Monograph:

Ortuzar, J.D. (1979) Testing the Importance of Fixing Exogenously Some Parameters in Aggregate Modal Split Models, by means of Sensitivity Analysis. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Working Paper 118

UNIVERSITY OF LEEDS
Institute for Transport Studies

ITS Working Paper 118

August 1979

TESTING THE IMPORTANCE OF FIXING EXOGENOUSLY SOME PARAMETERS IN AGGREGATE MODAL SPLIT MODELS, BY MEANS OF SENSITIVITY ANALYSIS

BY J D ORTUZAR

CONTENTS

ABSTRACT

ORTUZAR, J.D. (1979) Testing the importance of fixing exogenously some parameters in aggregate modal split models, by means of sensitivity analysis. Leeds: Univ. Leeds, Inst. Transp. Stud., Work. Pap. 118.

Aggregate modal split (and distribution) models currently need exogenously determined values for such key parameters as the value of in-vehicle time, the value of waiting time and the car occupancy factor.

Using hierarchical logit modal split models and data from the Garforth Corridor, to the east of Leeds, this paper set out to investigate the effects in the model aggrement to the data (and hence in its forecasting capabilities) of inputting different values for these parameters. To gain insight into the relative importance of each of these fixed parameters, the analytical point elasticities of the free parameters in the model with respect to them, were briefly examined. This exercise, together with some more practical post-hoc considerations led us to concentrate on the values of in-vehicle time and waiting time only.

The rather surprising outcome of the analysis was that the model fits were not statistically different, for different values of the fixed parameters, their variation being accommodated by changes in the values of the free parameters. The main conclusion was that provided the exogeneous parameters are reasonably accurate they should produce models that are capable of performing as well (or badly) as models calibrated entirely from the data, and at a much lower cost.

TESTING THE IMPORTANCE OF FIXING EXOGENEOUSLY SOME PARAMETERS IN AGGREGATE MODAL SPLIT MODELS BY MEANS OF SENSITIVITY ANALYSIS

1. INTRODUCTION

Hierarchical logit modal split models (Williams, 1977) for bus, rail and car morning peak work trips in a corridor to the east of Leeds have been calibrated as part of an SRC project on mixed-mode demand forecasting, using aggregate data provided by West Yorkshire Metropolitan County Council.

The Garforth Corridor has 117 zones and the data available consists of road and public transport networks and trip matrices disaggregated by household car ownership group (0, 1, 2+) and mode (Ortuzar, 1979b). Some preliminary problems with the data and implementation of the models have already been discussed (Hartley and Ortuzar, 1979).

The main objective of this paper is to highlight and discuss one particular issue of aggregate modal split models, which incidentally also applies to trip distribution. This is the need for exogeneously determined values for such key parameters as the value of in-vehicle time, the value of waiting time and average car occupancy. The ideal method should be, of course, to obtain estimates for these parameters from the data itself, as part of the calibration process (as currently occurs with the dispersion and modal penalty parameters), but unfortunately this is not possible with the current aggregate transportation tools and/or packages. This deficiency has been strongly criticised by advocates of disaggregate models. (Spear, 1977, among several others).

The first part of the paper introduces the modal split models and notation involved; then the point elasticities of the model parameters with respect to the fixed parameters under scrutiny are used to get a feeling of their relative importance; finally the findings of an extensive sensitivity analysis with respect to the more relevant parameters, namely the values of in-vehicle and waiting time, is reported.

2. NOTATION AND MODAL SPLIT EQUATIONS FOR A HIERARCHICAL LOGIT MODEL

Throughout the paper we will refer to the hierarchical logit model for bus, car and rail depicted in Figure 1. The secondary split (bus-rail) is in fact a binary logit model, whose parameters λ and δ are used in turn to compute composite costs for the 'public transport' mode that competes against car in the primary split. This has also a binary logit form and will yield parameters λ_1 and δ_1.

For the secondary split, the current aggregate methodology needs as inputs values for the in-vehicle, and waiting times(1). The primary split requires in addition, a value for the car occupancy factor(2). This last parameter would not present many problems if the data were grouped according to car availability (i.e. individuals with no car available do not make car trips and therefore do not need to have car costs specified(3)); however, in our case, for example, approximately 18% of non car owner trips are made by car, and this situation is not particularly uncommon.

We define:

D	=	distance (km)
FF	=	fixed fare (pence)
VF	=	variable fare (p/km)
F	=	fare = FF + VF·D (pence)
T	=	in-vehicle time (min.)
WT	=	waiting time (min)
N	=	weight of waiting time, i.e. WT is perceived as N times T.
V	=	value of in-vehicle time (p/min). Notice that the value of waiting time is therefore given by N·V
OP_c	=	perceived car operating costs i.e. fuel (p/km).
CO_c	=	car occupancy factor (travellers per car)
b,r,c	=	suffixes indicating bus, rail and car respectively
pt	=	suffix denoting public transport
Δ	=	operator denoting the difference between rail and bus attributes
Δ_1	=	operator denoting the difference between public transport and car attributes
λ,λ_1	=	secondary and primary modal split dispersion parameters (pence^{-1})
δ,δ_1	=	rail and public transport modal penalties (pence)
G	=	generalised costs (pence)
MS	=	modal split, i.e. proportion selecting the indexed mode.

1. As a matter of fact, a value is also needed for walking time, but we will not consider it in this analysis.

2. We are assuming that car operating costs can be determined more accurately than these other parameters.

3. For a discussion on simple methods to 'transform' car ownership data to car availability data, see Ortuzar (1979a).

then generalised costs are computed as:[4]

$$G_b = F_b + (T_b + N \cdot WT_b)\ V \qquad (1)$$

$$G_r = F_r + (T_r + N \cdot WT_r)\ V \qquad (2)$$

$$\Delta G = \Delta F + (\Delta T + N \cdot \Delta WT)\ V \qquad (3)$$

and the consistent composite costs are (Williams, 1977):

$$G_{pt} = -\frac{1}{\lambda} \ln \{\exp(-\lambda\ G_b) + \exp(-\lambda(G_r + \delta))\} \qquad (4a)$$

This latter expression can be represented, for our discussion, by a notional service with fares F_{pt} and times and waiting times T_{pt} and WT_{pt} respectively, such that:

$$G_{pt} = F_{pt} + (T_{pt} + N \cdot WT_{pt})\ V \qquad (4b)$$

Then if

$$G_c = \frac{OP_c \cdot D}{CO_c} + (T_c + N \cdot WT_c)\ V \qquad (5)$$

we have that

$$\Delta_1 G = F_{pt} - \frac{OP_c \cdot D}{CO_c} + (\Delta_1 T + N \cdot \Delta_1 WT)\ V \qquad (6)$$

With these definitions in mind, the modal split equations are simply:

a) Secondary split:

$$\left.\begin{aligned} MS_b &= \frac{1}{1 + \exp(-\lambda(\Delta G + \delta))} \\ MS_r &= 1 - MS_b \end{aligned}\right\} \qquad (7)$$

b) Primary split:

$$\left.\begin{aligned} MS_c &= \frac{1}{1 + \exp(-\lambda_1(\Delta_1 G + \delta_1))} \\ MS_{pt} &= 1 - MS_c \end{aligned}\right\} \qquad (8)$$

3. THE PROBLEM OF FIXED ENDOGENEOUS PARAMETERS

The normal information available in transportation studies comprises fares, operating costs, travel and waiting times, parking charges, etc. plus other important data such as car ownership, household structure, (sometimes) income, etc. To calibrate modal split models, one should

---

4. Notice that this is a much simplified definition of generalised costs; however, it retains its basic qualities and shortcomings and it is more useful for our purposes than more complext representations. The actual definitions used in the experimental work are described in Hartley and Ortuzar (1979).

write generalised cost equations as above and let the calibration program produce estimates, not only for the λ's and δ's, but also for V, N and CO_c. However, as we already mentioned, this is not yet possible with the current aggregate analysis tools.

