THE INFLUENCE OF PUBLIC TRANSPORT ON CAR OWNERSHIP IN LONDON

By M. H. Fairhurst*

The forecasting of car ownership is a key area of transportation modelling. Transportation studies have, however, generally ignored the possibility that the level of car ownership may be affected by the transport system itself and, in particular, by the quality of public transport. It has been assumed that the quality of public transport may influence modal split, and hence car use, but not car ownership. Indeed, there has been avoidance of any contact between the user cost factors that determine modal split and car ownership levels.

In terms of conventional capital investment theory, this seems an anomalous set of assumptions. Thus, a simple criterion for the purchase of a capital good is that the discounted sum of the associated costs and benefits

\[-A + \sum_{t=1}^{n} \frac{B_t - C_t}{(1+i)^t} + \frac{Z}{(1+i)^{n+1}}\]  

should be greater than zero. For the car-ownership decision:

- \(A\) is the initial purchase cost of the vehicle
- \(Z\) is its value when sold
- \(B_t\) are the benefits derived from the vehicle in the \(t\)th period
- \(C_t\) are the user costs (money and time) of the vehicle in the \(t\)th period
- \(i\) is the interest rate.

While one would not wish to imply that the individual car buyer makes his decision on the basis of discounted cash flow calculations, equation (1) does provide a comprehensive, logically impeccable framework within which to view the decision on purchase. It suggests that that decision can be affected by variations in user costs \((C_t)\). Thus, if:

(a) at any \(t\), \(C_t\) increases (because, say, of higher petrol prices) so that \(B_t < C_t\), then the associated trips will disappear and the value of (1) will fall.

(b) at any \(t\), \(C_t\) increases to become higher than the cost by an alternative mode, the trips will be transferred to the other mode and again (1) will fall.

(c) at any \(t\), the costs of an alternative mode fall to become less than \(C_t\), trips will be transferred from the car, reducing the value of (1).

---

*Economist, Business Planning Department, London Transport Executive. This paper represents the views of the author, which may not be those of the London Transport Executive.
This framework suggests, therefore, that, while changes in mode cost affect mode use initially, in the longer term they will also affect the decision on whether to purchase a car (thus inducing a second round of changes in mode use). The aim of the research described here was to try to find empirical confirmation of these ideas; in particular, to try to establish a statistical model of car ownership in London which included public transport quality as one of its explanatory variables. The development of this model is described in the rest of this paper.

THE PROBLEM

The variables that influence car ownership are so many and so interrelated that it is essential to consider them all together in a systematic way. Diagram 1 tries to do this, though it does not try to be fully comprehensive, but only to include stronger and more important relationships. Two broad sets of variables are distinguished: household factors and planning factors.

Household factors

Diagram 1 postulates that the two household characteristics likely to have a direct influence on car ownership are household size and household income.

Household size is important as an index of some of the benefits car ownership can bring: the larger a household, the greater the money cost of using public transport relative to that of using a car, and also the more significant the unquantifiable benefits of a car as a means of family transport—public transport being most suitable for the agile and unencumbered (see, for example, [1]).

Household income is a measure of the degree to which a household possesses the wealth to be able to purchase these benefits. The relationship has been extensively researched by Bates [2, 3] and Mogridge [4]. Diagram 1 postulates that together these two variables can capture most of the relevant variation in household conditions.

Thus, it is assumed that household income is closely related to

(a) the socio-economic group of the head of the household;
(b) the age of the head of the household;
(c) the number of economically active members of the household.

Household size is assumed to be related also to (a) and (b) and in addition to be linked to some extent to residential density. There is some tendency for one- and two-person households to live in single rooms and flats in Inner London, while larger households with children predominate in the lower-density suburbs of Outer London and in the Metropolitan Area beyond the G.L.C.’s boundaries.

