A METHOD FOR BREAKING DOWN AND MEASURING THE EFFECTS OF CORRELATIVE EXPLANATORY VARIABLES. AN APPLICATION TO THE EFFECTS OF URBAN SPRAWL, CAR OWNERSHIP AND TRANSPORT SUPPLY ON CHANGE IN THE MARKET SHARE OF PUBLIC TRANSPORT

Patrick BONNEL and Isabelle CABANNE
Laboratoire d'Economie des Transports
ENTPE, Université Lumière Lyon 2, CNRS

Most analysis of travel behaviour attempts to measure the individual impacts of a variety of explanatory factors on the investigated phenomenon. The same is true of modelling, which aims to identify the contribution of a number of factors in order to simulate how a change in these would affect the studied situation. However, both these types of studies encounter methodological problems. The different explanatory variables are frequently correlated, as are the ways they change. This raises difficulties for assessing and interpreting the respective impacts of the different factors.

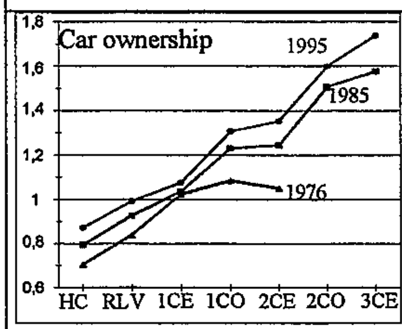
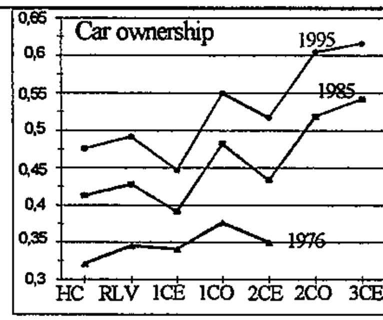
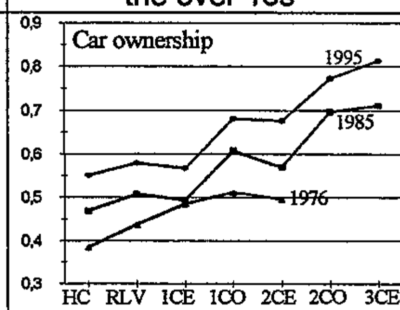
To overcome this difficulty this paper proposes a methodology for separating the impacts of the different variables. This methodology involves decorrelation of the fundamental variables and their change and a systematic study of the combined effects of the various factors. We shall then apply this methodology to a situation whose determinants are highly correlated in order to demonstrate the value of the technique.

We shall apply the technique to car/public transport modal split. The explanatory variables relate to locations, car ownership and transport supply. These are important input variables in most urban models. The Lyon conurbation in France has been selected, for which we have the results of three household travel surveys conducted between 1976 and 1995.

1. The problem

Most analysis or modelling of travel behaviour attempts to measure the individual impacts of a variety of explanatory factors on the investigated phenomenon. The different explanatory variables are frequently correlated, as are the ways they change. This raises difficulties for assessing and interpreting the individual impacts of the different factors.

To illustrate this let us take the case of car ownership and location. These two variables both have a considerable influence on urban travel and modal choice, as witnessed by their systematic inclusion in traffic forecasting models. They are, however, obviously correlated, as can be seen from Figure 1a. Furthermore, this correlation is tending to become stronger over time. Household car ownership is thus twice as great in the most distant part of the suburbs as in the centre. However, when one attempts to analyze the influence of this correlation on modal choice, our household car ownership definition is probably not the most appropriate. It is well known that household sizes have fallen greatly over the last twenty years, therefore at a constant rate of car ownership, car access within a household has become easier. We can therefore illustrate the correlation between car ownership and location in another way, as shown in Figure 1b, where car ownership is calculated on the basis of the number of cars per person in the household. The objection can be raised that a person needs to be 18 years old to drive, so defining car ownership on the basis of the number of cars per person of over 18 years of age in the household gives a better picture of access to a car. This gives us a third representation of the correlation between car ownership and location (Graph 1c).

household car ownership	individual car ownership	individual car ownership for the over 18s
		

Graph 1 : Change in car ownership according to location, for different definitions of car ownership. (Cf. Zonal breakdown map 1, section 4)

As our purpose is not to analyze car ownership we shall not pursue our analysis of these three graphs. However, they do allow us to see clearly the multiplicity of correlations between variables. Here, the high degree of correlation between location and household composition interferes with analysis of the correlation between location and car ownership as it alters the relationship between the two. It is therefore important to take such correlations into account and select definitions which are appropriate for the indicators used.

However, even with car ownership defined in terms of the number of persons over 18 years of age in the household, which seems a more appropriate way of describing an individual's access to a car, the correlation between location and car ownership remains. Forecasting the combined influence of change in car ownership and location therefore always faces difficulties as it cannot be formulated simply in terms of the effect of the change of each of the factors taken separately.

Let us return to our example to illustrate the difficulty. As car ownership is higher further from the centre, the urban sprawl of recent decades will tend to increase car ownership. When the combined influence of car ownership and urban sprawl is simulated just summing the effect of each of the variables is very likely to give rise to double accounting with regard to changes in car ownership.

To overcome this difficulty, this paper will propose a methodology for separating the impacts of the different variables. This methodology involves decorrelation of the fundamental variables and their change and a systematic study of the combined effects of the different factors. Section 2 will then propose a very simple multiplicative model of the individual contribution of each factor, once the combined effects have been cancelled out.

This method is then applied to modal split between the car and public transport in the Lyon conurbation. The explanatory variables we used were location, car ownership and transport supply, distinguishing between the passenger car and public transport. These variables are among the principal factors which affect the market share of public transport (Andan et al., 1988; Massot, Orfeuil, 1989, 1990, 1991; Stopher, Lee-Gosselin, 1997; Gärling et al., 1998; Ortuzar et al., 1998 ...). These four factors are unable in themselves to explain all the changes in modal split, a residue remains which corresponds to the set of "other factors".

