Monograph:
Willumsen, L.G. (1978) Estimation of an O-D Matrix from Traffic Counts – A Review. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Working Paper 99

Published paper

Willumsen, L.G. (1978) Estimation of an O-D Matrix from Traffic Counts – A Review. Institute of Transport Studies, University of Leeds, Working Paper 99

Working Paper 99

ESTIMATION OF AN O-D MATRIX FROM TRAFFIC COUNTS - A REVIEW

L G Willumsen

August 1978

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. This review is part of a research project on the estimation of trip matrices from traffic counts financed by the Science Research Council.

ABSTRACT

WILLUMSEN, L.G. (1978) Estimation of an O-D matrix from traffic counts - a review. Leeds: University of Leeds, Inst. Transp. Stud., WP 99.

Following a review of conventional methods for estimating a trip matrix, the possibility of using cheaper methods based on traffic counts is discussed. Three broad groups of models for such estimation have been identified. The first one assumes that trips follow a gravity type pattern and the problem is reduced to calibrating the parameters of such a model from the observed counts. Depending on the assumed form of the gravity model this technique leads to linear or non-linear regression solutions.

The second group of models attempts to estimate the O-D matrix through a network equilibrium approach based on Wardrop's first principle. The third group follows an entropy maximising approach in which the most likely trip matrix compatible with the observed flows is sought.

Finally, the advantages of each approach are discussed and the most promising technique and areas of application are identified.

CONTENTS

Page

ESTIMATION OF AN O-D MATRIX FROM TRAFFIC COUNTS - A REVIEW

1. THE PROBLEM OF ESTIMATING A TRIP MATRIX

1.1 Objectives

This review is part of a research project financed by the Science Research Council on the estimation of origin - destination trip matrices from traffic counts.

For a large number of applications the existence of a reliable and relatively inexpensive method for estimating O-D matrices would be very valuable. For example:

- modelling transport demand in towns
- modelling, assessment and design of traffic management schemes in rural and urban areas
- cross-checking other methods
- updating old O-D matrices
- modelling rural or inter urban transport demand in sparsely populated areas.
- modelling transport demand where other data is non existent, unreliable or out of date as perhaps in developing countries.

At this stage we are almost exclusively concerned with journeys by car, as are all the methods reviewed here. The plan of this report is as follows:

The remainder of Section one briefly reviews current methodologies for the estimation of O-D matrices. This is followed by a discussion of some basic concepts like trip generation, distribution and assignment models. The Section is closed with a comment on the practical and theoretical problems of estimating a trip matrix from counts.

Section two reviews the first group of techniques, namely linear models. Section three discusses an extension of these into non-linear models. Section four desribes methods based on the assumption of equilibrium assignment and Section five discusses entropy maximising models.

Finally Section six compares the theoretical and practical implications of these four groups of models and makes some recommendations for practical applications. Appendix I summarises the notation used in this paper.

1.2 Conventional methods

There are a number of methods for estimating an origin - destination trip matrix. All but one of them imply sampling and even if all trips on a day are recorded the problem of stability of the O-D matrix over time still remains. "Real" trip matrices then are never available; we only have more or less reliable methods for estimating them.

Some of the techniques available at present are:

(a) Road side interviews of drivers, usually at the exit of a cordoned area. Sampling fractions of 1 in 5 may be considered typical but they certainly depend on traffic levels and manpower available. They are usually asynchronic, that is, they are carried out on different days but at similar times. They tend to be expensive in manpower, delays to vehicles and processing.

(b) Home interview. Although usually not considered reliable enough for a detailed design of, for example, a traffic management scheme, they may be the result of a larger study and be available for cross-checking. The method is expensive on manpower and time consuming.

(c) Flagging methods. It is possible to identify vehicles at cordon and internal points by means of registration numbers, stickers or by asking drivers entering at a given point to switch their lights on. This last method has only been tried for very small areas (roundabouts), see for example Bebee (1959). The number plate method is of wider usage and also requires a good deal of manpower and processing effort. Usually drivers do not incur extra delays.

(d) Aerial photography. To a large extent this method is experimental, it requires a good deal of processing to identify cars and track them in a computer readable form. It is only successful for small areas but improvements in the automatic identification of vehicles may increase their applicability. It is the only method which does not necessarily require sampling.

(e) Car following. Suggested by Wright (1977) and tried in Westminster, this method has been claimed to be cheaper than other methods. Its application is probably restricted to the central areas of large cities.

In contrast, we are interested in this review in methods for the estimation of O-D matrices based on a description of the network and traffic counts. In addition some of the methods require some widely available zonal data, like population and employment. Traffic counts are relatively simple to obtain and further progress in automatic recording and classification is envisaged. We suspect that not enough use is made of this information.

1.3 Basic concepts and definitions

Most of the vocabulary in this review is borrowed from conventional transport demand models.

In a conventional transport modelling exercise the study area is usually divided into zones which are considered the generators and attractors of trips. The modelling process usually proceeds in a sequence of four submodels:

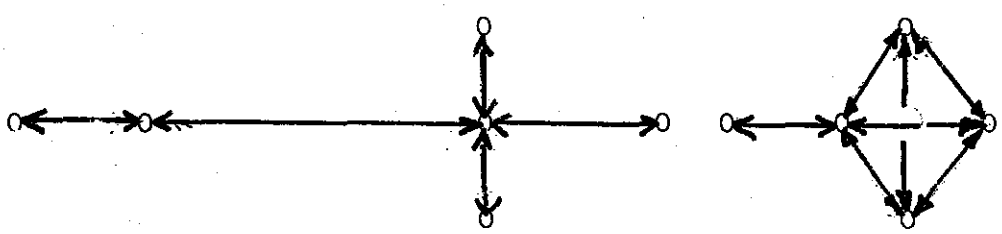
- Trip generation which uses zonal data to model the number of trips being generated from and attracted to each zone.
- Trip distribution which synthesises the origin-destination matrix, that is the number of trips from each zone to each zone. It is common practice to use here a gravity type model. This, in general, assumes that the number of trips between i and j are proportional to the generation and attraction power of each zone and proportional to some decreasing function of the cost (in time, distance or monetary terms) of travelling between them. This decreasing function is usually called the deterrence function of the model.
- Modal split in which the choice of mode made by each traveller is simulated. We shall not be concerned with this submodel in this review, as we restrict our attention to car traffic.
- Assignment, in which the route followed by each trip is modelled. The output from this stage includes link flows and a revised measure of the costs of travelling between each pair of zones.

As this assignment stage is very relevant to our problem we will discuss it in some detail.

1.4 Assignment

A road network is usually described in terms of a collection of points, called nodes and a collection of ordered pair of nodes called "real links" or simply "links". A link is usually deemed to represent a real section of a road, for example from one important intersection to another. Most of the delays in urban areas occur at intersections. Thus, in some cases it is worthwhile to define a network in such a way that certain links represent particular turning movements at an intersection. Figure 1.1 represents a way in which an intersection could be represented by links rather than modes.

Figure 1.1 "NODE" AND "LINK" REPRESENTATION OF AN INTERSECTION



intersection represented by one node.

intersection represented by 6 two way links and 4 nodes.

Of course with a "link" representation the number of nodes and links is greatly increased. On the other hand, it is customary to count traffic precisely at intersections recording each turning movement independently, thus this representation is of interest.

Origin and Destination zones may be considered as a sub-net of the nodes, generally referred to as "centroids". The arc linking a "centroid" with the "real link" network is usually called a "centroid connector". Each link has a travel cost associated with it (time, distance, money or generalised cost) which, due to congestion, depends on the total flow on that link expressed as

$$Ca = Ca(Va) \tag{1.1}$$

$$C_a = C\ (Va) \qquad (1.1)$$

where a is a link

C_a is a link cost

V_a is the traffic volume on link a

For a good discussion of the problem of traffic assignment the reader is referred to Van Vliet (1976). Here we shall only describe briefly three alternative approaches.

(a) All-or-nothing assignment via the minimum cost route. If we assume that all drivers travelling between two zones choose the same route it seems only logical to assume they travel via the (apparently) cheapest route (a sequence of links). In the absence of congestion or if there are no real alternative routes it is possible to determine the set of minimum cost routes before making any assumption about the number of trips using them.

(b) Proportional assignment. If we accept that not all drivers "perceive" a route cost to be the same we must allow in the assignment process for these differences. In this case the proportion of drivers choosing each of a reduced set of near optimum routes will be determined by assumed drivers and route characteristics. These methods currently do not consider congestion effects so these proportions are treated as independent of link flow levels.

Following Robillard (1975) it is possible to strictly define proportional assignment as a technique in which:

(i) The total assigned flow on a link equals the sum of the assigned flows obtained when the method is applied to each O-D pair separately.

(ii) If all the entries of the O-D matrix are multiplied by a constant factor λ all the assigned flows on each link will increase by that factor.

Clearly these conditions are not realistic when congestion plays an important role in assignment. On the other hand, the advantage of these assumptions is that it is possible to determine the proportion of the trips between i and j using a link before any guess is made about the O-D matrix. These proportions depend only on the network and the parameters of the assignment algorithm but not on flow levels at each link.

(c) Equilibrium assignment. Wherever congestion effects are thought to be more important than the differences in perceived travel costs a different type of assignment is considered realistic. This type tries to satisfy Wardrop's first principle which can be stated as:

Traffic on a network distributes itself in such a way that the travel costs on all routes used from any origin to any destination are equal while all unused routes have equal or greater costs.

In other words an assignment is said to have reached equilibrium when no driver can reduce his travel cost by switching to another route.

