



UNIVERSITY OF LEEDS

This is a repository copy of *Partial Matrix Techniques*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/2413/>

Monograph:

Kirby, H.R. (1979) *Partial Matrix Techniques*. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Working Paper 111

Reuse

See Attached

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



White Rose Research Online

<http://eprints.whiterose.ac.uk/>

ITS

[Institute of Transport Studies](#)

University of Leeds

This is an ITS Working Paper produced and published by the University of Leeds. ITS Working Papers are intended to provide information and encourage discussion on a topic in advance of formal publication. They represent only the views of the authors, and do not necessarily reflect the views or approval of the sponsors.

White Rose Repository URL for this paper:

<http://eprints.whiterose.ac.uk/2413/>

Published paper

Kirby, H.R. (1979) *Partial Matrix Techniques*. Institute of Transport Studies, University of Leeds, Working Paper 111

Working Paper 111

March 1979

PARTIAL MATRIX TECHNIQUES

Howard R. Kirby

ABSTRACT

KIRBY, H.R. (1979) Partial matrix techniques. Leeds: Univ. Leeds, Inst. Transp. Stud., Work. Pap. 111.

Partial matrix techniques are those in which gravity models are fitted to a partially observed matrix of trips and journey costs, and used to infer the trips in the unobserved cells. This paper reviews the theoretical basis from which such techniques have been developed, and demonstrates the need to pay careful attention to the underlying assumptions, which in effect require that the model be a good fit to the observed data (and also a good 'fit' to the unobserved data). Circumstances are described in which the estimates for the unobserved cells may not be uniquely determined, and the effects of data structure on the reliability of the estimates (assuming these to be unique) are discussed. Ways are suggested in which further theoretical and empirical research might demonstrate whether a given pattern of observations would lead to particularly error-prone estimates.

PARTIAL MATRIX TECHNIQUES

Howard R. Kirby

1. INTRODUCTION

Trip distribution models are often fitted to data in the form of origin-destination matrices of trips and generalised costs. For several years, it has been the practice to use matrices in which, by virtue of the survey design, not all origin-destination movements are observable. Such matrices are said to be partial as opposed to whole*. Cells not included in the partial matrix may be described as excluded, unobservable, or missing.

The phrase 'partial matrix techniques' generally refers to the practice of calibrating a gravity model to a partial trip matrix, and using the results of this calibration to infer something about the trip distribution for the whole matrix (including the missing cells). The practice was developed by Wootton (1972) and first used in Derbyshire, and subsequently applied by Neffendorf in Sheffield; see Neffendorf and Wootton (1974). It drew support from theoretical considerations first reported by Kirby (1972) and subsequently published partly in Kirby (1974) and partly in Beardwood and Kirby (1975).

Although partial matrix techniques have been widely used, very little has been reported in the published literature. Occasionally, however, statements are made - for example in Cunliffe and Nesbitt (1977) - that make it appear that these theoretical considerations are thought to have a wider validity or applicability than was claimed, with perhaps insufficient appreciation of the assumptions that have to be made when using the partial matrix technique.

Since failure to appreciate the theoretical issues or assumptions might cause the partial matrix technique to be used in conditions in which it is not appropriate, this paper has been prepared with three objectives:

* Note that a whole matrix does not necessarily mean that the observed values for each movement or cell are non-zero; indeed, whole matrices may contain zero entries that are zero by chance. If the proportion of observable cells that are zero is high (whether in a whole or partial matrix), the matrix is said to be sparse. Thus, a sparse matrix is not necessarily partial, and a partial matrix is not necessarily sparse - a distinction which has not always been observed in the literature (see for example Cunliffe & Nesbitt, 1977). The term full matrix is probably best reserved for whole matrices with no empty cells. In the statistical literature, it is more usual to use the terms incomplete and complete matrices rather than partial and whole ones.

- (a) to highlight and amplify the role of theory and assumption in the use of partial matrix techniques;
- (b) to suggest some practical tests for verifying the kinds of conditions under which the use of partial matrix techniques are most appropriate;
- (c) to report on some tests that have been carried out of the kind mentioned in (b), and to appeal for others to be reported.

2. THEORETICAL BASIS

The theoretical basis from which partial matrix techniques have been developed was that given in Beardwood and Kirby (1975), although, as we shall see in Section 3, an extension to the results there given should also have been appealed to in some circumstances. This extension is described in 2.2.

2.1 Basic result

The result given in Beardwood and Kirby relates to the synthesis of a trip matrix (using the two-way adjustment (or biproportional) procedure of Furness) such that it is biproportional to some starting matrix and agrees with prescribed row and column sums. The result demonstrates the equivalence between the solutions for bi-proportional (Furness) adjustments to suitably related whole and partial matrices. It is simply described in terms of an example, as in the three-zone example of Fig 1. The diagrams above the dotted line in Fig 1 define values (a, \dots, i) , for a starting matrix (which may be base-year observations, and contain some values that are zero by chance; or may be derived by applying some function to a cost-matrix); values (P, \dots, U) of the trip ends to which the starting values are to be adjusted pro-rata; and values (A, \dots, I) of the results of synthesising a whole matrix to agree with these trip-end totals. The diagrams below the dotted line define how the corresponding values for the synthesis of a partial matrix relate to those for the whole matrix. The cells shown shaded in the lower line are those excluded from the whole matrix in forming the partial matrix, and there may in general be several such cells in each row and column of the matrix. In general, the location of the excluded cells may be best expressed by an incidence matrix, in which '0' indicates an excluded cell, and '1' indicates a cell for which an observation is available. (See Fig 2 for some examples.) We note in Section 3 that it may be necessary to restrict the permissible locations of the excluded cells.

A whole base trip matrix:

| | | |
|---|---|---|
| a | b | c |
| d | e | f |
| g | h | i |

which is to be adjusted to agree with trip totals:

| |
|-------|
| P |
| Q |
| R |
| S T U |

gives

| | | |
|---|---|---|
| A | B | C |
| D | E | F |
| G | H | I |

A partial base trip matrix, say:

| | | |
|---|---|---|
| / | b | c |
| d | e | f |
| g | / | i |

which is to be adjusted to agree with the reduced trip totals

| |
|-----------|
| P-A |
| Q |
| R-H |
| S-A T-H U |

gives

| | | |
|---|---|---|
| / | B | C |
| D | E | F |
| G | / | I |

Fig 1. An illustration of the theoretical result which has been drawn on as a basis for partial matrix techniques.