The calibration process (in our case finding maximum likelihood estimates), will therefore produce the most likely values of λ, δ, λ_1 and δ_1 given input values for V, N and CO_c. We would normally expect that if we change these values, the calibrated parameters should change and more importantly, the goodness of fit statistics (e.g. modified log-likelihood) should also change.

Let us examine at this point some of the assumptions behind our modal split equations to put the problem into a better perspective. We are taking people as rational choice decision makers, i.e., they have perfect information about all possible options and they choose consistently the more convenient to them; notice that 'convenient' is not restricted to time and money considerations; we, as modellers, are only capable of observing these attributes, but there may be other attributes which we do not observe and which would explain otherwise apparent irrationalities (e.g. people choosing the slower and more expensive option). If we accept this basic and quite strong assumption, then in our linear in the parameters cost model we assume that individuals place a value on time and on waiting time in order to take a decision. It is quite clear then, that using extraneous information about these values could be rather misleading. For this reason, it is important to find the values of in-vehicle time and of the waiting time weighting that will produce the best fit to our data [5]. We will not be concerned here with the very complex problem of response and prediction with the model, where even stronger assumptions are required. (Williams and Ortuzar, 1979).

As mentioned in Section 2, the problem with the car occupancy factor arises from the fact that we need to specify car generalised costs for non car owners or more generally for car passengers. If we had car availability grouped data, the use of an average car occupancy, as surveyed for the system, would probably suffice, ie. it would not be stronger than the rest of our assumptions.

5. Notice again how simple are the models we are using. We are assuming a single value of in-vehicle time and waiting time, regardless of the choice. We are, however, allowing for different values for different car owning categories.

Not being able to estimate the 'fixed' parameters directly from the data, requires a 'second best' procedure. This clearly is to search for the values of V, N and CO_c, that optimise the goodness of fit statistics (which is actually equivalent to the first best, but following a more tortuous route). Unfortunately this can be extremely expensive as it is shown below.

In a system of only 300 zones (representing West Yorkshire) CPU times needed by the ICL 1906A computer at Leeds University, for the following calculations are:

Operation	CPU time (secs)
Bus generalised costs, standard fares	5300
Bus generalised costs, Metrocard	5100
Rail generalised costs, standard fares	470
Rail generalised costs, Bullseye tickets	460
Walking generalised costs	470
Bus minimum generalised costs	4
Rail minimum generalised costs, including walking	4
Rail only minimum generalised costs	4
Secondary modal split calibration	62
Total	11,874

and this is only for the secondary modal split and for only one pair of values of in-vehicle time and the waiting time weight.

Bearing this in mind it appears that a third best alternative is to explore the behaviour of the goodness of fit statistics for several values of the fixed parameters in a sensitivity analysis. This is neither simple nor inexpensive and for this reason we will start by looking at the elasticities of the calibrated parameters (λ's and δ's) with respect to the fixed parameters in order to gain some insight into their relative importance. The next section presents a brief account of our theoretical results.

It is worth mentioning here that a very detailed sensitivity analysis of a complete transport model with respect both to some parameters representing policy alternatives and parameters regarded as fixed inputs (as in our case) and even with respect to changes in modal form, has been reported recently (Bonsall et.al., 1977). Unfortunately for our purposes, their modal split model was part of a combined trip distribution-modal split model, quite different to the one presented here. Moreover, they were concerned with testing the sensitivity of model results with respect to the parameters and not with finding (as we are), the exogeneously fixed values of V, N and Co_c that should produce the best fit; for these reasons their important results do not help us at this stage.

4. GAINING INSIGHT FROM ANALYTICAL POINT ELASTICITIES

By definition if

$$y = f(X_1, X_2, \ldots, X_n) = f(\underline{X}) \tag{9}$$

then the elasticity of y with respect to X_i is given by

$$\varepsilon_{y/X_i} = \frac{\partial f(\underline{X})}{\partial X_i} \cdot \frac{X_i}{f(\underline{X})} \tag{10}$$

That is, it represents the percentage change in $f(\underline{X})$ for a one percent change in X_i, all other things being equal.

Table 1, shows the range of variation of some numerical estimates of the point elasticities of λ and λ_1, with respect to the fixed parameters (valued at the figures suggested by WYTCONSULT,[6] 1977), for alternative values of the other parameters and for some hopefully not unreasonable mean values of the variables in the model. The Appendix gathers together the analytical derivation of the point elasticities and the assumed mean figures used to work out their numerical values. Notice that we are using even simpler versions of our modal split equations, because we do not consider the scaling parameters δ and δ_1.

The medium values in the table, correspond to the estimates for the WYTCONSULT figures and the extremes represent their likely range of variation under our assumptions.

---

6. These values are: V = 0.52 pence/min; N = 2.3 and CO_c = 1.3 travellers/car.

Elasticity	Range		
	Minimum	Medium	Maximum
$\varepsilon_{\lambda/V}$	-0.19	-0.68	-0.86
$\varepsilon_{\lambda/N}$	-0.94	-1.08	-1.27
$\varepsilon_{\lambda_1/V}$	-0.56	-0.75	-0.89
$\varepsilon_{\lambda_1/N}$	-0.43	-0.54	-0.65
$\varepsilon_{\lambda_1/CO_c}$	-0.09	-0.27	-0.50

Table 1: Variations in the point elasticity estimates of the dispersion parameters

The numerical values presented in Table 1 have been calculated for just one point, taken as representing mean values for the Garforth-Leeds corridor, and for varying assumptions concerning the fixed parameters. The only claim we wish to make is that for these assumed conditions, a comparison of the point elasticity estimates can be revealing in terms of which are the more influential parameters. In this sense the conclusions seem to be as follows:

i) The car occupancy factor has the lowest elasticity values and therefore should be regarded as the least influential parameter. This view is reinforced by the fact that it only appears in one cost equation and because it is quite possible to argue that car drivers do not perceive that they share their costs with passengers, and also that full perceived costs provide a proxy for the difficulties associated with being a passenger and to lack control over the journey characteristics[7]

ii) Both the values of in-vehicle time and of the waiting time weight seem to be very important, perhaps the latter slightly more so in view of its greater effect on λ which in turn should affect λ_1 through equation (4a).

