose one of these broad regions. The second decision is to choose a specific city within the broad region. For example, an unemployed individual in the North East of England who has made a decision to move may be aware that better employment opportunities exist elsewhere in the United Kingdom, such as in the Midlands, the South West, the South East, etc. The individual first chooses one of these broad regions in which to concentrate his search for work and then chooses a specific destination within that region. Another example of the two-stage decisionmaking process is in choosing a vacation destination. Assume that an individual maximises the distance from his origin subject to some budget constraint. This results in an annulus of destinations being considered. For example, in the United States, an individual in Indianapolis may decide to vacation in New England, Florida, California, or the Pacific North West. His first locational decision is to choose one of these broad regions; the second locational decision is to choose a specific destination within the chosen region.

The above two examples of the two-stage decisionmaking process refer to interurban interaction. The theory may also apply to intraurban interaction. Consider the primary type of intraurban interaction—the journey-to-work. An individual's workplace is fixed: the decision which determines his journey-to-work is where to locate his home. The individual is likely to have predetermined criteria which his journey-to-work and housing must meet: for example, he may wish to be within a walking distance of twenty minutes of his place of employment or he may prefer to substitute travel convenience for more land by living in suburban areas or outside the city. In either case, the first decision made is that of choosing a particular ring of locations from his workplace. The second decision made is choosing a specific location within the ring of possible locations.

The implication of interactions being a result of a two-stage decisionmaking process, and not a one-stage process, can be seen with reference to the spatial systems given in figure 1. In these special cases, where all destinations lie at regularly spaced intervals from the origin, each ring may be considered a 'macrodestination'. Assume that people's preferences regarding these macrodestinations are the same in origins 1 and 2 [6]. Then, the proportion of the total outflow from each origin terminating at a given distance will be constant for the two origins. As the spatial structure of destinations varies, however, the volume of interaction terminating at a 'microdestination', one of the

(6) This assumption implies that competition between destinations is at a maximum. The increase in the attractiveness of a macrodestination resulting from the addition of a microdestination is completely offset by the increase in competition between the microdestinations. Competition effects are unlikely to be so extreme, but the assumption of maximum competition effects is unnecessary; it simply serves to illustrate the concept. Any degree of competition between destinations will produce the effects described herein; obviously, the intensity of the effects will vary with the intensity of competition between destinations.

individual destinations, must vary according to how many other destinations there are at that same distance. To demonstrate the relevance of this type of interaction decisionmaking behaviour to distance-decay parameters and gravity models, consider the following example. In figure 1, assume that the proportions of the total outflow from each origin terminating at destinations one, two, and three units away are x, y, and z, respectively. If O_i is the total outflow from origin i, the number of interactions over these distances are xO_i, yO_i, and zO_i, respectively. In figure 1(a), the number of interactions to specific destinations are then given by $I_{11} = I_{14} = I_{15} = xO_i/3$, $I_{12} = yO_i$, $I_{13} = zO_i$. In figure 1(b), the interactions are given by $I_{21} = xO_i$, $I_{22} = yO_i$, $I_{23} = I_{24} = I_{25} = zO_i/3$. Both sets of interactions are graphed against distance in figure 2. The principle which operates but which is ignored in gravity modelling is that *the more accessible a destination is to all other destinations in a spatial system, the less likely it is that that destination is a terminating point for interaction from any given origin, ceteris paribus.*

The existence of such a principle has the following implication for distance-decay parameters estimated from gravity models. Consider the population distance-decay parameter, which an attempt is being made to estimate, and which is equal for all origins; and consider also the predicted interactions from a gravity model with a distance-decay parameter equal to this actual value. For an origin which is accessible to destinations, there will be many destinations in close proximity to the origin and each of these destinations will be accessible with respect to all other destinations. Consequently, the actual volume of interaction terminating at any of these destinations in close proximity to the origin will be smaller than that predicted by a gravity model. If an origin is inaccessible to destinations, there will be few destinations in close proximity and the actual interaction terminating at each specific destination will be larger than that predicted by a gravity model. Conversely, there will be few destinations very distant from an accessible origin (the inaccessible centres) and actual interaction volumes to these destinations will be larger than those predicted by a gravity model: there will be more destinations which are very distant from an inaccessible origin (the accessible centres) and actual interaction volumes to specific destinations at these distances will be smaller than those predicted by a gravity model. Consequently the gravity interaction-distance relationship will be steeper than the actual relationship for accessible origins and shallower for inaccessible origins. For the gravity model to replicate reality accurately, the distance-decay parameter for the gravity model must be less negative than the actual value for accessible origins and more negative than the actual value for inaccessible origins. This pattern is also demonstrated in figure 2 where it is clear that in a gravity model in which interactions are simply a function of attractiveness and distance, $\hat{\beta}_1$ and $\hat{\beta}_2$ represent the slopes of the lines in figures 2(a) and (b), respectively. Consequently, $|\hat{\beta}_1| < |\hat{\beta}_2|$ even though $\beta_1 = \beta_2$. Obviously, there is another variable, destination accessibility, which

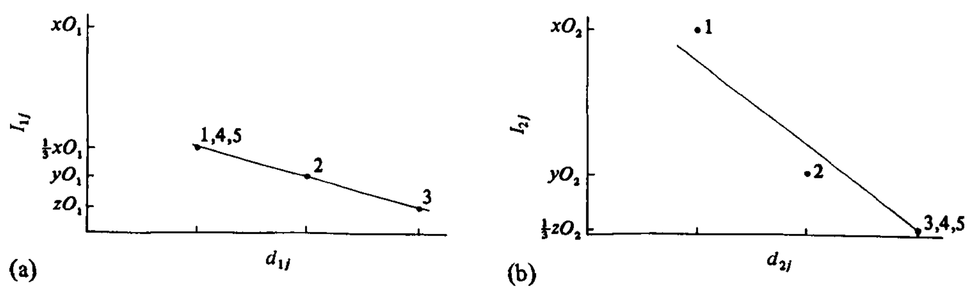


Figure 2: (a) and (b) give interaction-distance relationships for the spatial systems of figures 1(a) and (b), respectively. In figure 1(a), origin 1 is accessible; and in figure 1(b), origin 2 is inaccessible.

determines interaction patterns, but which is not included in the gravity model and which is not included in figure 2. Correct estimates of β_i are only obtained when this variable is accounted for.

The reason why several studies (Linneman, 1966; Frost, 1969; Fotheringham, 1981) report positive estimates of distance-decay parameters can be seen from figures 1(a) and 2(a). If $xO_i/p < zO_i/q$, where p and q are the number of destinations at one and three units from i respectively, $\hat{\beta}_i$ will be positive. Since the inequality will only occur when p is large compared to q, origin i would have to be very accessible for $\hat{\beta}_i$ to be positive. The positive $\hat{\beta}_i$ values reported in empirical studies are always for the most accessible origins.

4 A new set of interaction models: competing destinations models

The diagnostic feature of interaction behaviour resulting from a two-stage decision-making process is that as the accessibility of a destination to all other destinations increases, the interaction volume terminating at that destination decreases, ceteris paribus. If this two-stage decisionmaking process exists, gravity models are misspecified since they do not include a variable which explicitly measures the relationship between interaction and competition between destinations. Consequently, distance-decay parameters from such models will be, in part, a measure of spatial structure. It is a relatively simple matter, however, to relieve this potential misspecification by adding a variable to the gravity models which reflects the competition between destinations for interactions. Such a variable is the accessibility of a destination to all other destinations. The addition of this variable produces a new set of spatial-interaction models which are here called competing destinations models to distinguish them from gravity models. The competing destinations models which are equivalent to the gravity models given in equations (1) and (4) are written below.