In section 3 we shall apply this method to the 5 variables, which will require us to propose formulations for the factors and the functions linking them to modal split which are able to cancel out these combined effects. We have then, in section 4, applied the method to the data from the three most recent household surveys of the Lyon conurbation (1976, 1985, 1995) in order to measure the influence of each factor individually.

2. Methodology used to break down the effects

We shall begin by describing the mathematical principle used in our breakdown method, then interpret this and present the necessary hypotheses. For the sake of simplicity, we shall describe the method taking the case of just two factors. Extending it to a larger number of factors does not involve any difficulties apart from the need for longer equations.

2.1. Mathematical principle

Let us begin by explaining the notation:

Y is the factor to be explained;

a and b are two explanatory variables;

f is a function applying to the explanatory variables.

$$Y = \sum_{i} f(a_i, b_i)$$

This formula expresses the fact that Y is calculated by breaking the population down into statistical classes of individuals and then summing these. This breakdown is generally performed with reference to one (or more) of the explanatory variables. In our example, the breakdown relates to different locations, on the basis of the zoning we have applied to the study area.

To simplify notation we shall express the sum in the following way:

$$Y = a * b$$

The change in Y between two given dates is therefore expressed as follows:

$$\frac{Y_2}{Y_1}=\frac{a_2b_2}{a_1b_1}=\frac{a_2b_1}{a_1b_1}\times\frac{a_1b_2}{a_1b_1}\times\frac{\dfrac{(a_2b_2)}{(a_1b_2)}}{\dfrac{(a_2b_1)}{(a_1b_1)}}=E(a)\times E(b)\times E(a,b) \tag{1}$$

- where E(a) denotes the effect of factor a on its own. It is a rate of absolute variation obtained by varying this factor between 1 and 2 while keeping all other elements constant;
- E(b) denotes the effect of factor b on its own. It is a rate of absolute variation obtained by varying this factor between states 1 and 2;
- E(a, b) denotes the double effect of factors a and b. It is the ratio between two rates of absolute variation (the first rate corresponds to the effect of factor a when factor b is in state 2 and the second corresponds to the effect of factor a with factor b in state 1).

We are thus able to break down the overall rate of variation into 3 effects, two single effects caused by the variation of each factor in isolation and one double effect depending on the variation of the two factors. The objective is obviously to obtain a double effect that is equal to or very near 1 such that the change in Y can be regarded as being the product of each of the effects in isolation. We shall present an interpretation of the double effect which reveals the necessary conditions for obtaining this result.

2.2. Interpretation of the double effect

If the double effect is negligible (i.e. as this is a multiplicative model if it is equal to 1) we have (see equation (1)):

$$\frac{(a_1b_1)\times(a_2b_2)}{(a_1b_2)\times(a_2b_1)}=1$$

This can also be expressed as follows:

$\dfrac{(a_1b_2)}{(a_1b_1)}=\dfrac{(a_2b_2)}{(a_2b_1)}$ This result can be interpreted in the following way: the influence of change in factor b is identical for states 1 and 2 of factor a. It therefore does not depend on factor a.

$\dfrac{(a_2b_1)}{(a_1b_1)}=\dfrac{(a_2b_2)}{(a_1b_2)}$ Again this result can be interpreted in the following way: the influence of change in factor a is identical for states 1 and 2 of factor b. It therefore does not depend on factor b.

If the two terms are independent, that is to say if change in factor a is independent of change in factor b, the double effect will systematically be equal to 1. It will then be possible to interpret the different terms in our breakdown and state that change in Y is equal to the product of change in each of the factors taken separately, i.e. if all other factors are equal. On the other hand, if the two terms are not independent the influence of factor a is very likely to differ according to the state of factor b, as the influence of factor b is very likely to be dependent on the state of factor a. It is therefore very likely that the product of the two terms will not be equal to 1.

We therefore need to find a formulation of the factors a and b and the function f(a, b) where change in the two factors is independent, or at least where the dependency is reduced to the greatest possible extent. However, this presentation should be interpreted with caution. Just because a double effect is equal to 1 with a certain data set does not automatically imply that change in the two factors is independent. The result would need to be confirmed using other surveys. All such a presentation means is that the data do not disprove the hypothesis.

2.3. Extension to 5 factors

From among the factors which influence modal split we have selected four (locations, car ownership, transport supply distinguishing between road transport and public transport supply). These are not the only factors that can influence modal split: we have therefore added a fifth factor, representing all the other factors whether known or not, which we shall refer to as "other factors".

Extending the method explained above to five factors (denoted by a, b, c, d and e), leads to a breakdown consisting of:

- 5 single effects where only the studied effect varies with the others remaining in state 1;
- 10 double effects where only the two factors in question vary with the others remaining in state 1.

These double effects are notated in the following way:

$$E(a,b) = \frac{\dfrac{(a_2 b_2 c_1 d_1 e_1)}{(a_1 b_2 c_1 d_1 e_1)}}{\dfrac{(a_2 b_1 c_1 d_1 e_1)}{(a_1 b_1 c_1 d_1 e_1)}}$$

these are interpreted as the ratio between two rates of absolute variation, the first being variation in factor a with factor b remaining in state 2 and the second being variation in factor a with factor b remaining in state 1. The other double effects are written in the same way, permuting the role of the factors;

- 10 triple effects where only the three factors involved vary, the others remaining in state 1: These triple effects are expressed as follows:

$$E(a,b,c) = \frac{\dfrac{(a_2b_2c_2d_1e_1)}{(a_1b_2c_2d_1e_1)}}{\dfrac{(a_2b_1c2d_1e_1)}{(a_1b_1c_2d_1e_1)}} \div \frac{\dfrac{(a_2b_2c_1d_1e_1)}{(a_1b_2c_1d_1e_1)}}{\dfrac{(a_2b_1c_1d_1e_1)}{(a_1b_1c_1d_1e_1)}}$$ they are interpreted as the ratio between four rates of absolute variation: variation in factor a with the four possible states of factors b and c (b_1c_1, b_1c_2, b_2c_1, b_2c_2). The other triple effects are written in the same way, permuting the role of the factors;

- 5 quadruple effects where only the four factors in question vary, the last factor remaining in state 1. These quadruple effects are expressed using the same principle as the triple effects. They correspond to the ratio between eight rates of absolute variation: the variation of factor a with the eight possible states of factors b, c and d. The other quadruple effects are written in the same way, permuting the role of the factors;

- 1 quintuple effect where all five factors change. This is expressed using the same principle as the triple and quadruple effects. It corresponds to the ratio between sixteen rates of absolute variation: variation in factor a with the sixteen possible states of factors b, c, d and e.

