A Probability Model of Travel Mode Choice For the Work Journey

PETER R. STOPHER, Northwestern University, Evanston, Illinois

This paper describes the derivation of a probability model of travel mode choice for the work journey in terms of the differences in costs and times between the modes available for each individual's journey. The work is based on surveys carried out at two offices in central London, with some test data from a survey of workers in central Leeds. Regression techniques were used to establish simple linear relationships between the probability of using a car and the cost and time differences between a car and the best available public transport route. Reasonable correlation coefficients were obtained. Limited tests of the models have reproduced existing mode choices to a high degree of accuracy. Relationships were also established between the regression coefficients and income. A generalized model is derived from these relationships.

The basic model has a serious fault in that because it is linear, it permits values of probability to be obtained that exceed unity, or are less than zero. A logistic transformation is put forward to correct this and is used on the generalized model. The resulting model still appears to reproduce existing conditions to a similarly high degree of accuracy. The model appears to be reasonably behavioral, and could form the basis of a new and more accurate mode choice procedure.

•IN THE PAST, most models of travel mode choice have been developed as a stage in a traffic study for a particular area. To do this, the model devised has to conform to certain requirements. It must first fit into an existing forecasting package, or be able to be tailored to fit with existing trip distribution and assignment techniques. The model must then be able to be applied to a mass of people and to predict the volumes of travel on the various modes. Consequently, the model is often defined in terms of socioeconomic factors that serve to describe the population rather than the transport network available. Forecasting travel mode choice at a future date requires an assumption that people will change their socioeconomic classes, and that their travel desires and habits will conform with those of the class into which they have moved. This assumption is somewhat difficult to justify, particularly since it must depend on the ability of the transport network to accept an increased demand for travel.

More recently, attempts have been made to devise models using parameters that are descriptive of the transport network, e.g., San Francisco Bay Area Transportation Study (1) and Twin Cities Transportation Study (2). These parameters have usually been an attempt at devising a measure of accessibility to apply to an area, or some form of time ratio. However, the derivation of a model that is descriptive of the transport system, but that is applied to a mass of people still leaves something to be desired in terms of a behavioral model, although the model developed by Traffic Research Corporation (3) goes far toward achieving this.

The model attempts to describe mode choice in terms of each individual and the part of the transport system available to carry out the desired journey.

DESIGN OF THE MODEL

The first stage in the design of the model was to carry out a limited survey on a part of the population to which the model was intended to be applied. From the results of this survey, it was planned to choose the parameters to be used in the model and also possibly to make an initial tentative calibration of the model. For these reasons, the survey took the form of a questionnaire about the work journey, with a request for a number of socioeconomic factors required to define the sampled population that might be necessary to help explain mode choice. The survey also included two attitudinal questions about choice of mode. These were open-ended questions designed to elicit information on perceived reasons for choice of a particular mode.

This survey was designed as a questionnaire to be circulated without follow-up and with no time limit set on the returns. The location selected for the survey was University College London, and the questionnaires were circulated through the internal post system to all academic and administrative staff of the College. Apart from convenience, University College represented a useful location for this survey for two reasons. First, the model sought for at this stage was one that would predict the level of use of a car for the work journey. This required a fairly large number of car owners to be surveyed, but because it was not desirable to collect a large quantity of less useful information on non-car owners, a high proportion of car owners was desirable. This was expected to be the situation at University College. Second, the College is located within the Central Area of London as defined by the London Traffic Survey (4) and is well-served by public transport. It is also near the focus of the radial routes into London. The College also provides some free parking for staff, and although it is in an area of parking meters, it is less than a quarter-mile from the boundary of the controlled parking area (as of 1966, when this survey was carried out).

A total of 1308 questionnaires were distributed, of which 767 usable returns were received. This represents a response rate of 59 percent. Of the forms returned, 26 had to be rejected, representing only 2 percent of the dispatched questionnaires, or just over 3 percent of the returned questionnaires.