This equivalence of the two lines of calculations shown in Fig 1 means that if data for some movements is not obtainable (either as a result of the survey design or because some movements are physically impossible), the results of a fully constrained synthesis of the trip distribution for just the observed parts of the matrix would be the same as we would have synthesised for those parts had we been using the whole matrix, provided that the trip end data used in the partial matrix synthesis is consistent with that for the whole matrix.

In practice, therefore, attention needs to be paid to the proviso underlined, and the ramifications of this will be explored in Section 4, which discusses three ways in which the above theoretical result has been invoked in a practical context. But first we present a slight extension which is also relevant to that discussion.

2.2 An extension

The Beardwood & Kirby result applied to the case in which a two-way adjustment procedure is applied to a given two-dimensional matrix, (f_{ij}) say, in order to 'synthesise' a trip matrix (t_{ij}^*) that agrees with prescribed row and column sums (corresponding to trip totals (g_i) in generation zones and (a_j) in attraction zones). The result is expressed in the form:

$$t_{ij}^* = p_i q_j f_{ij} \quad (1)$$

in which the generation factors (p_i) and attraction factors (q_j) are such that

$$\sum_j t_{ij}^* = g_i \quad \text{and} \quad \sum_i t_{ij}^* = a_j \quad (2)$$

The above describes the process carried out in the prediction of certain types of trip matrix; an obvious extension of the result in 2.1 is to the calibration situation, in which we wish to estimate not only the parameters (p_i) and (q_j) but also the parameters of f as a function of the separation or generalised cost c_{ij} . In the case where the cost is divided into K ranges such that, if the cost falls in the k th interval, a factor r_k expresses the average value of $f(c_{ij})$ in this interval, we may define a three-dimensional matrix Δ_{ijk} of zeroes and ones which indicate in which interval a particular cost c_{ij} falls, and a maximum likelihood procedure for estimating the parameters of (p_i), (q_j) and (r_k) is such that, under certain conditions, the value

$$t_{ijk}^* = p_i q_j r_k \Delta_{ijk} \quad (3)$$

is such as to satisfy

$$\sum_{jk} t_{ijk}^* = g_i; \quad \sum_{ik} t_{ijk}^* = a_j; \quad \sum_{ij} t_{ijk}^* = s_k \quad (4)$$

where the (g_i), (a_j) and (s_k) are here to be understood as, respectively, the number of trips in each generation zone, attraction zone, and interval of separation as given by the corresponding summation over an observed trip matrix.

For such a three-dimensional situation a three-way balancing (or triproportional) procedure may be used. (See Kirby, 1974, Evans and Kirby 1974, and Kirby and Leese, 1978 for a discussion of the procedure and conditions).

The result for three dimensions that correspond to that given in Fig 1 for 2 dimensions is illustrated in Fig 2; the mathematical demonstration of the equivalence, and the conditions under which it holds, are similar to those given in Beardwood and Kirby (1975).

A whole base year trip matrix

| | | |
|---|---|---|
| a | b | c |
| d | e | f |
| g | h | i |

and the corresponding cost interval matrix:

| | | |
|---|---|---|
| k | l | m |
| m | k | l |
| l | m | k |

gives base year zone trip totals:

| | | |
|---|---|---|
| s | t | u |
| = | = | = |
| a | b | c |
| + | + | + |
| d | e | f |
| + | + | + |
| g | h | i |

| |
|-----------|
| p = a+b+c |
| q = d+e+f |
| r = g+h+i |

and base year trip cost interval frequency distribution

| | |
|---|-----------|
| k | x = a+e+i |
| l | y = b+f+g |
| m | z = c+d+h |

A triproportional calibration procedure that reaches agreement with the above zonal and interval trip totals gives a synthesised trip matrix:

| | | |
|---|---|---|
| A | B | C |
| D | E | F |
| G | H | I |

If we have only a partial matrix of base year trips:

| | | |
|---|---|---|
| / | b | c |
| d | e | f |
| g | / | i |

and corresponding cost-intervals

| | |
|---|---|
| l | m |
| m | k |
| l | k |

then, if the zonal trip totals are:

| |
|-------|
| p - A |
| q |
| r - H |

and trip cost interval totals are

| | |
|---|-------|
| k | x - A |
| l | y |
| m | z - H |

$$s-A \quad t-H \quad u$$

the triproportional calibration procedure that reaches agreement with the above zonal and interval trip totals gives a synthesised trip matrix:

| | | |
|---|---|---|
| / | B | C |
| D | E | F |
| G | / | I |

that is consistent with that for calibration to the whole matrix, (see above).

This clearly requires that, for the whole matrix, the sum of its estimates in the excluded parts agrees exactly with those given by the survey data for those parts. In this case: A=a and H=h.

Fig.2 Extension of the result demonstrated in Fig.1 to a calibration situation.

3 APPLICATIONS

There at least three ways in which the theoretical work described in section 2 has been invoked in a practical context: synthesis for a partial matrix, calibration and synthesis for a partial matrix, and calibration and synthesis for a whole matrix with data for a partial matrix.

3.1 Synthesis for a partial matrix

If one is interested only in synthesising trips for parts of a matrix given the cost matrix (c_{ij}) and cost function $f(c)$, then the result given in section 2.1 suggests that it is in order for the trip-end balancing (Furness) calculations to be done on the partial matrix, *provided that the trip-end estimating procedure yields zonal trip-end totals that properly exclude the trips that would have gone to the missing cells; that is, that these excluded trips would have been estimated reasonably accurately by the model had it been applied to the whole matrix with full trip-end totals.* Some typical partial matrix situations are illustrated by the incidence matrices given in Fig.3. One common situation is that in which intra-zonal movements are not estimated (Fig.3a): here, the trip generation relationships would have been developed only for inter-zonal movements. A second common situation is that in which there is no information about the external-external traffic (Fig. 3b).