---

7. In fact for these reasons we used a value of CO_c = 1 traveller per car, in all our calculations.

The next section describes how the tests were carried out and the last section will report on the results of the experimental sensitivity analysis for a wide range of values of V and N.

5. CARRYING OUT THE EXPERIMENTAL TESTS

Having decided to concentrate our analysis on the variation of the model parameters and fit statistics with respect to V and N, the first task was to decide what would be a sensible range of variation for these two parameters. Hartley and Ortuzar (1979) studied the likely range of variation of the value of time under several assumptions about its relation with household or workers income. They found that taking WYTCONSULT's value of 0.52 p/min (WYTCONSULT, 1977) as a mean area wide value, possible extremes for the distribution of V were 0.35 p/min (for non car owners) and 0.98 p/min (for 2+ car owners). Furthermore they found that under the assumption of V being linearly related with worker's income, while a value of 0.52 p/min was appropriate for both non car owners and one car owners, the corresponding value for members of household owning two or more cars should be 0.62 p/min. We decided to take these four values in our experiments.

In the case of the waiting time weight, the lowest value selected for analysis was 1.7, ie. WYTCONSULT's assumed walking time weight. The highest value considered was 4, which approximates to the values found in recent American studies (Spear, 1977).

The sensitivity analysis experiments were carried out in two stages, first for the lower hierarchy of the modal split model, ie. the bus-rail split, and then for the primary split, public transport-car.

In the first case, for each value of V and N calculations of minimum costs for each mode were performed as noted in Section 3. Prior to the calibration itself the modal split pattern versus cost differences was carefully analysed in order to define suitable first estimates of λ and δ.

Having found the maximum likelihood estimates of λ and δ we were able to calculate composite costs for the 'public transport' mode, from equation (4a), for each value of V and N selected for analysis.

Also for these values, private transport costs were calculated and parking charges added where applicable. The examination of the modal split pattern versus costs differences in this case, made apparent a deficiency in our data. In effect we have only available information on vehicular trips and this means that for short distance trips, the alternative to car, which is walking is not present. To avoid this problem we made the assumption that in the primary split, all trips of lengths less than a certain threshold distance (which itself was subjected to sensitivity analysis) were made by car and therefore did not enter the calibration program[8]. After taking care of this problem the modal split patterns versus cost differences looked very reasonable and suitable first estimates of λ_1 and δ_1 to enter the calibration were easily derived. It is worth noting that this procedure is partly responsible for public transport 'bonuses' rather than penalties, ie. in all cases the equiprobability choice occurred for cost differences (public transport less car) greater than zero. These bonuses, of course, reflect the fact that even in the car owning categories not everybody has a car available (this issue is discussed more fully in Ortuzar, 1979a)

The program used to calibrate the models (Hartley and Ortuzar, 1979) produces several indicators such as standard errors, t-ratios and an estimation of the variance-covariance matrix of the parameter estimates. In all the models calibrated, the t-ratios were such that the parameters were highly significant; however, due to the grossed-up nature of the data used[9], their validity is somewhat questionable and they were omitted from the presentation of results.

Our original intention was to compare the log-likelihood values of alternative calibrations (ie. for different fixed parameters) in order to find out the highest; this in turn should define the preferred set of fixed parameters. Unfortunately, the procedure was further complicated

---

8. The threshold distances finally used were 0.66 km. for non car owners and 1.15 km. for car owners. The public transport trips misclassified by this procedure were in every case less than 5%.

9. This problem is discussed at greater length by Hartley and Ortuzar(1979).

by the fact that it was not possible to correctly compare the different models in this way, the reason being that when using different fixed parameters, the number of relevant cost difference bins in which the data is grouped prior to calibration by the program varies. Also there is no guarantee that even if the number of groups is the same (which was never the case), the zone pairs involved would be the same. For this reason we decided to use another goodness of fit measure, namely the coefficient of determination (R^2) of the straight line given by:

$$\text{Modelled MS} = A + B \cdot \text{Observed MS} \tag{11}$$

As the degrees of freedom of the regression lines vary with the number of cost bins of each model, it is not possible to decide on the basis of R^2 measures alone if two models are statistically different. However, if ${R_1}^2$ is computed from N_1 groups and ${R_2}^2$ from N_2 groups we can calculate variables Z_1 and Z_2 given by

$$Z_i = \tfrac{1}{2} \ln \left(\frac{1 + R_i}{1 - R_i}\right), \quad i = 1, 2 \tag{12}$$

and test for 'no significant difference' between the correlations with

$$t' = (Z_1 - Z_2) / \left\{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}\right\}^{\frac{1}{2}} \tag{13}$$

where t' should be approximately distributed standard normal ($N(0,1)$) _if_ there is no statistical difference between the two values[10].

The parameters A and B in equation (11) should be zero and one respectively, for a perfect fit. We also looked for significant departures from these values, but in all cases they were either very close to the appropriate value or well within the error range allowed for by the regression.

---

10. I am grateful to Hugh Gunn for having suggested this procedure.

Finally we also looked at a measure ρ^2 defined as:

$$\rho^2 = 1 - \frac{\sum_{i=1}^{n} t_i \left(\frac{\pi_i}{t_i} - p_i \right)^2}{\sum_{i=1}^{n} t_i} \qquad (14)$$

where: n = total number of cost difference bins in the particular model

t_i = observed number of trips in cost bin i

π_i = observed number of trips by first mode in cost bin i

p_i = modelled probability of choosing the first mode in cost bin i.

However, the ρ^2 values were in all cases very similar to the R^2 values, so we do not report about them either.

6. RESULTS OF THE SENSITIVITY ANALYSIS

As mentioned in the previous section we started the sensitivity analysis with the bus-rail secondary split. The results of the experiments are shown in Tables 2. As it can be seen, although the dispersion parameter λ and the rail penalty δ show a wide variation (which is consistent with the findings of Section 4), the goodness of fit statistics R^2 remained almost unaltered[11]. It would appear then that the 'optimum optimorum', rather than being a point looks like a somewhat flat surface. To further show that the models do indeed seem indistinguishable, Figures 2, 3 and 4 depict the observed and modelled proportions of trips using bus, as a function of cost differences, for the 'best' and 'worse' cases of Tables 2a, 2b and 2c.