(a) *An origin-specific production-constrained competing destinations model* is given by

$$I_{ij} = Z_i O_i m_j A_{ij}^{\delta_i} d_{ij}^{\beta_i} \,, \tag{10}$$

where

$$Z_i = \left(\sum_{j=1}^{n} m_j A_{ij}^{\delta_i} d_{ij}^{\beta_i}\right)^{-1} , \tag{11}$$

which ensures that the constraint in equation (3) is met. A_{ij} represents the accessibility of destination j to all other destinations available to origin i as perceived by the residents of origin i and is defined as

$$A_{ij} = \sum_{\substack{k=1 \\ (k \neq i,\ k \neq j)}}^{w} m_k d_{jk}^{\sigma_i} \,. \tag{12}$$

The parameter σ_i measures the importance of distance in determining the perception of accessibility. Its relationship with β_i is discussed below. Notice that w and n [defined in equations (2), (3), and (5)] are not necessarily equal: w represents the total number of destinations available to origin i whether or not these destinations are included in the set of interactions used to calibrate the model, whereas n represents the number of destinations within a particular data set. Hence, $w \geqslant n$. The definition of accessibility (competition) in equation (12) has been well tested (for instance, Fotheringham, 1979), but it is somewhat arbitrary and other formulations for A_{ij} could be used. Sheppard (1979a) and Pirie (1979) discuss some of these alternatives. The arbitrariness of the definition, however, is no greater than that of the definition of the attractiveness and distance (cost) variables already included in the models.

(b) *An origin-specific doubly constrained competing destinations model* is given by

$$I_{ij} = Z_i O_i B_j D_j A_{ij}^{\delta_i} d_{ij}^{\beta_i} , \tag{13}$$

where

$$Z_i = \left(\sum_{j=1}^{n} B_j D_j A_{ij}^{\delta_i} d_{ij}^{\beta_i} \right)^{-1} , \tag{14}$$

and

$$B_j = \left(\sum_{i=1}^{m} Z_i O_i A_{ij}^{\delta_i} d_{ij}^{\beta_i} \right)^{-1} , \tag{15}$$

and A_{ij} is defined in equation (12). The balancing factors Z_i and B_j ensure that the constraints in equations (3) and (7) are met.

In both of the above models, the expected sign of $\hat{\delta}_i$ is negative—as the accessibility of a destination to all other destinations increases, the volume of interaction terminating at that destination decreases, ceteris paribus. The term A_{ij} measures the perception by individuals of origin i of the accessibility of destination j to all other destinations k. It is a measure of how destination j competes with all other destinations for interactions originating at i; and, as such, the accessibility of j is measured with respect to all other possible destinations—not necessarily simply to the particular destinations that have been sampled.

The interrelationship between the accessibility equation and each interaction equation has varying degrees of complexity. The two can be considered as completely independent equations. If the sample set of centres between which interaction is measured is not a good representation of the population set of centres, and if σ_i is estimated independently (for example, by the assumption that σ_i is equal to some constant for all i), then A_{ij} can be derived independently of the interaction equation. However, given the same poor sample of centres, σ_i and β_i can be estimated iteratively. An estimate of σ_i is made for all i; A_{ij} is calculated for all i and for all j; and the interaction equation calibrated. The estimates of β_is are then used as estimates of the σ_is and the A_{ij}s recalculated etc. The process continues until convergence of $\hat{\sigma}_i$ and $\hat{\beta}_i$ is reached or until the researcher is satisfied that accurate estimates of σ_i and β_i have been achieved. A further degree of interrelationship between the equation for A_{ij} and the interaction equation can be achieved if the sample set of centres is a good representation of the population set. The set of destinations used in the definition of accessibility can then be taken from the set of destinations between which interactions are modelled, and the two equations become further interrelated. The degree of interrelationship between the two equations chosen by a researcher must first depend upon the sample of centres between which interaction is measured and then on a decision as to the value of slightly better estimates of β_i at a cost of substantially more computing time.

If each centre in the spatial system under investigation is both an origin and a destination (such is usually the case in interurban interaction patterns, for example) and the set of such centres is a good approximation of the total set, then an approximation to the definition of A_{ij} given in equation (12) is:

$$A_{ij} \approx \sum_{\substack{k=1 \\ k \neq j}}^{m} m_k d_{kj}^{\sigma_i} , \tag{16}$$

where the summation in equation (16) now includes the term $m_i d_{ij}^{\sigma}$. A further approximation can be made if σ_i is assumed to equal σ for all i. Then,

$$A_{ij} \approx A_j = \sum_{\substack{k=1 \\ k \neq j}}^{m} m_k d_{kj}^{\sigma} . \tag{17}$$

The advantage of this latter formulation is that only n such terms need be calculated instead of $m \times n$. The approximation is only a good one, however, when the number of alternative centres is large, since

$$A_{ij} = A_j - m_i d_{ij}^{\sigma} , \tag{18}$$

and A_j has to be large relative to $m_i d_{ij}^{\sigma}$ in order to ignore this latter term. In the production-constrained competing destinations model, A_{ij} is simply substituted by A_j and the resulting model calibrated. The substitution of A_{ij} by A_j in the doubly constrained competing destinations models is more interesting, however. By making the approximation in equation (17), the competing destinations model can be written as:

$$I_{ij} = Z_i O_j B_j D_j A_j^{\delta_i} d_{ij}^{\beta_i} . \tag{19}$$

Since the addition of any x_i or y_j variable to the original gravity formulation will be accompanied by an equivalent decrease in Z_i or B_j, respectively, the doubly constrained competing destinations model is structurally identical to the doubly constrained gravity model.

Hence the conditions under which the doubly constrained gravity model is correctly specified are those for which A_j (or, equivalently, B_j) is a good approximation for A_{ij}. This occurs when each centre in a data set is both an origin and a destination, and if the set of such centres is a good approximation to the set of all possible destinations. The former condition is quite likely to occur in interurban interaction systems, but is unlikely to occur in intraurban interaction systems where origins and destinations are often quite separate. For example, in shopping-trip analysis where origins are defined as residential areas and destinations are defined as shopping plazas, each node in the system is either an origin or a destination, but not both. The misspecification of the doubly constrained gravity model is then clearly seen: B_j would measure the accessibility of a shopping plaza to all residential areas, whereas A_{ij} would measure the accessibility of a particular shopping plaza to all other shopping plazas. A similar situation arises with the journey-to-work data where employment locations and residential areas are often separate nodes.

The misspecification in the doubly constrained gravity model when the set of destinations included in the interaction matrix is not a good approximation to the complete set of all possible destinations, is also easily demonstrated. For example, if in a city with twenty-five shopping malls, a shopping-trip matrix describes interactions from residential zones to just ten of the malls, then the competing destinations variable would be summed over all twenty-five malls, whereas the gravity-model balancing factor would only be summed over the ten malls in the interaction matrix.

5 The derivation of the competing destinations models by entropy-maximising techniques

The constrained competing destinations models can be derived theoretically by entropy-maximising techniques in a similar manner to that which has been used by Wilson (1970b) to derive the constrained gravity models. In Wilson's formalism, the entropy function maximised is Shannon's (1948) measure of entropy (H) and which is defined as:

$$H = -\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} \ln p_{ij} , \tag{20}$$

where p_{ij} is the probability of an individual travelling between i and j. To obtain interaction models of the gravity format, equation (2) is maximised subject to the constraints:

$$\sum_{j=1}^{n} p_{ij} = \frac{O_i}{T}, \qquad \forall i, \tag{21}$$

$$\sum_{i=1}^{m} p_{ij} = \frac{D_j}{T}, \qquad \forall j, \tag{22}$$

and

$$\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} \ln d_{ij} = D, \tag{23}$$

where T is the total interaction volume and D is the expected logarithmic distance (or cost) travelled in the system. Maximising function (20) subject to constraints (21), (22), and (23) produces a doubly constrained gravity model with the same structure as that given in equation (4). If constraint (22) is omitted, a production-constrained gravity model results. The competing destinations models can be derived in the same manner by the addition of a constraint on destination accessibility, the form of which being identical to equation (23) with d_{ij} replaced by A_{ij}.

