

THE ESTIMATION OF AGGREGATE MODAL SPLIT MODELS

Patrick Bonnel

Laboratoire d'Economie des Transports, ENTPE, Univ. Lumière Lyon 2, CNRS, France

Tel: 04 72 04 70 48, Fax: 04 72 04 70 92

patrick.bonnel@entpe.fr

Abstract

In the last thirty years, there has been a considerable increase worldwide in the use of disaggregate models. These range from the seminal work of McFadden and Ben Akiva (Domencich, McFadden, 1975; McFadden, 2000; Ben-Akiva, Lerman, 1985) to more recent advances involving the use of mixed multinomial models (Bhat, 1997, 2000) whose application has been facilitated by the availability of a number of software packages. However, despite this undeniable progress, the disaggregate approach still requires a number of hypotheses and, in particular, a complex estimation process (Bonnel, 2002). Consequently, the aggregate approach is still widely used in many countries, in particular France where there has been little development of disaggregate modal choice models (there have been few really significant studies for urban areas: Abraham et al., 1961; CETUR, 1985; Daly, 1985; CETE de Lyon et al., 1986; Bouyaux, 1988; Hivert et al., 1988; RATP, Cambridge Systematics, 1982; Rousseau, Saut, 1997; CERTU, 1998b). Although the development of aggregate modal choice models is fairly simple as far as formalisation is concerned, the same cannot really be said to apply for their estimation. The samples in the travel surveys that are used for these estimations are generally too small for the needs of aggregate models. It is therefore usually necessary to conduct zonal aggregations which result in a high degree of uncertainty about the quality of the estimations. In response to this situation this paper proposes an estimation method which allows very small-scale zoning to be retained. The paper starts with a description of the estimation problem for aggregate models (Section 1) and then describes the model we have developed (Section 2). We then present an application of this method for estimating an aggregate modal choice model for the Lyon conurbation. This empirical analysis allows us to study the benefits of our method in comparison with more conventional methods (Section 3).

Keywords: Modal split model; Estimation; Aggregate modelling

Topic Area: D1 Passenger Transport Demand Modelling

1. The problem of estimating aggregate models

We shall present the logit functional form first (Ortuzar, Willumsen, 2001), as this is usually used for aggregate models (Section 1) and then consider the problem of calibration (Section 2).

1.1. Logit or logistical regression models

The theoretical basis for logit models is derived from neo-classical microeconomic theory and a probabilistic approach towards utility, whose deterministic component alone is defined. This functional form is nevertheless usually justified on empirical grounds. Analysis of modal split data shows that an S-shaped curve very closely fits the survey data.

Put simply, neo-classical microeconomic theory assumes that individuals are rational, and this rationality leads them to select the alternative with the greatest utility (Henderson, Quandt, 1980). In our deterministic (and aggregate) approach utility is expressed by:

$$V_i = \sum_k \beta_{ki} X_{ki} \quad (1)$$

where V_i is the utility of a good;

X_{ki} are the various random variables used to estimate utility;

β_{ki} is the coefficient of the variable X_{ki} .

Let us consider the simplest case of a choice between two transport modes, the car and public transport. The individual will select the mode with the highest utility. He will therefore select the car if:

$$V_{PC} \geq V_{PT}, \text{ therefore if } V_{PC} - V_{PT} \geq 0 \quad (2)$$

V_{PC} denotes the utility of the private car and V_{PT} the utility of public transport.

A deterministic utilisation of the model based on the average utility of the car and public transport leads to an all-or-nothing choice (0 or 100%) on each origin-destination pair. This obviously disagrees with the available empirical data which generally has an S-shaped distribution (Figure 1). To cope with this it is necessary to adopt a probabilistic approach. The aggregate approach causes us to consider an average individual travelling between a zone i and a zone j who is faced with a choice of transport mode. The next stage is to compute an average utility for our average individual for each of the modes in the light of his or her socio-economic characteristics. Of course, our average individual is fictional and in reality individuals' characteristics are distributed around his or hers. We can therefore use the probabilistic approach to constant utility developed by Luce and Suppes (1965), who were responsible for the theoretical justification of the use of the logistic functional form in discrete choice processes.