An analysis indicated a higher car ownership level and a higher average income than would be expected from a random sample in London. The comparative figures are given in Table 1, in which the London figures are from the London Traffic Survey (4). Analysis of the answers to the open-ended question on reasons for the choice of mode showed that the most important factors for people with a choice were speed, convenience, cost and comfort, in approximately that order. Speed was put forward by most people in all categories; i.e., non-car owners, car users, and car owners using public transport. Convenience came second to this and cheapness third for all public transport users, but fourth for car users. Comfort came third in importance to car users, but was mentioned by less than 10 percent of public transport users.

From this, it would appear that given a choice between any two modes, the most important factors that an individual perceives as influencing his choice are time, cost, and convenience, with the addition of comfort if the choice pair includes a car. Up to the present, it has not been possible to devise a quan-

TABLE 1

INCOME AND CAR OWNERSHIP DISTRIBUTIONS FOR
UNIVERSITY COLLEGE IN 1966 AND THE
LONDON TRAFFIC SURVEY AREA IN 1962

Income Range (pounds)	University College (%)	London Area (%)	Car Owners in Total Population (%)	
			University College	London Area
Under 500	2.0	14.0	0.2	0.5
500 to 999	21.0	39.5	3.1	10.3
1000 to 1499	18.0	27.5	6.8	13.0
1500 to 1999	16.0	10.5	10.8	7.2
2000 to 2999	23.0	5.5	16.7	4.3
3000 and over	20.0	3.0	15.0	2.7
Total	100.0	100.0	52.6	38.0

titative expression for comfort or convenience. It was therefore decided to concentrate on devising a model in terms of costs and times for the available modes only.

The basic model to be devised is one expressing the choice between car and public transport in terms of the time and cost for a particular journey by each mode. The form of model desired is one that allots a probability of using a car to each individual considered. In dealing with a given population, it is first necessary to exclude all those who have no car. Because the model is concerned with a real choice, it is then also necessary to exclude all captive car users, so that the public-private transport model is applied to the individuals in the remainder of the population.

The simplest form of the model would be a linear one between the probability of using the car and a function of costs and times. Both ratios and differences were considered for the functions of costs and times. To select one of these functions for the model, consideration was given to two factors: whether the function appeared to be well correlated with observed behavior, and whether the function appeared to reflect a person's attitude. The resulting data were grouped by total cost and time, in which an approximate probability was allotted to each group, based on the numbers of car and public transport users in each group. A very wide variation in probability occurred for any particular small range of cost and time ratios. It also seems more likely for relatively short journeys that cost and time savings are assessed in absolute terms not relative to the overall cost or time. For these reasons it was decided to use differences that show a much closer correlation with probability and that imply absolute valuation of any savings in costs and times.

The form of the model proposed is

$$p = a(c_2 - c_1) + b(t_2 - t_1) + d \qquad (1)$$

where d represents the probability of using the car when costs and times by public transport and car are the same. If choice depends on costs and times only, d would be 0.5. The departure of d from 0.5 gives an indication of a bias for or against car use on grounds other than cost and time. The value of b/a represents the implied value of travel time. The symbols c and t represent cost and time, respectively, and the subscripts 1 and 2 represent car and public transport, respectively. This model can be derived mathematically, and a possible derivation of it has been given previously (5).

An initial calibration was carried out on the sample of car owners from University College. The questionnaire had asked all respondents to give details of their usual work trip and details of the mode or modes that they considered to represent their best alternative. It was assumed, however, that all car owners would either use their car or specify car as their best alternative. For each car owner, costs and times were calculated for car and for the best public transport journey (i.e., either the public transport journey carried out or the one stated as alternative to using the car), and cost and time differences were calculated from these.