An alternative derivation is to minimise Kullback's information gain (G^{KI}) (Kullback, 1959) which is defined as:

$$G^{\mathrm{KI}} = \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} \ln\left(\frac{p_{ij}}{q_{ij}}\right), \tag{24}$$

where q_{ij} is the prior probability of an individual travelling between i and j given no knowledge of the constraints operating on the interaction system. The probability p_{ij} is then defined as the posterior probability of an individual travelling between i and j. Minimising Kullback's information gain minimises the gain of information as a result of knowing $\{p_{ij}\}$ instead of $\{q_{ij}\}$. It is well-known that if there is no prior knowledge of p_{ij} apart from that contained in the constraints, $q_{ij} = 1/mn$ for all i and for all j, and minimising expression (24) is equivalent to maximising function (20). However, if competition between destinations is an important explanatory variable of interaction, q_{ij} can be defined by

$$q_{ij} = A_{ij}^{\delta} \bigg/ \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^{\delta}, \tag{25}$$

since δ is expected to be negative. The denominator of expression (25) is a constant, and hence the new function to be minimised is given by

$$G^{\mathrm{KI}} = \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} \ln(p_{ij} A_{ij}^{-\delta}). \tag{26}$$

Minimising expression (26) subject to the constraints (21)–(23) produces a doubly constrained competing destinations model. Minimising expression (26) subject to constraints (21) and (23) produces a production-constrained competing destinations model. The formalism given in equations (20)–(23) can thus be seen as maximising entropy subject to a partial amount of known information on the interaction system: a misspecified model results. The addition of further knowledge about the system, in this case on destination accessibility, results in models which are more correctly specified.

6 An empirical comparison of gravity and competing destinations models

From the preceding theoretical discussion, competing destinations models are posited to be superior to gravity models, both in terms of yielding accurate parameter estimates and in terms of predictive ability. This superiority needs to be established empirically *since the theoretical arguments depend upon an alternative theory of interaction* decisionmaking whose validity can only be assumed after rigorous testing. For this purpose, a comparison of calibrated gravity and competing destinations models is undertaken.

The data used to calibrate the models are 1970 airline-passenger interaction data published by the United States Civil Aeronautics Board (1971). These data are a 10% sample of all airline journeys on domestic routes within the United States during 1970. The sample was taken continuously throughout 1970: each airline ticket sold with a number ending in zero was selected and the origin and final destination on the ticket were noted. Round-trip tickets were counted as two separate trips. The volume of interaction and airline (great circle) distances are given between every city with at least one commercial airport in the United States.

The data were chosen primarily for their accuracy and comprehensiveness. Many of the interactions in the 10% sample are of the order of 10^3 and some are as high as as 10^5. These sample figures are used throughout the analysis. The sampling error statistic given by the Civil Aeronautics Board indicates that the maximum percentage error of the sample is less than 10% at the 95% confidence level when the sample size is over 400 (7). The sampling statistic and further details of the data are given on pages v-xiii of the above reference.

A slight disadvantage of these data, however, for estimating origin-specific parameters is that the interaction given between i and j is the sum of the interaction from i to j and the interaction from j to i. For an average centre, 50% of the interactions result from decisionmaking behaviour by the inhabitants of that centre, and 50% result from decisionmaking behaviour by the inhabitants of all other centres. Consequently, the origin-specific parameters estimated for such data primarily reflect characteristics of each origin, but also reflect, to a very minor degree, characteristics of each of the other centres in the system. However, since this latter effect will be fairly constant for each origin, especially when the number of centres in the sample is large, the origin-specific parameters resulting from the data can accurately be considered as a reflection of interaction behaviour from each origin.

One hundred cities were selected as a basis for the analysis and interaction and distance matrices of dimension 100 x 100 were obtained from the published data set. The cities chosen were those in the 100 largest Standard Metropolitan Statistical Areas (SMSAs) in terms of their 1970 populations. The size of the sample was determined by two considerations: it had to be small enough to be manageable on the computing system, and it had to be large enough to produce a representative sample of origin-specific parameters(8). The attractiveness of a city was defined by its 1970 SMSA population.

Since each centre in the analysis is both an origin and a destination, and the number of such centres is large, the accessibility of a destination as perceived by residents of origin i is calculated with respect to the remaining ninety-nine centres (the destination itself is excluded from the calculation). This sample of alternative destinations is assumed to be a good representation of the population set of such centres. The error caused by the inclusion of the origin as an alternative destination will be minimal since the number of centres is large. The parameter σ_i ideally should be estimated iteratively with β_i. Initially, σ_i is set equal to some arbitrary value and the set of $A_{ij}(\sigma_i)$ is calculated. A set of $\hat{\beta}_i$ is obtained from the calibration of the

(7) For example, when the sample size is 500, the estimated population size is 5000, and 95 times out of 100 the true population figure will be in the range 5000 ± 416, where 416 is 8·3% of 5000.

(8) Calibrating an origin-specific doubly constrained interaction model with a 100 x 100 interaction matrix requires approximately twenty minutes of central memory time on a CDC Cyber 170/730.

competing destinations model, and this set is used to define a new set of destination accessibilities which in turn is used to recalibrate the model. The process continues until the parameter estimates converge. A disadvantage of this procedure, however, is that it is not known whether the parameter estimates will always converge. Also, the procedure becomes very costly in terms of computer time when the interaction matrix is large. Consequently, an alternative estimation procedure for σ_i was employed which involved setting σ_i equal to a constant, the mean $\hat{\beta}_i$ value from the production-constrained gravity model. The calibration results for the production-constrained competing destinations model were compared by means of the latter estimation procedure for σ_i, and by use of three iterations of the former procedure. There were only very slight differences in the results: the mean $\hat{\beta}_i$ value derived from estimating σ_i and β_i iteratively was $-1 \cdot 09$, and the mean value of the R_i^2 statistic was $0 \cdot 79$; the mean value of $\hat{\beta}_i$ derived from setting σ_i equal to a constant for all i was $-1 \cdot 07$, and the mean value of the R_i^2 statistic was $0 \cdot 80$. Consequently, in all reported model calibrations, σ_i was set equal to a constant for all i—this constant being equal to the mean value of $\hat{\beta}_i$ derived from the production-constrained gravity model.

Given the above assumptions concerning the set of destinations and given that σ_i equals a constant for all i, the doubly constrained gravity model and the doubly constrained competing destinations model are equivalent for this particular data set. Only one set of calibration results will thus be reported for the doubly constrained models, and the expectation is that these results will be very similar to those from the production-constrained competing destinations model.

The parameter δ_i in the production-constrained competing destinations model was set equal to $-1 \cdot 0$ for all i since an inverse relationship between interaction and destination accessibility is hypothesised. With δ_i allowed to vary, increased goodness-of-fit could be gained at a cost of increased computing time. If δ_i is not allowed to vary, however, the potential problem of predicting variations in δ_i when the model is used in the prediction mode is eliminated. The calibration of the competing destinations model thus consists of estimating the parameter β for each origin.

The calibration of each model was undertaken by maximum likelihood estimation, with the use of the distance constraint that

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \hat{I}_{ij} d_{ij}^{*} = \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij} d_{ij}^{*} , \tag{27}$$

where '^' denotes a predicted value, as before, and '*' denotes a logarithmic value. The production-constrained models were calibrated by a Newton–Raphson iterative procedure, whereas the doubly constrained model was calibrated by means of a first-order iterative procedure: both techniques are outlined in Batty and Mackie (1972). For the doubly constrained model, the first-order iteration procedure is possibly a faster calibrating procedure than the Newton–Raphson technique when the number of origins is large, since the latter procedure consists of inverting an $m \times n$ matrix of derivatives which becomes very time-consuming when m and n become large[(9)].