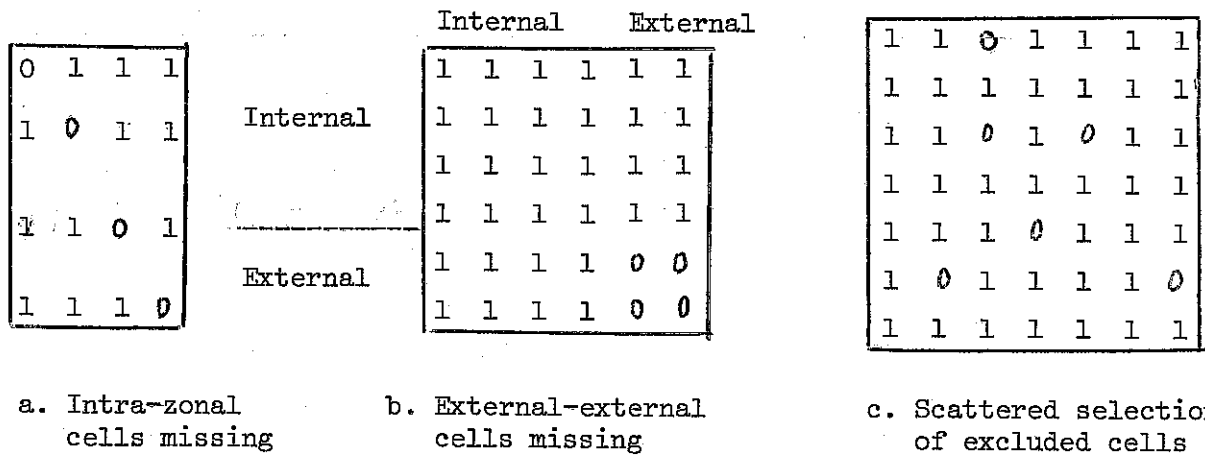


Fig. 3 Incidence matrices for some typical partial matrix situations

A third situation is where the excluded cells are scattered throughout the matrix (Fig.3c). One way in which this might occur would be if certain pairs of zones were in a special relationship. That relationship would have to be modelled separately and those zone-pairs excluded from the usual trip distribution calculations. An example is the link between an RAF base and a housing estate containing mainly RAF personnel; or that which sometimes happens in the planning of new towns in which certain estates may be (initially) earmarked for the employees of certain firms.

3.2 Calibration and synthesis for a partial matrix

In Bearwood and Kirby (1975), it was suggested in the conclusions that "The analyst need not worry too much if, when he wants to do a calibration, there is information missing about some interzonal transfers. He may omit completely from his calibration all cells for which information is missing, and rest assured that, *had the missing data conformed to his (calibrated) model*, the trips he synthesises *for the partial matrix* would be the same as those he would have obtained by synthesising the whole matrix.

Note firstly that these remarks are addressed to the problem of calibrating and synthesising for the partial matrix, not to that of calibrating and synthesising for the whole matrix, given data only for parts of it. The latter problem we discuss in 3.3. We again see the need for ensuring that the proviso in italics is reasonably adhered to, although, if the analyst is not concerned about the extent to which his model agrees with the model that he would have fitted in the whole matrix, he need not be concerned if the proviso does not hold; the model fitted to the observed data will still be the best fitting model for that data.

3.3 Calibration and synthesis for a whole matrix, from partial data

The partial matrix techniques which Neffendorf and Wootton (1974) developed build on the suggestions reproduced in 3.2 concerning the calibration for partial matrices. The essential difference between the situations of 3.2 and that now discussed is as follows.

- a. In 3.2 we infer something about those parts of the matrix for which we have information, seeking only to be satisfied that this is not far different from what we might infer for those parts had we had the data for the whole matrix.

- b. In 3.3 we infer something not only about the parts of the matrix for which we have information, but also those parts for which we do not have information.

The inferences made about the parts of the matrix for which no information is available require the calibration resulting from the partial matrix to be applied to the whole matrix, and are of two kinds.

- i. Estimates of the zonal totals of trips (summed over both observed and unobserved cells), as a basis for deducting (or checking) trip generation relationships
- ii. Estimates of the inter-zonal transfers in the unobserved cells.

It is now very much more necessary than in 3.2 to ensure that the proviso in italics in 3.2 holds. Since the model fitted to the observed data is, in a sense, being extrapolated, the consequences of a departure from the conditions of the proviso are more severe, as we shall see in section 4. For (i) to be a reasonable procedure, it is necessary only to ensure that, in total for each of the rows and columns over the excluded cells, the model reproduces the number of trips that would have been observed in those cells. For (ii) to be a reasonable procedure, rather more is required: namely, that each estimate in the excluded cells would be in general agreement with what observation would have shown; or, at least, the agreement in the unobserved cells would be no worse than the agreement between model and data in the observed cells. This implies that the model fitted to the data in the observed cells is not only a good one but is also, in some sense, representative of the unobserved cells as well. This question is discussed further in section 5.

Note that there is no suggestions in the theoretical literature that the parameters of the separation function (for example, β in $f(c_{ij}) = \exp(-\beta c_{ij})$) are, as Cunliffe and Nesbitt (1977) claimed, the same as those that would have been obtained by calibrating to the whole matrix.

4. THEORETICAL CONSIDERATIONS

It will be evident from the description of the theoretical underpinning of the partial matrix technique given in sections 2 and 3 that, for the technique to be successful, there are essentially two issues:

- a. the model needs to be a good fit to the data that one has got;
- b. the parameters that represent the data that one has got need also to be representative of the data one has not got.

Strangely, the question of whether the gravity model is a good fit to the data is rarely discussed in the literature. Some empirical tests have been described by Haskey (1972, 1979). The standard statistical tests (such as χ^2) are biased towards rejection of the model, and therefore suitable techniques need to be developed that are appropriate to the kinds of data typically to be found in trip matrices (Leese, 1977). Although it is particularly important in the context of section 3.3 to be assured of the model's appropriateness, we shall discuss this question no further here.

The second issue, that of representativeness, is discussed in section 7.