---

11. In fact the biggest difference detected was in Table 2b, for one car owners and corresponded to the cases V=0.52, N=2.3 and V=0.98, N=3.0, where the values of R^2 were 0.9885 and 0.9770 respectively. Calculating t' from equation (13) yielded the value 1.379 which means that the observed difference could occur by chance in between 15% and 20% of cases even if the two R^2's were truly identical. Therefore we cannot conclude that the observed difference between the R^2's in these two cases is statistically significant

V \ N	1.7	2.3	3.0	4.0
0.35	λ=0.3218 δ=-0.8125 R^2=0.9990 n=19	λ=0.2857 δ=-1.3290 R^2=0.9987 n=20		
0.52	λ=0.2535 δ=-2.7198 R^2=0.9983 n=24	λ=0.2233 δ=-3.3767 R^2=0.9970 n=26	λ=0.1625 δ=-8.3325 R^2=0.9950 n=31	λ=0.1478 δ=-6.1568 R^2=0.9951 n=32
0.62		λ=0.1850 δ=-5.0193 R^2=0.9972 n=30		
0.98		λ=0.1230 δ=-10.6119 R^2=0.9937 n=44	λ=0.1013 δ=-12.9649 R^2=0.9949 n=48	

NOTATION: λ=dispersion parameter (pence^{-1}) δ=rail penalty (pence)
R^2=coefficient of determination of line defined by equation (11)
n=number of relevant cost bins

Table 2a: Sensitivity analysis results for non car owners in the bus-rail secondary split

Although all the differences are statistically insignificant, it is encouraging to find a slight tendency in the non-car owners results of Table 2a towards smaller values of both V and N. No clear pattern in terms of tendency emerges from the results for one car owners though (Table 2b) except for a very subtle preference for the medium values of both V and N.

The less encouraging results in terms of a tendency shown, are those for members of households with two or more cars, because it would appear, contrary to expectations, that higher values of V and N are not preferred. In fact the results seem to show a slight preference for the smaller values of V and N, as was the case for non car owners. However the amount of data was rather small in this case (around 20% of the data for either of the other two groups) and therefore the results must be understandably more suspect.

V \ N	1.7	2.3	3.0	4.0
0.35	λ=0.2264 δ=-0.1703 R^2=0.9814 n=20	λ=0.2053 δ=-0.6091 R^2=0.9838 n=21		
0.52	λ=0.1843 δ=-1.7819 R^2=0.9836 n=25	λ=0.1672 δ=-2.3248 R^2=0.9885 n=27	λ=0.1472 δ=-6.3285 R^2=0.9882 n=32	λ=0.1248 δ=-3.8871 R^2=0.9864 n=34
0.62		λ=0.1494 δ=-3.3038 R^2=0.9879 n=31		
0.98		λ=0.1072 δ=-7.5175 R^2=0.9873 n=42	λ=0.0902 δ=-9.1754 R^2=0.9770 n=48	

NOTATION: λ=dispersion parameter (pence^{-1}) δ=rail penalty (pence)
R^2=coefficient of determination of line defined by equation (11)
n=number of relevant cost bins

Table 2b: Sensitivity analysis results for one car owners in the bus-rail secondary split

V \ N	1.7	2.3	3.0	4.0
0.35	λ=0.4572 δ=-1.8762 R^2=0.9982 n=19	λ=0.3685 δ=-2.6270 R^2=0.9978 n=20		
0.52	λ=0.3737 δ=-4.1301 R^2=0.9975 n=23	λ=0.3283 δ=-4.7888 R^2=0.9949 n=25	λ=0.1912 δ=-9.7726 R^2=0.9910 n=30	λ=0.1822 δ=-7.8449 R^2=0.9873 n=31
0.62		λ=0.2704 δ=-6.4249 R^2=0.9962 n=27		
0.98		λ=0.1810 δ=-12.1311 R^2=0.9922 n=38	λ=0.1304 δ=-15.0681 R^2=0.9869 n=41	

NOTATION: λ=dispersion parameter (pence^{-1}) δ=rail penalty (pence)
R^2=coefficient of determination of line defined by equation (11)
n=number of relevant cost bins

Table 2c: Sensitivity analysis results for 2+ car owners in the bus-rail secondary split

With the values of λ and δ found in this part of the analysis it was possible to work out public transport composite costs and to calibrate alternative versions of the primary split model. Because the exercise is quite expensive and time-consuming and because of the lack of sensitivity shown in the secondary split calibrations, we decided to test only those cases which appear to have the greatest chance of producing significant differences. The results are summarized in Table 3.

V \ N	1.7	2.3
0.35		λ_1=0.0741 δ_1=44.8530 R^2=0.9673 n =19
0.52	λ_1=0.0743 δ_1=46.6144 R^2=0.9556 n =23	λ_1=0.0725 δ_1=48.9687 R^2=0.9408 n =26
NOTATION: λ_1=dispersion parameter (pence $^{-1}$) δ_1=public transport bias (pence) R^2 and n, as in Table 2.		

Table 3a: Sensitivity analysis results for non car owners in the car-public transport split

Several issues are worth noting from these results. First the size and sign of the public transport bias (because of the positive sign it cannot be called penalty), which gives an indication of how difficult it is for non car owners to travel by car [12]. Another important fact is that again the difference between the models is not statistically significant [13]

12. This was confirmed by Ortuzar, 1979(a) which found a definite relation between the value of δ_1 and that of a parameter ψ representing car availability. For $\psi = 1$ (everybody has a car available), δ_1 is of the order of 50p (ie. the proportion of people choosing car equals that choosing public transport, when the cost of public transport less the cost of car is 50p) and for $\psi = 0.2$ (only 20% of the population has a car available) δ_1 decreased to roughly 19p.

13. In fact t' from equation (13) is less than 1 for the 'worse' case, ie. the difference could appear by chance in more than 30% of the cases.

and moreover, the differences between the parameter themselves are not marked either. The slight tendency towards lower values of V and N is consistent with the findings of the secondary split. Table 3b presents the results for one car owners.

V \ N	2.3	3.0
0.52	λ_1 = 0.0837 δ_1 =12.8837 R^2 = 0.9860 n =21	λ_1 = 0.0591 δ_1 =15.9319 R^2 = 0.9805 n =27
0.98	λ_1 = 0.0432 δ_1 =22.0939 R^2 = 0.9584 n =32	
NOTATION:	λ_1 = dispersion parameter (pence^{-1}) δ_1 = public transport bias (pence) R^2 and n, as in Table 2	

Table 3b: Sensitivity analysis results for one car owners in the car-public transport split.

In this case the variation in the parameter estimates is more marked, but still there are no statistical differences between the goodness of fit statistics(14). The positive still, but smaller, values of δ_1 reflect the car availability problem in one car owning households. If there is a tendency, it seems to show a preference

14. Although in this case the value of t' = 1.834 for the 'worse' case means that only in between 6% or 7% of cases the differences would have been observed for truly identical R^2's. However the values for the alternative measure ρ^2 were almost identical in the three cases.

for the values of V and N recommended by WYTCONSULT (1977), which is roughly consistent with the results of the secondary split. This view is reinforced by the smaller size of the public transport bias, an always welcome feature.

V \ N	1.7	2.3
0.52	λ_1 = 0.1276 δ_1 = 3.9532 R^2 = 0.9876 n = 16	λ_1 = 0.1191 δ_1 = 4.9123 R^2 = 0.9789 n = 17
0.62		λ_1 = 0.0982 δ_1 = 5.1316 R^2 = 0.9800 n = 20
NOTATION:	λ_1 = dispersion parameter (pence^{-1}) δ_1 = public transport bias (pence) R^2 and n, as in Table 2	

Table 3c: Sensitivity analysis results for 2+ car owners in the car-public transport split.