Below we shall restate, in simplified form, the presentation of this by Ben-Akiva and Lerman (1985). Let $P\langle i | C_n \rangle$ be the probability of individual n selecting the alternative i from the set C_n of the alternatives available to him or her. This probability must obviously satisfy a number of properties:

$$0 \leq P\langle i | C_n \rangle \leq 1, \forall i \in C_n \quad (3)$$

which is the equality for the deterministic case.

$$\sum_{i \in C_n} P\langle i | C_n \rangle = 1 \quad (4)$$

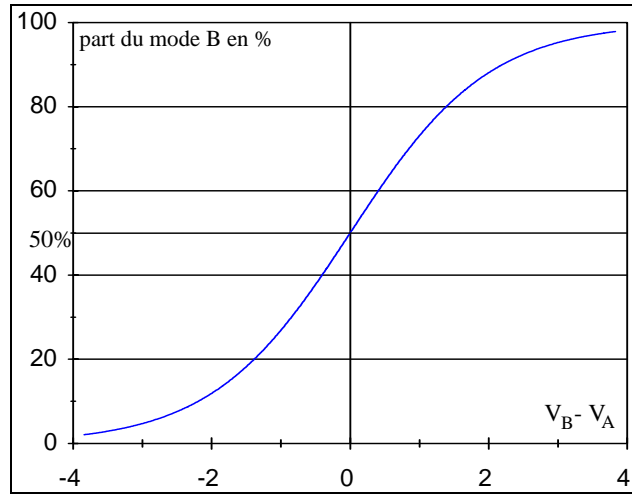


Figure 1: The general S-shaped modal split curve

For each subsample $\tilde{C}_n \subseteq C_n$, the probability of \tilde{C}_n in C_n , is given by:

$$P\langle \tilde{C}_n | C_n \rangle = \sum_{i \in \tilde{C}_n} P\langle i | C_n \rangle \quad (5)$$

From which we can deduce the conditional probability of i in \tilde{C}_n :

$$P\langle i | \tilde{C}_n \subseteq C_n \rangle = \frac{P\langle i | C_n \rangle}{P\langle \tilde{C}_n | C_n \rangle} \quad (6)$$

on condition that at least one of the conditional probabilities of an alternative that belongs to \tilde{C}_n is not null such that $P\langle \tilde{C}_n | C_n \rangle$ is not null.

Luce (1959) constructed the simplest model that is based on a probabilistic constant utility approach, according to Ben-Akiva and Lerman (1985). It is based on the following hypothesis known as the “*choice axiom*”: The “*choice axiom*” stipulates that for any alternative that belongs to the subset \tilde{C}_n in the set C_n such that $i \in \tilde{C}_n \subseteq C_n$, we have:

$$P\langle i | \tilde{C}_n \subseteq C_n \rangle = P\langle i | \tilde{C}_n \rangle \quad (7)$$

“*In other words, if some alternatives are removed from a choice set, the relative choice probabilities from the reduced choice set are unchanged*” (Ben-Akiva, Lerman, 1985).

From Equations 6 and 7 we can deduce the following:

$$P\langle i | C_n \rangle = P\langle i | \tilde{C}_n \rangle * P\langle \tilde{C}_n | C_n \rangle \quad \text{for all cases where } i \in \tilde{C}_n \subseteq C_n \quad (8)$$

from this axiom we can deduce the *Independence from Irrelevant Alternatives property* also known as IIA, which is familiar to logit model users:

$$\frac{P\langle i | \tilde{C}_n \rangle}{P\langle j | \tilde{C}_n \rangle} = \frac{P\langle i | C_n \rangle}{P\langle j | C_n \rangle}, \text{ for all cases where } i, j \in \tilde{C}_n \subseteq C_n \quad (9)$$

Luce (1959) has shown that in the case where this axiom is satisfied, the probability of alternative i being selected can be written very simply. To begin with, we shall restrict ourselves to a set C_n which contains only two alternatives i and j :

$$P\langle i | C_n \rangle = \frac{1}{1 + e^{-\mu(V_{in} - V_{jn})}} = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}} \quad (10)$$

as μ is a constant it is generally integrated within the formulation of utility. The equation then becomes:

$$P\langle i | C_n \rangle = \frac{1}{1 + e^{-(V_{in} - V_{jn})}} = \frac{e^{V_{in}}}{e^{V_{in}} + e^{V_{jn}}} \quad (11)$$