The manner in which we have presented the double, triple, quadruple and quintuple effects concentrates on factor a. It would be possible for the notations to emphasize each of the factors in exactly the same way. Consequently, the interpretation of the double effect described in section 2.2 can be extended without any difficulty to each of the combined effects however many factors are brought into play.

As we have now established the principle of breakdown, we shall now present the modelling of the factors and the relationships that link them with modal split.

3. Application to the study of the explanatory factors for modal split

We shall now use the above method to analyze modal split in the Lyon conurbation.

The explanatory variables we shall use for modal split are as follows: the location of trip flows, car ownership and transport supply, distinguishing between the passenger car and public transport.

Initially, we shall define the variables used while attempting to minimize the correlations between them. This will lead us to define car ownership in an original way, which we shall refer to as origin-destination matrix of car ownership. We shall then formalize our models of the different effects.

3.1. Modelling the explanatory factors

We shall divide the study zone into n zones.

3.1.1. Modelling location

For the location term we shall use the spatial distribution of trips. The location term therefore corresponds to the origin-destination matrix for trips. The manner in which the location term changes does not therefore correspond exactly to urban sprawl in the normal sense of a spreading of the location of activities within the urban area. It corresponds more to the consequences of this spreading on the geographical distribution of flows and the destinations selected by individuals. This definition has the virtue of being operational and allowing us to perform very straightforward mathematical modelling while expressing changes in the locations of trip origins and destinations rather than changes in the locations of activities.

We shall formalize this mathematically using the matrix of the weights of each origin-destination pair, i.e.:

$$l_{ij} = \frac{N_{ij}}{N}$$

where l_{ij} is the proportion of all car or public transport trips made by respondents that is on the origin-destination pair i/j;

N_{ij} is the number of public transport trips for the origin-destination pair i/j.

N is the total number of car or public transport trips made by respondents.

We can then express the market share of public transport trips according to location with the following formula:

$$P_{TC} = \sum_{i,j} l_{ij} * P_{TCij}$$

where P_{PT} is the market share of public transport over the whole conurbation;

l_{ij} is the proportion of all car or public transport trips that is on the origin-destination pair i/j;

P_{PTij} is the market share of public transport for the origin and destination pair i/j.

3.1.2. Modelling car ownership

In order to overcome the problem of the correlation between change in location and car ownership, we shall propose a model for car ownership that is independent of location. We do this by transferring the car ownership of the household or the individual to the trips made by the household or the individuals. It then becomes possible to calculate a level of car ownership for each origin-destination by averaging the car ownership for each trip on the origin-destination pair.

$$m_{ij} = \frac{\sum_{k=1}^{N_{ij}} t_k}{N_{ij}}$$

where mot_{ij} is the car ownership for trips on the origin-destination pair i/j;

N_{ij} is the number of trips on the origin-destination pair i/j;

t_k is the car ownership of individuals over 18 years of age, i.e. the number of cars in the household divided by the number of respondents over 18 years of age in the household.

This model allows us to vary location and car ownership independently. We have also, in a manner of speaking, rendered the two factors orthogonal. We shall refer to this unconventional definition of car ownership as the origin-destination matrix of car ownership. Changes in this matrix between two surveys express, in a certain manner, changes in car ownership with a constant spatial structure of origin-destination pairs as each matrix cell represents the mean level of car ownership for a given origin-destination pair. This definition has the advantage of being operational and allowing us to perform modelling that is mathematically very simple.

3.1.3. Modelling transport supply

To model transport supply we have followed normal practice and used the generalized cost. This has been calculated for each origin-destination pair in the case of public transport (denoted by pt_{ij}) and the passenger car (denoted by pc_{ij}).

For the passenger car, the generalized cost consists of the journey time plus fuel costs. The generalized cost for public transport has been determined on the basis of the generalized journey times (journey time by the mode of transport considered + an access and regress time weighted by a factor of 2 + waighting time weighted by a distress factor of 1.8 + access penalty of 5 minutes for the underground and 10 minutes for surface transport) and the mean revenue for a public transport journey. In order to guarantee comparability with the generalized cost for the passenger car, the cost of public transport has then been divided by a factor of 1.8 (Lichère, Raux, 1997). The same value-of-time has been used for both modes.

The parking constraint is not included in transport supply as we lack the necessary data. It has therefore been included among the unexplained "other factors".

Mathematically, we therefore have the following expression:

$$P_{TC} = \sum_{ij} l_{ij} * P_{TCij} = \sum_{ij} l_{ij} * g(m_{ij}, pc_{ij}, pt_{ij}, a_{ij})$$

where l_{ij} is the proportion of all car and public transport trips that is on the origin-destination pair in question;

P_{PTij} is the market share of public transport for the origin and destination pair i/j;

mot_{ij} is car ownership for trips between i and j;
pc_{ij} is the generalized cost by car of trips on the origin-destination pair i/j;
pt_{ij} is the generalized cost by public transport of trips on the origin-destination pair i/j;
a_{ij} is the "other factors" term for the origin-destination pair i/ j.