There are, however, two other issues that assume particular importance in dealing with partial matrix techniques. These are:

- c. a solution of the form (1), satisfying conditions (2) (for the situation described in 3.1), or of the form (3), satisfying conditions (4) (for the situation described in 3.2) must exist
- d. the matrix which is being synthesised, or to which a model is being calibrated, cannot be split (or disconnected) into two or more independent parts. (The matrix would be a two dimensional one for the situation of 3.1, and a three-dimensional for the situation of 3.2)

Until recently, it had been thought likely that these conditions would not be encountered in practice, and so it is for whole matrices. The conditions are however more likely to be encountered with partial matrix applications; in particular, it has recently been realised that disconnectedness can occur in the calibration (triproportional) problem in such a way as to cause the inferences about the form of the separation function and hence the estimates for unobserved parts of the matrix to be particularly unreliable. This is the subject of the companion paper by Hawkins and Day (1979). The conditions for a solution to exist and the meaning and implications of disconnectivity in the matrix are described in sections 5 and 6 respectively.

5. UNIQUENESS OF THE TRIP ESTIMATES

If we wish to synthesise a trip matrix, as in 3.1, we need to be satisfied that a two-dimensional matrix of the form (1) that satisfies conditions (2) exists. The condition for this is that a solution exists if and only if some two-dimensional matrix exists that satisfies the constraints (2) and contains zeroes where the matrix (f_{ij}) is zero, and is strictly positive elsewhere. Similarly, when calibrating (as for 3.2 and 3.3), the existence of a three-dimensional matrix of the form (3) that satisfies (4) is proven if some three-dimensional matrix exists that satisfies the constraints (4), and contains zeroes where the matrix (Δ_{ijk}) is zero, and is strictly positive elsewhere. (Evans & Kirby, 1974, p.115 and Beardwood & Kirby, 1975, pp.366, 367). Thus these conditions involve considering both the location and number of zeroes in the matrices (f_{ij}) and (Δ_{ijk}) respectively. We note that, in the case of a prediction, if a growth factor method is being used the (f_{ij}) would be a (partial) matrix of observed trips, and thus the zeroes present in it might be either 'structural zeroes' (that is, due to the movement being impossible to observe in the base year), or 'sampling zeroes', (that is, observed as zero by chance). But in the calibration situation, the (Δ_{ijk}) matrix is a defining function, and all the zeroes in it are structural ones.

To illustrate these above conditions, we show in Fig. 4(a) a starting matrix with two empty cells (shown shaded); the remaining cells would have strictly positive entries (it does not matter what their values are) which are to be adjusted (using the Furness, or biproportional procedure) to agree with the row and column totals shown. Fig 4(b) shows one of several arrays of strictly positive entries that can be made in all the other cells so as to satisfy the row and column conditions. Therefore a solution to this gravity model prediction problem exists.

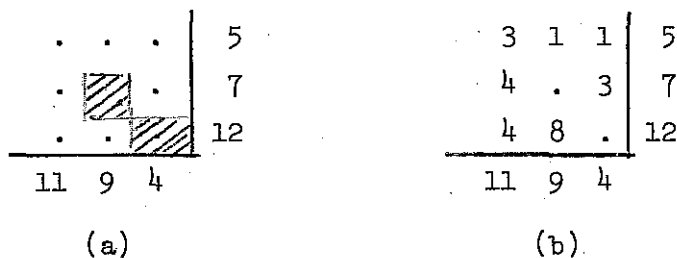


Fig 4. Demonstration of the conditions for a solution to the biproportional problem to exist

For completeness, we should add that the above assumes that the constraints (2) and (4) are consistent; that is, that

$$\sum_i g_i = \sum_j a_j \quad \text{in (2)}$$

$$\text{and } \sum_i g_i = \sum_j a_j = \sum_k s_k \quad \text{in (4).}$$

Since, in calibration, values of (g_i) (a_j) and (s_k) in (4) are determined from a base-year trip matrix, we know that these constraints are consistent. Bacharach (1970, Theorem 3, p.51) also formulate a condition for the convergence of the bi-proportional problem for the situation in which the base matrix (f_{ij}) contains zero terms; but this too is simply a check on the consistency of the constraints.

6. UNIQUENESS OF THE MULTIPLYING FACTORS

In the preceding section, we considered the conditions under which one would obtain unique estimates for the trips in a given cell - the t_{ij}^* or t_{ijk}^* values. On considering instead the estimates of the row (or generation) factors (p_i) , column (or attraction) factors (q_j) and level (or separation) factors (r_k) , which combine to form the trip estimates, a more complicated situation emerges as to the conditions affecting their uniqueness; this we illustrate first by reference to the two-dimensional situation.

6.1 Basic concepts: the two-dimensional solution

It is well known that, in synthesising a matrix of the form (1) satisfying (2), the factors (p_i) and (q_j) are only unique up to an arbitrary multiplying factor: one can multiply each p_i by some scaling factor α , and divide each q_j by the same factor, without affecting the trip estimates t_{ij}^* . Thus, to identify the (p_i) and (q_j) factors uniquely, one of those has to be set arbitrarily to some value. Or rather, at least one: for situations can arise in which more than one factor has to be set to identify the others uniquely. Consider the situation shown in Fig 5 (i), in which opposite quadrants of the (f_{ij}) matrix contain zero terms. On applying a bi-proportional procedure to estimate a model of the form (1) satisfying (2),

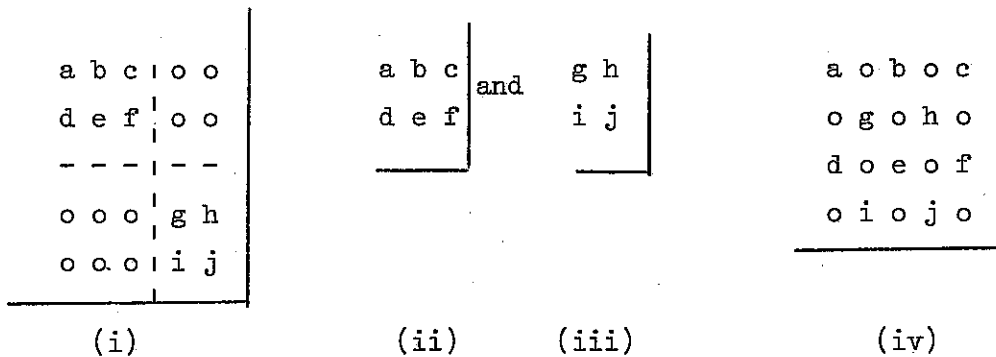


Fig. 5 A disconnected two-dimensional matrix

we find that the trips synthesised in the two non-empty quadrants are independent of each other; we could obtain the same results by separating matrix (i) into two independent parts (ii) and (iii) and synthesise for each separately. For each part, one of the row or column factors needs to be set to identify the others uniquely; that is, in the original matrix (i), two factors (taken from different quadrants) rather than one factor need to be set.