This is the equation for the logistic distribution. This theoretical justification is usually ignored in favour of the empirical observation of survey data combined with an analytical simplification. Empirical observation gives an S-shaped curve that resembles the logistic distribution. Likewise, the logistic functional form, by the IIA property, results in a very simple equation even in the case of a very large number of alternatives. It is simply a generalisation of equation 11:

$$P\langle i | C_n \rangle = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \quad (12)$$

The IIA property results in a very simple analytical form, which is the strength of the logit formulation in comparison with the others. But this property is also the model's weakness, as it results in the well-known blue bus/red bus paradox (Ben Akiva, Lerman, 1985). This property is actually a direct consequence of the hypotheses that are present in the "choice axiom", which resembles a hypothesis of independence between the utility of each alternative. This hypothesis is frequently acceptable when there is only a small number of alternatives. However, it becomes much less so when there are more alternatives, particularly when some of the alternative modes are of the same type (for example underground, bus, train, etc.). It was in order to cope with these difficulties that nested logit models or other more complex functional forms were developed (Bhat, 1997, 2000; Bonnel, 2002).

While the analytical form is extremely simple, estimation is much less straightforward because of the small size of the samples that are generally available for estimating origin-destination matrices.

1.2. Calibration of logit models

Calibrating a logit model involves estimating the unknown coefficients, i.e. the coefficients of the utility function for each transport mode (the coefficients β_{ki} if we keep the notation used in equation 1). To perform this estimation, data is required to construct

origin-destination matrices for each mode. Estimation aims to obtain β_{ki} coefficients which provide the “best” reproduction of these modal demand matrices, for the selected convergence criteria.

Data from household travel surveys are frequently used to construct these matrices. Consequently, both travel data and socio-economic data concerning the individuals are available for inclusion in the utility functions. The sample sizes of these surveys are nevertheless generally too small to provide matrices in which the data for each origin-destination pair is sufficiently statistically significant.

To provide a simple illustration of the problem we shall take the case of the Lyon conurbation. A transposition to other cities elsewhere in the world would not pose any problems, as the conclusions depend on the number of surveyed trips that are available for estimating the demand matrix for each mode. The most recent household travel survey in Lyon involved 6,000 households, corresponding to 14,000 respondents and a total of 53,000 trips. In addition, the models developed for the conurbation are mostly based on a division of the conurbation into between 100 and 500 zones. Even if the largest-scale zoning is selected, the matrix already contains 10,000 cells. These cells will be filled by just 53,000 trips. Furthermore, it is necessary to construct one matrix for each mode that is considered. The matrices will therefore contain a large number of empty cells and very few cells which contain enough trips to provide an acceptable level of statistical accuracy (Figure 2 provides an indication of the confidence interval for the hypothesis of simple random selection with no refusals for a mode with an 18% share of the market. These hypotheses are obviously more favourable than those that actually pertain for the surveys which would therefore give even larger intervals (Richardson et al., 1995)).

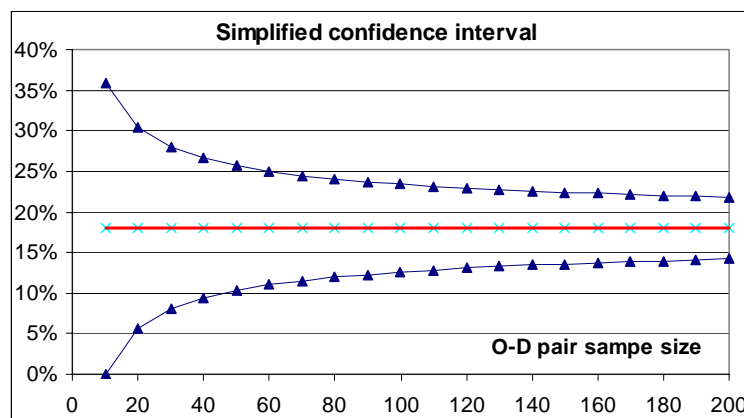


Figure 2: Simplified estimation of the confidence interval in the case of an 18% modal share and simple random selection

Combining household travel survey data with data from other sources, such as traffic counts or data provided by public transport operators can improve the quality of trip matrix estimation. However, the problem of statistical reliability still remains for many origin-destination pairs. If the reference matrix is not sufficiently reliable the problem is automatically carried through to the estimation of the utility function coefficients for each of the modes. Generally, one of two solutions is adopted in an attempt to overcome this difficulty.