Planning factors

Among the planning factors in Diagram 1, car ownership is assumed to depend directly on:

(a) car prices;
(b) the attractiveness of public transport relative to private, which is itself a
DIAGRAM 1: The factors that influence car ownership

Probability that a household will be car owning

1. Car Prices
2. Congestion
3. Traffic Management
4. Public Transport's Attractiveness relative to the private car
5. Fares, petrol prices etc.
6. Residential Density
7. Household socio-economic group
8. Household Income
9. No. of economically active members
10. Age of household head
11. Household size

HOUSEHOLD FACTORS

PLANNING FACTORS

strong relationships weak relationships
function of the money costs of the alternative modes, traffic management policies and residential density. (The quality and attractiveness of public transport will tend to be a function of density, since the more compact and populous an area is the more intense will be the demand for public transport). The level of congestion is assumed to be related to car ownership itself and to residential density. This has the effect of creating a simple feedback loop between car ownership and the attractiveness of public transport, via traffic management.

The key postulate here is, of course, the relationship between car ownership and the relative user costs of making trips by public and private transport. It is now generally accepted that this relativity determines car use, and it is no great step to extend this relativity into the decision on ownership—indeed, in view of equation (1) it would seem perverse not to do so.

**Modelling implications**

The consequences for model building of these behavioural relationships are clear. The use of a technique such as multiple regression to try to calibrate a single equation directly relating to car ownership the complete set of variables 1 to 11 in Diagram 1 is likely to produce strange results; and selections, to try to maximise “explanation” with a minimum number of variables, are likely to be unsatisfactory. For example, setting up a regression equation that includes both household income and socio-economic group is likely to show that one or the other is “insignificant”; perhaps socio-economic group will emerge as the single most significant factor because of its links with both household income and household size. Yet plainly socio-economic group will be an unreliable predictor over a period of changing household incomes and household sizes.

Problems like this have to some extent brought regression analysis into disrepute—often falsely, since the problems may well have been due not to failings of the technique but to failure to evolve a plausible behavioural framework in which the technique could be deployed effectively.

**ANALYSIS**

**The approach**

Much has been published concerning the explanation and prediction of car ownership. Two basic statistical approaches have been taken:

(a) time series: analysis of changes in car ownership over time;

(b) cross-sectional: analysis of differences in car ownership at one point in time between different areas or socio-economic groups.

The present paper is based on the latter approach: it analyses data collected as part of the London Transportation Study in 1962, aggregated to the level of the Traffic District (of which there were 186). Mean household incomes and public transport access indices were, in some cases, only calculated for groups of districts; this made
necessary some aggregation of districts, so the final analysis was carried out on some 149 districts or groups of districts.

The analytical task was thus to explain the differences in car ownership between these 149 areas (hereafter referred to simply as districts) by relating them to differences in the variables which seemed likely to be influential.

Data

By no means all the data called for by Diagram 1 were available from the LTS. On the household side the data were virtually complete, but on the planning side the situation was much less satisfactory. What are really needed here are, perhaps, areal indices of the relative generalised costs of journeys by public and private transport. To formulate and calculate such indices at a district level would plainly be an intricate and time-consuming process. Nonetheless, such estimates seem vital to formulate a true predictive model of the effects on car ownership of policies related to public transport and traffic restraint.

The only step taken by the LTS in this direction was to calculate, for bus and rail, public transport access indices (referred to hereafter as p.t.i.'s). These were formulated as:

\[ \sum_{i} \frac{\sqrt{N_{ij}}}{\sqrt{A_{j}}} \]

where \( j \) refers to district \( j \)
\( A_{j} \) refers to the acreage of \( j \)
\( N_{i} \) refers to the midday frequency of the \( i \)th bus service.

These indices fall far short of what is needed. They attempt to estimate simply the availability of public transport in an area, though the measure chosen seems rather arbitrary. For the present exercise, however, they had to be taken largely as given. One anomaly was corrected, however, by substituting residential acreage for total acreage. Thus the adjusted index calculates the availability of public transport per residential acre. The original index tended to produce very low values for rural areas, ignoring the fact that the rural population is found mainly concentrated close to the main roads and bus routes.

More general problems are caused by the use of district data rather than data that refer to individual households. If a variable varies more within a district than it does between districts, its full effect will be obscured and, for example, a regression model that omits the variable will produce an over-optimistic impression of the amount of behaviour actually being explained (see, for example, Button [5]).