Clearly, in general, it may be possible to separate a matrix into several independent parts. Moreover, it is not always very apparent whether a given matrix has such a structure, as Fig 5(iv) illustrates; yet it has the same structure as Fig 5(i), as may be seen by re-ordering its rows and columns.

A number of names have been used to describe this situation. Thus, if a matrix can be separated into two or more independent parts, it may be said to be separable (Bishop, Fienberg & Holland, 1975, p.182) or disconnected (Bacharach 1970, p.47). The notion of connectedness is also well used by Bishop et al (1975, p.182), and is further explained below. The term separability is not widely used (and we used it in a different sense in Beardwood & Kirby, 1975), so we shall not use it here. Another useful term, that focusses attention on the essential uniqueness properties of the row, column (and level) factors is that of identifiability. A disconnected matrix has an extra degree of freedom for each of the parts into which it can be separated.

6.1.1 Detecting disconnectivity in the two-dimensional situation

A test for disconnectivity in a two-dimensional matrix (or for the non-identifiability of the row and column factors) may be described as follows. Suppose, for some matrix (f_{ij}) with given row and column constraints,

a matrix (t_{ij}^*) is found of the form (1) satisfying those constraints (2). The row factors (p_i) and column factors (q_j) are (we hypothesise) uniquely identified by specifying one of these factors at a particular value. Should we choose to change the value of that (or some other) specified factor, a 'chain' of pro rata multiplications would be set up in the matrix, the effect of which would be to leave the original predictions unchanged. For example, if the first generation factor (p_1) were changed by a multiplier α_1 , the first attraction factor (q_1) would have to be changed by a multiplier $1/\alpha_1$. The difference between a connected and disconnected matrix is that, in the former, the multiplication chain reaches all parts of the matrix, whereas in the latter it cannot. For a matrix to be connected, it must be possible for all cells with non-zero f_{ij} to be linked in a chain, any two consecutive members of which are either in the same row or the same column.

6.3 The three-dimensional situation

The two-dimensional situation described above, in which the factors (p_i) , (q_j) may not be uniquely identifiable, is in fact of little interest in practice, because these factors are not used in their own right: it is their product, in the form $p_i q_j f_{ij}$ that is used in synthesis, and for this uniqueness conditions were discussed in 5.

But in calibration, when we are dealing with a situation in which a set of empirical factors are determined, one for each interval of separation, the factors (r_k) are of primary interest, because they are related to the separation or cost of travelling, and used in forecasting separately from the values of (p_i) and (q_j) . Moreover, with the partial matrix technique it is not sufficient to be satisfied that the product $p_i q_j r_k \Delta_{ijk}$ is unique for the observed parts of the matrix (as was demonstrated in Evans and Kirby 1974); we also want the product to be unique for the unobserved parts of the matrix as well, and this can mean that all the factors (p_i) , (q_j) and (r_k) need to be identified.

Thus, when calibrating to a whole matrix, and even more when calibrating to a partial matrix with subsequent synthesis of the whole matrix, we need to be sure that the three-dimensional matrix (Δ_{ijk}) is not disconnected. In fact, this is rather more likely to occur than with two-dimensional matrices since, for a single mode calibration, every zone-pair has but one cost-interval associated with it; and, for partial matrices of course there are even more zeros.

The phenomenon of a disconnected matrix is rather more complex and difficult to demonstrate in the three-dimensional situation than it is in the two-dimensional one, however. Indeed, it was only when Hawkins of the Department of Transport, was exploring the application of partial matrix techniques to a hypothetical situation that the phenomenon was encountered in the partial matrix situation and interpreted by Day. (Hawkins and Day, 1979). Bishop et al (1975, p 212 et seq) show that the definition of the appropriate measure of connectedness is linked with the definition of the model that is being fitted.

That circumstances might occur in which more than two of the generation, attraction and separation factors (taken from different sets) might need to be specified to uniquely identify them all was recognised by Evans and Kirby (1974, pp 116, 117). As it happened, the main proofs in that paper, on the uniqueness of the trips estimated in the observed cells and on the convergence of the tri-proportional process, were valid whether or not these circumstances held, and unfortunately they expressly excluded further consideration of, for example, the effects on the estimates in the unobserved cells should such circumstances occur.

The effect of disconnectedness on the calibration for a particular matrix (whether whole or partial), is that we could have two or more independent sets of separation factors; within each set, the relative values of the separation factors are correctly determined, but the relative values between factors drawn from different sets is arbitrary. This would mean that, on looking at all the separation factors together, as a function of separation, the shape of the function would in general not be correctly interpreted. Thus, in forecasting, the relative effects on the amount of travel of changes in zone to zone journey costs may be inadequately depicted.

Hawkins and Day encountered this effect on examining the situation in which the parameters (p_i) , (q_j) and (r_k) were not only estimated from a partial matrix of observations, but also used to synthesise trips in the unobserved parts. They found that a change in the values for the separation factors assumed at the start of the iterative process led to differences in the values for the trips synthesised for the unobserved cells.

6.2.1 Detecting disconnectivity in the three-dimensional situation

The identifiability or otherwise of the factors (p_i) , (q_j) and (r_k) cannot be detected by examining the performance of the calibration process as such; the trip generation and attraction constraints for the

zones, and the constraints for the cost bands, are all well met (if the process converges). Hawkins and Day (1979) suggest that the effect could be detected in practice by doing as they did, applying the partial matrix technique to an idealised situation. This requires the synthesis of a trip distribution for the whole matrix, using some arbitrary but non-trivial separation function; followed by a calibration with an empirical function to those parts for which observations exist.

A systematic procedure for detecting non-identifiability is similar in principle to that described in 6.1.1, but is more complex because at least two factors may now be arbitrarily set. Murchland (1978) has provided a rigorous procedure for detecting the effect, dealing with the more comprehensive case in which all possible combinations of i , j and k might occur. Some simplification of this procedure is probably possible, for the usual single-mode situation in which there is only one value of k for a given i , j pair. All that is needed for that situation is a two-dimensional matrix indicating for each observed zone-pair its corresponding cost-interval, with unobserved zone-pairs indicated by a 'zero'.