The car owning population was then grouped into a number of classes representing small ranges of cost and time differences. The total numbers of car users and public transport users within each class were counted, and the probability of using a car in each class was taken to be the number of car users in the class divided by the total population of that class. A regression technique was then used to find a relationship between this probability and the actual cost and time differences for each individual. The result of this process was the following model:

$$p = 0.00278(c_2 - c_1) + 0.00695(t_2 - t_1) + 0.394 \qquad (2)$$

for which the correlation coefficient R was 0.63 (R^2 = 39.7 percent). The implied value of time from this model is 2.5 pence per minute. The average wage rate of the sample was 2400 pounds a year, so that this value of time represents 0.42 of the wage rate. For salaries in a similar range to this, Beesley (6) found values of time between 0.42 and 0.50 of the wage rate; and in the Toronto survey (7), a value of 0.41 of the wage rate. However, it will be seen later that further analysis produced some lower time values, and it appears that the above estimates are somewhat inflated.

Resubstitution of the University College data in the model gave a prediction of 135 car journeys out of 311 car owners.   The observed figure was 136 car journeys.

TEST OF THE BASIC MODEL

Having set up a basic model form and having carried out a tentative calibration of the model, it was decided to collect further data to test the model and to allow some recalibration if necessary.   A second work trip survey was carried out, this time at County Hall London, the central office of the Greater London Council.   This office block is situated on the south side of the Thames nearly opposite the Houses of Parliament and close to Waterloo Main Line Terminus station and Waterloo Underground station.

Approximately 6500 people work at County Hall and, as at University College, a questionnaire survey was designed to be circulated to all people working in the block. Again, no interviewing or personal follow-up was carried out.   County Hall is not situated in a meter-controlled parking area, but practically no free parking is provided for employees.   Cars must either be parked in side streets or at a pay car park adjacent to County Hall.   It was hoped that the population of County Hall would exhibit much less bias than the University College sample, and that in terms of both income and car ownership it would much more closely resemble a random sample from the London area.

Because the model dealt with car owners only, the survey at County Hall was designed to eliminate non-car owners at an early stage and request journey details from car owners only.   Information was collected from everybody on income, car-owning status of themselves and their household, and availability of a car for use for the work journey.   Non-car owners were then eliminated.   This allowed computation of statistics of car ownership and income, but avoided the collection and coding of a vast quantity of irrelevant information.

A total of 6395 questionnaires were sent out, with a time limit of $2\frac{1}{2}$ weeks set for the return of completed forms.   Of these, 5166 forms were returned, from which 258 had to be rejected as incomplete.   The 4908 usable replies represent a response rate of 77 percent that is exceptionally high.

Details of car ownership and income, compared with values from the 1962 London Traffic Survey, are given in Table 2.   By applying an adjustment to the lower income groups recorded in the London Traffic Survey to allow for the $4\frac{1}{2}$-yr difference between the two surveys, so that the proportion in the lowest group is assumed to decrease while the next two groups increase, it can be seen that the distribution from County Hall is very close to what might be expected from a random sample from the London area.   Thus, one of the aims in selecting County Hall for the second survey was achieved.

Of the 4908 replies, 2871 were from non-car owners, and 2037 from car owners. All the remaining analysis refers to these latter.   Only 456 of the car owners used

TABLE 2

INCOME AND CAR OWNERSHIP DISTRIBUTIONS FOR COUNTY HALL
IN 1966 AND THE LONDON TRAFFIC SURVEY AREA IN 1962

Income Range (pounds)	County Hall (%)	London Area (%)	Car Owners in Total Population (%)	
			County Hall	London Area
Under 500	6.5	14.0	0.5	0.5
500 to 999	46.5	39.5	10.5	10.3
1000 to 1499	26.5	27.5	14.0	13.0
1500 to 1999	11.5	10.5	8.8	7.2
2000 to 2999	7.0	5.5	6.0	4.3
3000 and over	2.0	3.0	1.7	2.7
Total	100.0	100.0	41.5	38.0