3.1.4. Modelling the "other factors"

It now remains for us to model the other factors, the nature of which we are unaware. Analyses of travel (Andan et al., 1988) have revealed that modal choice is influenced by a large number of factors. We shall therefore not attempt to formalize these other factors on the basis of the individual variables which comprise them. However, the available data allow us to calculate the market share of public transport, the level of car use and the generalized costs on each origin-destination pair.

Once we have selected and calibrated a function g which minimizes the double, triple, quadruple and quintuple effects between the different factors we are able to deduce a_{ij} by solving the equation:

$$P_{TCij} = g(m_{ij}, pc_{ij}, pt_{ij}, a_{ij})$$

where a_{ij} represents the unknown factors.

We shall return to this formalization later.

3.2. Modelling the relationship between the market share of public transport and the studied factors.

To establish value for the function g we shall use a logit model to express the market share of public transport on an origin-destination pair according to car ownership, generalized costs and other factors. This has the advantage of providing a theoretical basis to modelling via the theory of utility (Manheim, 1984) while also being the formulation which is most commonly used to model modal split.

$$P_{TCij} = \frac{1}{1 + \exp(u)}$$

where u represents the difference in utility between the two modes of transport as measured using the parameters mot_{ij}, pc_{ij}, pt_{ij}, a_{ij}.

The next section will cover the selection of an appropriate utility function.

As we have chosen to use a logit model, the five factors are no longer connected by a multiplicative formulation. However, it can easily be demonstrated that the breakdown principle described in section two applies in an identical manner. The term E(mot) (change in the origin-destination matrix of car ownership) accurately expresses for each origin-destination pair the effect of change in the market share of public transport when car ownership is in states 1 and 2 and all the other factors are in state 1. The same applies by transposition to each factor and for all the combined effects.

The function u is calibrated by a logarithmic conversion (Ortuzar, Willumsen, 1994):

$$u = \ln\left(\frac{1}{P_{TCij}} - 1\right).$$

We now need to test different (mot_{ij}, pc_{ij}, pt_{ij}, a_{ij}) formulations. However, as a_{ij} is unknown we are not able to calibrate u with respect to it.

In order to calibrate u we shall express it in the following form: $u = f(mot_{ij}, pc_{ij}, pt_{ij}) + \varphi(a_{ij})$
and assume that $\varphi(a_{ij}) = 0$ for each origin-destination pair i-j. This hypothesis is obviously debatable to the extent that there are other factors that influence modal split. We nevertheless feel that we have identified the principal factors. Furthermore, this hypothesis is essential for us to be able to calibrate the function u. It is of the same type as the hypotheses which are usually made concerning the distribution of residues which always assume that the mean is null.

We then calibrate the coefficients for mot_{ij}, pc_{ij} and pt_{ij} using the expression:

$$f(m_{ij}, pc_{ij}, pt_{ij}) = \ln\left(\frac{1}{P_{TCij}} - 1\right).$$

We then calculate the other factors (or more precisely $\varphi(a_{ij})$) with:

$$\varphi(a_{ij}) = \ln\left(\frac{1}{P_{real\,TCij}} - 1\right) - f_{calculated}(mot_{ij}, pc_{ij}, pt_{ij}).$$

If we have knowledge of part of these "other factors" and are able to include more factors in f than mot_{ij}, pc_{ij} and pt_{ij}, the calibration coefficients will change. However, the quality of the results is highly dependent on the relevance of the variables used. If we omit some important explanatory variables, the model almost completely ceases to be accurate and the values of the parameters of u may become uncertain. In this way, in the next section we have included parking as an additional variable in f in order to increase the reliability of the parameters of mot_{ij}, pc_{ij} and pt_{ij}. The variables we have selected thus correspond to those used in the majority of modal split models.

Determination of the utility function u

We shall begin by testing a conventional additive formulation (formulation A1). We can nevertheless consider the nature of the mechanism involved in the choice between the car and public transport. It is likely that comparisons between these modes are more concerned with performance in relative as opposed to absolute terms. We shall therefore test a formulation in which the generalized costs of public transport and the car are expressed as a ratio (formulation P1).

In these formulations, the origin-destination matrix of car ownership is present as an additive factor. Some research (Lichère, Raux, 1997) has demonstrated that car ownership can also be interpreted as a factor that influences perception of the two motorized modes. The higher the level of car ownership the less the generalized cost of car travel is perceived. Conversely, the higher the level of car ownership, the more negative is the perception of the generalized cost of public transport. This interpretation of car ownership is expressed in the formulations A1 and P1 which lead to the formulations A2 and P2.

We have obtained the formulations below (others were also tested, but they gave less satisfactory results) (Bonnel, Cabanne, 2000):

$$\text{A1: } u = a\ mot + b\ pc + c\ pt + d + b_1 B_1 + b_2 B_2 + \varphi(a_{ij})$$

$$\text{P1: } u = a\,mot + b\frac{pt}{pc} + b_1 B_1 + b_2 B_2 + c + \varphi(a_{ij})$$

$$\text{A2: } u = a\frac{pc}{mot^m} + b\,pt\,mot^m + c + b_1 B_1 + b_2 B_2 + \varphi(a_{ij})$$

$$\text{P2: } u = a\frac{pt\,mot^m}{pc} + b_1 B_1 + b_2 B_2 + c + \varphi(a_{ij})$$

where:

mot is the value of the origin-destination matrix of car ownership on the pair i/j;

pc is the generalized cost by passenger car on the pair i/j;

pt is the generalized cost by public transport on the pair i/j;

B_1 and B_2 are two variables that represent parking constraints:

B_1 such that:

$B_1 = 1$ if the origin or the destination is in the hypercentre;

$B_1 = 0$ if not;

B_2 such that:

$B_2 = 1$ if the origin or the destination is in the rest of Lyon and Villeurbanne;

$B_2 = 0$ if not.