Any areally averaged variable will suffer from this effect to some extent, but particularly one that does not vary primarily between areas. For example, household size is likely to vary far more within any single district than the size of the average household is likely to vary between districts. Such problems do not invalidate aggregated analysis, but they must be kept in mind when interpreting the results.

Finally, some of the variables in Diagram 1 essentially vary over time—car and petrol prices and fares, for example. Cross-sectional analysis can say little or nothing about the effects of changes in these variables, but their state constitutes the backdrop of the analysis—again this needs to be borne in mind.
Diagram 2: Plot of Po against LTS bus access index for all 186 LTS districts.

- Example of an outlying data point that was based on a bus access index calculated for a group of districts.

NB as part of LTS all access values greater than 60 were arbitrarily set to 60.

NB; because they contain total district acreage, the indices understate bus access at the lower end of the scale.
Diagram 3: Plot of Po against Average household income in L.T.S. districts.
The form of relationship chosen

Diagram 2 plots the proportion of households in each district that do not own a car against the unadjusted LTS bus p.t.i.'s. It is interesting to compare this with Diagram 3, which plots the same variable against mean household income. It will be seen that, while the latter relationship seems to be fairly linear, the former is definitely not. This is because the range of variation of household income has been curtailed by aggregation, while the p.t.i.'s range from zero to 60 (the Central London maximum).

Thus, it seems clear that the underlying relationships are in some way non-linear, and also probably multiplicative as well. This latter point implies that the effect of an increase in household income on the likelihood of car ownership will depend on

(a) the household’s previous income;
(b) the state of other variables such as the size of the household.

These two features are incorporated by Bates [3] in the most intellectually satisfying form of relationship postulated in this subject. This is:

$$P_o = \frac{e^X}{1 + e^X}$$

(2)

where $P_o$ is the probability that a household will not own a car and $X$ is an influential variable. A crucial property of equation (2) is that

$$\frac{dP_o}{dX} = \frac{bP_o}{X} (1 - P_o)$$

This is behaviourally attractive: when $X = 0$, $P_o = 1$; as $X$ tends to infinity, $P_o$ to zero.

The curve indicated by equation (2) is plotted in Diagram 4 for different values of $b$. The changing slope of the line implies that when car ownership is high or low a given absolute change in $X$ produces a relatively small change in $P_o$.

Equation (2) generalises easily to the multivariate form

$$P_o = \frac{e^{X_1 b_1 X_2 b_2 \ldots X_n b_n}}{1 + e^{X_1 b_1 X_2 b_2 \ldots X_n b_n}}$$

which for estimating purposes can be transformed to

$$\log \left( \frac{P_o}{(1 - P_o)} \right) = c + b_1 \log X_1 + b_2 \log X_2 + \ldots + b_n \log X_n$$

(3)

There are, of course, theoretical objections to transforming a relationship of the type of equation (2) into a form like equation (3) which is amenable to estimation by ordinary least squares, and it may well be that a non-linear curve-fitting technique of the hill-climbing type might have produced somewhat more accurate results. However, in the present case, an equation (3) type relationship proved

---

1As mentioned earlier, some districts were grouped for the calculation of mean incomes and p.t.i.'s. Though the groups were fairly homogenous in income terms, access to public transport appears to have varied markedly within them. Diagram 2 shows some of the outlying data points that result if the p.t.i.'s are referenced to individual districts rather than to the groups of districts on which they are based.
THE INFLUENCE OF PUBLIC TRANSPORT ON CAR OWNERSHIP

M. H. Fairhurst

Diagram 4: Hypothesised relationship between $P_0$ (the probability that a household will not own a car) and an influential variable $X$ for differing values of $b$.

satisfactory—a demonstration of the flexibility of the linear regression technique, which in the car ownership modelling field sometimes seems to be confused with the linear, additive model (see, for example, Douglas [6]).

RESULTS

Income

The first relationship tested was

$$\log \left( \frac{P_e}{1 - P_o} \right) = \epsilon + b \log I$$

where $I$ = household income in £ per week.