As a measure of the accuracy with which each calibrated model replicated the existing interaction data, an R^2 statistic was calculated using actual and predicted flows from each origin in the system.

[(9)] Batty and Mackie indicated that for a small matrix (30 x 30), the Newton–Raphson procedure was more efficient than the first-order iteration procedure when a single distance-decay parameter was estimated. No tests investigated, however, the effects of matrix size or origin-specific parameter estimation on calibration efficiency.

7 The problem of short-distance interactions

Distance-decay parameters estimated from airline interaction data contain a measurement of spatial structure which is peculiar to this type of data. Consider two centres between which interaction is taking place. Few people would travel by commercial aircraft if the two centres are separated by only a very short distance, and up to some distance, say x, interaction by air will increase as the distance between the two centres increases. Beyond this distance, interaction by air will decrease as distance between the centres increases because of the friction of distance effect. In estimating distance-decay parameters, where a measure of the friction of distance is required, the interest is only in interactions which occur over distances greater than x. If all interactions are used in estimating β, then the resulting estimate will be misleading, since it would also include a measure of the rate at which interaction increases as distance increases up to distance x. This would have the effect of reducing the absolute value of $\hat{\beta}$. It is evident that this effect is related to spatial structure when origin-specific distance-decay parameters are estimated. The absolute value of $\hat{\beta}_i$ will be less for an origin having destinations at distances less than x than the equivalent value for an origin not having destinations at distances less than x, ceteris paribus. Different values of $\hat{\beta}_i$ would not necessarily reflect varying perceived travel disutilities between centres, but merely that some centres have destinations in close proximity, whereas others have not. This spatial-structure effect needs to be removed from the data before the set of distance-decay parameters is estimated since it is peculiar to this type of data[(10)]. The spatial-structure effect that is general to the calibration of gravity models with any interaction data needs to be identified as a separate entity.

To remove the potential bias in $\hat{\beta}_i$ described above, it is necessary to eliminate from the model calibrations all interactions which take place over distances less than x. Hence, n in all model formulations is replaced by n_i, the number of destinations greater than distance x from origin i. In the case of the 100×100 airline-passenger interaction matrix, the distance x was found to be 160 miles so that interactions over distances less than 160 miles were excluded from the model calibrations. It is interesting to note that this distance is similar to that given by Iklé (1954) who suggested that the substitution of ground transportation by air transportation begins at distances between 150 and 200 miles.

8 Calibration results

Origin-specific parameter estimates were derived and compared for the interaction models described in equations (1), (4), and (10). Since the emphasis is on the spatial pattern and spatial variation of these estimates and there are 100 estimates for each model, the results are mapped in a series of figures. Quantitative evidence is also presented on the relationship between the sets of parameter estimates and the set of origin accessibilities, as a measure of spatial pattern. To obtain a visual impression of the varying degrees of spatial pattern exhibited by the sets of parameter estimates, the accessibility of each origin to all other centres in the analysis is given in figure 3. Lines joining origins of approximately equal accessibility are drawn to clarify the spatial pattern of these data[(11)]. The pattern is one in which origins in the North East are very accessible, and accessibility generally decreases as distance from the North East increases. The pattern is interrupted only around California where there

[(10)] Actually, all interaction data exhibit the relationship between interaction and distance described above since the volume of interaction over zero distance is always zero. However, in most cases, the distance x is so small that it causes no spatial bias in estimated parameters and can be ignored.
[(11)] In this, and in each of the subsequent figures where isolines are drawn, there exist exceptions to the general trend these lines represent.

is a slight increase in accessibility. It is interesting to compare the spatial pattern of each set of parameter estimates with the spatial pattern of origin accessibilities.

For clarity, the calibration results concentrate on the spatial variation, spatial pattern, and the interpretation of the parameter estimates. This concentration is justified given the a priori theoretical expectation that the gravity models produce parameter estimates which are biased by spatial structure, and the variation in such estimates is due to variation in spatial structure.

The parameter estimates obtained from calibrating the production-constrained gravity model are mapped in figure 4. Several characteristics of these estimates indicate that the estimates are determined in part by spatial structure, and consequently

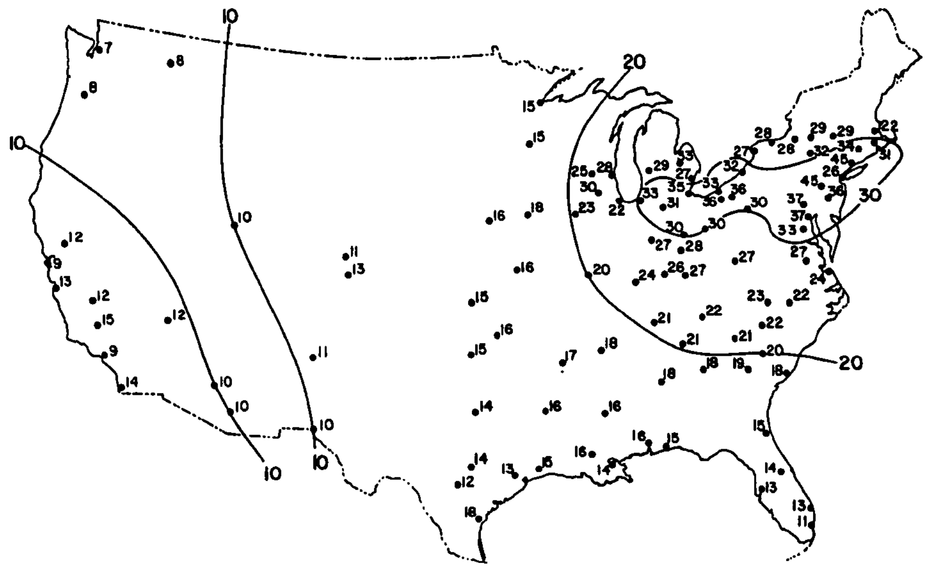


Figure 3. The spatial variation of origin accessibility ($A_i = \sum m_j d_{ij}^{-1}$, with summation over $j = 1\text{-}n$, $j \neq i$; and figures are in units of 10^4).

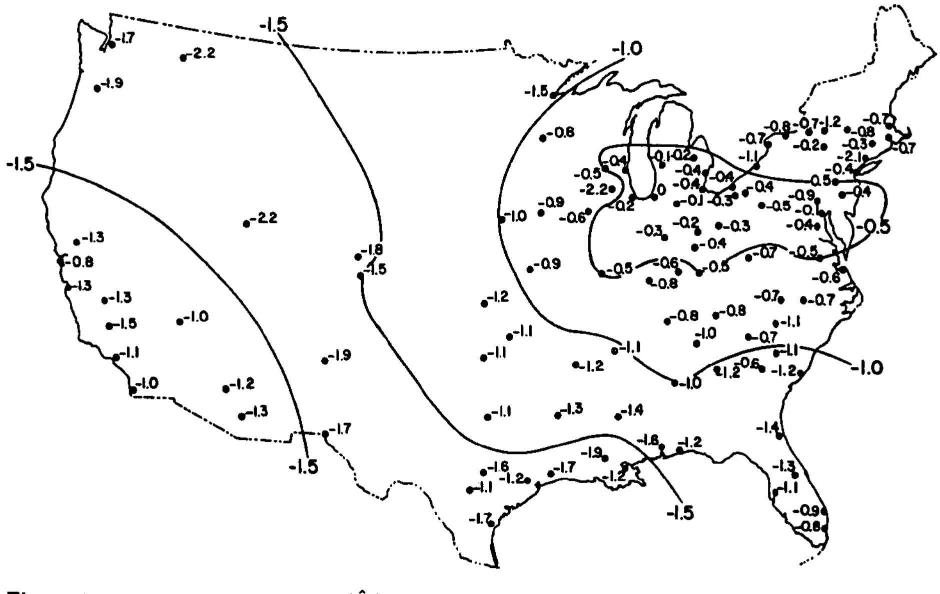


Figure 4. The spatial variation of $\{\hat{\beta}_i\}$ derived from the production-constrained gravity model.