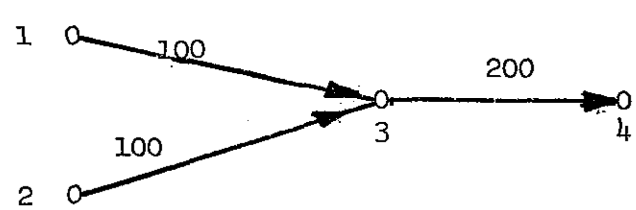
Methods based on both heuristic and mathematical programming techniques are used to obtain this type of assignment. In this case the proportion of trips between an O-D pair using each route depends on the actual flows on each link.

1.5 Reliability of traffic counts.

There are a number of problems associated with traffic counts and we shall review some of them.

(a) Independence. Not all traffic counts contain the same amount of "information". For example in Figure 1.2 traffic link 3-4 is made up of the sum of traffic on links 1-3 and 2-3. Counting traffic on link 3-4 is then redundant and only two counts there can be said to be independent.

Figure 1.2 DEPENDENT COUNTS



In mathematical terms this can be expressed by writing for each link in the network:

$$\left.\begin{aligned} V_1 &= \textstyle\sum_{ij} T_{ij} p_{ij}^{\,1} \\ V_2 &= \textstyle\sum_{ij} T_{ij} p_{ij}^{\,2} \\ &\text{-------------------} \\ V_\ell &= \textstyle\sum_{ij} T_{ij} p_{ij}^{\,\ell} \end{aligned}\right) \qquad (1.2a)$$

Dependence in the flows occurs where it is possible to express one of these equations as a linear combination of some of the other equations.

This is likely to happen in situations such as that depicted in Figure 1.2 where a flow continuity equation applies

$$V_{13} + V_{23} = V_{34} \qquad (1.2b)$$

The existence of a source or sink of flow (centroid connector) at 3 will immediately remove this dependence.

(b) Inconsistency. Counting errors and asynchronous counting are likely to lead to inconsistencies in the flows. In other words, the expected continuity relationships will not be met. As we will see some models can readily cope with this problem while others require these inconsistencies to be removed beforehand.

(c) Counting periods. Depending on the counting method it is possible to obtain traffic flows for periods as short as 15 minutes or as long as 24 hours. An O-D matrix derived on this basis will only be relevant if the trip time within the study area is significantly shorter than the counting period. For inter urban studies it will usually be possible to estimate the average Annual Daily Traffic (AADT) and use it in the models.

(d) Symmetry. Flows are in general asymmetric, that is $V_{ij} \neq V_{ji}$. In most inter urban applications, though, it is possible to assume symmetry in the counts and in the O-D matrix thus reducing the complexity of the models.

1.6. Estimability of a trip matrix from counts.

Let Tij be the number of trips from zone i to zone j, and p_{ij}^{a} be the proportion of trips from i to j travelling on a link. For each link we can now write the equation

$$Va = \sum_{ij} T_{ij} p_{ij}^{a} \tag{1.3}$$

where p_{ij}^{a} is the proportion of trips from i to j using link a.

In a study area with N centoids (zones) and assuming p_{ij}^{a} to be known we would need N^2 independent equations of the type (1.3) to uniquely define the whole set of $[Tij]$ If we assume that Tij = 0 for i = j we would only need N(N-1) equations. However, in most cases the number of independent link counts will be less than this figure. The most important implication of this is that traffic counts will normally be insufficient to uniquely determine an O-D matrix $[Tij]$. Thus, in order to obtain an estimation of the O-D matrix some assumption will have to be made about trip making behaviour.

As we will see when reviewing the proposed models, a frequent approach is to assume some sort of gravity type behaviour in trip making. In this way the actual number of unknowns is reduced to the number of parameters for calibration in the assumed model.

If the number of unknowns is smaller than the number of independent equations (link counts) errors have to be assumed in the model or the observations in order to solve the system. For example

$$Va = \sum_{ij} (T_{ij} p_{ij}^{a}) + \varepsilon_a$$

where ε_a is an error term with assumed mean 0. The solution is then found by minimising the supposed errors, usually by means of multiple (linear or otherwise) regression analysis.

Most computer implementations include software to solve this type of problem.

1.7 Criteria for selecting a method of estimation.

For short term applications, (when the assumption of a fixed trip matrix is realistic), the main criterion is probably accuracy.

In order to test the accuracy of a model the ideal case would be to compare the estimated O-D matrix with the "real O-D matrix" which produced the counts used by the method. However, we did not find any such test in the literature, no doubt because a "real" O-D matrix is difficult to come by. Instead, a more common indicator of the accuracy of a model, is to compare observed and predicted link flows via, for example

$$\text{RMS error (Root mean square error)} = \left[\frac{\sum (\text{OBSERVED} - \text{PREDICTED})^2}{n-1} \right]^{\frac{1}{2}}$$

expressed either in absolute value or as a percentage of the average observed value. This measure could also be used to test the accuracy of each cell in the O-D matrix and this test would be much more significant.

In addition, we would like to know how acceptable in each case are the assumptions implicit in each method. For example, is it possible to expect a gravity type of travel behaviour? Is it sensible to disregard congestion effects and assume route choice to be independent from flow levels?

Finally, if we are interested in medium to long term implications we would like to know if the implicit travel demand model is sufficiently policy sensitive for our purposes and if it is possible to have a reliable forecast figure for the planning parameters in the model (Population, employment, etc.).

We shall return to these considerations in Section six.

1.8 Overview of traffic count based methods.

We have only found a restricted number of publications in the field. We suspect, though, that other researchers and practitioners have developed and used similar methods without reporting them in technical journals.

We have grouped the proposed methods into three broad groups:

The first assumes that a gravity model is capable of explaining most of the travel behaviour in the study area. The underlying gravity model can be presented with several degrees of sophistication and this has led us to consider two sub-groups:

- Gravity models leading to linear equations on the links
- Gravity models leading to non-linear equations on the links.

Most of the proposed methods and all the applications with real data fall into this broad first category.

The second group of models attempts to estimate an O-D matrix through a network equilibrium approach based on Wardrop's first principle. These are described in Section 4.

The third group of models considers methods for estimating the likely trip matrix compatible with the observed traffic counts. These methods could be referred to as entropy maximising or information minimising approaches and are described in Section 5.

2. LINEAR MODELS

2.1 Introduction

The idea of using a simplified model for travel demand is not new. Probably the first such model calibrated from link flows was proposed by Low (1972). The basic idea is to use a simple linear model of the gravity type like

$$T_{ij} = \alpha P_i E_j d_{ij}^{-2} \tag{2.1}$$

where P_i, E_j = population, employment in zones i and j.

d_{ij} = distance or travel time between i and j.

α = parameter for calibration.

In this model α plays a combined role, as trip generation-attraction parameter. Then, in order to calibrate α the modelled trips must be translated into link flows to compare them with observed counts. This implies an assumption about the type of trip assignment considered to be realistic. The simplest assumption is to accept all-or-nothing assignment, that is

$$p_{ij}^{a} \begin{cases} 0 & \text{if trips from i to j do not travel on link a.} \\ 1 & \text{if they do.} \end{cases}$$

Then

$$Va = \sum_{ij} T_{ij} p_{ij}^{a} = \alpha \sum_{ij} (P_i E_j d_{ij}^{-2}) p_{ij}^{a} \qquad (2.2)$$

in which all parameters except α are known. It is then necessary to determine α in such a way that the difference (or the sum of the squares of the differences) between modelled and observed flows is minimised. This can be achieved here using standard linear regression techniques.

This chapter reviews different linear models as suggested by Low, Overgaard, Lamarre and Jensen et al. They differ mainly with respect to:

- The independent variables
- The deterence function
- The type of assignment.

This is the family of models with the greatest number of applications, mostly in rural or inter-urban areas.

2.2 D.E.Low's Model

2.2.1 Description. The main objective of Low's approach is to "effectively combine into one process what is usually handled in a series of three or four sub models, each with its own set of errors." (Low 1972) Attempts to develop a model whose direct output is traffic volumes and where modelling errors, which appear in the same terms, can be easily described and assessed.

Low sees his model as applicable, in principle, to a large range of study areas but especially to small ones (around 50,000 inhabitants).

1.8 Overview of traffic count based methods.

We have only found a restricted number of publications in the field. We suspect, though, that other researchers and practitioners have developed and used similar methods without reporting them in technical journals.

We have grouped the proposed methods into three broad groups:

The first assumes that a gravity model is capable of explaining most of the travel behaviour in the study area. The underlying gravity model can be presented with several degrees of sophistication and this has led us to consider two sub-groups:

- Gravity models leading to linear equations on the links
- Gravity models leading to non-linear equations on the links.

Most of the proposed methods and <u>all</u> the applications with real data fall into this broad first category.

The second group of models attempts to estimate an O-D matrix through a network equilibrium approach based on Wardrop's first principle. These are described in Section 4.

The third group of models considers methods for estimating the likely trip matrix compatible with the observed traffic counts. These methods could be referred to as entropy maximising or information minimising approaches and are described in Section 5.

2. LINEAR MODELS

2.1 Introduction

The idea of using a simplified model for travel demand is not new. Probably the first such model calibrated from link flows was proposed by Low (1972). The basic idea is to use a simple linear model of the gravity type like

$$T_{ij} = \alpha P_i E_j d_{ij}^{-2} \qquad (2.1)$$

where Pi,Ej = population, employment in zones i and j.

d_{ij} = distance or travel time between i and j.

α = parameter for calibration.

His approach consists of three steps:

(a) Assignment of external trips. Current external trips as obtained from external cordon roadside interviews are assigned to the existing network to produce estimates of "current external volumes" throughout the network.

(b) Estimation of "current internal volumes". The so called "current internal volumes" are then obtained from the actual counted flows less the assigned "current external volumes". These "current internal volumes" are the result of trips "wholly explainable in terms of area characteristics."

(c) Internal volume forecasting model.

(i) Inter zonal trip opportunity matrices of the form $O_i D_j$ are developed. For the journey to work Population and Employment are used.