The first is to select only those origin-destination pairs whose statistical accuracy is judged to be satisfactory. Consequently, the utility function coefficients are only estimated for part of the matrix. The main problem with this method is that it eliminates part of the

available data. Furthermore, there is no guarantee that the selected origin-destination pairs are representative of all the trips. It is even likely that they will not be, for the following reasons. The origin-destination pairs with the largest number of trips are usually central or radial and there are usually considerably fewer trips for peripheral flows. Calibration will therefore be conducted on the basis of flows for which public transport and environmentally friendly modes play a greater role. This selection procedure is therefore likely to reduce the variance of the initial data set, which will necessarily reduce the quality of the estimation.

The second is to perform zonal aggregation to increase the number of trips on each origin-destination pair. The aggregation is generally performed on the basis of geographical proximity with the attempt to combine zones whose characteristics do not differ excessively. In order to obtain adequately sized samples, the number of zones must be reduced dramatically. If we return to the example of Lyon, even if the number of zones is reduced to 25 (i.e. 625 O-D pairs) there are still many O-D pairs for which the accuracy is very low. The main problem with this aggregation relates to computation of the time and cost data for each O-D pair. When the size of the zone increases, it becomes difficult to estimate public transport access and regress times at the origin and destination. There are a number of possible routes between the origin zone and the destination zone depending on the exact location of the origin and destination within the zones. It therefore becomes difficult to compute a reasonable value for each explanatory variable. The generally adopted solution is to produce a weighted average of the data on the basis of the weight of each of the smallest zones that make up the macro zone. However, the size of the zones means that the accuracy of this measurement is uncertain as there is considerable variation in trip duration for a given origin-destination pair depending on the exact location within the origin zone and the destination zone. There has thus been an improvement in the accuracy of the flow estimates, but at the cost of a loss of accuracy in the estimation of certain explanatory variables which play a role in defining the utility of each mode.

The zonal aggregation process prior to calibration must therefore comply with two contradictory objectives:

- the number of zones must be as small as possible in order to increase the number of surveyed trips that can be used during estimation and hence the accuracy of the O-D matrices for trips conducted by each mode;
- the zones must be as small as possible in order to produce accurate data for the explanatory variables that are included in the utility functions.

To cope with this contradiction we shall propose another data aggregation method which conserves all the information and the smallest-scale zoning.

2. A new calibration method

This method uses a different process to aggregate O-D pairs in a way that conserves the smallest-scale zoning. Because for many O-D pairs the number of trips is too small, it is necessary to develop an aggregation procedure which is as accurate as possible. As this method is much simpler to implement in the case of a choice between two modes we shall limit our presentation to this case. In this two-mode situation, the probability of mode i being chosen can be written as follows (equation 11):

$$P\langle i | C_n \rangle = \frac{1}{1 + e^{-(V_{in} - V_{jn})}}$$

Modal choice is thus made on the basis of the difference in utility between the two modes. It is logical to use this quantity for the aggregation procedure. The principle is

straightforward; O-D pairs are aggregated on the basis of their closeness in terms of the difference in utility between the two modes. This procedure is justified by the fact that in general the statistical precision of the measurement of the difference in utility is much higher than that of a market share observation based on household travel survey data. This is because, generally, the variables used to compute the difference in utility are either zonal variables (the number of trips at zonal level is necessarily greater than at O-D pair level), or level of service variables whose accuracy depends on how the network is coded and how small the zones are. In practice, therefore, what is required is to achieve the right balance between having small-scale zoning for service level data and zones that are sufficiently large for the zonal data.