Ideally, the parking variable ought to be included among the explanatory variables for modal split. Unfortunately we have no measurement of parking supply at the different survey dates and are therefore unable to evaluate its impact. However, introducing a variable that represents parking when calibrating u is of value as this increases the reliability of the parameters of u for other variables. Introducing the binary variables B_1 and B_2 thus allows us to approximate the difficulty of parking in the most central zones (in the other zones we can consider that the parking constraint in public zones is very low).

4. Results for the Lyon conurbation

We have applied the methodology we have developed to the Lyon conurbation, which is the second largest in France with a population of 1.2 million inhabitants. We have used data from the three most recent household surveys in the conurbation.

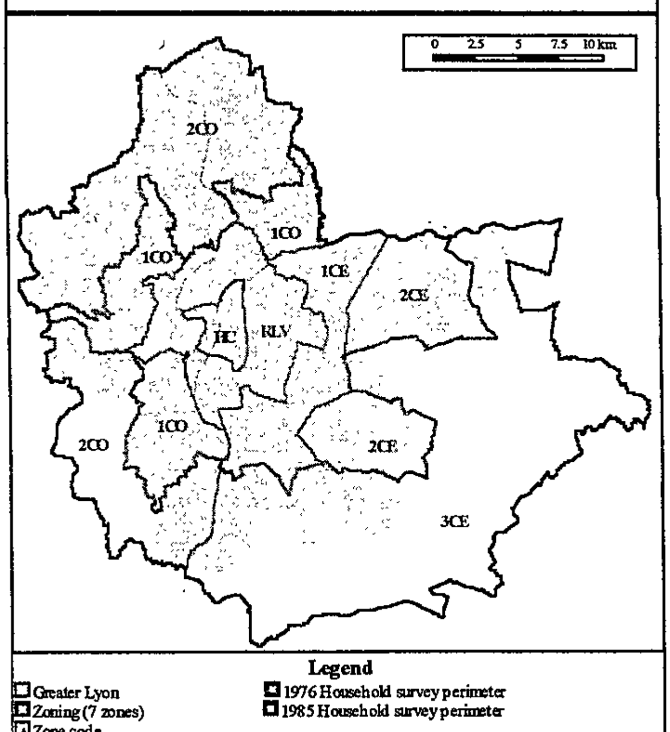
4.1. The data

The first survey was conducted between September 1976 and February 1977. For convenience we shall refer to this as HS 76. The second took place between November 1985 and March 1986 (HS 85). The last was conducted between November 1994 and April 1995 (HS 95). These three surveys therefore provide an overview of travel in the Lyon region over a twenty year period.

The information they contain make the surveys particularly suited to our purpose. They share a common methodology, under guidance from the CERTU (Centre d'Etudes sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques) (CETE de Lyon, 1977, 1986, 1995; CERTU, 1998). The surveys were conducted at the homes of the respondents, and all individuals of over 5 years of age were interviewed individually. Information on all the trips conducted the day before the survey day was collected, in particular the mode or modes used. These surveys used a random sample of households which was selected after geographic stratification within the survey perimeter. The surveys in Lyon involved 3,700 households in 1976, 5,000 in 1985 and 6,000 in 1995.

Only persons living within the survey perimeter were eligible for the survey. Trips made by visitors and exchange and through traffic were excluded from the survey, as were deliveries or freight trips. Our working data were therefore representative of weekday trips made by persons living within the survey perimeter.

Map 1: Study perimeter and zoning system for 7 zones



Although the methodologies used in the three surveys are comparable, their perimeters are not. The perimeter of the 1976 survey is slightly smaller than the COURLY (Communauté Urbaine de Lyon, Greater Lyon), that of the 1985 survey corresponds to that covered by the Lyon conurbation master plan (a zone a little larger than the COURLY), and the 1995 survey returns to the 1985 perimeter with the addition of a few communes on the banks of the Ain river. We have constructed two study perimeters: that of the 1976 survey (hereafter referred to as the 1976 perimeter, and that of the 1985 survey hereafter referred to as the 1985 perimeter, which also applies to the 1995 survey (see map 1).

We have divided up the study zone in two different ways: into 7 zones, only 5 of which are inside the 1976 perimeter, and into 25 smaller zones. The division into 25 zones permits interzonal journey times to be specified more accurately and increases the number of observation points. However, the flows on many origin-destination pairs are very small. These two divisions have been made with reference to concentric rings supplemented by an East-West segmentation in the case of the 7 zone division and a catchment area analysis in the case of the 25 zone division (see map 1).

From the survey data we have been able to calculate the matrix of locations, the origin-destination matrix of car ownership and the matrix of modal split. The generalized cost matrices for the three survey dates have been calculated from the journey times generated by the strategic model for the Lyon conurbation developed by the SEMALY and the LET (Lichère, Raux, 1997). This model was mainly calibrated using the data from the three household surveys.

4.2. Analysis of the correlation between the factors

Elimination of the combined effects requires the correlations between the factors in question to be either nil or very small. We have attempted to reduce these correlations by the manner we have constructed our formulations. The correlations can also be measured using our data. This has been done using the 7 zone division. Within the 1976 study perimeter there are 5 zones, that is to say 25 origin-destination pairs for each of the three household surveys.

We have calculated the coefficients of correlation between the different variables. For this purpose we have combined the results from the three surveys, thereby obtaining 75 origin-destination pairs. The coefficient of correlation between car ownership and location has been defined as the coefficient of correlation between the 75 mot_{ij} and the 75 l_{ij}. The 9 other correlation coefficients were defined in a similar manner.

We have also calculated the coefficients of correlation between changes in the different variables. In the case of each origin-destination pair we have calculated the changes in each variable between 1976 and 1985 and then between 1985 and 1995. We thus obtained 50 values of change (25*2) for each variable. The coefficient of correlation between changes in location and changes in car ownership is defined as the coefficient of correlation between the 50 vectors of changes in the values of mot_{ij} and the 50 vectors of changes in the values of l_{ij}. The 9 other correlation coefficients were defined in a similar manner.