(ii) Each element of the trip opportunity matrix is then multiplied by a friction factor of the type C_{ij}^{-m}. In general time is suggested as measure of cost. The product $O_i D_j C_{ij}^{-m}$ is called then an "inter zonal trip probability factor" and the full set of factors is the trip probability matrix. Several trip probability factors, say for different journey purposes and person types, may be so developed.

$$X_{nij} = O_i O_j C_{ij}^{-m} \qquad (2.3)$$

Where X_{nij} = trip probability factor between i and j for person type or journey purpose n.

(iii) Trip probability matrices are then assigned separately to the current network (in practice an all-or-nothing assignment) just as if they were trips.

For a link a

$$X_{an} = \sum_{ij} p_{ij}{}^{a}X_{nij} \qquad (2.4)$$

where p = (1 if the least cost route from i to j
(passes through link a
(0 otherwise.

(iv) Multiple regression techniques are used to develop equations of the following form

$$V^1 = b_o + b_1X_1 + b_2X_2 + --- + b_nX_n \qquad (2.5)$$

where V^1 is the internal traffic volume on link a and b_k are the constants to be obtained.

(v) Low suggests that separate equations can be developed for different types of roads or areas or that additional parameters defining the characteristics of the link may be included.

2.2.2 Scope and applications. It is suggested that the model could be used for forecasting purposes as well as replicating the present pattern of trips. To this end the value of the socio economic parameters should be obtained for the design year as well as the alternative networks that would be tested.

Low also suggests that more sophisticated versions of his model can be developed, especially in relation to the way in which the trip probability matrix is obtained.

No provision is made for public transport trips or for modal split.

The model was used during the Monongalia County Transportation Study in West Virginia in 1970-71. Conventional techniques were used in parallel to this model.

The basic information used was

- Base year population and employment by zone (52 zones)
- Base year road traffic assignment network
- Base year traffic counts
- Base year external trips table

Finally, tests were made with multipath assignment techniques to see if it was possible to improve the fit of the model. The model used was the standard UMTA multipath assignment program UROAD. Only marginal improvements were obtained with this approach.

A conventional model was not applied to the area so it was not possible to compare their accuracies. Nevertheless, the level of accuracy of the IVF model in reproducing base year link volumes was "certainly within the limits of conventional models." This is to some extent surprising as only population and zone to zone travel times are included in the model. The reported correlation coefficient was $R = 0.87$ and the RMS values were 209 in absolute terms and 53% in percentage terms.

In order to evaluate the forecasting ability of the model empirically the model was applied to "forecast" 1960 traffic volumes using the 1960 Fond du Lac network and census data. The results of this exercise showed that overall, the "absolute error of the 1960 estimate of assigned volumes was less than that of the 1970 estimates."

The researchers, though favourably impressed by the accuracy of the IVF model, were critical of its theoretical basis. Smith and McFarlane pointed out three basic mis-specification errors in the model.

(a) Changes in the trip making propensity of the study area population over time cannot be considered in the model. For instance, the impact of car ownership is not reflected in the model.

(b) The unconstrained nature of the model. If population is doubled traffic flows will be multiplied by a factor of four instead of just doubling as would be expected. This criticism applies mainly to the model used in the forecasting mode. This objection is easily overcome by the inclusion of a "normalising factor" equal to $1/\sum_j O_j$ in equation (2.3).

(c) The different measures of the trip probability factors are likely to be co-linear in which case only one probability factor should be included in the equation.

The researchers considered several possible improvements to the model, some of which are described later in this working paper as they appear in the literature. They finally concluded that the inclusion of these improvements would probably made the IVF model far too similar to conventional techniques and its original attractiveness would be lost.

2.3 Overgaard's Model

2.3.1. Description Overgaard's, Low's and the Danish Road Directorate methods were discussed in the OECD Report on possibilities for the simplification of urban traffic models OECD (1974).

Overgaard's model, though probably developed independently, can be considered as an improved version of Low's approach. As it includes a proxy measure for trip making propensity it goes some way to answer some of the criticisms of Smith and McFarlane. This model has been applied to Silkeborg, a Danish town of 44,000 inhabitants.

External trips are obtained in the same way as in Low's method. The main change is regarding trip generation:

Car ownership levels were known for Silkeborg so that trip generation equations were stated in terms of trips per car per day (It has to be remembered these methods are only concerned with car movements). Furthermore, as a proxy for socio economic level, the type of dwelling ("single family house" or "apartment") was also introduced in the trip generation term.

The general trip generation (and attraction) term was then expressed in the form

$$O_i = b_1 E_i + b_2 P_i + b_3 (P_i \ \delta_i) \tag{2.7}$$

where O_i = generation/attraction force in zone i

E_i P_i = employment and population in zone i

δ_i = percentage of population in "one family houses"

This trip generation model was calibrated independently at two different zones.

The other main difference with Low's approach is that Overgaard uses two different deterrence functions for motorised trips of less and more than 90 seconds of duration.

2.3.2 Applications. When this model was applied in Silkeborg it was found that

b_1 = 1.75 [trips per workplace]

b_2 = 0.7 [trips per inhabitant]

b_3 = 0.008 [trips per inhabitants in "one family houses"]

gave the best results.

Also, an exponent of -1.8 for travel time was calibrated for trips with t > 90 secs. The "accuracy" of the model in terms of the comparison of observed and predicted link flows was measured by an RMS (percentage) error. This was found to be around 20% and better for higher flows.

2.4 Lamarre's Model.

2.4.1. Description. In general linear models are restricted in terms of the number of parameters for calibration. For example, the calibration of the trip generation - attraction and distance exponent parameters is made in steps. A suitable, and sounder, alternative method would be to use a non-linear regression technique as we shall see in Section 3.

Lamarre (1977) developed and tested a simple linear model incoroporating calibratable parameters affecting socio-economic and travel time indicators. Lamarre developed two alternative formulations. The first one is an "O-D based" formulation as follows

$$Tij = bo + b_1(Oi + Oj) + b_2\, C_{ij} \tag{2.8}$$

and

$$Va = bo \sum_{ij} p^a_{ij} + b_1 \sum_{ij} (O_i + O_j)\, p^a_{ij} + b_2 \sum_{ij} p^a_{ij}\, Cij \tag{2.9}$$

where Oi, Oj are indicators of the generation and attraction "power" of each zone and c_{ij} is the travel time between zones. His second model is a "link based" formulation where

$$Va = bo + b_1 \sum_{ij} (Oi + Oj) \; p^a_{ij} + b_2 \sum_{ij} p^a_{ij} . c_{ij} \qquad (2.10)$$

It can be seen that the main difference between the two formulations lies in the assumption of a constant term bo in either each O-D pair or each link. Lamarre developed this model for the analysis of inter urban trips. He used data from cities at the interior of Quebec and the relevant network description and traffic counts of that area. Lamarre apparently performed most calculations without using an existing computer programme (for example his shortest route is determined by hand and the number of trips between certain O-D pairs are more or less arbitrarely assigned 50% to each of two paths).

2.4.2. Results. Lamarre tested several socio-economic indicators as independent variables including population, employment, total income, car ownership, French speaking population, etc. He also tested his model for summer and winter trip making. His main findings can be summarized as follows

(a) The "link based" formulation always gave a better fit than the "O-D based" one.

(b) The travel time elasticities b_2 were found to range between -2.0 for the "O-D based" and -0.4 for the "link based" formulations.

(c) The elasticities associated with population and car ownership were not satisfactory (wrong sign) and this made Lamarre doubt the soundness of this linear model. His best coefficient of determination was $R^2 = 0.72$.

He acknowledges that the results, on the same network, of a non-linear model by Wills (see section 3.4) were superior to his two linear models.

2.5 Commonwealth Bureau of Roads model

2.5.1 Description. This approach was developed for the Commonwealth Bureau of Roads (CBR) in Australia by Symons, Wilson and Paterson (1976). The objective of the exercise was "to devise a rigorous methodology for disaggregating the Australian National System of Urban Centres into distinct zones of economic activity; then, to use, this zonation to construct a model of traffic generation for the National Highway System". This approach is only applicable to inter-city travel.

The main innovations of this approach resides at the trip generation stage. A relationship is assumed between traffic generation and the urban hierarchy which is determined by the provision of services to lesser urban centres. A hierarchy of urban centres is produced using Central Place Theory considerations and it is assumed that trips will be generated from a low order centre to a higher order centre only. The trip generation rates per head of population are expected to vary from level to level. Each of the Australian urban centres was then assigned to one of seven broad ranks using population and employment type statistics. As explanatory variables both population and a recreational attractiveness index were used. The model could then be expressed as a group of sub models.

Trip Purpose	Functional Form (F_r)
Seeking level 7 services	$P_i \, C_{ij}^{-k}$
Seeking level 6 services	$P_i \, C_{ij}^{-k}$
Seeking level 5 services	$P_i \, C_{ij}^{-k}$
Seeking level 4 services	$P_i \, C_{ij}^{-k}$
Recreation	$P_i \, R_j \, C_{ij}^{-k}$
Inter-capital	$P_i \, P_j \, C_{ij}^{-k}$

Where P_i = Population urban centre i.

R_i = Recreational index urban centre i.

C_{ij}^{-k} = Deterence function.

F_r = Functional form for journey purpose r.

The general model is then

$$T_{ij} = \sum_r F_r b_r \qquad (2.11)$$

and $$V_a = \sum_{ij} T_{ij} P_{ij}{}^a$$

that is

$$V_a = \sum_{ij} \sum_r b_r F_r P_{ij}{}^a$$

$$V_a = \sum_r b_r \sum_{ij} F_r \; P_{ij}{}^a \qquad (2.12)$$

where equations (2.12) can be used to determine the calibration parameters b_r for each journey purpose. Of course an implicit assumption in this model is that route choice does not depend on journey purpose and that proportional assignment is sufficiently realistic.