To apply this method it is necessary to know the values of the utility function coefficients. However, these values are obtained during the model calibration phase. To remove this contradiction we propose an iterative process whose principal stages are as follows:

- selection of initialisation values for the utility function coefficients for the two transport modes in order to initialise the iteration process. These values can, for example, be obtained from a previous study or from calibration performed with the classical method described in Section 1.2. Selection of these values is not of great importance as application of the method has shown that if an adequate number of iterations is performed the initialisation values do not affect the final results;
- the utility of each mode is computed on the basis of the above coefficients, followed by the difference in the utility for each O-D pair in the smallest-scale zoning system for which explanatory variables with an adequate level of accuracy are available (therefore with very limited zonal aggregation or even none at all);
 - the O-D pairs are classified on the basis of increasing utility difference;
 - the O-D pairs are then aggregated. This is done on the basis of closeness of utility difference. The method involves taking the O-D pair with the smallest utility difference then combining it with the next O-D pair (in order of increasing utility difference). This process is continued until the number of surveyed trips in the grouping is greater than a threshold value that is decided on the basis of the level of accuracy that is required for trip flow estimation. When this threshold is reached the construction of the second grouping is commenced, and so on and do forth until each O-D pair has been assigned to a group;
 - for each new class of O-D pairs it is necessary to compute the values of the explanatory variables that are included in the utility functions for each mode. This value is obtained from the weighted average (with weights assigned on the basis of the number of trips made on each O-D pair) of the values for each O-D pair in the class. It is likewise necessary to compute the market share of each mode for this class. This is obtained very simply by summing the trips by each mode on all the O-D links in the class;
 - a new estimation of the utility function coefficients.

This process must obviously be repeated many times to achieve satisfactory convergence in the estimation of utility function coefficients. This method has been tested on the Lyon conurbation in order to check its ability to reach convergence and its reliability when compared with empirical data.

3. A Test of the method on the Lyon conurbation

Before presenting the results of an application of this new calibration method we shall describe the model we have tested (Section 3.1) and the data used (Section 3.2).

3.1. The modal choice model for the Lyon conurbation

We have selected a model developed in a previous study by the LET and the SEMALY (Lichère, Raux, 1997ab). This model aims to estimate the market split for motorised transport between private cars and public transport. It uses a logit functional form, even though the utility expression is not entirely additive for the selected explanatory variables:

$$\%TC = \frac{1}{1 + \exp \left(k_m + \tau pt_m * \left(tpt_{ij} * mot_i \right) - \tau pc_m * \frac{tpc_{ij}}{mot_i} - \delta_m * d_j \right)} \quad (13)$$

where tpt_{ij} is the generalised time by public transport between zones i and j ;

tpc_{ij} is the generalised time by private car between zones i and j ;

mot_i is the car ownership rate in zone i ;

d_j is the density of zone j , expressed in terms of population + jobs per hectare;

k_m , τpt_m , τpc_m and δ_m , are the modal split parameters for purpose m .

This formulation requires some explanation concerning the components of the utility functions for the private car and public transport:

$$- \tau pt_m * \left(tpt_{ij} * mot_i \right)$$

is the generalised time taken on public transport to travel between zones i and j . It is weighted by a time perception factor which depends on the car ownership rate in the origin zone. The higher the car ownership rate the less favourably time spent travelling will be viewed (and therefore the more it will be increased). τpt_m is the coefficient of this variable which must be determined when the logit model is calibrated;

$$- \tau pc_m * \frac{tpc_{ij}}{mot_i}$$

is the generalised time taken to travel by private car between zones i and j . It is weighted by a time perception factor which depends on the level of car ownership in the origin zone. The higher the car ownership level the more favourably time spent travelling will be viewed (and therefore the more it will be reduced). τpc_m is the coefficient of this variable which must be determined when the logit model is calibrated;

$$- \delta_m * d_j$$

is the density of zone j , expressed in terms of population + jobs per hectare. This term expresses pressure on parking. It was introduced because accurate parking data is not available for the Lyon conurbation. The higher the density of the zone (in terms of population and jobs) the higher the pressure on parking. We know from experience that this density frequently provides an accurate idea of parking pressure. In another study on Lyon in which parking constraint for central zones was introduced as a dummy variable the

coefficient ratios were found to be identical to the density ratios (Bonnel, Cabanne, 2000). However, while the introduction of this variable is important in order to calibrate the model and avoid biasing the coefficients of the other variables, this definition poses problems for forecasting. This variable would not allow us to simulate directly a modification of parking supply in the zone. At most, it would enable us to simulate changes in percentage terms. δ_m is the coefficient of this variable which must be determined when the logit model is calibrated.