Table 1 shows the different correlation coefficients. The "other factors" are those calculated using formula A1. Abbreviations have been used for the variables: l for locations, mot for matrix of car ownership, pc for the generalized cost by car, pt for the generalized cost by public transport, a for the "other factors".

We can see from this that a high degree of correlation exists between locations, car transport supply and public transport supply (correlation coefficients of between 0.6 and 0.7). This is as one would expect. The greater the distances between origins and destinations the longer the transport times by car or public transport. This correlation between the variables reduces the goodness of fit during calibration. However, when we examine the changes in the variables rather than the variables themselves the correlation coefficients are lower (for example a correlation coefficient of 0.10 between change in locations and change in both types of transport supply). The combined effects should therefore be smaller so it should be possible to isolate the single effects generated by each variable.

*Table 1: Coefficients of correlation between the variables calculated for the 5*5 zones in the HS76 perimeter*

Variables	l, mot	l, pc	l, pt	l, a	mot, pc	mot, pt	mot, a	pc, pt	pc, a	pt, a
Coefficient of correlation between variables	-0.18	-0.60	-0.66	0.10	0.21	0.34	0.05	0.73	-0.09	-0.02
Coefficient of correlation between changes in variables	-0.12	-0.10	0.10	-0.16	0.23	0.04	0.02	0.43	0.42	0.28

4.3. Calibration

Calibration was performed for the different origin-destination pairs by linear regression using Excel in the cases where f was linear (formulations A1 and P1), and using JMP software with a nonlinear formulation calibration algorithm in the cases where f was not linear (formulations A2 and P2).

Calibration nevertheless posed problems. If the surveyed flows on a given origin-destination pair are too low the estimated modal shares will be unreliable. Under these circumstances the O-D pair in question must be rejected for fear that it will bias the estimates of coefficients. On the other hand, when the number of O-D pairs considered is too small, the estimated coefficients will be unreliable due to an insufficient number of degrees of freedom in the regression. In our study (Bonnel, Cabanne, 2000), we performed calibration with the conurbation divided up into zones of various sizes and for different thresholds for acceptance of the surveyed flows on each O-D pair, thereby testing the robustness of our coefficients.

We consequently obtained the following formulations:

A1: $u = a\ mot + b\ pc + c\ pt + d + b_1 B_1 + b_2 B_2 + \varphi(a_{ij})$
Calibration method: linear regression; $R^2 = 0{,}93$
Value and significance of parameters:

	Constant	a	b	c	b_1	b_2
Coefficient	-0.13	3.26	-0.02	0.02	-1.11	-0.50
t statistic		-0.25ll 4.20	-3.48	3.45	-10.60	-6.12

P1: $u = a\ mot + b\ \frac{pt}{pc} + b_1 B_1 + b_2 B_2 + c + \varphi(a_{ij})$
Calibration method: Linear regression; $R^2 = 0{,}93$
Value and significance of parameters:

	Constant	a	b	b_1	b_2
Coefficient	-1.17	3.47	0.52	-1.05	-0.45
t statistic	-1.99	4.83	4.49	-10.26	-5.59

A2: $u = a \frac{pc}{mot^m} + b\ pt\ mot^m + b_1 B_1 + b_2 B_2 + c + \varphi(a_{ij})$
Calibration method: Newton-Raphson algorithm.
Value of parameters:

Variable	m	c	a	b	b_1	b_2
Value of coefficient	1.32	1.97	- 0.01	0.03	- 1.13	- 0.52

P2: $u = a \frac{pt\ mot^m}{pc} + b_1 B_1 + b_2 B_2 + c + \varphi(a_{ij})$
Calibration method: Newton-Raphson algorithm
Value of parameters:

variable	m	c	a	b_1	b_2
value of coefficient	2.82	1.16	0.65	- 1.23	- 0.55

The signs of the coefficients were reasonable from the economic point of view for all the formulations we have retained.

Furthermore, in the case of the formulations which were calibrated by linear regression the Student t test values for the coefficients of the variables are higher than 2, so we can reject the hypothesis that the coefficients are nil with a confidence level of 95%.

For the parking variables B_1 and B_2, the Student t test values are particularly high (of the order of -10 for b_1 and -6 for b_2) which confirms the value of including these variables in the utility formulations. Moreover, the values of b_1

and b_2 are very stable for all formulations: b_1 is between roughly –1.1 and -1.2 and b_2 between roughly –0.45 and –0.55. These coefficients are remarkably in proportion to urban density (population density + jobs) which is used in some models to express pressure on parking (Lichère, Raux, 1997). Thus, the density of zone 1 (the hypercentre) is 17,000 inhab/km², that of zone 2 (the rest of Lyon and Villeurbanne) is 6,900 inhab/km², i.e. a density ratio of 0.4.

We propose various indicators to evaluate the distance between observed and calculated modal shares:

- the coefficient of correlation between calculated and observed modal shares: Cor;
- mean error A': $$A' = \sum_{ij} l_{ij} \times \left| P_{TCij\ calculated} - P_{TCij\ observed} \right|$$
- mean percentage error B': $$B' = \sum_{ij} l_{ij} \times \left| \frac{P_{TCij\ calculated} - P_{TCij\ observed}}{P_{TCij\ observed}} \right|$$

We have applied these indicators to the different formulations (table 2).

Table 2: The values for the indicators of the proximity between the calculated and observed market share matrices

	A1	P1	A2	P2
Cor	0.93	0.94	0.93	0.94
A'	3.6%	3.4%	3.6%	3.1%
B'	18.8%	15.3%	20.2%	14.6%

The accuracy of formulations A1, P1, A2 and P2 seems satisfactory when judged by a variety of criteria (the sign of the coefficients, statistical tests and indicators of the proximity between calculated and observed modal shares).

N.B. in addition to the formulations A1, P1, A2 and P2 described above, we tested others which we were obliged to reject either because the optimization algorithm did not achieve convergence or because there were problems concerning the sign or significance of the regression coefficients.