Again in this case the variation in the parameter estimates is not too marked and again there are no statistical differences between the goodness of fit statistics [15]. The small values of δ_1 reflect the increased chance of having a car available in households which own two or more cars.

15. The value of t' was again less than 1.

7. CONCLUSIONS

We have argued throughout this note against the practice of inputting fixed exogenous values for such key parameters as the value of time, the value of waiting time and the car occupancy factor in aggregate modal split modelling. Because the current aggregate tools do not allow us to obtain estimates for these parameters directly from the data and because a full optimization search is prohibitively expensive, we proposed to test the importance of exogenously fixing the parameters using sensitivity analysis.

Analytic estimation and numerical valuation of mean point elasticities of the model parameters (λ's and δ's) with respect to the parameters under scrutiny (V, N and CO_c), led to the conclusion that the sensitivity analysis should be centred on the first two. This conclusion was supported post-hoc by other more practical considerations and because the car occupancy factor only affects car costs, while the others affect all costs.

The extensive sensitivity analysis, summarized in Tables 2 and 3 of the previous section showed no significant improvements to the models fits when varying the fixed parameters. We are fairly certain that our data base is typical of aggregate modelling. For this reason our main and rather unexpected conclusion is that the current procedure of fixing the values of V, N and CO_c, exogenously to the calibration process, does not seem to be of any significance to the model agreement to the data. Therefore, provided the exogenous values are roughly on target they should produce models that perform as well (or badly) as models calibrated entirely from the data, with the added advantage of being a less costly procedure.

REFERENCES

BONSALL, P.S., A.F. CHAMPERNOWNE, A.C. MASON and A.G. WILSON (1977) Transport modelling: sensitivity analysis and policy testing, Progress in Planning, Vol. 7, pt.3, pp. 153-237, Pergamon Press, Oxford.

HARTLEY, T.M. and J.D. ORTUZAR, (1979) Modal split models for the journey to work in West Yorkshire by car, bus and rail: some comments. Technical Note 9, Institute for Transport Studies, University of Leeds.

ORTUZAR, J.D. (1979a), Simplified methods to transform car ownership data into car availability data: a discussion, Technical Note (forthcoming), Institute for Transport Studies, University of Leeds.

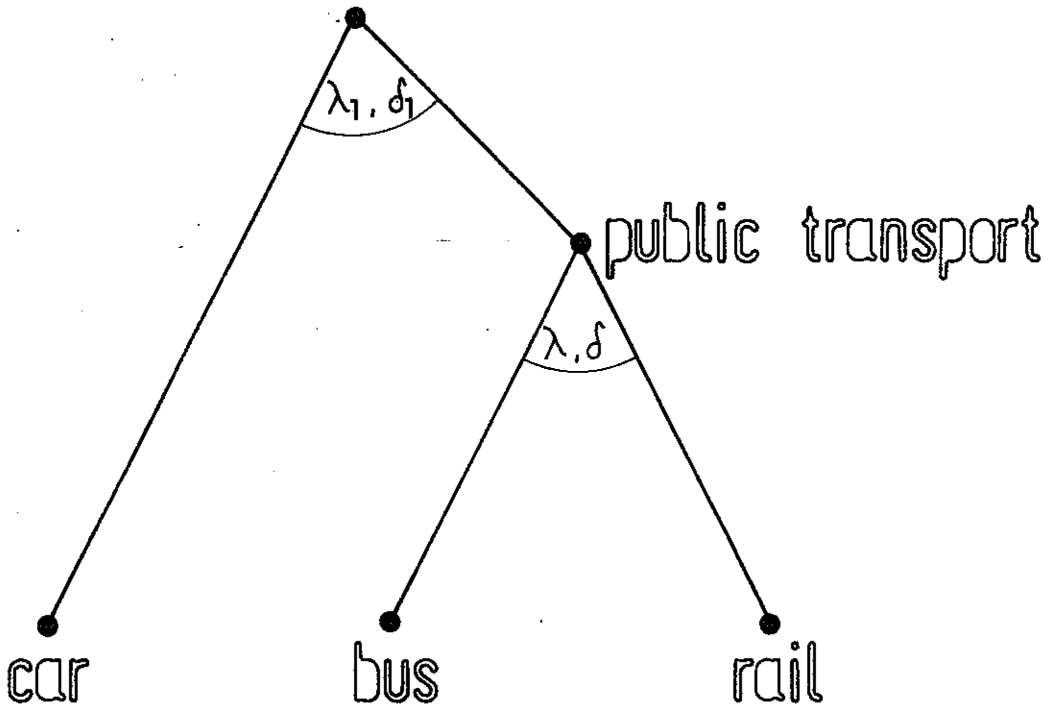
ORTUZAR, J.D. (1979b), The Garforth Transportation Corridor: description and aggregate data available, Technical Note (forthcoming), Institute for Transport Studies, University of Leeds.

SPEAR, B.D. (1977) Applications of new travel demand forecasting techniques to transport planning: A study of individual choice models, Urban Planning Division, Federal Highway Administration, U.S. Department of Transportation.

WILLIAMS, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit, Environment and Planning A, vol. 9, pp.285-344.

WILLIAMS, H.C.W.L. and J.D. ORTUZAR (1979), Behavioural theories, model specification and the response error problem, PTRC Summer Annual Meeting, University of Warwick. (Also available as Working Paper 116, Institute for Transport Studies, University of Leeds).