TABLE 3

RESULTS OF REGRESSION ANALYSIS BY INCOME GROUP FOR
THE COUNTY HALL SAMPLE

Income Range (pounds)	Value of Time (pence per min)	Coefficient of Cost Difference	Constant Term	Correlation Coefficient	Sample Size
Under 1000	0.45	0.00998	0.357	0.77	415
1000 to 1499	0.65	0.00801	0.336	0.80	514
1500 to 1999	0.80	0.00698	0.308	0.77	321
2000 to 2999	1.10	0.00550	0.300	0.66	214
3000 and over	0.20	0.00376	0.218	0.49	73

their cars to travel to work, 1529 used public transport and the remaining 52 used a cycle, or walked to work, or traveled as a passenger in someone else's car.

The basic model calibrated at University College was then used on the County Hall data. The standard error of estimate of the model applied to the County Hall data is ±10 percent. The model predicted 554 car journeys by 1983 car owners compared with 456 who actually used their cars. The difference between prediction and observation is just under 5 percent. Considering the large differences in income and car ownership between University College and County Hall, this prediction error seems to be remarkably small. The County Hall data were split into ten groups based on income, and it was found that the University College model predicted too many trips in each group. All the misclassifications in these groups were additive.

MODIFICATION OF THE MODEL

It may be expected that valuation of time and probability of using the car will vary with income. The population of car owners obtained at University College was too small to allow any analysis to be carried out on variation of the model with income. However, the much larger population at County Hall does allow analysis of this sort. It may be anticipated that a very different value of time would be found for the whole population at County Hall, because the 2.5 pence per minute found at University College represents 0.74 of the average wage rate at County Hall.

The same technique was used on the County Hall data, which was split into five income groups, as was used on the University College data to perform a regression analysis. The results of the regression analyses are given in Table 3.

With the exception of the highest income group, the values of the coefficients appear to vary consistently. The constant term and the cost coefficient each decrease with increasing income, although the value of time increases. However, Table 4 shows that although the value of time increases with income, as a proportion of the wage rate it falls. Although the value of time found in the University College model does not correspond with that at County Hall for a comparable income level, the coefficients of time difference are similar, being 0.00695 for a mean income of 2400 pounds at University College and 0.00695 for the 2000- to 2999-pound income group at County Hall. Also, the values of time found here are very similar to values obtained by Quarmby (8) at Leeds.

The new models were again tested by resubstitution of the data used to calibrate them. The overall result was a prediction of 382 car journeys for a sample of 1537 car owners (some car owners had to be dropped from the analysis owing to insufficient data to calculate times and costs on an alternative mode), of whom 384 actually made car journeys. The total misclassification by income groups was 8 trips, the misclassifications having a tendency to be self-cancelling when the results for all income groups were summed up.

TABLE 4

VARIATION OF VALUE OF TIME WITH INCOME
FOR THE COUNTY HALL SURVEY

Income Range (pounds)	Value of Time as a Proportion of the Wage Rate
Under 1000	0.32
1000 to 1499	0.26
1500 to 1999	0.23
2000 to 2999	0.21

GENERALIZED MODEL

It has already been observed that a consistent variation of model constants appears to occur with income group. An investigation was carried out to determine the mathematical form of the variation. The analysis of these variations must not be relied on to too great an extent, because the regressions carried out use only four or five observations. However, all the correlation coefficients are significant. The following relationships were established for the model constants:

$$a = 0.0337 - 0.0036 \log_e (I) \tag{3}$$

$$b = 0.00125 \log_e (I) - 0.00375 \tag{4}$$

$$d = 0.3825 - 0.000038 \, I \tag{5}$$

where a, b, and d are the constants in Eq. 1, and I is the income in pounds a year. The correlation coefficients obtained for these equations gave R^2 values of 0.99, 0.98, and 0.90 respectively. From Eqs. 3 and 4 the value of time is found to be given by

$$v = \frac{0.00125 \log_e (I) - 0.00375}{0.0337 - 0.0036 \log_e (I)} \tag{6}$$

From these equations, overall models were computed for the County Hall and University College samples in terms of the mean incomes of the two samples:

$$\text{County Hall:} \quad p = 0.0076(c_2 - c_1) + 0.00529(t_2 - t_1) + 0.331 \tag{7}$$

$$\text{University College:} \quad p = 0.0057(c_2 - c_1) + 0.00599(t_2 - t_1) + 0.291 \tag{8}$$