4.4. Weak combined effects

In the case of formulations A1, P1 and A2, the combined effects (i.e. the double, triple, quadruple and quintuple effects) were negligible irrespective of which study perimeter was used and the way the study zone was divided up (Table 3). It is therefore possible for us to talk of the separate effects of the different factors for formulations A1, P1 and A2. However, for P2, the double effect of transport supply and car ownership was close to 1.04 and the product of the different effects was equal to 1.038. For P2 the effects of single variables are not well isolated and are difficult to interpret. We shall therefore retain only A1, P1 and A2 in order to calculate and analyze single effects.

Table 3: Values of the combined effects

formulation	Difference between the product of all the combined effects and unity (absolute value)	Maximum difference (absolute value) between each combined effect and unity
A1	0.3%	E(mot, pt) - 1 = 0.6%
P1	0.1%	E(pt, pc) - 1 = 1.8%
A2	0.4%	E(mot, pt) - 1 = 1.4%
P2	3.8%	E(mot, pt) - 1 = 4.1%

4.5. Estimation of the single effects of locations, car ownership, transport supply and "other factors"

The effects of locations were identical irrespective of the formulation as they are not dependent on the calibration of the utility function.

Defining the impact of locations y(l) in the following manner: y(l) = E(l) - 1, where E(l) is the effect of locations as defined in section 2, and y(l) is expressed as a percentage gives the values shown in Table 4:

Table 4: Impact of locations

	HS76 to HS85	HS85 to HS95	HS76 to HS95
Impact of location	-3.9%	-6.5%	-8.6%

The other single effects are quite sensitive to the formulation. Nevertheless, their order of magnitude remains comparable (Table 5).

Over the period 1976 - 1995, the effect of locations can be estimated at -8.6%. The locations were defined as the matrix of the proportion of trips on each origin-destination pair. The "location" effect is therefore different from the impact of urban sprawl.

Table 5: The Impact of simple effects other than locations

	76-85				85-95				76-95			
	mot	pc	pt	a	mot	pc	pt	a	mot	pc	pt	a
A1	-11.7%	1.1%	7.1%	21.6%	-1.1%	3.9%	2.5%	-8.7%	-12.7%	5.2%	10.1%	11.7%
P1	-12.4%	1.6%	11.3%	17.8%	-1.1%	4.0%	3.4%	-9.4%	-13.5%	6.5%	15.6%	7.2%
A2	-7.0%	1.1%	6.9%	15.7%	-1.0%	3.3%	2.5%	-8.3%	-8.1%	5.0%	9.8%	6.3%

The effects of the origin-destination matrix of car ownership, transport supply and "other factors" are more difficult to measure as they do not involve a straightforward calculation but depend on the validity of the modal split formulation used (the relevance of the variables, the precision of the formulation and the reliability of the parameters). However, using our most accurate formulations, we can estimate the effect of car ownership at between roughly -8% and -14% over the first period. With regard to the second period (1985-1995), the origin-destination matrix of car ownership has remained almost stable, leading to a very small effect. The effect of car ownership corresponds to the definition of the origin-destination matrix of car ownership. Therefore we have not measured the full effect of the change in car ownership over the entire conurbation. What we have done, in a way, is break down the change in car ownership into two factors: the change in car ownership in the context of a given spatial structure and the change in car ownership linked to change in the spatial structure of flows. The effect of car ownership that we measured is therefore limited to the influence of the first factor, while the second is included among the effect of locations.

We estimate the impact of transport supply at between 15 and 20% over the whole period. Transport supply has been defined on the basis of the generalized cost of public transport and the car. The impact of parking difficulties has been excluded from this as we do not have data for the three survey dates. As a result this has been included among the "other factors". The effect of transport supply can be split between the two modes. The considerable improvement in public transport supply during the period (the construction of four metro lines, restructuring of the bus network, large increase in vehicle/kilometres, etc.) has led to a reduction in the generalized costs of public transport which explains the 10 to 15% increase in its market share. Increased car use combined with increased distances travelled has been responsible for a fairly moderate increase in the generalized costs of car transport as a result of improvements to the network and traffic control measures. This limited its increase to between roughly 5 and 6% in the period 1976-1995.

We have also shown the importance of the impact of the "other factors" in the case of the formulations we have discussed. The values of the different effects should be considered in terms of orders of magnitude, but the "other factors" would seem to have as much influence as the various explanatory factors and seem to have been responsible for altering the direction of a trend. The increase in the modal share of public transport would seem to be greater than expected in the first period, but in the second other factors seem to have restrained this increase. Several possible reasons can be given for this. The "marketing" effect of the opening of the metro in 1988 could have increased the attractiveness of public transport to a greater extent than the increase in services would justify - extension of the network after 1985 failed to have a similar effect. Between 1985 and 1995 policies aimed at increasing parking supply in the centre of the conurbation could have helped reduce the modal share of public transport, as could the increasing complexity of journeys as compared with straightforward home-based return journeys. Finally, perhaps a change in behaviour is taking place stemming from a perception of transport modes which tends to favour the car, which the stable generalized cost functions over the period as a whole has not been able to express.

5. Conclusions

With regard to methodology, it is worth returning to the interpretation of these results. First of all, we should be careful not to consider that they express a unidirectional causal link. "Effect" is in fact an inappropriate term, as it refers to a causal relationship we do not wish to consider here, as our research does not allow us to deal fully with the issue. However, we do show that a certain change in the spatial distribution of trips can be linked to a certain change in the market share of public transport and we are able to quantify this link. The same can be stated with regard to car ownership, transport supply and other factors.

The second comment we shall make refers to independence between the factors. Our findings do not mean that the five factors are independent - it would be easy to prove otherwise. However, we have managed to find a model that allows us to reduce (if not eliminate) the dependency between change in the factors we have considered.