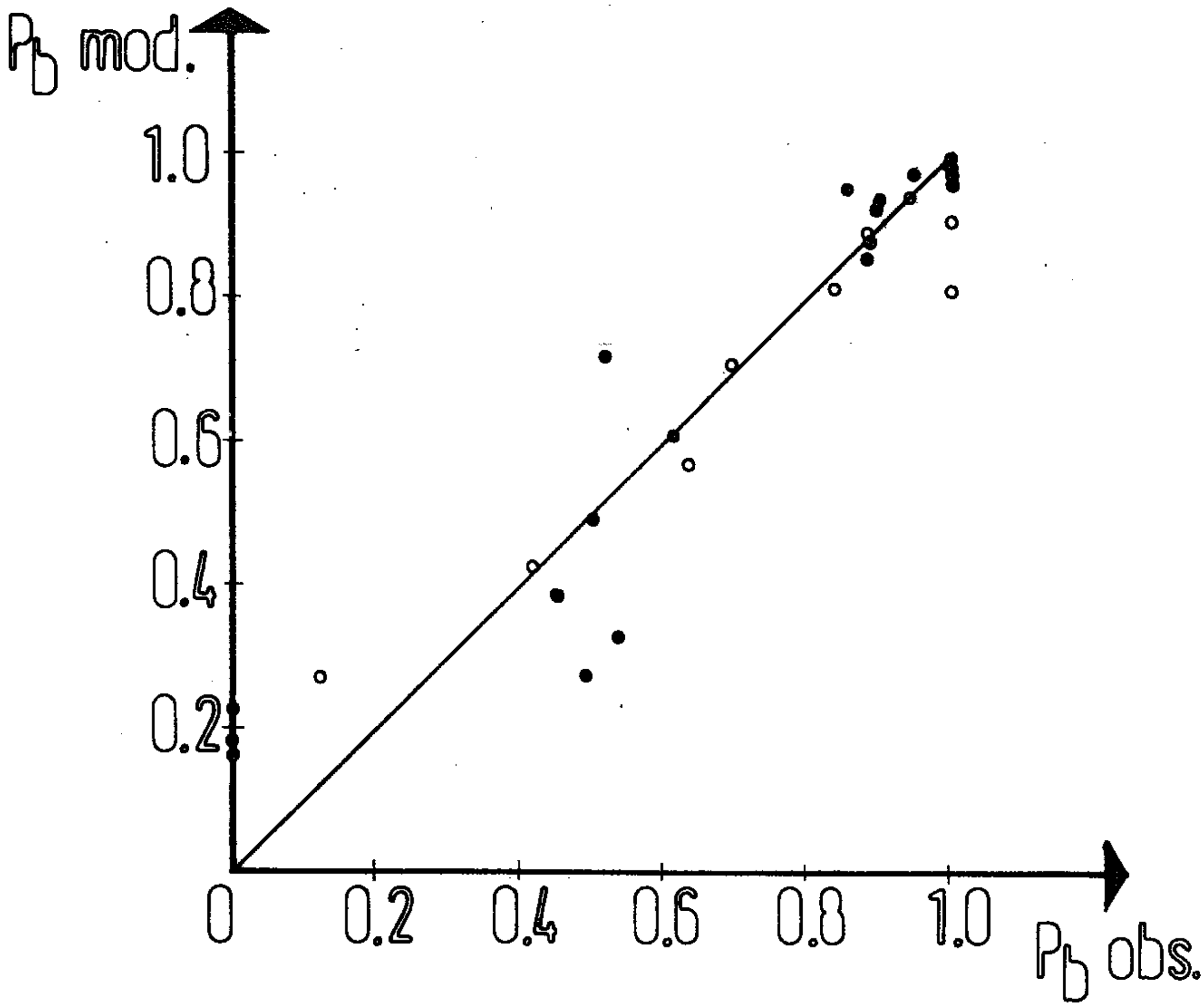
WYTCONSULT (1977) The structure and calibration of the Leeds and Bradford Transport Model , Document 313 , vol.1 West Yorkshire Transportation Studies, Leeds.



NOTATION:

λ , λ_1 = secondary and primary split dispersion parameters.
δ , δ_1 = rail and public transport modal penalties.

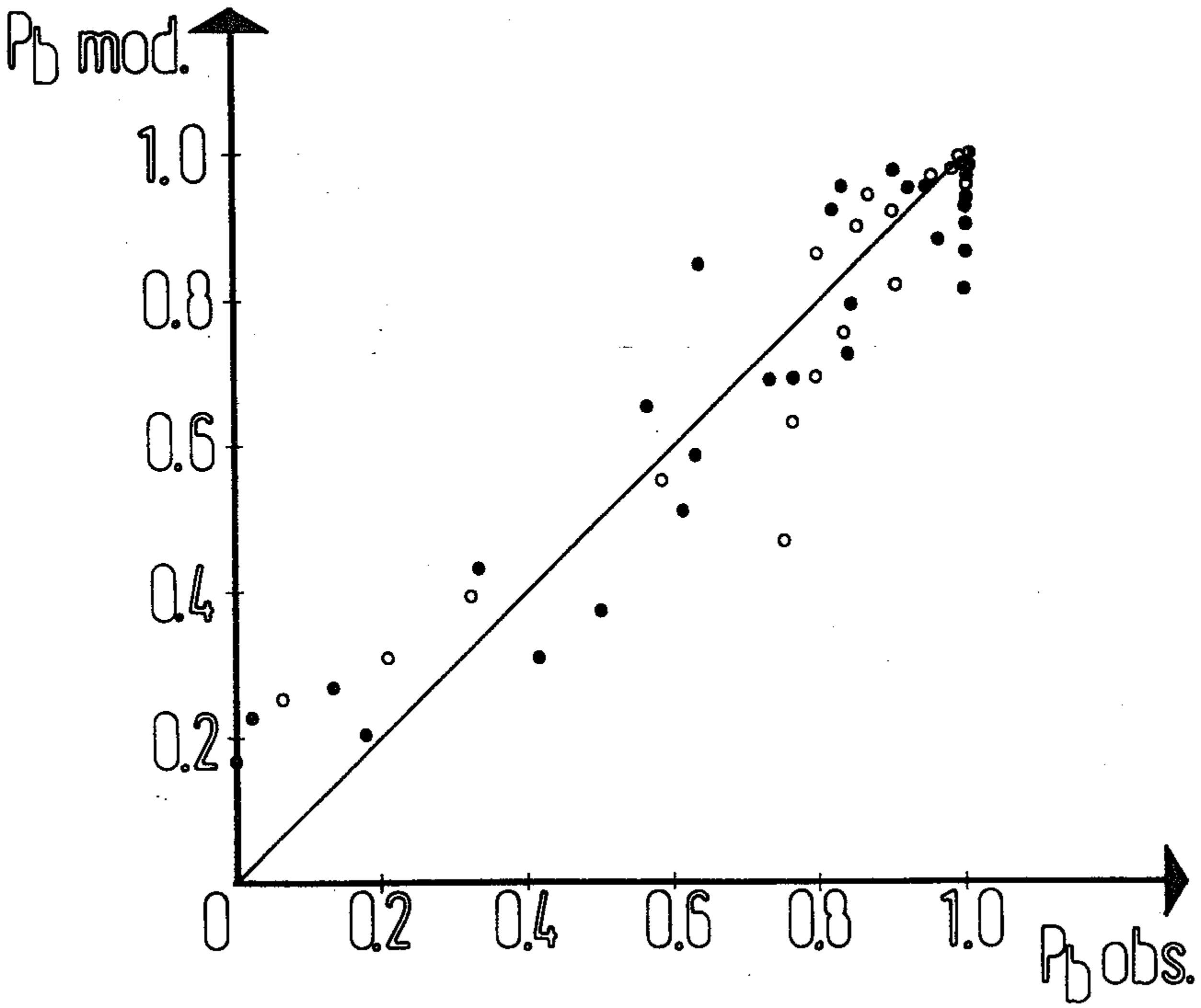
FIGURE 1: Representation of a hierarchical logit modal split model for car, bus and rail trips.



NOTATION:

P_b	=	proportion using bus.
o	=	best fit, for V = 0·35 and N = 1·70.
•	=	worse fit, for V = 0·98 and N = 2·30.
V	=	value of in-vehicle time (p /min).
N	=	waiting time weight.