In order to simplify calculations only trips between State capitals and trips between a centre and the nodes in its market area were considered. It was felt that this set of trips would capture the bulk of inter urban traffic.

2.5.2 Application. In the Australian case assignment was performed manually and multiple linear regression techniques were used to calibrate the model. In addition different exponents k for the deterrence function for each journey purpose were tested and those giving best fit were adopted. They found that the smallest exponent was that of trips to level 5 and suggested that perhaps the most essential and irreplaceable goods and services were supplied by these centres. The researchers also found a high exponent for inter-capital trips. They explained this by the impact of competition from other modes (air and rail services).

Symons et al concluded that the model was to be of great interest to State Highway Authorities for it indicates how much information may be extracted from census information and road counts. They were satisfied that the results were encouraging and statistically sound. They envisaged no conceptual difficulties in transforming the model into a "robust predictive tool". The researchers mentioned a number of desirable improvements in that direction mainly in terms of automating processes within the model.

2.6 Danish Road Directorate approach.

2.6.1. Description. This is probably the method with the greatest number of reported applications. At least two full scale applications in rural areas, Jensen & Nielsen (1973) and Holm et al (1976), have been reported in English. The method has been improved since its first conception but only the version described by Holm et al (1976) will be discussed here.

In general the method is again seen as a way of directly calibrating a trip-generation-distribution-assignment model from traffic counts. Probably the main point of interest is the use of an iterative assignment technique instead of the simpler all-or-nothing via minimum cost path approach of previous researchers.

In general "external trips" are not obtained independently but as a result of the use of the model. In effect, the "external zones" are coded in the same way as the "internal" ones and teated in the same way.

The basic trip generation-distribution model is:

$$T_{ij} = bO_i O_j c_{ij}^{-m} \qquad (2.13)$$

where b is the trip generation factor.

The problem is then to determine a value for b consistent with the iterative assignment and which minimises the difference between observed and modelled link flows.

An iterative assignment has been chosen, hopefully converging towards a Wardrop's equilibrium, and consequently the p_{ij}^a values change at each iteration. The Danish Road Directorate chose Smock's (1968) iterative algorithm for assignment and developed a method for calculating b as part of the iterative process.

2.6.2 The iterative process. The "cost-flow" relationship for a link a is assumed to be

$$Ca = Ca(Va)$$

Then all or part of the traffic between each pair of zones is assigned for routes in such a way that, for each link, the flow at iteration n is

$$Va^n = (1 - \frac{1}{n})\, Va^{n-1} + \frac{1}{n}\, Va' \qquad (2.14)$$

where $V'a$ = traffic volume resulting from all-or-nothing assignments to the link after the nth iteration.

We can now create an auxiliary variable X_a (analogous to Low's Xa) such that

$$X_a = \sum_{ij} O_i \, O_j \, Cij^{-m} \cdot p_{ij}^{a} \qquad (2.15)$$

where p_{ij}^{a} = proportion of traffic T_{ij} which uses link a.

At the same time

$$Va^n = bXa^n = b(1-\frac{1}{n})Xa^{n-1} + \frac{1}{n}bXa \qquad (2.16)$$

by making

$$Xa^{n} = (1-\frac{1}{n})Xa^{n-1} + \frac{1}{n}Xa \qquad (2.17)$$

At this stage, it would be possible to calculate, after each iteration, the value of b using linear regression, and we could accept that convergence has been achieved when the estimations of b do not differ significantly in two consecutive iterations.

The Danish Road Directorate however preferred a maximum likelihood method. It is assumed that the observed flows $\hat{V}a$ are mutually independent, normally distributed variables with mean Va^n, that is

$$\hat{V}a = Va^n + \varepsilon_a \qquad (2.18)$$

where ε_a is a normally distributed variable with mean 0 and variance $[(Va^n)^{\gamma}\sigma^2]$ †

† γ = o means that the standard deviation is independent from the mean.

γ = 1 means that mean and variance are in proportion.

γ = 2 means that average flow and standard deviation are proportional.

Using the maximum likelihood method it is possible to find the best estimator for b as

$$b^* = \frac{\sum_a \hat{V}a(X_a{}^n)^{1-\gamma}}{\sum_a (X_a{}^n)^{2-\gamma}} \qquad (2.19)$$

and also

$$\sigma^* = \frac{1}{n-1} \sum_a \frac{(\hat{V}a - Va^n)^2}{(Va^n)^{\gamma}} \qquad (2.20)$$

The derivation of these relationships is shown in Appendix II.

The algorithm developed by the DRD can be summarised as:

Step 1 = Calculate free flow travel times.

Step 2 = Determine minimum cost routes for all O-D pairs.

Step 3 = Assign following Smock's algorithm, the χ'_{ij} = $O_i O_j\ c_{ij}^{-m}$ factors to the network.

Step 4 = Estimate b* with the maximum likelihood formulae.

Step 5 = Calculate traffic volumes on all links as $b^*\chi a$.

Step 6 = Calculate new travel times.

Step 7 = If b* has not converged proceed to step 2, otherwise stop.

This algorithm is tested with different values of the exponent γ and m in a manner similar to Smith and McFarlane (1978).

The model is intended mainly as a replacement for more sophisticated and expensive forecasting models.

2.6.3 Applications. The model has been applied, in slightly different versions, in the Aarhus and in the South Zealand rural areas in Denmark. In this last case the data used consisted of a road network, with seven types of roads (and speed-flow relationships), parish populations and about 40 counting points. The network included 73 zones and 334 links. It was found that $\gamma = 1$ and $m = 3.25$ gave the best fit.

The researchers found that the model tended to over estimate large flows and under estimate small volumes. This is certainly not very desirable from a planning view point. They thought the source of this error was the fact that intrazonal trips were not considered by the model.

The researchers estimated the percentage error in flows on the links was around 17% and that this did not deteriorate much when the model was calibrated with only half the counts.

The researchers also prepared confidence intervals for the curves. They found that the half-width of the 95% confidence interval was, for example, 5200 pcu/day at 10.000 pcu/day and about 3000 pcu/day at 5.000 pcu/day levels. They concluded that this accuracy was no worse than that obtained through the use of more expensive traditional modes.

In a personal communication one of the researchers involved informed the author that the Danish Road Directorate was now working on a non-linear regression approach, similar to those described in the next section.

3. NON-LINEAR REGRESSION APPROACH

3.1 Introduction.

Several techniques for solving non-linear regression problems are available and implemented in most computers. A non linear regression problem normally takes longer to solve than a linear one with the same number of variables. However, the non linear regression approach, allows the use of more realistic and generally accepted implicit gravity models in the estimation of Origin-Destination matrices.

We will discuss here three such methods. The first one by Hogberg (1976) assumes that an all-or-nothing minimum cost assignment is realistic enough but incorporates a very general deterrence function and up to three "journey purposes" in the gravity type distribution model. The second approach by Robillard (1975), uses a generalized gravity model and accepts proportional assignment as realistic. The third model has been developed by Wills (1977) and also includes a generalized gravity model but of a different kind.

It is possible to extend all linear and non-linear models (Sections 2 and 3 in this work) to the proportional assignment case but even then they fail to consider congestion as an important factor in route choice. This seems to indicate that these methods could be more appropriate to rural or uncongested urban areas.

3.2 Hogberg's Model

Hogberg assumes that the joint generation-distribution model is of the form

$$Tij = b_1 O_i^1 A_i^1 D_j^1 f(c_{ij}) + b_2 O_i^1 A_i^2 D_j^2 f(c_{ij}) + b_3 Oi^2 A_i^2 D_j^2 f(c_{ij}) \tag{3.1}$$

where

O_i^1 and O_i^2 are different types of trip generation parameters

D_j^1 and D_j^2 are different types of trip attractions parameters

A_i^1 and A_i^2 are balancing factors of the form $A_i^1 = \left[\sum_j D_j^1 f(c_{ij})\right]^{-1}$ etc.

and $f(c_{ij}) = c_{ij}^{b4} \exp\left(b5 \left(\log_e (C_{ij})^2\right)\right)$

Using the indicator variable p_{ij}^a* we can state the flow at each link as

$$Va = \sum p_{ij}^a T_{ij} \tag{3.2}$$

---

* In the original work p_{ij}^a could only take values 0 or 1. Here, it can take any value p with $0 \le p \le 1$.

For a subset of links we can then compare the observed with the modelled flows and minimise the square of the difference.

$$(\hat{V}a - \sum p_{ij}{}^{a}\ Tij)^2$$

Hogberg carried out a desk study of his method on a synthetic network with 16 modes and 22 two way links using population and employment as O_is and D_js. He assumed the three elements in the distribution model stood for home-work, home-home and work-work trips. Hogberg then used an algorithm for non-linear regression suggested by Marquant (1963) and from a synthetic O-D matrix he sampled flows from half of the links. After introducing an artificial error component he used this sample to obtain a minimum squared difference between observed and modelled traffic. He then compared the flows predicted for the rest of the 22 links and found the model "acceptable".

Hogberg (1975) has carried out some theoretical work on the "contribution" of each link to the accuracy of the model. He found that after the inclusions of a few links in the calculations there was little gain in the accuracy of the model. In a personal communication to this author Mr. Hogberg has confirmed that practical applications are being carried out in Sweden with this model but they have not yet been reported in English.

3.3 Robillard's Model

Robillard (1975) assumes a generalized gravity model leading to

$$Va = \sum_{ij} R_i S_j f(C_{ij})\ p_{ij}^{a} \quad + \mathcal{E}a \qquad (3.3)$$

where $\mathcal{E}a$ is an error term and

R_i, S_j are the parameters to be estimated by the non-linear regression approach.

$f(C_{ij})$ is a known cost function.

It is noted that the estimation of R_i and S_j need not be unique; what matters is that their product is. It is always possible to solve this problem so that the R_i and S_j are positive.