Six purposes were selected: work, primary education, secondary education, university level education, shopping/services and other purposes. For this application we have restricted ourselves to work trips (Ferey, 2002 contains an analysis of all the purposes).

3.2. Data description

Calibration was conducted using the trip origin-destination matrices for the two modes considered in this analysis, namely the private car and public transport. These matrices were constructed using data from the most recent household travel survey for the Lyon conurbation, which was conducted in 1995 (CETE de Lyon et al., 1995). In the course of this survey all the individuals aged 5 years and over from more than 6,000 households were questioned about the trips they had made on the day before the survey (Table 1). The methodology complied with a specification which is used for all household travel surveys conducted in French cities (CERTU, 1998a). The sample was obtained from random selection after geoFigureical stratification over 87 areas.

Table 1: Principal data from the Lyon household travel survey 1995

	Number of households	Number of individuals	Number of trips	Number of work trips
survey sample size	6,001	13,997	53,213	8,123
weighted sample size	536,317	1,195,189	4,659,777	728,818

(source: LET based on the Lyon household travel survey)

The generalised time data for public transport was produced by the SEMALY public transport assignment model (SEMALY, 2000), which is a shortest path assignment model. Modelling was carried out using a description of the public transport network as it was at the time of the household travel survey (Ferey, 2002).

Lastly, the generalised time data for the private car was produced by the CETE de Lyon using the DAVIS private car assignment model (PTV-Isis, 2001). Assignment was conducted using the Wardrop equilibrium assignment model (Ferey, 2002).

These two models have to be used because no multimodal model exists for the Lyon conurbation. As two different programs are used we cannot be certain that the generalised time data for the two modes is completely consistent. This would be detrimental if our aim was to calibrate a forecasting model for the Lyon conurbation, but our study only aims to test the calibration method. The effect is therefore marginal. It is nevertheless obvious that we cannot re-use the calibration values we produce for forecasting purposes. Our use of two programs raises another problem, which results from the fact that they do not both use the same zoning. We therefore had to construct a passage matrix to move from one zoning to the other. Once again this situation would be detrimental if our aim was to estimate calibration coefficients, but it is not really for the test which we shall perform. Consequently, we have worked with the smallest-scale zoning which is common to both

the SEMALY and CETE de LYON programs, and then with more aggregated zonings. The available zonings are summarised below:

- D196: division into 196 zones used for SEMALY's TERESE program;
- D87: division into 87 zones used for geoFigureical stratification during selection of the sample for the household travel survey;
- D25: division into 25 zones on the basis of concentric rings and catchment areas (Lichère, Raux, 1997ab);
- D7: division into 7 zones on the basis of concentric rings and an East-West separation along the Rhône (the river which crosses the conurbation and which is responsible both for physical severance and sociological differentiation).

By examining the calibration for each of these divisions it will be possible to perceive the influence of zonal aggregation on estimation of the coefficients.

3.3. Application of the estimation methods

As we are considering just two transport modes, the logit equation (equation 13) can be written differently:

$$\log\left(\frac{1}{\%PT} - 1\right) = k_m + \tau p t_m * \left(t p t_{ij} * mot_i \right) - \tau p c_m * \frac{t p c_{ij}}{mot_i} - \delta_m * d_j \quad (14)$$

This brings us back to the linear form, which permits the use of linear regression to estimate the model instead of maximum likelihood (Ortuzar, Willumsen, 2001). We have therefore used linear regression for the various estimations of the modal split model. This means that it is possible to calibrate the model using any spreadsheet program. This led us to include the two estimation methods presented in Section 1.2 and Section 2 that are contained in transportation modelling courseware published by LET, IMTRANS and MVA (Bonnell et al., 2002).