The coefficients were estimated as $\epsilon = 6.08$

$$b = -1.81 \ (t \, \text{ratio} \, -9.93).$$

The $R^2$ (% of variation explained) was 40 per cent.
TABLE 1

Value of c and b in equation

<table>
<thead>
<tr>
<th>Source</th>
<th>Year</th>
<th>c</th>
<th>b</th>
<th>S.E.(b)</th>
<th>$-c/b$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FES</td>
<td>1964</td>
<td>9.146</td>
<td>-2.716</td>
<td>0.2808</td>
<td>3.367</td>
<td>92.56%</td>
</tr>
<tr>
<td></td>
<td>1965</td>
<td>7.023</td>
<td>-2.098</td>
<td>0.1523</td>
<td>3.347</td>
<td>96.43%</td>
</tr>
<tr>
<td></td>
<td>1966</td>
<td>6.774</td>
<td>-2.030</td>
<td>0.06044</td>
<td>3.336</td>
<td>99.38%</td>
</tr>
<tr>
<td>1967*</td>
<td>7.752</td>
<td>-2.268</td>
<td>0.2027</td>
<td>3.417</td>
<td>92.60%</td>
<td></td>
</tr>
<tr>
<td>1968*</td>
<td>6.142</td>
<td>-0.821</td>
<td>0.1550</td>
<td>3.372</td>
<td>93.24%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1969</td>
<td>5.814</td>
<td>-1.780</td>
<td>0.1571</td>
<td>3.266</td>
<td>92.78%</td>
</tr>
<tr>
<td>NTS</td>
<td>1965</td>
<td>6.906</td>
<td>-2.159</td>
<td>0.08466</td>
<td>3.198</td>
<td>98.78%</td>
</tr>
<tr>
<td></td>
<td>1966</td>
<td>6.882</td>
<td>-2.193</td>
<td>0.08784</td>
<td>3.138</td>
<td>98.43%</td>
</tr>
</tbody>
</table>

Source: [3].

Income is not adjusted for inflation.

For FES, the first 3 years were calculated on 9 income ranges, and the last 3 years on 12.

NTS 1965 was calculated on 10 income ranges, and NTS 1966 on 12.

*Car ownership definition was changed for these years.

It is interesting to compare this result with those obtained by Bates [3] by analysis of the Family Expenditure Survey and National Travel Survey (see Table 1, reproduced from [3]). It is evident that the $R^2$ on the LTS data (40 per cent) is much lower than those found by Bates (90 per cent). This would seem to demonstrate the dangers of aggregation. Bates’s equations, although based on a sample of several thousand households, are obtained by aggregating all the households into, at the most, 12 income groups, and then regressing the group average values of $P_o$ against the average values for income. This is likely to have had the effect of eliminating any variance due to household size or public transport access. The LTS data give a different impression of the importance of income because the areal aggregation has

(a) eliminated extreme income values
(b) not eliminated variance due to public transport access
(c) only partially eliminated variance due to variations in household size.

The true importance of income is undoubtedly less than suggested by Bates’s 90 per cent $R^2$s and, since (a) and (c) will tend to cancel, may not be significantly different from the 40 per cent suggested by the LTS data.

Bates demonstrated that the average income level at which more than 50 per cent of households own cars was apparently falling over time (note the decreasing value of $-c/b$ in Table 1). This was so even before any adjustment for inflation: if adjustment was made, the decline was even more marked. To explain this, Bates developed an ingenious argument relating to changes in the used car market and changes in taste.

That is, because of the aggregation, household size and public transport access are likely to be very similar in each group.
THE INFLUENCE OF PUBLIC TRANSPORT ON CAR OWNERSHIP

M. H. Fairhurst

On the basis that the income relationship apparently explains over 90 per cent of car ownership variation, such refinement is obviously attractive. However, given that the true explanatory power of income is less than half this, it seems possible that the observed downward drift of the function may be due to changes in the other variables, not included in equation (4), whose importance has been masked by the form of the aggregation.

**Household size**
The relationship tested was

$$\log \left( \frac{P_o}{1 - P_o} \right) = \epsilon + b \log I + d \log H$$

where $H =$ persons per household.
The co-efficients were estimated as $\epsilon = 10.5$

$$b = -2.2 \quad (t = -17.42)$$

$$d = -3.0 \quad (t = -12.79)$$
The $R^2$ (% of variation explained) was 72 per cent.
The correlation between $I$ and $H$ was 0.23: not high enough to cause concern.
The increase in the $R^2$ is perhaps surprising in view of the way the data have been aggregated—taking the mean household size in a district effectively eliminates extreme values. The quite large changes in the co-efficients $\epsilon$ and $b$ are worthy of note.