Robillard indicates that the method suggested by Lawton and Sylvester (1971) can be used to solve this model in order to minimise the difference between observed and synthesized flows. Of course the number of independent links with traffic counts available should be greater than the number of the variables R_i and S_j. In terms of Hogberg's previous example it means that at least 33 of the 44 one way links should be sampled. The problem of minimising

$$(\hat{V}a - \sum_{ij} R_i S_j \; f(C_{ij}) \; p_{ij}^{a} \;)^2 \tag{3.4}$$

can be solved by replacing the S's by their linear regression solution in terms of the R's and then solve by the R's only. Of course other algorithms are also possible but it is claimed that this approach should reduce computer time. It is interesting to note this method does not require any information regarding population or other trip generating parameter. These are all embodied in the R's and S's. The model is, in effect, only a generalised gravity model. One can compare the standard doubly constrained gravity model.

$$T_{ij} = O_i A_i D_j B_j \; f(C_{ij}) \tag{3.5}$$

with

$$T_{ij} = R_i S_j \; f(C_{ij}) \tag{3.6}$$

to find the equivalences

$$R_i \equiv O_i A_i$$

$$S_j \equiv D_j B_j$$

It must be mentioned that Robillard's method is applicable only where p_{ij}^{a} is known, that is when there are no congestion effects. In other words the applications of this model are restricted to cases in which a proportional assignment technique is considered realistic.

3.4 Wills' Model.

3.4.1. Description An alternative non-linear model for the estimation of an O-D matrix in inter urban cases has been proposed by Wills (1977).

His starting point is a generalised gravity model of the form.

$$T_{ij} = b_o \prod_k (A_i^k \, A_j^k)^{b_k} \tag{3.7}$$

where

A_i^k = The k^{th} variable corresponding to origin i, for example population at i.

and b_o---b_k parameters for calibration from traffic counts.

As usual these trips are translated into modelled link flows as

$$Va = \sum_{ij} b_o \prod_k (A_i^k . A_j^k)^{b_k} \, p_{ij}^a \tag{3.8}$$

The parameters b_o ---b_k are calibrated so as to minimise

$$\varepsilon^2 = (\hat{V}a - Va)^2 = (\hat{V}a - \sum_{ij} b_o \prod_k (A_i^k A_j^k)^{b_k} . p_{ij}^a)^2 \tag{3.9}$$

There are several algorithms which could be used to minimise (3.9). Wills in fact used a quasi-Newton optimizing method due to Fletcher and others.

The main innovation in Will's approach is the wide range of variables that can be introduced in this model. He tried, for example, three groups of variables for an inter urban system:

- Socio economic variables like population and employment
- Accessibility variables; he used a centrality index assuming that towns in the periphery will have a different propensity to generate trips than those near the "centre".
- Separation, time or distance, variables.

Provision can also be made for a non-symmetric version of the model. This takes the form

$$T_{ij} = b_o \prod_k (A_i^k)^{b_k} (A_j^{k+1})^{b_{k+1}} \tag{3.10}$$

replacing equation (3.6).

3.4.2 Applications   Two applications of this model are known to this author. The first one by Wills to a road network in British Columbia including 37 centroids and only 3 parameters. As traffic counts for several years were available it was possible to compare the performance of the model in the forecasting mode.

The second application has been carried out by Lamarre (1977) using Wills' computer programs. This exercise was aimed at comparing the performance of Wills' non-linear model with Lamarne's linear one. As reported in 2.4 Wills' model gave a better fit.

In both cases the comparisons were made against observed flows and not against "real" O-D matrices. In both cases the predictions of the model were found "satisfactory". In the British Columbia case an R^2 of around 0.82 was reported. Lamarne reports an R^2 of 0.88 for Wills' model against an R^2 of 0.72 for the linear approach.

4. NETWORK EQUILIBRIUM APPROACH

4.1 Introduction

With the sole exception of the Danish Road Directorate method all other approaches described so far are restricted to the proportional assignment case, that is where p_{ij}^{a} is independent from T_{ij}.

This assumption may be good for a large number of applications, especially where the number of alternative routes between two points is small and where congestion does not play an important role in route choice. Most inter urban transport networks probably satisfy these conditions.

Sang Nguyen (1977) has extended some of Robillard's ideas for the equilibrium assignment case which is more relevant to urban areas. Nguyen suggests two methods. The first one is for the case in which consistent traffic counts are available for all links. The second method requires only costs for all O-D pairs to be known and does not explicitly ask for traffic counts at all.

4.2 Nguyen's method, case I.

According to Nguyen (1977) the first method can be described as follows:

Let $\hat{V}_a$ observed flows link a and $\hat{c}_{ij}$ the observed costs of travelling between i and j by any used route. For each O-D pair only one $\hat{c}ij$ will exist since $\hat{V}_a$ is assumed to be in Wardrop's equilibrium. The cost-flow relationship in any link is $C_a(\hat{V}_a)$.

If $\hat{T}_{ij}$ is the trip matrix corresponding to the $\hat{V}_{ij}$ then the equilibrium state is expressed by

$$\text{b} \qquad \sum_{ij} \hat{C}_{ij} \hat{T}_{ij} = \sum_a C_a(\hat{V}_a)\, \hat{V}_a \tag{4.1}$$

Hence the <u>necessary</u> conditions for a trip matrix $[T_{ij}]$ to be identical to $[\hat{T}_{ij}]$ are

$$\text{a)} \quad \sum_{ij} \hat{C}_{ij} Tij = \sum_a C_a (\hat{V}_a)\hat{V}_a \tag{4.2}$$

and

$$\text{b)} \quad C_{ij} = \hat{C}_{ij} \text{ for all O-D pairs} \tag{4.3}$$

where C_{ij} is the cost on all used routes between i and j when $[T_{ij}]$ is assigned onto the network according to Wardrop's first principle.

Nguyen goes on then to prove that the Tij matrix satisfying (4.2) and (4.3) can be obtained by solving P1 where

$$(\text{P1}) \min F(Va) = \sum_a \int_0^{Va} C_a(x)dx \tag{4.4}$$

subject to

$$C_1 \left(\begin{array}{lll} Tij - \sum_p t_{pij} = o & \text{for all ij} & (4.5) \\ t_{pij} > o & \text{for all p, ij} & (4.6) \\ T_{ij} > o & \text{for all ij} & (4.7) \\ V_a = \sum_{ij} P_{ija} T_{ij} & & (4.8) \end{array} \right.$$

and

$$C_2 \qquad \sum_a C_a(\hat{V}_a)\, \hat{V}_a - \sum_{ij} \hat{C}_{ij} T_{ij} = 0 \tag{4.9}$$

Where t_{pij} = flow on route p connecting i with j.

Using Kuhn and Tucker conditions Nguyen (1977) establishes the equivalence between P1 and equations (4.2) and (4.3). Nguyen then describes an algorithm to solve P_1.

The problem P1 without the constraint C2 has the same structure as a network equilibrium problem. Accordingly, the following algorithm is suggested by Nguyen.

Step 1 - Select an initial feasible $[Tij]$, for example $Tij = K/\sum_{ij}\hat{C}_{ij}$ where $K = \sum_a Ca(\hat{V}a)\hat{V}a$. Determine an initial flow pattern Va using Tij.

Step 2 - Determine the shortest route between each O-D pair and let C_{ij} be the travel cost on this route.

Step 3 - Find the O-D pair rs for which

$$C_{rs}/\hat{C}_{rs} = \min_{i,j} C_{ij}/\hat{C}_{ij}$$

and load $K/\hat{C}_{ij}$ onto the shortest route from r to s.

Let $[\tilde{V}a]$ be the resulting traffic pattern.

Step 4 - Test convergence.

$$\text{If} \quad \sum_a C_a(Va)\ (\tilde{V}a-Va)/_{F(Va)} < \varepsilon \qquad (4.10)$$

the solution is ε optimal and stop.

Step 5 - Find an optimum combination between previous flow pattern and the new one. Determine λ minimising

$$F(Va + \lambda\ (\tilde{V}a-Va)) \qquad (4.11)$$

subject to o<λ<1.

Step 6 - Revise the trip matrix and flows as follows

$$T_{ij} = T_{ij} - \lambda\ T_{ij} \quad \text{for all } ij \neq rs \qquad (4\tfrac{1}{2}\text{-}2)$$

$$T_{rs} = T_{rs} + \lambda \left(\frac{k}{\hat{C}_{rs}} - T_{rs}\right) \quad 4.13)$$

Return to step 2.

4.3 Nguyen's method, Case II.

The Case I method is only suitable where it is possible to obtain traffic counts for all links. This would restrict its application to small networks. In addition, the algorithm is likely to be slow for a large number of O-D pairs. Nguyen then proposed a second method which reduces the data requirements.

He starts by proving that the solution to the following problem P2 also is a solution to the initial model.

$$\text{P2} \quad \text{Min } F^{1}(Va_{1}Tij) = \sum_{a}\int_{o}^{Va} Ca(x)dx - \sum_{ij}\hat{C}_{ij}Tij \quad (4.14)$$

subject to

$$C'_{1} \left\{ \begin{array}{llll} T_{ij} - \sum_{p} t_{pij} & = & 0 \text{ for all } ij & (4.15) \\ T_{pij} & > & 0 \text{ for all } p,i,j & (4.16) \\ Tij & > & 0 \text{ for all } ij & (4.17) \\ V_{a} & = & \sum_{ij} P_{ij}^{\ a}\ Tij & (4.18) \end{array} \right.$$

It can be noted that here only the observed costs $\hat{C}_{ij}$ are required and not the traffic counts.

It should also be noted that P2 has the same form as a formulation of a traffic assignment problem with elastic demand. By analogy Nguyen considers that the increase of the demand function for an O-D pair ij is constant and equal to $\hat{C}_{ij}$. He then suggests that an algorithm as the one stated in Ngugen (1976) could be used to solve the problem.