3.4. Analysis of results

We shall begin our presentation with an analysis of the "classical" calibration method presented in Section 1.2 in order to illustrate the limits we have already described. We shall then describe the results from the new calibration method we are proposing.

3.4.1. "Classical" calibration method

The first stage of calibration is to select the appropriate zoning for this analysis. This requires us to reconcile two contradictory objectives (see Section 1.2):

- the number of zones must be as small as possible in order to increase the number of surveyed trips that are considered for the purposes of estimation and therefore the accuracy of the origin-destination matrices for each mode;
- the size of the zones must be as small as possible in order to produce data which are accurate with regard to the explanatory variables included in the utility functions.

We therefore begin our analysis with the smallest-scale zoning, i.e. 196 zones. Only 8,123 trips (unweighted samples) were surveyed for the work purpose (Table 1). As there are 38,416 origin-destination pairs, it is obvious that most of the flows are null and that very few will have enough trips to allow us to estimate public transport market share with acceptable precision. We have to accept a threshold of 10 surveyed trips to obtain a sufficient number of origin-destination pairs to be able to perform regression (Table 2). However, in this case, the origin-destination pairs only include 9% of all the trips and Figure 2 shows that for an 18% market share (which was the proportion given by the Lyon household travel survey for the work trip purpose) the accuracy is extremely low (the

confidence interval of the percentage at the 95% confidence threshold is between -0.5% and 35%).

Table 2: Number of origin-destination pairs with a number of trips above a given threshold, for the D196 division

D196, Threshold	30	25	20	15	10
Number of origin-destination pairs with a number of trips above the threshold	3	4	8	17	48
These origin-destination pairs as a percentage of all surveyed trips for the work trip purpose	1.3%	1.6%	2.6%	4.6%	9.1%

(source: LET based on the Lyon household travel survey)

Zonal aggregation is therefore essential. The position is nevertheless identical for the 87-zone division (Table 3). We therefore propose a new zonal aggregation with a 25-zone division (Table 4). Even though the results pose fewer problems than in the case of the previous divisions, the number of origin-destination pairs with enough trips is still quite small. With a threshold of 40 surveyed trips, which nevertheless leads to a quite a large confidence interval ($[8,6\% - 25,4\%]$ at the 95% threshold), only 52 origin-destination pairs have a sufficiently large surveyed number of trips. Furthermore, half of the surveyed trips do not belong to these origin-destination pairs. Based on these criteria, the last 7-zone division leads to much more satisfactory results (Table 5).

Table 3: Number of origin-destination pairs with a number of trips above a given threshold, for the 87-zone division

D87, Threshold	30	25	20	15	10
Number of origin-destination pairs with a number of trips above the threshold	13	23	38	59	164
These origin-destination pairs as a percentage of all surveyed trips for the work trip purpose	7.5%	10.8%	14.7%	18.9%	33.6%

(source: LET based on the Lyon household travel survey)

Table 4: Number of origin-destination pairs with a number of trips above a given threshold, for the 25-zone division

D25, Threshold	100	80	60	40	20
Number of origin-destination pairs with a number of trips above the threshold	12	19	29	52	132
These origin-destination pairs as a percentage of all surveyed trips for the work trip purpose	21.2%	28.8%	37.6%	51.1%	77.3%

(source: LET based on the Lyon household travel survey)

Table 5: Number of origin-destination pairs with a number of trips above a given threshold, for the 7-zone division

D7, Threshold	100	80	60	40	20
Number of origin-destination pairs with a number of trips above the threshold	23	23	30	34	40
These origin-destination pairs as a percentage of all surveyed trips for the work trip purpose	88.6%	88.6%	94.4%	96.8%	99.1%

(source: LET based on the Lyon household travel survey)

The selection of the level of zoning and the desired level of accuracy for the estimation of the market share matrix for each mode must therefore involve a compromise between four dimensions which are to a considerable extent contradictory:

- the selected threshold for the number of surveyed trips must be as high as possible to guarantee an acceptable level of precision for the market share matrix for each mode;
- the selected number of regression classes must be as high as possible in order to increase the number of degrees of freedom in the regression and very probably the initial variance of the data set;
- the number of zones in the division must be as great as possible in order to ensure that the measurement of the explanatory variables for the origin-destination pair is sufficiently accurate;
- the number of trips for the selected origin-destination pairs must be as high as possible so as to make the best use of the available data.