6.2.2 Avoiding disconnectivity in the three-dimensional situation

An extreme way of avoiding a calibration situation in which disconnectivity occurs is to use an analytic form of separation function rather than an empirical set of separation factors. Thus the calibration process becomes one of estimating, say, the parameters α and β in a function of the form $f(c) = e^{-\alpha c} c^{-\beta}$, rather than with the factors (r_k) ; the three-dimensional situation is itself avoided. Essentially, the analytic function makes the link between the different cost intervals. Whether other, related, kinds of estimation problem might occur is not however yet known for this situation; especially if different parameters are assumed to apply to different parts of the matrix.

Assuming, however, that one wants to explore the empirical shape of the separation function before prescribing an analytic form for it, one might stay with the three-dimensional representation. If disconnectivity is detected, it may then be removed by redefining one (or more) of the cost intervals so as to overlap the costs occurring in the two unconnected portions of the matrix. Such links could be made at different places; and, indeed, this is advisable, to guard against having a loosely-connected structure (see 7).

Alternatively, a modified calibration procedure could be adopted, in which the separation factors are essentially linked with one another by a smoothing process automatically. A method has been suggested by Murchland (1979) which achieves this by associating with each (i, j) pair not just the interval k in which the cost c_{ij} falls, but also (say) the intervals k-1 and k+1. In the calibration process, a weighted average of the factors r_{k-1} , r_k , r_{k+1} would be used to estimate the trips in a given cell. It is understood that such a technique is used by Wootton. Whilst the technique will undoubtedly reduce the risk of disconnectedness occurring, it has yet to be demonstrated whether it will avoid it in all cases.

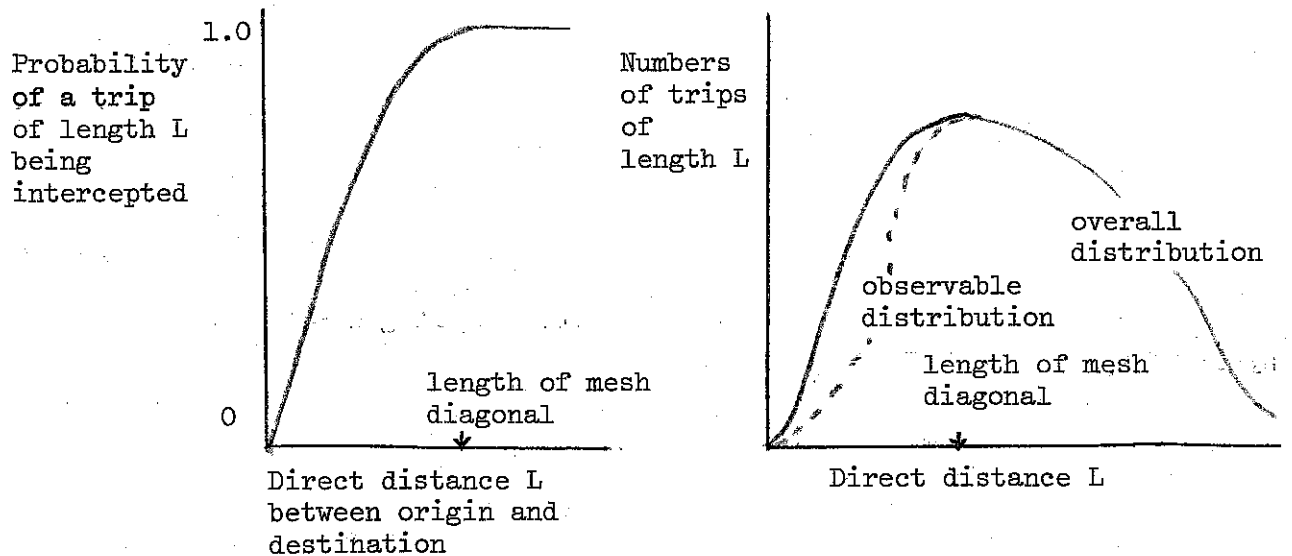
7. REPRESENTATIVENESS AND RELIABILITY

Even for situations in which non-identifiability is not a problem, the estimates for the model parameters may be such as to make the estimates of the trips in the unobserved cells more unreliable than the estimates for the trips in the observed cells*. In other words, the model might be a good fit to the data one has got, but a poor fit to the data one has not got! This is being illustrated in practice by some studies which have found that the estimates of the trip-end totals obtained by application of the partial matrix techniques can be very different from those obtained by the application of trip generation and attraction relationships. Of course, in those studies, there may be incompatibilities in the data sources and relationships which in part account for these differences. But other studies have found that the estimates for the unobserved cells can be very sensitive to changes in the values for the observed trip end totals (Branston, 1978). We shall here review just those issues which might make the application of the partial matrix technique itself unreliable. We discuss these issues in a tentative way, since a clear understanding of all the factors affecting the reliability of the estimates in the unobserved cells has yet to be reached.

*Note We include here both models with an analytic form of separation function, such as $e^{-\alpha c_{ij}}$ $c_{ij}^{-\beta}$, and an empirical set of separation factors (r_k); and by parameters we mean not only the α, β or (r_k) values, but also the (p_i) and (q_j) values.

First of all, we note that, in many partial matrix applications, the data obtained is not necessarily representative of that for the whole matrix. For example, intrazonal trips are generally shorter than, and external-external trips generally longer than those in other parts of the matrix, so that, were no data to be available for either of these movements, the trip length frequency distribution will be distorted from the shape appropriate to the whole matrix.

We illustrate the effect by an idealised situation. Suppose the survey data is collected by roadside interviews where roads cross a square grid of screen-lines. All journeys whose direct distance between origin and destination is greater than the length of the diagonal of the mesh will be intercepted; but, less than this, the shorter the journey, the smaller the probability of being intercepted (Fig.6 a). Therefore, if the trip length frequency distribution for all journeys in the area looked something like the solid line in Fig 6b, the trip length frequency distribution for journeys intercepted by the roadside interview stations will look like the dashed line in Fig 6b, that is, it will under-represent short-distance movements.