**Public transport**
The relationship test was

$$\log \left( \frac{P_o}{1 - P_o} \right) = \epsilon + b \log I + d \log H + f \log (B + 1) + g \log (R + 1)$$

where $B$ and $R$ are the bus and rail p.t.i.’s (modified as described above).
The co-efficients were estimated as $\epsilon = 5.02$

$$b = -1.58 \quad (t = -11.69)$$

$$d = -0.705 \quad (t = -1.9)$$

$$f = 0.227 \quad (t = 4.13)$$

$$g = 0.125 \quad (t = 3.42)$$
The $R^2$ (% of variation explained) was 80 per cent.
The correlation matrix between the variables was as follows:

<table>
<thead>
<tr>
<th></th>
<th>log $I$</th>
<th>log $(B + 1)$</th>
<th>log $(R + 1)$</th>
<th>log $H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>log $I$</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log $(B + 1)$</td>
<td>-0.26</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log $(R + 1)$</td>
<td>-0.15</td>
<td>0.82</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>log $H$</td>
<td>-0.23</td>
<td>-0.64</td>
<td>-0.74</td>
<td>1.0</td>
</tr>
</tbody>
</table>

203
It will be observed that, in addition to a high level of correlation between the bus and rail p.t.i.'s, there is a high correlation between them and household size. (Diagram 1 postulates that both may be associated with density.)

At first sight, the relatively small increase in the $R^2$ from 72 to 80 per cent between equations (5) and (6) suggests that the p.t.i.'s are of little importance. Closer examination of equation (6) tells a different story, however. The $t$-ratios of the variables are interesting: the $t$-ratio of the household size variable has fallen from 12.8 in equation (5) to only 1.9 in equation (6); that is, it is only marginally significant statistically in equation (6). The p.t.i.'s, on the other hand, have $t$-ratios of 4.0 and 3.5—considerably higher values. Again, if the proportion of the mean $P_o$ "explained" by the various variables is considered, it can be shown that the p.t.i.'s are twice as important as household size.

The increase in $R^2$ produced by equation (5) is thus possibly spurious, the household size variable picking up the effects of the absent public transport variables with which it is so correlated. Omitting household size from equation (6), in fact, only reduced the $R^2$ by 4 per cent.

This is not to say that household size is only a marginal factor in explaining car ownership, but that in this particular analysis, given the way the data have been aggregated, it would be surprising if household size were to prove statistically very significant.

Thus equation (6) seems a behaviourally valid result: income, household size and public transport availability together determining car ownership, and the $t$-ratios of the co-efficients according fairly well with a priori reasoning.

Diagram 5 plots the actual and fitted values of $P_o$, using the co-efficients from (6) in an equation like (2). Regressing the actual against the fitted values gives an $R^2$ of 76.68 per cent and a line standard error of 5.8 per cent.

This virtually completes the analysis that can be made with the LTS data of the factors which Diagram 1 suggests may be most directly related to car ownership. One other variable remains which is often supposed to have an effect on car ownership and which is available from LTS: this is net residential density.

**Density**

The problems associated with density are apparent if Diagram 1 is considered. This indicates that density has an influence on or association with almost every other variable that can be considered relevant in car ownership forecasting (for example, regressing the modified bus and rail p.t.i.'s on density yields $R^2$ of 70 to 80 per cent).

It can, of course, be argued that density is not merely a proxy for the variables included in equation (6) but is also a measure of physical limits on car ownership, brought about by increasing congestion. Density, however, seems unsuitable for this role because of its strong behavioural links with so many of the other variables. Further, the omission of, for example, a restraint variable from equation (6) is perhaps not of great importance, since

(a) the LTS was carried out in 1962—when parking meters were first being introduced in Central London;

(b) the area covered by LTS was very wide and included many rural districts;
Diagram 5: Actual and Fitted Values of Po using equation (6)
(c) a recent study [7] of the growth of car ownership in London between 1966 and 1971 indicates that the rates of growth in G.L.C. boroughs have been remarkably similar.