We found only one test of both methods on a synthetic network with 4 centroids and 18 one way links as in Figure 4.1. This is a system with 4 T_{ij} to estimate from 18 equations. The errors obtained were of about 16% for the T_{ij} cells, and about 3% on link flows. Both methods gave an answer in this case with approximately the same precision.

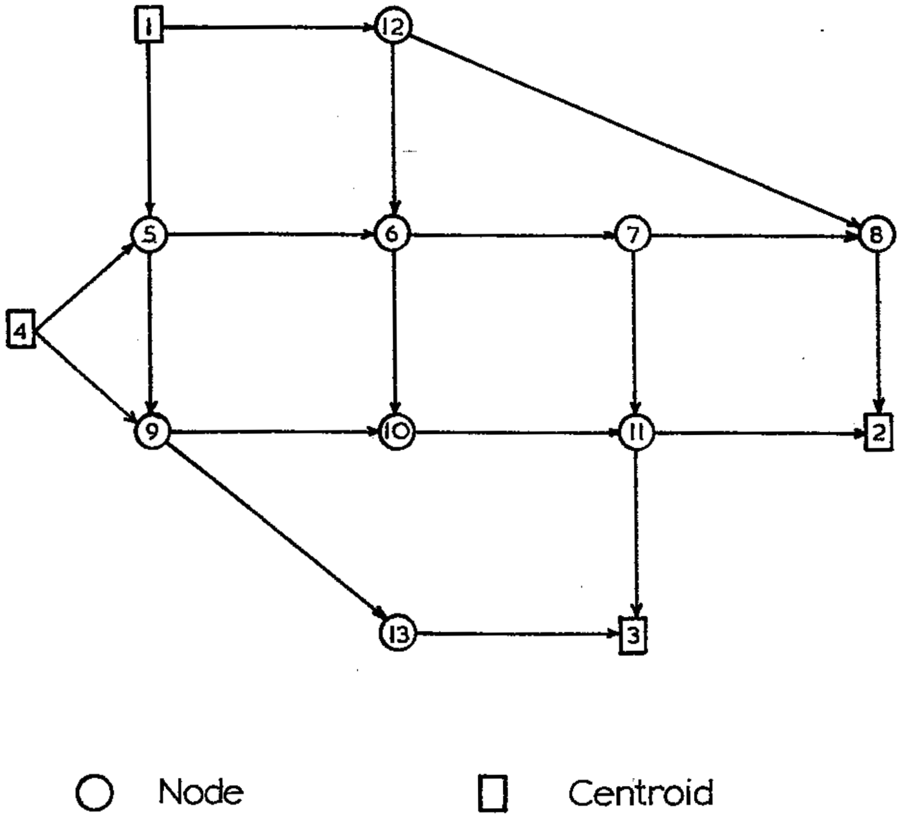


Figure 4.1 : NGUYENS TEST PROBLEM NETWORK

It is interesting to note that Nguyeu has only identified the necessary conditions to solve the problem and not the sufficient ones. That is, the solutions found with his method meet the equilibrium conditions but they will not be normally unique solutions. This issue will be further discussed later.

5. ENTROPY MAXIMISING APPROACH

5.1 Background

The concept of entropy originated in physics. In a closed physical system its elements tend to an arrangement which can be organised in as many ways as possible compatible with the system constraints (energy, mass, etc). This arrangement is the most likely one and also that of the greatest "disorder".

The idea of entropy is also linked to that of "information". It can be seen intuitively at least, that a state of maximum disorder is also one containing a minimum of information. The potential information content of a message grows as the sequence of symbols departs from a purely random (high disorder) sequence.

The use of the concept of entropy (or information) has found several applications in transport, urban and regional systems, see for example Wilson (1970). Its best known application is in the derivation of a fully constrained gravity model.

The idea behind the use of an entropy maximising (or information minimising) approach to our problem can be described as follows:

We would like to find the most likely origin - destination matrix compatible with the available set of link counts. In other words, we would like to "exploit" all the information contained in the observed link flows to determine the most likely O-D matrix compatible with them.

Wilson and MacGill (1977) have observed that the entropy minimising method is particularly suited to this type of system (systems with large numbers of components with apparent disorganized complexity).

It is possible to argue that when one is assuming an underlying transport demand model (i.e. gravity type) and seeks to calibrate it from traffic counts one is probably not fully using the information content of the flows. This is to some extent confirmed by a theoretical study by Hogberg (1975). He studied the contribution of each extra link count on the accuracy of his non-linear model (for his 16 nodes 44 one way links network). He observed that after including the 4 most important one way links "the gain in precision is very small".

Two very similar entropy maximising models will be discussed here. The first one put forward by van Zuylen (1977) uses a particulsr measure of information by Brillouin. Van Zuylen also extended his solution to the inconsistent flows case. The second one by the author of this paper follows a more traditional approach but reaches a very similar formal solution. This one is also extended for the incomplete counts case.

5.2 Van Zuylen's model

5.2.1 Description. Van Zuylen (1977) approaches the problem from the information minimising view point. He is trying to develop an algorithm to obtain the trip matrix without losing more information than necessary.

Among the several information measures discussed by Walsh and Weber (1977) he chooses Brillouin's definition. This assumes that an "a priori" solution t_{ij} is known. Only a brief summary of van Zuylen's analysis will be given here. A more complete description appears in van Zuylen (1977).

If we have a set of observed flows $\hat{V}a$ the modified Brillouin's measure of information for each counted link is given by:

$$I_a = -\log_e \hat{V}a! \cdot \frac{\prod_{ij} (t_{ij} p_{ij}^a)^{(T_{ij} p_{ij}^a)}}{\prod_{ij} (T_{ij} p_{ij}^a)!} \qquad (5.1)$$

using Sterling's approximation

$$\log_e X! \simeq X \log X - X + \log\sqrt{2\pi} \qquad (5.2)$$

and neglecting the constant term

$$Ia = \sum_{ij} T_{ij}\ p_{ij}{}^{a}\ \log_e \frac{T_{ij}}{\hat{V}a\ t_{ij}} \qquad (5.3)$$

for all the links in the network

$$I = \sum_{a} \sum_{ij} T_{ij}\ p_{ij}{}^{a} \log_e \frac{T_{ij}}{\hat{V}a\ t_{ij}} \qquad (5.4)$$

This has to be maximised subject to the link counts constraints

$$\hat{V}_a = \sum_{ij} p_{ij}{}^{a}\ T_{ij} \qquad (5.5)$$

The formal solution can be obtained by maximising the Lagrangean

$$L = \sum_{a} \sum_{ij} T_{ij}\ p_{ij}{}^{a} \log_e \frac{T_{ij}}{\hat{V}a\ t_{ij}} + \sum_{a} \lambda a\left(\sum_{ij} p_{ij}{}^{a}\ T_{ij} - \hat{V}a\right) \qquad (5.6)$$

where λa are Lagrange multipliers corresponding to the flow constraints.

According to van Zuylen differentiating L the solution is found to be

$$Tij = t_{ij}\ . \ X_a{}^{p_{ij}{}^{a}/q_{ij}} \qquad (5.7)$$

$$\text{where} \quad Xa = \hat{V}ae^{-(1+\lambda a)} \quad \text{and } q_{ij} = \sum_a p_{ij}{}^{a} \qquad (5.8)$$

and the Xa can be calculated from the constraints (5.5)

Formally equations (5.7) and (5.5) are a case of the multiproportional problems as discussed by Murchland (1977). He has shown that this model converges to a unique solution and has developed a simple algorithm for it.

5.2.2 Inconsistent constraints. In the inconsistent constraints case the flow equations can be rearranged as

$$\sum_{ij} T_{ij}\, p_{ij}^{a} = Va + \mathcal{E}_a \tag{5.9}$$

where $\mathcal{E}a$ is an error term for each link. Van Zuylen then suggests that the algorithm should now be modified so that the Xa factors for equation (5.7) must be such that minimize the sum of squares

$$\left[\sum_a \left(\sum_{ij} p_{ij}^{a}\, t_{ij} \prod_a X_a^{pij^a/q_{ij}} - \hat{V}a \right)^2 \right] \tag{5.10}$$

This is a problem of more difficult solution.

In a recent paper Van Zuylen (1978) suggests that a better approach would be to remove inconsistencies in the counts first and then to solve the problem using Murchland's algorithm. He used a maximum likelihood method as adapted by Hamerslag and Huisman (1978) to remove inconsistencies in the observed flows in a small example. Van Zuylen found then that the quality of the solution depended a great deal on how good the "a priori" estimation of the trip matrix $[tij]$ was. He concluded his method would be ideal to update an old Origin-Destination matrix.

5.3 An alternative entropy maximising model.

5.3.1 Description. The author has developed a model similar to that of Van Zuylen, but which does not require an "a priori" estimation of the Origin-Destination matrix. As this derivation has not been published it will be described in some detail here.

According to the usual entropy assumption, the most probably arrangement of trips $[Tij]$, is formed by the configuration with the greatest number of states associated with it. The number of ways in which individual trips can be assigned to a particular trip matrix is

$$w(T_{ij}) = \frac{(\sum_{ij} T_{ij})!}{\prod_{i\,j} T_{ij}!} \qquad (5.11)$$

The most likely trip matrix will be that maximising w(Tij). Since w(Tij) is in variant with a monotonically increasing transformation, the problem can be described as

$$\max \log w(Tij) = \log (\sum_{ij} T_{ij})! - \sum_{i,j} \log Tij! \qquad (5.12)$$

and using Stirling's approximation

$$\max \log w(T_{ij}) = \log (\sum_{ij} T_{ij})! - \sum_{i,j} (T_{ij} \log T_{ij} - T_{ij}) \qquad (5.13)$$

Since $\log (\sum_{ij} T_{ij})!$ is a constant dependent only on the total number of trips, it need not be considered in the maximisation problem.