(a) The probability of interception

(b) Effects on the trip length frequency distribution

Fig 6. The under-reporting of short distance movements using data from interviews on roads which cross a square grid of screen-lines.

Now this effect is not necessarily important in itself, since the model is fitted only to the observed cells, and it is of course appropriate to make it such as to agree with the trip length frequency distribution for those observed cells. What we are essentially trying to do though is to estimate from a partial matrix values of the parameters that are applicable to the whole matrix. We therefore need to ensure if possible that neither the data structure, nor the model structure, nor the estimation procedure used, introduce particular distortions to the estimates of the parameters. We discuss these in turn.

7.1 Data structure

Clearly, the first consideration is that all the cost-intervals (k) that occur in the unobserved cells also occur in the observed cells, which must include at least one cell with a non-zero observation. But if for a given cost interval (or zone for that matter) the number of cells with non-zero observations is a low proportion of the number of cells observed, and the numbers of cells in the observed part of the matrix is a low proportion of the number in the matrix as a whole, we might have reason to think that errors in the data and thus in the separation factors might be magnified in their effect when used to estimate the values in the missing cells. This effect might be all the more serious if the excluded cells in a given cost interval (or zone) were such that they might be estimated to contain large numbers of trips, but the included cells in that interval contained only small numbers of trips. Such considerations suggest that some simple exploratory analyses of the data might provide helpful insights.

7.2 Model structure and data structure

For the issues discussed here, there are essentially two kinds of problem: well-conditioned ones, that is, those for which small changes to the data input lead to small changes in the estimates; and ill-conditioned problems, in which small changes in data input lead to large changes in the estimates. Branston (1979) of Greater Manchester Council has provided an example of the latter situation. It is of course ill-conditioned problems which we need to be able to detect and if possible cure.

The considerations of 7.1 were suggestive of some exploratory analyses; but these may not adequately identify if and where problems might occur,

or their magnitude or how they might be resolved. Three formal ways of investigating this might be as follows:

- a. Estimate the variances and covariances of the model parameters as well as the means, and thus estimate the variances of the estimates of the number of trips in the observed and unobserved cells.
- b. Perturb the data input and see what effect it has on the calibration.
- c. Analyse the linkages between the different components of the model for the given data structure, and the data itself, in order to identify those parts of the matrix which most affect the trip estimates made for the unobserved cells.

Both (b) and (c) are forms of sensitivity analysis, but with methods of the type (c) - assuming that they can be developed - it should become possible to detect whereabouts more data might be needed in those situations for which the technique seems unreliable; and in due course to learn what kinds of partial matrix structure ought to be avoided.

For (a), methods of estimating the variances and co-variances of the parameters are well-developed in the statistical literature on the analysis of contingency tables, particularly in the field of log-linear models (which is the form that the gravity model takes). Computer packages such as GLIM (Nelder, 1974) and CATLIN (Grizzle, Starmer & Koch, 1969) exist for making such estimates and may be suitable for use on smaller scale problems (Hutchinson, 1977); but it is probable that some adaptation and approximation will be required before the variance estimates can be made for problems of typical transportation study size. It may, however, in certain cases be possible to express the variances in the estimates of the numbers of trips simply in terms of the variances and means of the parameter estimates (Murchland, 1978).

For (b), it is very easy in principle to see what happens if the input data is changed. But it can be time-consuming and costly in practice to do this unless either one perturbs all the data at once (in which case one is unlikely to know which changes had greatest effect) or one has a strategy for selecting those items of data which one suspects might have the greatest effect. It will also be important in practice to have an adequate means of isolating where the main effects of such perturbations are apparent. Perturbations to the data might be best made in such a way as to reflect their relative errors.

For (c), there are at present no procedures known to the author which have been developed, but there are at least two suggestions that have been advanced which might indicate ways forward for the situation in which a set of empirical separation factors are to be estimated. Hawkins and Day (1979) point out that, although a cost-interval matrix may be connected, the connections between different parts of the matrix may be weak; for example, a single occurrence of a particular cost-interval may alone act as the 'bridge' between two parts of the matrix. We can refer to such a matrix as being loosely or weakly connected. Then the relative values inferred for the sets of separation factors that are, in a manner of speaking, on either side of this bridge, are particularly dependent on the number of trips in this cell. If this number is small - and of course it could be zero - the effect of data error on the relative values might be large, leading to a relative instability in the estimates obtained for the unobserved cells.

A second suggestion for a situation which might give rise to a relative instability between the synthesised values for the unobserved cells and those for the observed cells was made by Kirby in the course of discussing the Hawkins and Day problem. It is best described by an example. Suppose that the rows and columns of the cost-interval matrix have been so re-ordered that the unobserved cells are contained in the quadrant as in Fig 7. (We do this for the sake of clarity; it is in general neither necessary nor possible to so re-order the matrix).

| | | | | | |
|---|---|---|---|---|---|
| ③ | ⑤ | 7 | 4 | 7 | 6 |
| ⑤ | 2 | 1 | ③ | 4 | 2 |
| 7 | 1 | 4 | 6 | 3 | 5 |
| 2 | 7 | 2 | 4 | 7 | 3 |

unobserved cells

Fig.7 Cost interval matrix showing differences in the coupling of the factors applying to the observable and unobservable portions

The situation shown in Fig 7 is such that the costs occurring in the unobserved cells (bottom right hand quadrant) occur only in the opposite quadrant. Thus, in calibration, the separation factors for these will be determined in association with generation and attraction factors that do not apply to the unobserved quadrant. This suggests a looseness in the coupling of the different factors that may affect the reliability of the elements synthesised for the unobserved cells. But it should be stressed that, at the present time, it is only a matter of conjecture that a lack of association between these factors might have such an effect.

7.3 Estimation procedure

Finally, we note that, since partial matrix techniques in practice are usually such as to reduce the numbers of observations that we have of shorter-distance movements, the errors associated with those observations are correspondingly greater than they would have been. If the calibration process takes proper account of the presence of sampling variability then there is likely to be greater chance of consistency between the results of a calibration on a partial matrix and a calibration on a whole matrix. Kirby and Leese (1978) showed that assuming (amongst other things) that the survey method is homogeneous, one appropriate method would be to carry out the calibration on ungrossed up data instead of the grossed up data that is normally used.