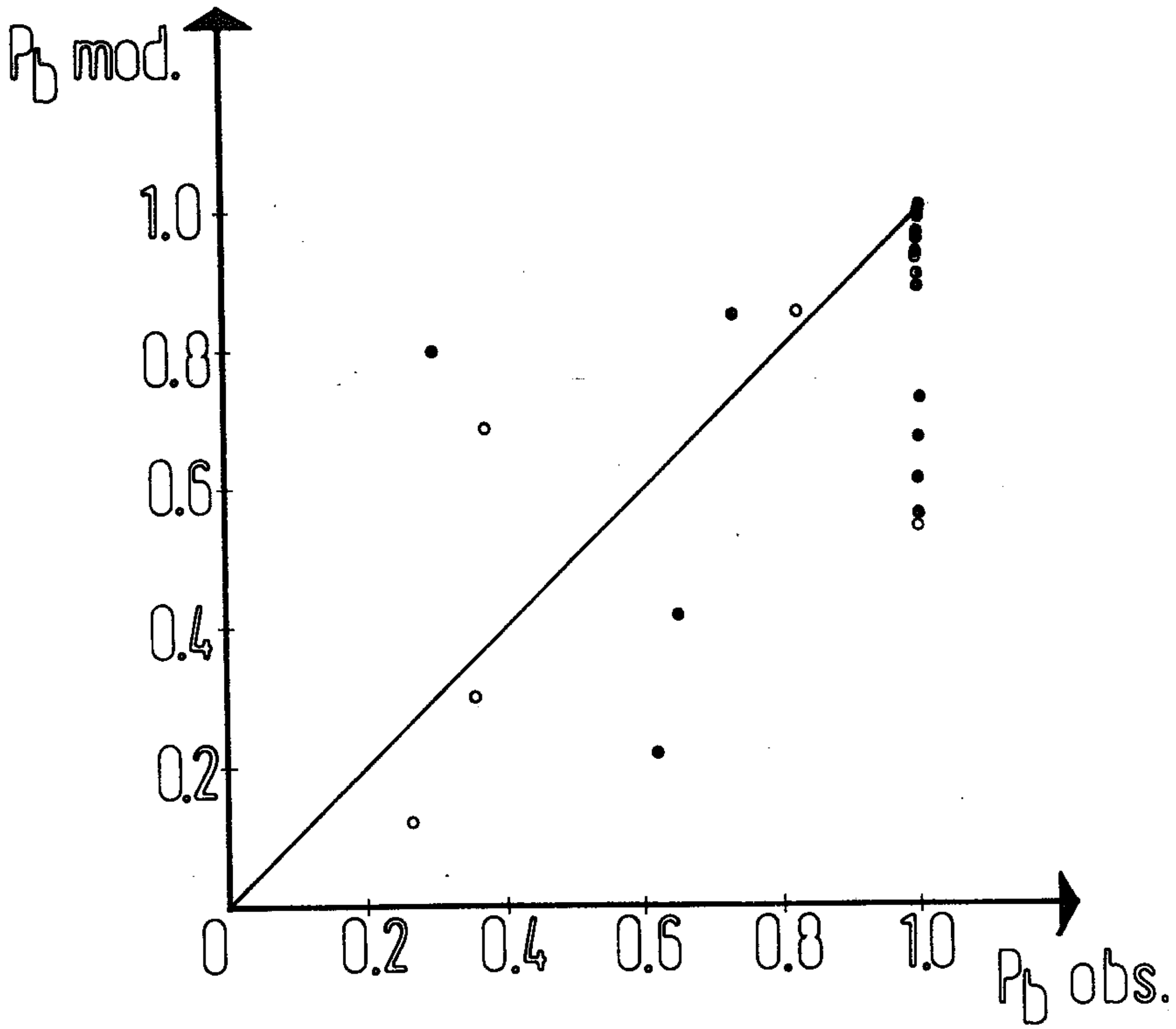
FIGURE 2: Non car owners. Comparison of best and worse model fits in the bus-rail split.



NOTATION:

P_b = proportion using bus.
○ = best fit, for V = 0·52 and N = 2·30.
● = worse fit, for V = 0·98 and N = 3·00.
V = value of in-vehicle time (p/min).
N = waiting time weight.

FIGURE 3: One car owners.
Comparison of best and worse model fits in the bus-rail split.



NOTATION :

P_b	=	proportion using bus.
o	=	best fit, for V=0·35 and N=1·70.
•	=	worse fit, for V=0·98 and N=3·00.
V	=	value of in-vehicle time.
N	=	waiting time weight.

FIGURE 4 : Two or more car owners. Comparison of best and worse model fits in the bus-rail split.

ACKNOWLEDGEMENTS

This work would not have been possible without the data for 1975 peak journeys in West Yorkshire, provided by West Yorkshire Metropolitan County Council.

I wish to thank my colleagues Trevor Hartley, Dirck Van Vliet, Huw Williams and Luis Willumsen for useful discussions and amendments to the original manuscripts. Special thanks are due to Hugh Gunn for his patience and knowledge.

Finally, I would like to thank the team of secretaries of ITS for deciphering the manuscript into type and Margarita Greene for skilfully producing the figures.

APPENDIX

1. Derivation of analytic point elasticities

For the purposes of this analysis we will examine an even simpler version of the equations presented in Section 2. We will not consider here the modal penalties, which are in fact, just scaling parameters. Therefore, equation (7) can be rewritten as:

$$MS_b = \frac{1}{1 + \exp(-\lambda \Delta G)}$$

$$MS_r = 1 - MS_b \tag{A1}$$

Similarly, equation (8) can be rewritten as:

$$MS_c = \frac{1}{1 + \exp(-\lambda_1 \Delta_1 G)}$$

$$MS_{pt} = 1 - MS_c \tag{A2}$$

Now the following identities are easily derived from the equations presented in Section 2.

$$\frac{\partial \Delta G}{\partial \Delta V} = \Delta T + N.\Delta WT \tag{A3}$$

$$\frac{\partial \Delta G}{\partial N} = \Delta WT.V \tag{A4}$$

$$\frac{\partial \Delta_1 G}{\partial V} = \Delta_1 T + N.\Delta_1 WT \tag{A5}$$

$$\frac{\partial \Delta_1 G}{\partial N} = \Delta_1 WT.V \tag{A6}$$

$$\frac{\partial \Delta_1 G}{\partial CO_c} = \frac{OP_c.D}{CO_c^{\ 2}} \tag{A7}$$

For the purposes of working out elasticities for the parameters in the secondary split equation we need to assume that the modal split remains constant; this is in fact equivalent to say, looking at equation (A1), that:

$$\lambda \, \Delta G = \text{constant} = K \tag{A8}$$

For the primary split case we can similarly have that:

$$\lambda_1 \Delta_1 G = \text{constant} = K' \tag{A9}$$

From these two identities, we can easily derive the following equations:

$$\frac{\partial \lambda}{\partial \Delta G} = - \frac{K}{\Delta G^2} \tag{A10}$$

$$\frac{\partial \lambda_1}{\partial \Delta_1 G} = - \frac{K'}{\Delta_1 G^2} \tag{A11}$$

Therefore, the elasticity of λ with respect to V can be worked out as:

$$\begin{aligned} \varepsilon_{\lambda/V} &= \frac{\partial \lambda}{\partial V} \cdot \frac{V}{\lambda} \\ &= \frac{\partial \lambda}{\partial \Delta G} \cdot \frac{\partial \Delta G}{\partial V} \cdot \frac{V}{\lambda} \end{aligned} \tag{A12}$$

and from equations (A10), (A3) and (A8), this simply reduces to:

$$\varepsilon_{\lambda/V} = \frac{(\Delta T + N \, \Delta WT) \; V}{\Delta G} \tag{A13}$$