This leads us to the last comment. Our reference to "the effect of car ownership", for example, is a misuse of language. Change in car ownership can influence location choice and in this study any results are considered as the effect of location not car ownership, in the same way that choice of location can affect car ownership. What we refer to as the "effect of car ownership" does not therefore include all the "effects of car ownership" on use of modes. Rather, it expresses the "effect" of change in car ownership in the absence of any structural change in trip-making. That is to say that in a manner of speaking we have broken down change in car ownership over the conurbation as a whole into two components. The first is related to changes in the location of households (urban sprawl thus leads inevitably to increased car ownership as car ownership is higher in the outskirts than in the centre). The second is related to the growth in car ownership in each of the zones used to compute the car ownership matrix. We shall thus quantify the relationship between change in the origin-destination matrix of car ownership and change in the market share of public transport. In the same way, what we refer to as the effect of location in fact corresponds to the effect of change in the spatial distribution of trip origins and destinations and not the effect of urban sprawl directly. What we quantify is therefore the relationship between change in the trip origin-destination matrix and change in the market share of public transport. The same also applies to the matrices of transport supply and of other factors.

The value of this method lies in the fact that the combined change in the five matrices signifies a change in the market share of public transport which is equal to the product of the five relationships, with nil combined effects. This can be expressed by means of the following mathematical formula:

$$\Delta P_{PT} = f(\Delta \text{location matrix}) * g(\Delta \text{car ownership matrix}) * h(\Delta \text{car supply matrix}) * k(\Delta \text{public transport supply matrix}) * l(\Delta \text{other factors matrix})$$

where ΔP_{PT} is the variation in the market share of public transport between two dates and f, g, h, k and l are five functions.

It is these five functions which for the sake of ease we refer to as the effect of location, car ownership, car supply, public transport supply and other factors, even if in fact what is involved is a relationship between a certain formulation of each of these factors and the market share of public transport.

With this method it is therefore possible to determine the way each of the factors affects change in the market share of public transport. Change in location explains a 8.6% reduction in the market share of public transport over the 20 year period covered by the study. Change in the origin-destination matrix of car ownership gave rise to a fall of between 7 and 12% over the period 1975-86 and almost no change during the subsequent period. On the other hand, changes in transport supply explain an increase of between 15 and 20%, essentially caused by an increase in public transport supply (which explains between 10 and 15% of the rise), while the rather limited deterioration in traffic conditions for cars explains an increase of between 5 and 6%. Finally, the "other factors" which group together all the unexplained factors, are responsible for a 6 to 12% rise over the 20 year period, with, however, a quite marked difference between the two periods (there was a sharp rise in the first period and a smaller fall in the second).

This method provides a means of overcoming the problem of correlation between explanatory variables. Implementing the method requires appropriate modelling of each variable in order to reduce, or even eliminate, the correlation between them. Once this has been achieved, the combined effect of each variable with the others is negligible or non-existent. The effect of the combined change of each variable can therefore be expressed very simply as the product of each of the effects taken in isolation.

Bibliography

Andan O, Bonnel P, Raux C (1988), *Les analyses des comportements de mobilité individuelle quotidienne, une synthèse bibliographique*, Laboratoire d'Economie des Transports, pour le compte du SERT, Ministère des Transports, Lyon, 118p.

Bonnel P (1995), Urban Car Policy in Europe, in *Transport policy*, Vol 2, N°2, pp. 83-95.

Bonnel P, Gabet P (1999), *Mesure de l'effet de l'évolution des localisations et de la motorisation sur la part de marché des transports collectifs*, LET pour le compte de la DRAST programme PREDIT, Lyon, 160p.

CERTU (1998), *L'enquête ménage déplacements " méthode standard "*, Collections du CERTU, éditions du CERTU, Lyon, 295p.

CETE de Lyon, D.E.U. (1977), *Enquête Ménages de Lyon 1976,* document technique, Lyon : CETE de Lyon, 49p.

CETE de Lyon (1986), *Enquête Ménages Lyon 1985,* document technique, Lyon : CETE de Lyon, 156p.

CETE de Lyon, SYTRAL, INSEE (1995), *Enquête " Déplacements auprès des ménages de l'agglomération lyonnaise " 1994 - 1995,* document technique, Lyon : CETE de Lyon, 119p.

Gärling T, Laitila T, Westin K (Eds.) (1998), *Theoretical foundations of travel choice modeling*, Pergamon, Oxford, 498p.

Lichère V, Raux C (1997), *Développement d'un modèle stratégique de simulation des déplacements, présentation générale*, SEMALY, LET, Lyon, 28p. + annexes.

Manheim ML (1984), *Fundamentals of transportation systems analysis, volume 1: basic concepts*, MIT Press, Cambridge, Massachusetts, USA, 658p.

Massot M-H, Orfeuil J-P (1989), *Offre et demande de transport en commun dans les villes françaises sans métro - Usages et usagers des services de transport en commun urbain*, Rapport INRETS n° 100, premier tome, Arcueil : INRETS, 63p.

Massot M-H, Orfeuil J-P (1990), *Offre et demande de transport en commun dans les villes françaises sans métro - Structures de l'offre de transport*, Rapport INRETS n° 121, deuxième tome, Arcueil : INRETS, 73p.

Massot M-H, Orfeuil J-P (1991), *Offre et demande de transport en commun dans les villes françaises sans métro - Sensibilité de l'usage à l'offre*, Rapport INRETS n° 148, troisième tome, Arcueil : INRETS, 76p.

Ortuzar J de D, Willumsen LG (1994), *Modelling transport*, John Wiley & Sons, UK.

Ortuzar J de D, Hensher D A, Jara-Diaz S R (Eds.) (1998), *Travel Behaviour Research: Updating the State of Play*, Pergamon, Elsevier, Oxford.

Stopher P, Lee-Gosselin M (eds.) (1997), *Understanding travel behaviour in an era of change*, Pergamon, Oxford, 583p.