8. EMPIRICAL QUESTIONS

So far as the author is aware, there has been very little work done to test out the empirical validity of the partial matrix technique. For example, to what extent is under or over estimation in the unobserved cells likely to happen in practice, after calibrating on the observed cells? Are there ways of determining what is a good pattern of roadside interview stations and what is a bad pattern (from the point of view of being able to estimate unobserved movements from this model)? Are there ways of determining where it might be helpful to collect extra data to improve the estimation procedure? Finally, what evidence is there that the model is any good anyway?

An example of the kind of empirical work that would be helpful is a systematic comparison of the estimates for the unobserved cells obtained using the partial matrix technique with those given by the survey data, when different patterns (and numbers) of cells are excluded from a whole matrix (obtained for example from home interview data). Cunliffe and Nesbitt (1977) said (but did not report) that they had made one such comparison, and Hardcastle (1978) has done another. It is not known whether his result is typical, but Hardcastle found that, on excluding intrazonal cells, the separation factors obtained from a calibration to the partial matrix were under-estimated in the short-distance range compared with those found in calibrating the whole matrix.

9. DISCUSSION

In this report we have reviewed the theoretical background to the use of the partial matrix technique, the conditions under which it fails to work at all, and the examination of the circumstances in which it may be particularly prone to error. Some of the problems discussed have only been recently realised and the most appropriate ways of resolving them are very much the subject of current research and debate. It is evident that there is a dearth of literature and research that have investigated the basic properties of the technique, hitherto, and readers with experience in the use of this technique are invited to let the author know of their findings - good or bad! - in the use of this technique - particularly if they have undertaken some of the kinds of analyses suggested in section 8.

10. ACKNOWLEDGEMENTS

The first draft of this article was prepared in response to issues raised by D. Hardcastle, formerly of West Sussex County Council. Its present form owes much to the discussions that have taken place over the problems encountered with the technique by A.F. Hawkins and M. Day of the Department of Transport and by D.M. Branston of Greater Manchester Council. Particular acknowledgement is made of the comments of A.F.Hawkins and Dr. J.D. Murchland; H.F. Gunn, J. Wootton and W.Wyley have also contributed helpfully to consideration of the issues involved.

11 REFERENCES

- BACHARACH, M. (1970) Biproportional matrices and input-output change. Cambridge: University Press.
- BISHOP, Y.M.M., S.E. FIENBERG and P.W. HOLLAND (1975) Discrete multivariate analysis: theory and practice. Cambridge, Mass., London: MIT Press.
- BEARDWOOD, J.E. and H.R. KIRBY (1975) Zone definition and the gravity model: the separability, excludability and compressibility properties. Transpn. Res. 9, 363-369.
- BRANSTON, D.M. (1979) The sensitivity of trip matrices produced by the partial matrix technique to errors in the assumed trip end constraints. Manchester: Greater Manchester Council, County Engineer's Department, Transportation Planning Group, TPG Note 54.
- CUNLIFFE, J.M. and P.G. NESBITT (1977) Bi-modal sparse matrix expansion - an Irish application. Traff. Engng. Control, 18 (12), 572-574.
- DAY, M.J.L. and A.F. HAWKINS (1979) Partial matrices, empirical deterrence functions and ill-defined results. Traff. Engng. & Control,
- DELTRAN ANALYSIS LTD (1977) Report to Kent County Council on the Deal Trip Matrix tests, May 1977.
- van EST, J and J. van SETTEN (1977) Calibration methods: application for some spatial distribution models. Delft: Planologisch Studie Centrum TNO, Working Paper 6.
- EVANS, S.P. and H.R. KIRBY (1974) A three-dimensional Furness procedure for calibrating gravity models Transpn. Res. 8, 105-122.
- GRIZZLE, J.E., C.F. STARMER and G.G. KOCH (1969) Analyses of categorical data by linear models. Biometrics, 25, 489-504
- HARDCASTLE, D.J. (1978) Private communication
- HASKEY, J.C. (1972) Zonal trip flow error analysis. London: Department of the Environment, Math.Adv.Unit., MAU Note 241.
- HASKEY, J.C. (1973) Empirical methods for judging the efficacy of mathematical models for transportation studies. Proc.PTRC Seminar on Urban Traffic Models, held at Univ. Sussex, June, 1973. Vol.1, Paper F 2(i). London: Planning & Transport Research & Computation Co. Ltd.
- HUTCHINSON, T.P. (1977) Application of CATLIN to trip distribution. London: University College London, Transport Studies Group, (unpublished).

- KIRBY, H.R. (1972) Theoretical requirements for the calibration of gravity models. Proc. PTRC Seminar on Urban Traffic Model Research, week 1, London, Feb./March 1972. London: Planning and Transport Research and Computation Co. Ltd.
- KIRBY, H.R. (1974) Theoretical requirements for calibrating gravity models. Transpn. Res. 8, 97-104.
- KIRBY, H.R. and M.N. LEESE (1978) Trip distribution calculations and sampling error: some theoretical aspects. Environ. & Plann., A.10, 837-851.
- LEESE, M.N. (1977) A statistical test for trip distribution models. 9th Ann. Conf. Univ. Transp. Stud.Group., Univs. Glasgow and Strathclyde (unpublished).
- MURCHLAND, J.D. (1978) Perturbation of the bi-proportional model. London: University College London, Transport Studies Group, JDM - 305.1.
- MURCHLAND, J.D. (1979) Non-identifiability in triproportional problems. Univ. Leeds, Inst. Transp. Stud.,JDM-307.1 (unpublished)
- NEFFENDORF, H and H.J. WOOTTON (1974) A travel estimation model based on screenline interviews. Proc. PTRC Summer Annual Meeting, Univ. Warwick, July 1974. Seminar N: Urban Traffic Models. London: Planning and Transport Research and Computation Co.Ltd.
- NELDER, J.A. (1974) Log linear models for contingency tables: a generalisation of classical least squares. Appl.Statist, 23 (3), 323-329.
- NICHOLL, J.P. (1979) A note on the estimation of error in the calibration of full and partial matrices. London: Univ.Coll.London, Transp.Stud.Group (unpublished)
- WOOTTON, H.J. (1972) Calibrating a gravity model and estimating trip ends from a partially observed trip matrix. Memorandum of 21 Dec.1972.