the model from which they are derived is a poor device for analysing interaction behaviour. For instance, there is a marked spatial pattern to the estimates which is very similar to that of origin accessibility described in figure 3. Origins in the North East and Mid West have very small negative parameter estimates, whereas those in the South and West have large negative estimates(12). The correlation coefficient between A_i and $\hat{\beta}_i$ is 0·64. The spatial variation of parameter estimates in figure 4 is large, the coefficient of variation being 52·2. Although no positive estimates are derived, twenty-seven origins have parameter estimates whose absolute values are less than or equal to 0·5, and the behavioural interpretation given to the estimates is suspect. With a mean value of $\hat{\beta}_i$ of $-0\cdot9$, most Californian cities have parameter estimates more negative than the mean and appear more parochial than average, whereas most of the Mid Western cities have parameter estimates very close to zero, which apparently indicates an almost complete disregard for distance by the inhabitants of these cities. It is interesting to note that these results are virtually identical to the calibration results of an unconstrained gravity model using the same data (Fotheringham, 1981).

Figure 5 represents the set of parameter estimates resulting from the calibration of the production-constrained competing destinations model. The relationship between $\hat{\beta}_i$ and origin accessibility is eliminated (the correlation coefficient between the two variables is $-0\cdot01$) and the parameter estimates are remarkably constant over space. There are only three estimates outside the range $-0\cdot5$ to $-2\cdot0$, whereas there are twenty-five such estimates for the gravity model. The coefficient of variation for the competing destinations estimates is 30·2, which is a significant decrease (at $\alpha = 0\cdot05$) from 52·2, the value obtained in the calibration of the gravity model(13).

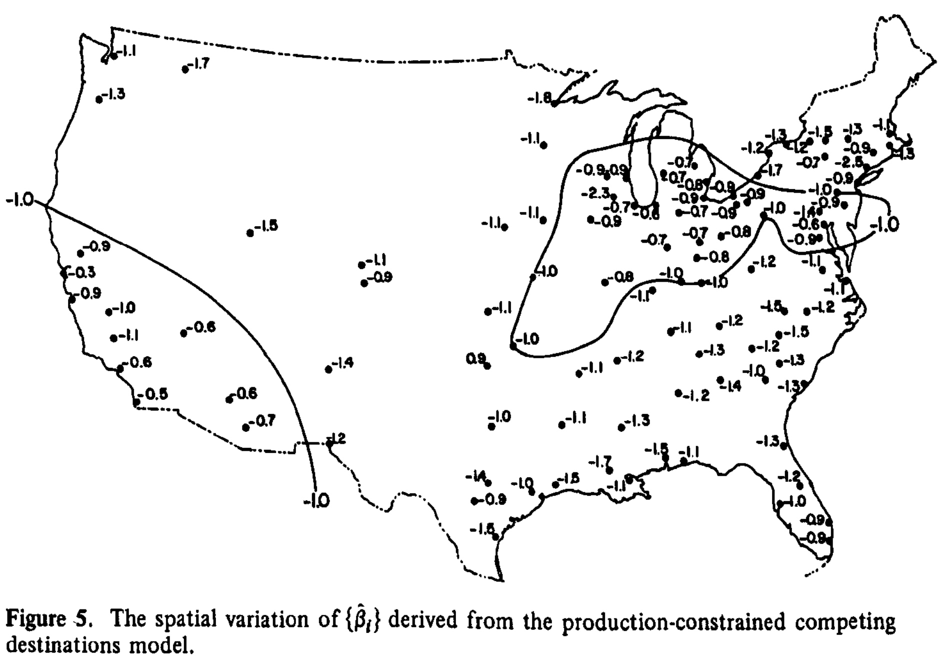


Figure 5. The spatial variation of $\{\hat{\beta}_i\}$ derived from the production-constrained competing destinations model.

(12) There are two parameter estimates which are obvious exceptions to this generalisation. Rockford, Illinois, and Bridgeport, Connecticut, both very accessible origins, have parameter estimates of $-2\cdot2$ and $-2\cdot1$, respectively. These estimates probably reflect the fact that the airports in both origins are overshadowed by the far larger airports in nearby Chicago and New York City, respectively. The airports in Rockford and Bridgeport are likely to be used for local commuter flights only. This behavioural interpretation is possible since the parameter estimates are also very negative when derived from a competing destinations model.

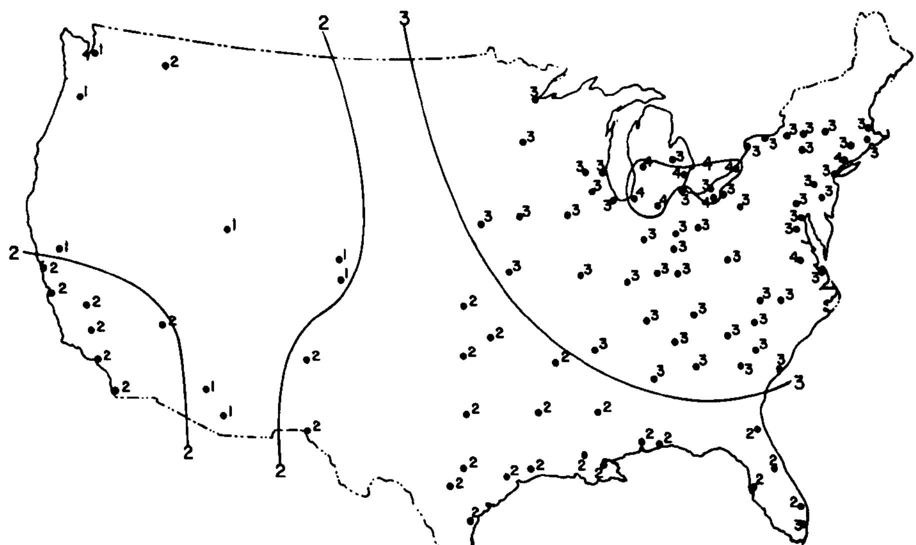
(13) $Z = 8\cdot25$, where Z is defined as $|V_1 - V_2|/[(V_1^2/2n_1)+(V_2^2/2n_2)]^{1/2}$.

The interpretation of the parameter estimates given in figure 5 as measures of interaction behaviour is also more reasonable. With the mean value of $\hat{\beta}_i$ equal to $-1 \cdot 1$, the 'jet-setter' cities, cities with parameter estimates less negative than the mean, can be identified as San Francisco, Las Vegas, Los Angeles, San Diego, and Phoenix, whereas the 'parochial' cities can be identified as Spokane, Salt Lake City, Corpus Christi, Baton Rouge, and Duluth.

The pattern of changes in $\{\hat{\beta}_i\}$ when A_{ij} is added to the gravity model is mapped in figure 6, and it closely resembles the pattern of origin accessibility mapped in figure 3. The greatest changes in β_i occur for origins of extremely high or low accessibility, which indicates that the misspecification bias in the gravity model is greatest for these origins. Origins of high accessibility have parameter estimates that are biased upwards by spatial structure, whereas origins of low accessibility have parameter estimates that are biased downwards by spatial structure. These biases are removed when the competing destinations model is calibrated.