Then the problem to be solved is to maximise

$$\log_e w(T_{ij}) = -\sum_{i,j} (T_{ij} \log_e T_{ij} - T_{ij}) \qquad (5.14)$$

subject to the link flow constraints

$$Va = \sum_{i,j} p_{ij} \cdot T_{ij} \qquad (5.15)$$

for all counted links a.

This is easily obtained by means of the Lagrangean

$$L = -\sum_{i,j} (T_{ij} \log_e T_{ij} - T_{ij}) + \sum_{a} \lambda a \, (Va - \sum_{i,j} p_{ij}^{a} T_{ij}) \qquad (5.16)$$

Taking the derivatives with respect to Tij we have

$$\frac{\delta L}{\delta Tij} = - \log Tij - \sum_a \lambda a \; p_{ij}^{\;a} = 0 \qquad (5.17)$$

$$Tij = e^{-\sum_a \lambda a \; p_{ij}^{a}} \qquad (5.18)$$

by making $e^{-\lambda a} = Xa$

the formal model is found to be

$$Tij = \prod_a Xa^{p_{ij}^{\;a}} \qquad (5.19)$$

where again the Xa are calculated from the restrictions

$$Va = \sum_{ij} \prod_a X_a^{p_{ij}^{\;a}} \; p_{ij}^{\;a} \qquad (5.20)$$

It can be seen this is van Zuylen's model if we make all his "a priori" estimations $[\; tij \;]$ equal to 1.

5.3.2 Extensions and limitations. It has not been yet possible to find an adequate interpretation to the variables Xa. They are certaimly associated with the contribution the flow on each counted link makes to explain each particular O-D pair.

The conventional gravity model can be derived by maximising (5.14) subject to

$$\sum_j T_{ij} = O_i \qquad (5.21)$$

$$\sum_i T_{ij} = D_{ij} \qquad (5.22)$$

$$\sum_{i,j} C_{ij} T_{ij} = C \qquad (5.23)$$

(5.21) and (5.22) being the origin and destination constraints and (5.23) the total cost constraint. If in our model we count traffic on all links, including centroid connectors, we are forcing our solution to meet at least the constraints (5.21) and (5.22) as well. In a cordoned urban area this would imply counting traffic at car parks and other major sources and destinations of trips.

If, in addition, one assumes that all transport costs are only trip end and link costs (this excludes things like modal penalties or parking costs which depend on the length of the stay for example) it can be said that the cost constraint (5.23) is also met by the model. But it must be remembered that there are only 2n + 1 constraints for the gravity model, when n is the number of zones. Our model has as many constraints as independent counts, each one increasing the information content of the solution.

When only a partial set of traffic counts is available the model is likely to underestimate the total number of trips. If one has an independent estimate of this number the solution can be improved. In effect, it is possible to include a new constraint to the model imposing

$$\sum_{i,j} T_{ij} = T \qquad (5.24)$$

where T is the total number of trips.

Maximising (5.14) subject to (5.15) and (5.24) leads to the solution

$$T_{ij} = e^{-\beta} \prod_a X_a^{p_{ij}^a} = X_0 \prod_a X_a^{p_{ij}^a} \qquad (5.25)$$

where β is the Lagrange multiplier for constraint (5.24).

This may be interpreted as the inclusion of an extra (notional) link which all trips use and whose flow equals the total number of trips.

At the time of writing the author is experimenting with this model on a synthetic network on which encouraging results have been obtained. These will be reported later.

Finally, it is worth noting that both models assume p_{ij}^{a} to be constant thus restricting their applications to the proportional assignment case.

6. DISCUSSION

6.1 Introduction

The main advantages and disadvantages of each group of models will be discussed here. We will consider applications in three types of areas

- Inter urban or rual areas
- Free standing urban areas
- Sub or inner areas inside an urban area.

We shall discuss each group of models in term but we must admit, from the outset, that little can be said about the relative accuracy of the proposed methods of estimating a trip matrix. In effect, very few tests have been made with real data and in these cases all comparisons have been made against observed link flows and not against a "real/ observed" O-D matrix.

Nevertheless, something can be said about the most promising models and areas of application from theoretical considerations. In particular we shall discuss each model in terms of

- the conditions under which the assumptions are likely to be acceptable
- the amount of extra data (in addition to traffic counts) required
- internal consistency
- their eventual use for forecasting purposes and not just short term estimation of a trip matrix.
- their flexibility for incorporating information already available about trip making behaviour.

6.2 Gravity type approaches

This group includes the models reviewed in chapters 2 and 3. It must be said that most planners are familiar with the idea that travel demand may follow a "gravity type" of behaviour. This assumption is more likely to be valid in large areas, where trip cost or trip length is an important factor in the formation of a travel demand pattern.

Of this group, very general approaches like the ones suggested by Hogberg and Wills, are in general more theoretically sound. On the other hand, they tend to require more data and greater effort in calibration. If we assume here that mispecification errors disqualify the use of a model for medium term planning we will have to accept these extra costs as a reasonable price to pay in this case.

The extra data used in the model will depend on the type of data readily available or obtainable with reasonable effort. Robillard's model, for example, passes most specification tests, requires no extra data, but implies a very large calibration effort (The minimum number of parameters to determine by minimum squares techniques is twice the number of zones plus one).

It is not by chance that most applications of this approach have been in inter urban areas. The method seems ideally suitable for problems where each town represents a single zone.

Applications to free standing towns will usually imply a more detailed zoning system and an independent roadside interview to determine "external trips". Moreover, congestion effects are likely to be more important in this case and an all-or-nothing assignment will not suffice.

We do not think this approach to be appropriate to inner city areas. There seems to be two reasons for this. One is the short trip lengths involved and the other the fact that probably most of a trip will take place outside the study area. Hence, a gravity type of trip making behaviour is not likely to be a sensible assumption.

Finally, this type of model can cope perfectly well with inconsistencies in the traffic counts.

6.3 Equilibrium assignment approach

We have already mentioned that we think both Nguyen's methods will determine a trip matrix but, with the exception of peculiar circumstances, the solution offered will not be unique. It will actually depend on the initial, feasible, solution assumed in each case.

Nevertheless, Nguyen has rightly pointed out the importance of introducing an allowancd for congestion effects in assignment. We will discuss in section 6.5 some of the ways in which this might be achieved.

6.4 Entropy maximising approach

This approach appears as the one of greater generality so far. In principle it could be applied to the three areas of interest mentioned before. We have already pointed out, though, that the model appears sensitive to the total number of trips, a variable usually not available in most cases. Failing this, counts on links to major trip productions or attractions are required and these are more likely to be available in urban areas (car parks, roads representing only links with major residential areas, etc).

Furthermore, this approach, being independent of planning parameters like population and car ownership, does not seem suitable for medium to long term planning applications.

Computationally speaking the algorithm for solving this type of problem appears faster and more convenient than the corresponding methods for solving non linear regression models. Nevertheless, inconsistencies in the traffic counts must be removed previous to an application of the algorithm.

6.5 Extensions to include congestion

We have seen that both gravity type and entropy maximising models can be solved for the proportional assignment case. There are cases in which congestion effects are unimportant compared with differences in driver's perceived route costs.

The Danish Road Directorate has developed their gravity model so that it can introduce some congestion effects via Smock's iterative approach. It is possible to extend this type of approach to all the other methods as well (with the obvious exception of Nguyen's models).

The following generalised algorithm might be used

Step 1. Assume traffic volumes in the network to be zero and calculate travel costs for all O-D pairs.

Step 2. Find the set of minimum cost routes implicitly producing a new set of p_{ij}^{a}, $[\tilde{p}_{ij}^{a}]$. Set n = n+1.

Step 3. Combine this $[\tilde{p}_{ij}^{a}]$ with the previous one $[p_{ij}^{a}]^{n-1}$ in the following way

$$[p_{ij}^{a}]^{n} = [p_{ij}^{a}]^{n-1} (1 - \frac{1}{n}) + [\tilde{p}_{ij}^{a}] \frac{1}{n}$$

Step 4. Estimate an O-D matrix under proportional assignment assumptions (using $[p_{ij}^{a}]^{n}$) and any selected approach.

Step 5. Assign $[Tij]^{n}$ using $[p_{ij}]^{n}$ and find new route costs.

Step 6. Test for convergence using a suitable convergence criterion. For example test if $[C_{ij}]^{n}$ is sufficiently similar to $[C_{ij}]^{n-1}$. If algorith has "converged" stop, otherwise proceed to step 2.

The main problem with this approach is that there is no guarantee of achieving convergence. Other approaches are possible but their analysis escape the purpose of this report.

There is also an extra cost associated with any method incorporating congestion effects. All of them require good flow-cost relationships for the links. These are not easy to obtain and may involve considerable ad hoc data collection. There are two other alternative sources:

- The standard DTp speed flow curves which can only cater for very coarse "types" of links but may provide a simple and inexpensive allowance for congestion effects. They are probably quite good for inter urban areas.

- For urban areas most delays occur at intersections and hence any model giving a good estimate of delays under different conditions there could be adapted. Typical examples would be models like TRANSYT or SATURN. Bolland et al (1977).

7. References

1. BEBEE, M.A. (1959) Novel traffic survey method utilises vehicle lights. Bulletin of the Highway Research Board 224.

2. BOLLAND, J., HALL, M., VAN VLIET, D., WILLUMSEN, L., (1977) A model for the simulation of traffic management schemes PTRC Summer Annual Meeting, Warwick.

3. HAMERSLAG, R., and HUISMAN, H.C., (1978) Binaire Calibratie, Verkeers Kunde 166-168.

4. HOGBERG, P. (1975) Planning of traffic counts. University of Goteborg, Department of Statistics Research Report 1975:3

5. HOGBERG, P. (1976) Estimation of parameters in models for traffic prediction: a non-linear regression approach. Transportation Research Vol.10, pp. 263-265.