The two sets of data were substituted in these models, and predictions of 383 and 105 car journeys, respectively, were obtained. The observed numbers of car journeys were 384 and 136. The relationships in Eqs. 3 to 6 do not appear to hold for incomes over 3000 pounds a year. Because this category represented only 4.5 percent of the County Hall sample, but 16 percent of the University College sample, this may account for the poorer result obtained at University College using the generalized model.

It was noted that the County Hall sample consisted of almost entirely car-train choices, there being only 92 car-bus choices for the 1537 car owners in this sample. It has been stated previously that the constant term of the model (d in Eq. 1) indicates a bias for or against car travel on grounds other than cost and time. It seems probable, on the basis of the replies to the attitude question in the University College survey, that much of this bias would be based on the grounds of comfort and convenience. If this is so, one would expect different values of the constant term according to whether one is considering car-bus or car-train choices. To investigate this, the general County Hall model was applied to the 92 car-bus choices with the result that a prediction of 40 car users was obtained. It was observed that 61 of the 92 people in this sample actually used their cars, so that the model is 22.8 percent in error. Without further data to check, it was then assumed that a correlation factor of 0.228 should be added to d for application to car-bus choices alone.

Data from Quarmby's survey at Leeds (8) were then used to test the model. These data consisted of trip details for 542 car owners, all of whom had a simple car-bus choice. The mean income of the car owners was 1400 pounds a year, which gives the following model for this population:

$$p = 0.0076(c_2 - c_1) + 0.00529(t_2 - t_1) + 0.559 \tag{9}$$

using the addition of 0.228 to d for car-bus choices. Of the 542 car owners, 112 were found to be captive to car or bus, so the model was applied to the remaining 430 car owners. The model predicted 270 car users, and the observed number was 264, an error of just less than 2 percent.

FURTHER MODIFICATIONS

In its present form the model is found to have one serious fault. Because it is a purely linear relationship, it is possible for very large positive or negative values of cost and time difference to yield probabilities in excess of 1 or less than 0, neither of which is permissible. This suggested that modifications should be made to prevent the occurrence of such values. The simplest method would be to define the model as a step function so that the linear form held true for values of cost and time differences that yield a probability between 0 and 1, but that for all other values of cost and time difference, the value of p becomes independent of cost and time differences, and tends to 0 or 1 according to whether the differences are negative or positive. However, this is not a particularly satisfactory treatment of the model.

A relatively simple mathematical form that would yield the desired form of relationship is the simple logistic where $p = e^y/(1 + e^y)$, and y is a linear function of various parameters. In this instance, y would be a linear function of cost and time differences, transformed directly from the original models. The logistic form can be written as

$$\log_e \frac{(p)}{(1 - p)} = y = a'(c_2 - c_1) + b'(t_2 - t_1) + d' \tag{10}$$

It is convenient to assume that the value of time remains unaltered in this transformation, so that b'/a' retains the same value as b/a. In that case, it is found that a' equals $\log_e [d/(1 - d)]$. Thus, each of the models already produced can be altered into the logistic form by transforming the coefficients. This same transformation can also be applied to the generalized formulas for the model constants with the following results:

$$a' = 0.1348 - 0.0144 \log_e (I) \tag{11}$$

$$d' = \log_e \frac{(0.3825 - 0.000038 \, I)}{(0.6175 + 0.000038 \, I)} \tag{12}$$

where v is as in Eq. 4. Application of these transformed models to each income group in the County Hall data yielded a prediction of 397 car journeys instead of the 384 observed. Deriving a general model similar to that of Eq. 7, a prediction of 398 car journeys is obtained. Similarly, devising a logistic model for the University College sample, a prediction of 113 car journeys is obtained compared with 136 observed for the 311 car owners.