As mentioned, with the interaction data used here it is reasonable to assume that the doubly constrained gravity model is identical to the doubly constrained competing destinations model, and hence the results from the two models can be considered simultaneously. These results are expected to be free from any spatial structure effect that arises from model misspecification. The set of parameter estimates is mapped in figure 7 and is very similar to that derived for the production-constrained competing destinations models. There is very little spatial pattern evident in the estimates (the correlation coefficient between $\{\hat{\beta}_i\}$ and $\{A_i\}$ is $0 \cdot 34$) and there is very little spatial variation (the coefficient of variation of $\{\hat{\beta}_i\}$ is $37 \cdot 5$). With a mean value of $\hat{\beta}_i$ of $-1 \cdot 1$, the estimates conform to expectations: the 'jet-setter' cities include San Francisco, Los Angeles, Miami, Washington, and New York, and the 'parochial' cities include Spokane, Salt Lake City, Baton Rough, Mobile, and Duluth.

As a measure of goodness-of-fit in the two production-constrained models, an R^2 statistic was calculated when a single β parameter was estimated for the whole system.



1—increase $0 \cdot 51 - 1 \cdot 00$ } $\hat{\beta}_i$ becomes less
2—increase $0 \cdot 00 - 0 \cdot 50$ } negative

3—decrease $0 \cdot 01 - 0 \cdot 50$ } $\hat{\beta}_i$ becomes more
4—decrease $0 \cdot 51 - 1 \cdot 00$ } negative

Figure 6. Changes in $\hat{\beta}_i$ when A_{ij} is added to the production-constrained gravity model.

The value of this statistic was 0·73 for the gravity model and 0·76 for the competing destinations model. Since n in each case is equal to 9900, this indicates a significant increase at the 95% confidence level ($Z = 4{\cdot}968$). For the origin-specific versions of the models, however, there was considerable spatial variation in the improvement of goodness-of-fit when the competition variable was added to the gravity model. The smallest increases in R_i^2 (and, in some cases, decreases in R_i^2) occurred for origins of extremely high or extremely low accessibility, whereas the largest increases in R_i^2 occurred for origins of medium accessibility. This is the opposite of the pattern of changes in $\hat{\beta}_i$ described in figure 6. Both patterns, however, can be explained by the spatial relationship between d_{ij} and A_{ij}.

For the accessible origins in the North East and Mid West, A_{ij} and d_{ij} are highly negatively correlated: as distance from origin i increases, the accessibility of the destination decreases. Hence, the addition of A_{ij} to the original gravity model adds little or nothing to the explanatory power of the model, but it does correct the positive bias in the estimated distance-decay parameter. Consequently, the change in R_i^2 is low, whereas the change in $\hat{\beta}_i$ is large and negative. Where origins are inaccessible, such as in the West and Florida, A_{ij} and d_{ij} are highly positively correlated, so the addition of A_{ij} to the gravity model again adds little or nothing to the explanatory power of the model, but it does correct the negative bias in the estimated distance-decay parameters. Consequently, the change in R_i^2 is low, whereas the change in $\hat{\beta}_i$ is large and positive. For origins of medium accessibility, A_{ij} and d_{ij} are weakly correlated: in one direction as d_{ij} increases, A_{ij} increases, whereas in the other direction as d_{ij} increases, A_{ij} decreases. Hence, the addition of A_{ij} to the gravity model adds significantly to the explanatory power of the model, but it does not alter the relationship between distance and interaction which, since A_{ij} and d_{ij} are only weakly correlated, is not a spurious one. Consequently, the change in R_i^2 is large and positive, whereas the change in $\hat{\beta}_i$ is minimal.

The errors in gravity-model predictions for an individual origin of medium accessibility are seen in figure 8 which describes the spatial pattern of errors in predicted interactions from Kansas City. The model from which the interactions are predicted is the production-constrained gravity model, and the errors given are

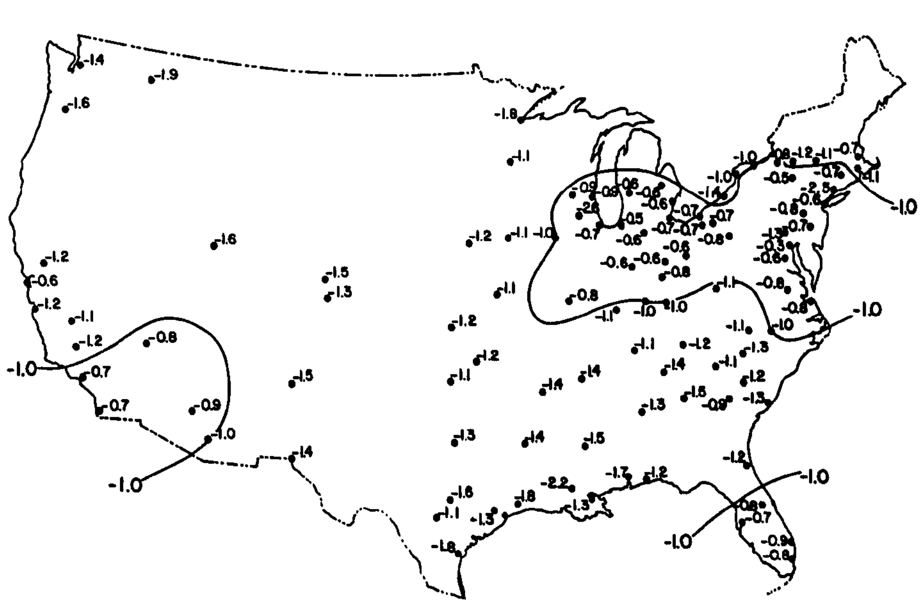


Figure 7. The spatial variation of $\{\hat{\beta}_i\}$ derived from a doubly constrained model.

percentage errors, which are defined as $100(\hat{I}_{ij} - I_{ij})/I_{ij}$ [14]. Hence, negative errors denote underprediction by the gravity model and positive errors denote overprediction. It is obvious that the gravity model overpredicts interactions to most of the accessible centres in the North East and underpredicts interactions to the inaccessible centres in the rest of the country. The model fails to account for the effect of competition between destinations. Clearly, the addition of an inverse relationship between destination accessibility and interaction will reduce the magnitudes both of the positive and of the negative errors reported in figure 8.

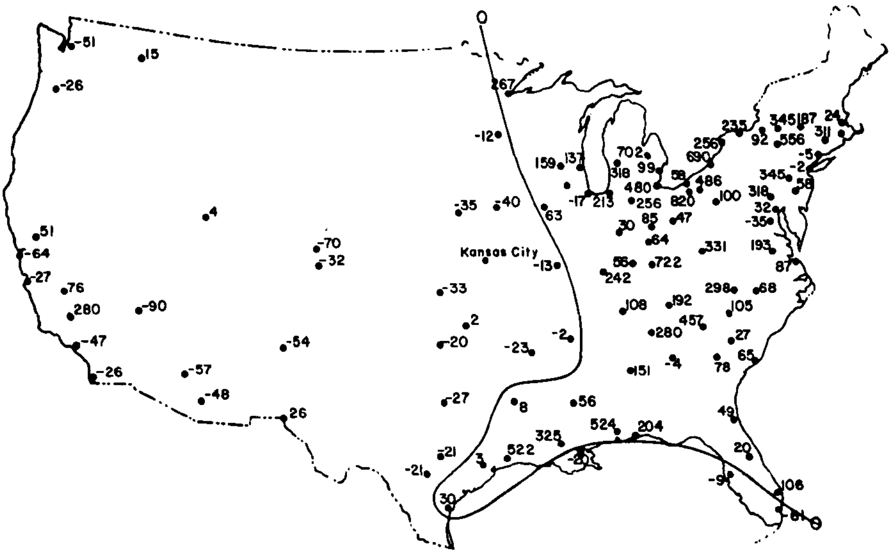


Figure 8. Percentage errors in predicted interactions from a production-constrained gravity model. Interactions are from Kansas City ($R_i^2 = 0 \cdot 88$); negative figures indicate underpredictions by the model, and positive figures overpredictions.