6. HOLM, J., JENSEN, T., NIELSEN, S., CHRISTENSEN, A., JOHNSEN, B., RONBY, G (1976) Calibrating traffic models on traffic census results only. Traffic Engineering and Control, Vol.17, April 1976.

7. JENSEN, T., NIELSEN, S. (1973) Calibrating gravity models and estimating their parameters using traffic volume counts. 5th Conference of Universities Transport Study Group. London.

8. LAMARRE, L. (1977) Une methode lineaire simple d'estimation de la matrice des origines et des destinations a partir de comptages sur les liens d'un reseau: une application au reseau routier du Quebec. Centre de Recherche Sur Les Transport, Universite de Montreal, Publication No. 76.

9. LOW, D. (1972) A new approach to transportation systems modelling. Traffic Quarterly, Vol. 26, 391-404.

10. MURCHLAND, T.P. (1977) The multiproportional problem. London: University College, London, Transport Studies Group, Notes JDM263 (unpublished).

11. NGUYEN, S. (1976) Procedures for equilibrium traffic assignment with elastic demand. Centre de Recherche sur les Transports, Universite de Montreal, Publication No. 39.

12. NGUYEN, S. (1977) Estimating an O-D matrix from network data: a network equilibrium approach. Centre de Recherche sur les Transports, Universite de Montreal, Publication No. 87.

13. OECD Road Research Group (1974) Urban Traffic Models: Possibilities for Simplification Paris; OECD.

14. ROBILLARD, P. (1975) Estimating an O-D matrix from observed link volumes. Transportation Research 9, 123-128.

15. SMITH, R., McFARLANE, W. (1978) Examination of a simplified travel demand model Transportation Engineering Journal of ASCE 104, (TE1).

16. SMOCK, R. (1962) An iterative assignment approach to capacity restraint on arterial networks. Highway Research Board Bulletin No. 347, 60-66.

17. SYMONS, J., WILSON, R. and PATERSON, J. (1976) A model of inter-city motor travel estimated by link volumes. Australian Road Research Board Proceedings, Vol.8 Session 20.

18. VAN VLIET, D (1976) Road assignment; parts I, II and III Transportation Research 10, pp. 137-143.

19. VAN ZUYLEN, H.J. (1977) Some remarks on the information minimising method Verkeersakademie, Tilburg, Mimeographed notes, Oct.1973.

20. VAN ZUYLEN, H.J. (1978) A method to estimate a trip matrix from traffic volume counts PTRC Summer Annual Meeting, University of Warwick, July 1978. London: PTRC Education & Research Series Ltd. (unpublished)

21. WALSH, J., WEBBER, M. (1977) Information theory: some concepts and measures. Environment and Planning, A.9, 395-417.

22. WILLS, M. (1977) Non linear estimation of origin-destination flows and gravity model parameters from traffic counts on links. Department of Geography, University of British Columbia mimeographed notes.

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

24. WILSON, A.G., MacGILL, S.M. (1977) A systems analytical framework for comprehensive urban and regional model building. Leeds: University of Leeds, School of Geography Working Paper No. 209.

25. WRIGHT, C. ORROM, H. (1976) The Westminster route choice survey: a new technique for traffic studies. Traffic Engineering and Control 17. (819).

8. APPENDIX 1.

Notation.

We have tried to keep notation uniform throughout this paper. Nevertheless, some inevitable but minor departures from this norm will be apparent to the reader.

In general

T_{ij} trips between i and j, where i and j are origin-destination points, that is zone centroids.*

$[T_{ij}]$ = The whole trip matrix

a a link

Va total flow on link a

Ca(Va) Cost flow relationship on link a

O_i A trip end characteristic for zone i, in particular

P_i, E_j Particular population and employment at i and j.

C_{ij} = cost of travelling between i and j

r_{ij} = a route or path from i to j, defined by

$$\delta^a_{r_{ij}} \begin{cases} 0 & \text{if link a is not on } r_{ij} \\ 1 & \text{if link a \underline{is} on } r_{ij}\text{, that is } r_{ij} \text{ uses link a.} \end{cases}$$

t_{rij} = number of trips from i to j on path r_{ij}; in general $T_{ij} = \sum_r t_{rij}$

t^a_{rij} = number of trips from i to j on path $_{rij}$ using link a

$p_{ij}{}^a$ = $\sum t^a_{rij} / T_{ij}$ = proportion of trips between i and j using link a

in general $0 \leq p_{ij}{}^a \leq 1$ and the extreme values are taken when there is only one path between i and j.

---

* We do not differentiate between person and car trips in the understanding that a multiplication by an index corresponding to an average car occupancy rate is the only operation required to convert one into the other.

A^ as in ($\hat{V}a$) is used to represent observed data (flows)

A* as in (r_{ij}*) is used to represent an optimum figure (like optimum path between i and j)

9. APPENDIX 2.

A least square estimation of the trip factor for the DRD model

9.1 Method

The method is closely linked to the normal distribution. In fact, for the normal distribution the maximum likelihood solution is found when the variance is minimised.

If we make $\varepsilon = X - \text{mean}$ it is possible to represent the normal distribution by

$$f(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\varepsilon^2/2\sigma^2}$$

and $dP = f(\varepsilon)d\varepsilon$

For n observations we get the corresponding joint probability density

$$P = dP_1 \cdot dP_2 \text{ --- } dP_n = \frac{1}{\delta\sqrt{2\pi}} e^{-\varepsilon^2/2\sigma^2} d\varepsilon_1 \text{ ---- } \frac{1}{\delta n\sqrt{2\pi}} e^{-\varepsilon_n^2/2\sigma^2} d\varepsilon_n$$

if we make $p_1\sigma^2 = p_2\sigma_2^2 = \text{---} = pn\,\sigma n^2 = \sigma^2$

p_i represents a scalar defining the relationship between the anticipated (a priory) variances

then

$$P = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left[\frac{(p1)^{\frac{1}{2}}.(p2)^{\frac{1}{2}} \text{---}(pu)^{\frac{1}{2}}}{\sigma\,\sigma\,\sigma\,\sigma}\right] e^{-\sum_1 p_i\varepsilon^2/2\sigma^2} * d\varepsilon_1 * \text{---} d\varepsilon_n$$

if $C^n = (p_1\, p_2 \text{ ---} p_n)^{\frac{1}{2}}$

$$P = \left(\frac{C}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_1 p_i\varepsilon^2/2\sigma^2} d\varepsilon_1 . d\varepsilon_2 \text{---} d\varepsilon_n$$

taking $\log_e$

$$L = -n\log_e\sigma - n\log_e\sqrt{2\pi} - \frac{\sum p\varepsilon^2}{2\sigma^2} + n\ln c$$

Where L is the likelihood function. The maximum likelihood is obtained for $\sum p^2$ = minimum.

$$\frac{\partial L}{\partial \sigma} = \frac{\partial}{\partial \sigma}\left(-n \log_e \sigma - n \log_e \sqrt{2\pi} - \frac{\sum p\varepsilon^2}{2\sigma^2} + n \ln c\right)$$

$$\frac{\partial L}{\partial \sigma} \Rightarrow -\frac{n}{\sigma} + \frac{\sum p\varepsilon^2}{\sigma^3} = 0$$

$$\text{so } \sigma^2 = \frac{\sum p\varepsilon^2}{n}$$

where p is normally called "weight".

9.2 Application to the DRD model

The model estimates the flow on a link as

$$\hat{V}a = b\,Xa + \varepsilon a \quad \text{where}$$

$$Xa = \sum_{ij} O_i\, O_j\, C_{ij}^{-m}\, p_{ij}^{a} \qquad \text{with } \underline{m} \text{ known}$$

and the error term εa is normally distributed with mean zero and variance

$$(bXa)^{\gamma}\sigma$$

For example $\gamma = 1$ means that mean and variance are in proportion and
$\gamma = 2$ means that the standard deviations of the error is proportional to traffic flow.
γ will be chosen independently of the method.

In the absence of better information an initial guess can be made regarding γ and that improved by trial and error in order to obtain the best fit.

In our case we may define

$$\frac{\hat{V}a}{Xa} = Y_a = \sigma + \frac{\sum \varepsilon a}{Xa}$$

and $\varepsilon'_a = \frac{\varepsilon_a}{Xa}$ with $\varepsilon'_a \in \bar{N}(0, (b^{\gamma}X^{\gamma-2})^2)$

Now, our best unbiased linear estimate would be that which minimises $\Sigma p\varepsilon^2$

In our case

$$\gamma_{a-b} = {}_a$$

we compute $\sum_i (\gamma_i - b)^2 p_i = \sum_i p_i \varepsilon_i^2$

$$\sum_i p_i \varepsilon_i = \sum_i \gamma_i^2 p_i + b^2 \sum_i p_i - 2b \sum p_i \gamma_i$$

Making the first derivative equal zero.

$$\frac{\partial(\sum p_i \varepsilon_i^2)}{\partial b} = 2b\sum_i p_i - 2\sum_i p_i \gamma_i = 0$$

$$b = \frac{\sum_i pi \gamma i}{\sum_i pi}$$

but our weights were $pi = \frac{\gamma}{\gamma i} \quad \frac{\gamma}{(b^{\gamma} X_i^{\gamma - 2})\gamma}$

so $pi = b^{-\gamma} X_i^{2-\gamma}$

$$b = \frac{\sum_i b^{-\gamma} X_i^{2-\gamma}\, Vi/Xi}{\sum b^{-\gamma} X^{2-\gamma}}$$

$$b = \frac{\sum_i X_i^{1-\gamma}\, V_i}{\sum_i X_i^{2-\gamma}}$$

which does coincide with DRD's solution.