The effect of this logistic transformation can be seen in Figures 1 and 2, which show the original linear models and the new logistic models for the County Hall and University College samples. The main result of this exercise is to yield a model more behavioral and more mathematically satisfactory than the original simple linear model.

FUTURE DEVELOPMENTS

The model presented here represents the first stages in the development of a new mode choice procedure for work journeys. The survey at University College suggests that the same parameters used in this model may apply to the choices between public transport submodes for captive riders. If this is so, it should prove possible to build models for the choice between any pair of modes in an overall system containing more than two modes. This could lead to a modal split procedure involving the isolation of captive car users and captive public transport users. The captive public transport users could be split between available public transport modes by use of public transport submodels, and the remainder of the population could be split between available modes by a procedure involving either the use of certain submodels representing all the options available, or by a public-private model with submodels for the public transport users.

It is hoped that the development of such a procedure as that outlined above will be pursued. This could be allied to a new distribution and assignment procedure that in-
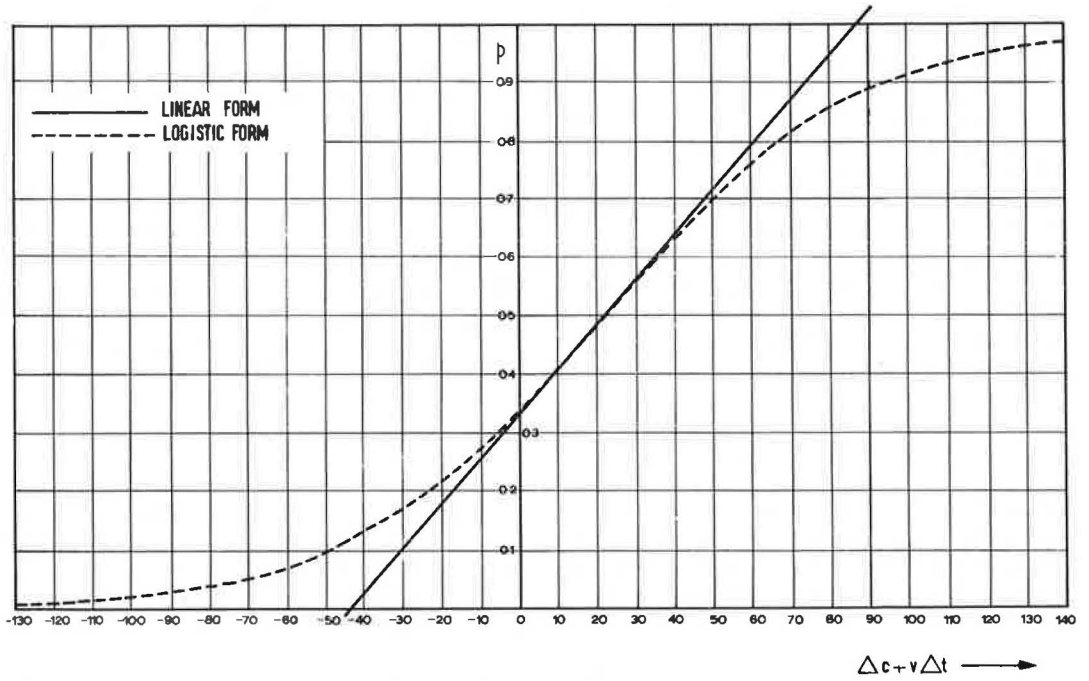
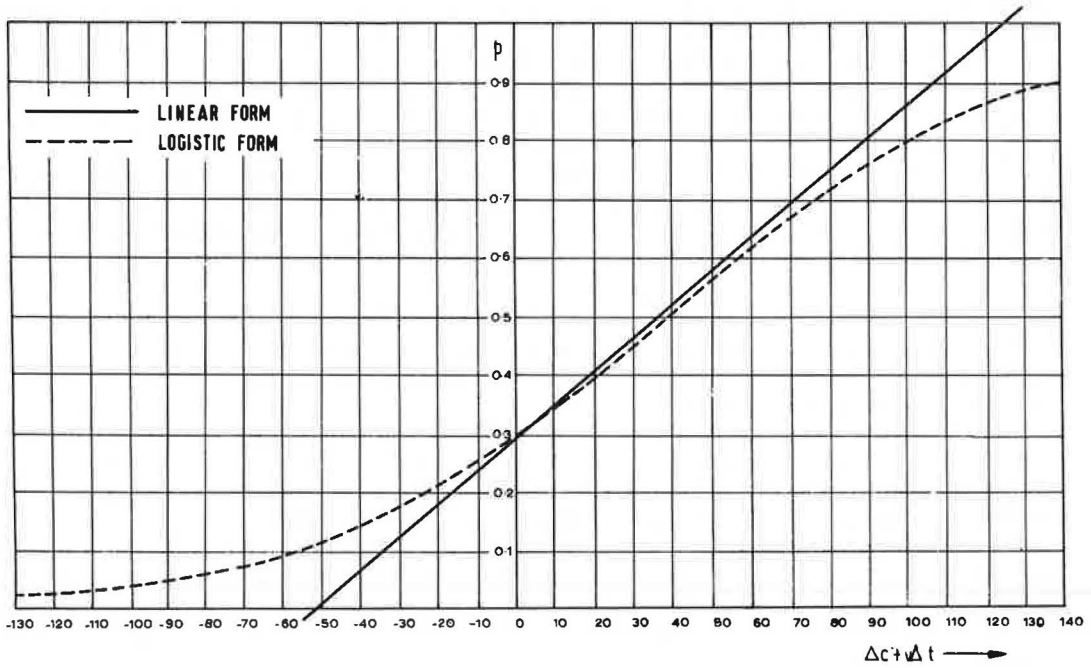Figure 1. Comparison of linear and logistic model forms for the County Hall sample.