From a policy point of view, inclusion of a physical restraint variable in equation (6) would plainly be most useful. However, it seems that more recent data are needed to calibrate the relationship with any degree of realism.

Nonetheless, it is interesting to test the relationship

$$\log \left( \frac{P_o}{1 - P_o} \right) = c + b \log I + h \log D$$  \hspace{1cm} (7)

where $D$ is net residential density.

This gives an $R^2$ of 78 per cent and co-efficients $c = 1.32$

$$b = -1.02 \hspace{0.5cm} (t = -8.48)$$

$$h = 0.59 \hspace{0.5cm} (t = 15.68)$$

However, when net density was introduced into equation (6) and the following equation tested:

$$\log \left( \frac{P_o}{1 - P_o} \right) = c + b \log I + h \log D +$$

$$d \log H + f \log (B + 1) + g \log (R + 1)$$  \hspace{1cm} (8)

the co-efficients obtained were $c = 3.3$

$$b = -1.3 \hspace{0.5cm} (t = -8.05)$$

$$d = -0.358 \hspace{0.5cm} (t = -0.92)$$

$$f = 0.159 \hspace{0.5cm} (t = 2.62)$$

$$g = 0.087 \hspace{0.5cm} (t = 2.24)$$

$$h = 0.22 \hspace{0.5cm} (t = 2.50)$$

The $R^2$ is 81 per cent. Generally, this seems a less satisfactory model than equation (6), for a number of reasons:

(a) the increase in $R^2$ is very small, and, if an attempt is made to explain the residuals from equation (6) by density, the resulting $R^2$ is virtually zero. Thus density does not appear to contribute new information to equation (6).

(b) the co-efficient of household size in equation (8) is only half that in equation (6), and household size appears to be insignificant. Behaviourally this seems quite unreasonable.

In sum, it seems likely that the performance of density in equations (7) and (8) is due to its intimate connections with the other variables that directly influence car ownership. Equation (6) seems preferable, therefore. However, it should be noted that even in equation (8) the p.t.i.'s remain significant statistically—the effect of density on them being comparable to its effect on income.

Of course, it can be claimed that household size and the bus and rail p.t.i.'s contribute little extra to equation (7). But this is to overlook the point that the causal link between car ownership and the equation (6) variables is closer than that between car ownership and density. Thus, equation (6) is a more plausibly predictive, behaviourally valid model.
That an equation like (7) will be an unsatisfactory predictive tool in a period of declining household size and public transport attractiveness has been demonstrated by the work of Quarmby and Bates [8] at the Ministry of Transport. They calibrated an equation identical to (7) on National Travel Survey data. However, attempts to validate the model by back forecasting to 1959 showed that the equation seriously underestimated the increase in car ownership over the period. This proved so intractable that the model was abandoned. The tendency for an income-density model to underestimate increases in car ownership over time may be due to the fact that density in a cross-sectional analysis serves as a proxy for the attractiveness of public transport, which in recent years has been diminishing more quickly than density. This is easily demonstrated by the facts that in London, between 1953 and 1970, real bus and rail fares rose by over 70 per cent and bus miles fell by 40 per cent. Over this period, when the real generalised cost of using public transport must have increased by around 50 per cent, real car prices fell by about one-third.

CONCLUSIONS

It has been shown that household income, household size and public transport access provide as satisfactory a statistical explanation of variations in car ownership between LTS districts as does a model that also includes residential density. Even in the latter model, public transport access is a statistically significant factor. Behaviourally, the simpler model (excluding density) provides a more satisfactory explanation and seems capable of explaining, at least in part:

(a) the tendency for income/density models to underpredict the increase in car ownership during the 1960s and

(b) the instability of the income models of Bates [3].

This result has important consequences for transport planning. More interaction needs to be allowed between the various stages of the process. The findings here imply that policies favourable to one mode will have not only an initial effect on modal split but also significant second order effects as marginal households consider whether to own a car. The consequences of their assessments will be to produce second order changes in the usage and financial viability of public transport.

REFERENCES


207