9 Implications and conclusions

Gravity models are misspecified when they do not explicitly model the relationship between interaction and the accessibility of a destination to all other destinations. The degree of this misspecification varies with the constraints operating upon the model. Gravity models which have no constraint on destination flow totals (that is, no β_j balancing factor), such as unconstrained and production-constrained models, are misspecified in any situation. Gravity models which do constrain destination flow totals such as attraction-constrained and doubly constrained models, can be correctly specified under certain conditions. These conditions are that every centre in the system under investigation is both an origin and a destination and that the set of such centres accurately represents the set of all possible destinations available to each origin.

The misspecification of gravity models causes the estimates of distance-decay parameters obtained from these models to be biased by spatial structure so that they cannot be interpreted as measures of interaction behaviour. Accessible origins have parameter estimates that are biased upwards by spatial structure, and inaccessible origins have parameter estimates biased downwards by spatial structure. Consequently, the set of origin-specific distance-decay parameters resulting from a gravity-model

(14) The percentage errors are not given for Bridgeport and Rockford since they are highly inflated because of the known interactions being so much lower than expected given their location and size. The reason for this has been discussed previously.

calibration exhibits a strong spatial pattern and a large variance, both of which are spurious. Empirical studies which discuss the spatial variation of gravity-model parameter estimates in terms of varying interaction behaviour are thus misleading. Except when spatial structure is constant, information can only be gained on the perception of distance as a deterrent to interaction when estimates of distance-decay parameters are obtained from the calibration of competing destinations models.

Studies which have compared gravity-model parameter estimates through time are also suspect. Since spatial structure is likely to vary over time, distance-decay parameter estimates will vary even if the perception of distance as a deterrent to interaction remains constant. For example, a general finding from investigations of the temporal stability of distance-decay parameters is that $\hat{\beta}_i$ becomes less negative over time (for instance, Hägerstrand, 1957; Clark and Ballard, 1980). Since these estimates are derived from gravity models, it is impossible to determine whether the increase in $\hat{\beta}_i$ over time is due to changing perceptions of distance or variations in spatial structure.

A major implication of the results of this study concerns the prediction of interactions. Two types of prediction can be considered. One occurs when a spatial-interaction model is calibrated in a spatial system where interaction data are known and then it is used to predict interactions in spatial systems where such data are not available: there is a macrochange in spatial structure. The second occurs when a spatial-interaction model is calibrated in a spatial system and then is used to predict interactions in the same system given a change in the spatial structure of the system: there is a microchange in spatial structure. The parameters of the calibrated model are assumed to remain constant in both cases, but distance-decay parameters derived from gravity models will not be constant when there is a change in spatial structure. Consequently, there will be misleading predictions of interaction resulting from the use of gravity models or similar misspecified interaction models. Southworth (1980) indicates the magnitude of the errors in prediction which result from assuming parameter estimates to be constant when there is variation in spatial structure[(15)].

It has long been a problem that gravity models replicate the data used to calibrate them well but fail to replicate other data equally well (for instance, see Hyman and Wilson, 1969; Taaffe and Gauthier, 1973, page 98; and Southworth, 1980). Such a problem may be due to the spurious spatial variation in parameter estimates resulting from the misspecification of the models. Since competing destinations models are free from this misspecification, they are likely to perform much better in spatial systems other than the one in which they are calibrated.

In gravity modelling, the errors in assuming constant distance-decay parameters when there are macrochanges in spatial structure are evident in the calibration results given here (for example, compare the parameter estimate for Seattle with that for Boston in figure 4). The errors in gravity modelling which arise from assuming constant distance-decay parameters when there are microchanges in spatial structure are also quite evident. Consider shopping trips in the spatial system given in figure 9. In each spatial system, the perception of distance as a deterrent to interaction and the size of destinations are assumed to be constant over space. In figure 9(a), the accessibilities of each residential area to the three shopping plazas (A_i) are ranked $A_2 > A_1 > A_3$. If an origin-specific gravity model were calibrated in this system, the distance-decay parameter estimates would be ranked $\hat{\beta}_2 > \hat{\beta}_1 > \hat{\beta}_3$. Suppose a new shopping plaza is located in close proximity to origin 1 and that the new origin

(15) Southworth's results concern the temporal instability of distance-decay parameter estimates. When 1962 estimates are used to forecast 1971 interactions, the errors in prediction are approximately two to three times larger than when 1971 estimates are used.

accessibilities (A_i') are ranked $A_1' > A_2' > A_3'$. The spatial structure of this system is given in figure 9(b). If a gravity model were calibrated for this system, the ranking of the distance-decay parameter estimates (β_i') would be $\hat{\beta}_1' > \hat{\beta}_2' > \hat{\beta}_3'$ and thus predicting interactions using the original parameters, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, would give misleading results. Consider, for example, interactions from origin 1. The relationship between interaction and distance for the spatial systems given in figure 9 is shown in figure 10. The distance-interaction relationship indicated by the gradient $\hat{\beta}_1$ is the prediction of the 'true' distance-interaction relationship described by $\hat{\beta}_1'$. It is evident that if $\hat{\beta}_1$ is assumed to be a constant even though there is a change in spatial structure, misleading predictions result. Short-distance interactions would be overpredicted whereas long-distance interactions would be underpredicted.

A competing destinations model reflects changes in spatial structure, and the effects of these changes on interaction patterns, via an accessibility variable in the model and not through the distance-decay parameter. In gravity modelling, because variations in spatial structure are reflected by parameter variations, much of the poor predictive performance of gravity models may be attributed to the spurious spatial variation of parameter estimates resulting from model misspecification, and perhaps not because of, as Hyman and Gleave (1976) conclude, the existence of intrinsically different spatial behaviour.

It is not only simple gravity models that will produce misleading predictions when $\hat{\beta}$ is assumed to be invariant to changes in spatial structure. Any multiequation model which incorporates a gravity model will produce suspect results, and especially if the gravity model is used in an iterative procedure. An example of such a misuse of gravity models is the Lowry model (Lowry, 1964) which forms the basis for much urban land-use modelling (see for instance, Batty, 1976). In the Lowry model, $\hat{\beta}$ is predetermined and assumed to be constant. An attraction-constrained gravity model is used to predict the distribution of residential population within an urban system and a production-constrained gravity model is used to predict the distribution of retail employment. Both models are used iteratively until the convergence of each distribution is achieved. Consequently, the errors in assuming $\hat{\beta}$ to be invariant to the changes in spatial structure which are an integral part of the model are compounded and this produces potentially large errors in prediction.

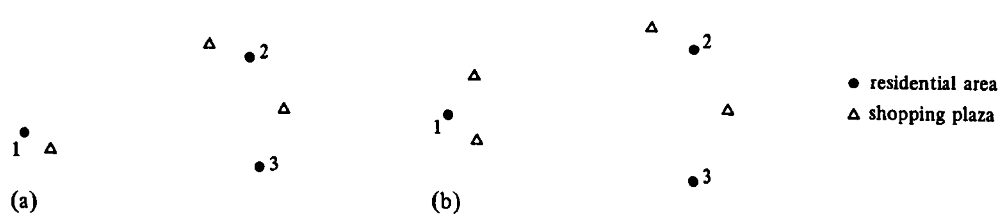


Figure 9. Two urban spatial systems: in (a) origin 2 is most accessible, and in (b) origin 1 is most accessible to the shopping plazas.