Now recalling that

$$\Delta G = \Delta F + (\Delta T + N . \Delta WT) \; V$$

we can rewrite equation (A13) as:

$$\varepsilon_{\lambda/V} = - \frac{1}{1 + \dfrac{\Delta F}{V\ (\Delta T + N.\Delta WT)}} \tag{A14}$$

Proceeding similarly for the others we get:

$$\varepsilon_{\lambda/N} = - \frac{1}{1 + \dfrac{\Delta F + \Delta T.V}{N.\Delta WT.V}} \tag{A15}$$

$$\varepsilon_{\lambda_1/V} = - \frac{1}{1 + \dfrac{F_{pt} - \dfrac{OP_c.D}{CO_c}}{V\ (\Delta_1 T + N.\Delta_1 WT}} \tag{A16}$$

$$\varepsilon_{\lambda_1/N} = - \frac{1}{1 + \dfrac{F_{pt} - \dfrac{OP_c.D}{CO_c} + \Delta_1 T.V}{N.\Delta_1 WT.V}} \tag{A17}$$

$$\varepsilon_{\lambda_1/CO_c} = - \frac{1}{\dfrac{F_{pt} + V\ (\Delta_1 T + N.\Delta_1 WT)}{OP_c.D/CO_c} - 1} \tag{A18}$$

2. Producing numerical values

In order to have some numerical values which will allow us to make some comparisons, we will boldly postulate from our data and common sense, that not unreasonable mean values for the Leeds-Garforth corridor are as follows:

$D = 3$ Kms

$T_b = 11.25$ min (speed of 10 mph)

$T_r = 4.50$ min (speed of 25 mph)

$WT_b = 7$ min

$WT_r = 15$ min

$T_c = 4.50$ min (speed of 25 mph)

$WT_c = 1.0$ min

Apart from these assumed values, the standard fares for bus and rail and the car operating costs in 1975 were as follows, according to WYCONSULT (1977):

$F_b = 4.63$ pence + 0.97 p/km

$F_r = 7.50$ pence + 0.95 p/km

$OP_c = 1.77$ p/km

With all these values we can work out the following mean differences:

$\Delta F = 2.81$ pence

$\Delta T = -6.75$ min

$\Delta WT = 8.00$ min(*)

Because the split bus-rail for the whole of West Yorkshire is roughly 9:1 for the journey to work, a not unreasonable representation of the composite public transport costs is:

$F_{pt} = 4.9$ pence + 0.97 p/km

$T_{pt} = 10.5$ min

$WT_{pt} = 7.8$ min

therefore

$\Delta_1 T = 6$ min

$\Delta_1 WT = 6.8$ min(*)

$OP_c D = 5.31$ pence

(*) Notice that we are taking ΔWT and $\Delta_1 WT$ as being of the same order of ΔT and $\Delta_1 T$. As the former are further amplified by N in the model, if our assumption is incorrect it could lead to very misleading results. For this reason we will check what happens if these values are reduced substantially (ie people 'time' their arrivals).

Using these mean values we can get an idea of the magnitude of the elasticities derived above. Notice in the formulae, that the elasticities with respect to V depend on N and vice versa. Notice also that the elasticities of λ_1 are more complex than those of λ because they depend indirectly on λ through the values for the fares and times of the composite public transport mode.

The rest of the Appendix is a collection of tables showing the sort of variations it is possible to get in the point elasticity estimates under several assumptions. We will be interested both in the range and in the absolute magnitude of the values.

N	$\varepsilon_{\lambda/V}$ for V = 0.52 p/min and ΔW = 8 min.	$\varepsilon_{\lambda/V}$ for V = 0.52 p/min and ΔW = 4 min.
1.7	- 0.56	- 0.009
2.3	- 0.68	- 0.310
3.0	- 0.76	- 0.490
4.0	- 0.82	- 0.630
5.0	- 0.86	- 0.710

Table A1: Elasticities of λ with respect to V for different values of N and ΔWT.

V	$\varepsilon_{\lambda/N}$ for N = 2.3 ΔWT = 8 min.	$\varepsilon_{\lambda/N}$ for N = 2.3 ΔWT = 4 min.
0.35	- 0.94	- 0.88
0.52	- 1.08	- 1.17
0.62	- 1.14	- 1.32
0.98	- 1.27	- 1.73

Table A2: Elasticities of λ with respect to N for different values of V and ΔWT.

$\varepsilon_{\lambda_1/V}$		N				
		1.7	2.3	3	4	5
	1.0	-0.73	-0.82	-0.85	-0.87	-0.89
	1.3	-0.64	-0.75	-0.79	-0.82	-0.85
CO_c	1.5	-0.61	-0.73	-0.76	-0.80	-0.83
	2.0	-0.56	-0.69	-0.73	-0.77	-0.80

Table A3: Elasticity of λ_1 with respect to V for different values of N and CO_c.

$\varepsilon_{\lambda_1/N}$		V			
		0.35	0.52	0.62	0.98
	1.0	-0.54	-0.59	-0.61	-0.65
	1.3	-0.48	-0.54	-0.57	-0.61
CO_c	1.5	-0.46	-0.52	-0.55	-0.60
	2.0	-0.43	-0.50	-0.52	-0.58

Table A4: Elasticity of λ_1 with respect to N for different values of V and CO_c

$\varepsilon_{\lambda_1/CO_c}$		N				
		1.7	2.3	3	4	5
	0.35	-0.50	-0.36	-0.32	-0.27	-0.23
	0.52	-0.39	-0.27	-0.23	-0.19	-0.17
V	0.62	-0.35	-0.24	-0.20	-0.17	-0.14
	0.98	-0.25	-0.16	-0.14	-0.11	-0.09

Table A5: Elasticity of λ_1 with respect to CO_c for different values of V and N.

Table A6 summarizes our results presenting for the dispersion parameters (λ and λ_1) point elasticities with respect to V, N and CO_c, their mean estimates for the WYTCONSULT values and their likely range of variation under our assumptions.

	RANGE			
	Minimum	Medium	Maximum	
$\varepsilon_{\lambda/V}$	-0.56	-0.68	-0.86	See Table A1
$\varepsilon_{\lambda/N}$	-0.94	-1.08	-1.27	See Table A2
$\varepsilon_{\lambda_1/V}$	-0.56	-0.75	-0.89	See Table A3
$\varepsilon_{\lambda_1/N}$	-0.43	-0.54	-0.65	See Table A4
$\varepsilon_{\lambda_1/CO_c}$	-0.09	-0.27	-0.50	See Table A5

Table A6: Variation in the point elasticity estimates of the dispersion parameters.

If we accept the trends shown by these values it would appear that V and N are both very important, perhaps the latter slightly more so in view of its effect on λ which in turn would affect λ_1. It also appears quite clearly that CO_c is the less worthwhile factor to consider, a not surprising finding.