Figure 2. Comparison of linear and logistic model forms for the University College sample.

cludes allowances for congestion on the system, and uses times derived from the loaded network to produce input to the modal split phase and the assignment phase.

The existing model is at present being used at the College of Aeronautics, Cranfield, Bedford, to assist with the economic evaluation of alternative public transport systems for the projected new town of Milton Keynes in Buckinghamshire. The College is involved in examining a number of possible urban layouts and their implications in terms of trip distribution and modal split with various alternative transport systems. The proposed town will have a population of 250,000 by the year 2000, and it is necessary to decide upon the best transport system and layout for the town before development proceeds. No results are available yet from these tests.

REFERENCES

1. Quinby, H. D. Traffic Distribution Forecasts—Highway and Transit. Traffic Engineering, Vol. 31, No. 5, 1961.
2. Forbord, R. J. Twin Cities Modal Split Model. Twin Cities Joint Program for Land Use and Transportation Planning, 1966.
3. Hill, D. M., and Von Cube, H. G. Development of a Model for Forecasting Travel Mode Choice in Urban Areas. Highway Research Record 38, p. 78-96, 1963.
4. London Traffic Survey, Volume I. Greater London Council, July 1964.
5. Stopher, P. R. Predicting Travel Mode Choice for the Work Journey. Traffic Engineering and Control, Vol. 9, No. 9, p. 436-439, 1968.
6. Beesley, M. E. The Value of Time Spent in Travelling: Some New Evidence. Economica, Vol. 32, p. 174-185, 1965.
7. Modal Split Analysis. Metropolitan Toronto and Region Transportation Study, Traffic Research Corporation, Dec. 1965.
8. Quarmby, D. A. Choice of Travel Mode for the Journey to Work: Some Findings. Jour. of Transport Economics and Policy, Dec. 1967.