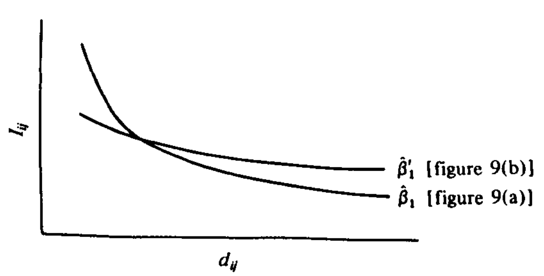


Figure 10. Distance-interaction relationships for origin 1 in figure 9.

A final implication of these results concerns the aggregation problem in gravity modelling. It is well-known that estimated distance-decay parameters are, in part, a function of the level at which the interaction data are aggregated (Openshaw; 1977). This relationship may be caused by changing patterns of destination accessibility associated with changes in zonal boundaries. The addition of the accessibility variable could thus account for some of the aggregation effect since A_{ij} is redefined when the level of aggregation changes. The inclusion of the accessibility variable obviously introduces greater flexibility into gravity models. Changes in spatial structure outside the area in which interactions are measured can affect interaction patterns within that area. These external effects are not accounted for in gravity models but they are in competing destinations models since A_{ij} is defined outside the interaction system.

Thus, gravity models are shown to be misspecified interaction models since they ignore the relationship between destination competition and interaction. This misspecification results in parameter estimates which are a function of spatial structure and renders the models less useful for prediction. The extent of this misspecification, however, in terms of different types of interaction remains to be determined. It is shown here to be a serious problem in intraurban interaction. Whether the misspecification is as serious for interurban interaction, or indeed for particular types of intraurban interaction, can only be determined by further empirical analyses. The theoretical development of the competing destinations models, however, suggests that it is as serious and that the calibration of any spatial-interaction model which does not include a distance variable and a competing destinations variable produces estimates of distance-decay parameters which are behaviourally meaningless and predictively useless.

Acknowledgements. The research for this paper was undertaken while the author was a graduate student at McMaster University and the advice of Mike Webber is gratefully acknowledged. The author also acknowledges the aid of a Summer Faculty Fellowship awarded by Indiana University.

References

Batty M, 1976 *Urban Modelling* (Cambridge University Press, Cambridge)

Batty M, Mackie S, 1972 "The calibration of gravity, entropy and related models of spatial interaction" *Environment and Planning* 4 205-233

Black W R, Larson R, 1972 "A comparative evaluation of alternate friction factors in the gravity model" *The Professional Geographer* 24 335-337

Chisholm M, O'Sullivan P, 1973 *Freight Flows and Spatial Aspects of the British Economy* (Cambridge University Press, Cambridge)

Clark G L, Ballard K P, 1980 "Modelling out-migration from depressed regions: the significance of origin and destination characteristics" *Environment and Planning A* 12 799-812

Curry L, 1972 "A spatial analysis of gravity flows" *Regional Studies* 6 131-147

Curry L, Griffith D A, Sheppard E S, 1975 "Those gravity parameters again" *Regional Studies* 9 289-296

Fotheringham A S, 1979 "Polarised growth within a multi-growth-centre environment: a case study of the United States 1920-1970" *Environment and Planning A* 11 193-208

Fotheringham A S, 1981 "Spatial structure and the parameters of spatial interaction models" *Annals of the Association of American Geographers* 71 425-436

Fotheringham A S, Webber M J, 1980 "Spatial structure and distance-decay parameters" *Geographical Analysis* 12 33-46

Frost M, 1969 "Distribution costs as a factor in the location of industry policy" DP-34, Department of Geography, London School of Economics and Political Science, London

Gordon I R, 1976 "Gravity demand functions: accessibility and regional trade" *Regional Studies* 10 25-37

Gould P, 1975 "Acquiring spatial information" *Economic Geography* 51 87-99

Hägerstrand T, 1957 "Migration and area" in *Migration in Sweden: A Symposium* Eds D Hannerberg, T Hägerstrand, B Odering, Lund Studies in Geography, series B, volume 13 (C W K Gleerup, Stockholm) pp 27-158

Hyman G M, 1969 "The calibration of trip distribution models" *Environment and Planning* 1 105-112

Hyman G M, Gleave D, 1976 "A reasonable theory of migration" CES-RP-22, Centre for Environmental Studies, London
Hyman G M, Wilson A G, 1969 "The effects of changes in travel costs on trip distribution and modal split" *Journal of High Speed Ground Transportation* 3 79-85
Iklé F C, 1954 "Sociological relationship of traffic to population and distance" *Traffic Quarterly* 8 123-136
Johnston R J, 1973 "On frictions of distance and regression coefficients" *Area* 5 187-191
Johnston R J, 1975 "Map pattern and friction of distance parameters: a comment" *Regional Studies* 9 281-283
Johnston R J, 1976 "On regression coefficients in comparative studies of the friction of distance" *Tijdschrift voor Economische en Sociale Geografie* 67 15-28
Kullback S, 1959 *Information Theory and Statistics* (John Wiley, New York)
Leinbach T R, 1973 "Distance, information flows and modernisation: some observations from West Malaysia" *The Professional Geographer* 25 7-11
Linneman H V, 1966 *An Econometric Study of International Trade Flows* (North-Holland, Amsterdam)
Lowry I S, 1964 "A model of metropolis" RM-4035-RC, Rand Corporation, Santa Monica, CA
Openshaw S, 1977 "Optimal zoning systems for spatial interaction models" *Environment and Planning A* 9 169-184
Pirie G H, 1979 "Measuring accessibility: a review and proposal" *Environment and Planning A* 11 299-312
Porter R, 1956 "Approach to migration through its mechanism" *Geografiska Annaler* B38 317-343
Shannon C F, 1948 "A mathematical theory of communication" *Bell System Technical Journal* 27 379-423; 623-656
Sheppard E S, 1979a "Geographic potentials" *Annals of the Association of American Geographers* 69 438-447
Sheppard E S, 1979b "Gravity parameter estimation" *Geographical Analysis* 11 120-132
Sheppard E S, Griffith D A, Curry L, 1976 "A final comment on mis-specification and autocorrelation in those gravity parameters" *Regional Studies* 10 337-339
Southworth F, 1980 "Temporal versus other effects on spatial interaction model parameter values" paper presented to the seventy-sixth Annual Meeting of the Association of American Geographers, Louisville; a copy of the paper may be obtained from the author, Department of Civil Engineering, University of Illinois at Urbana-Champaign, Urbana, IL
Stillwell J C H, 1977 "Some historical tests of spatial interaction models for inter-area migration. Part 1: inter-county migration" WP-185, School of Geography, University of Leeds, Leeds, England
Stillwell J C H, 1978 "Interzonal migration: some historical tests of spatial-interaction models" *Environment and Planning A* 10 1187-1200
Taaffe E J, Gauthier H L Jr, 1973 *Geography of Transportation* (Prentice-Hall, Englewood Cliffs, NJ)
Taylor P J, 1975 "Distance decay in spatial interactions" *Concepts and Techniques in Modern Geography* 2 23
Thomas R W, 1977 "An interpretation of the journey-to-work on Merseyside using entropy-maximising methods" *Environment and Planning A* 9 817-834
United States Civil Aeronautics Board, 1971 "Domestic origin-destination survey of airline passenger traffic, 1970, summary data" compiled by the Bureau of Accounts and Statistics, Air Transport Association of America, Washington, DC; published by the Airline Finance and Accounting Conference, a division of the Air Transport Association of America, Washington, DC 20036
Wilson A G, 1970a "Advances and problems in distribution modelling" *Transportation Research* 4 1-8
Wilson A G, 1970b *Entropy in Urban and Regional Modelling* (Pion, London)
Wilson A G, 1971 "A family of spatial interaction models, and associated developments" *Environment and Planning* 3 1-32