An examination of Tables 2 to 5 shows that it is extremely difficult to reconcile these criteria. In addition, the selected origin-destination pairs are mostly central or radial flows. Even with the largest-scale 7-zone division, the selected origin-destination pairs contain few peripheral flows. The variance of the data set is inevitably reduced as a consequence, and with this the quality of the estimation.

Table 6: Results of calibration for different surveyed trip thresholds for the 25-zone division

Threshold	Number of O-D pairs	R ²	Coefficient (Student's t value in brackets)				PT's estimated %
			Density δ_m	τ_{pc}_m	τ_{pt}_m	constant	
40	52	75%	-0.0044 (-4.96)	-0.038 (-3.44)	0.052 (6.64)	2.70 (8.24)	16.71%
50	37	70%	-0.0050 (-3.76)	-0.038 (-2.20)	0.045 (4.48)	2.93 (6.43)	16.89%
60	29	72%	-0.0063 (-3.99)	-0.037 (-2.14)	0.030 (2.33)	3.47 (6.25)	17.36%
70	25	73%	-0.0072 (-4.46)	-0.054 (-2.57)	0.017 (1.20)	4.29 (5.96)	22.25%
80	19	84%	-0.0066 (-4.82)	-0.042 (-2.09)	0.024 (2.09)	3.69 (5.93)	19.40%
90	15	70%	-0.0046 (-2.06)	-0.068 (-1.85)	0.027 (2.19)	4.16 (4.14)	23.14%
100	12	65%	-0.0045 (-1.40)	-0.080 (-1.60)	0.023 (1.38)	4.52 (3.24)	26.24%

(source: LET based on the Lyon household travel survey)

For the purposes of calibration therefore, we will select the configurations which seem the most appropriate. This leads to the elimination of the 87 and 196-zone divisions in favour of the most aggregated divisions. The coefficient estimates for these two divisions are set out in Tables 6 and 7. Several problems are apparent from an analysis of the 25-zone division:

- above the threshold of 70 surveyed trips, some coefficients are no longer significant at the 5% threshold (Student's t value < 2.1). With the same threshold, the number of classes also becomes very low, leading to a small number of degrees of freedom. Lastly, the origin-destination pairs retained for regression contain too small a percentage of all the trips (Table 4);

- for thresholds between 40 and 60 surveyed trips, the estimated proportion is quite close to that which was observed, i.e. 17.9%. However, the estimated coefficients exhibit quite a high degree of instability with regard to the selected threshold, with the exception of the coefficient τ_{pc_m} . Furthermore, the percentage of trips which involve the selected origin-destination pairs is still quite low (between a third and half of all trips are considered for the regression).

These results are considerably more satisfactory for the most aggregated 7-zone division (Table 7). The estimated percentage of trips that is made by public transport is in all cases close to that observed in the survey (17.9%). The R^2 value is always high with satisfactory Student's t values for the estimated coefficients. Lastly, the estimated coefficients are generally stable, irrespective of the threshold that is employed. These results lead to prefer the more aggregated zoning. However, there is some uncertainty about the accuracy of the generalised time measurements both for the private car and public transport when such large zones are used. The computed average values conceal extremely high disparities.

Table 7: Calibration results for different surveyed trip thresholds for the 7-zone division

Threshold	Number of O-D pairs	R^2	Coefficient (Student's t value in brackets)				PT's estimated %
			Density δ_m	τ_{pc_m}	τ_{pt_m}	constant	
40	34	85%	-0.0041 (-4.79)	-0.035 (-4.64)	0.064 (9.33)	2.05 (6.62)	17.67%
50	31	84%	-0.0039 (-4.22)	-0.038 (-4.44)	0.065 (8.47)	2.16 (6.15)	17.50%
60	30	84%	-0.0041 (-4.38)	-0.038 (-4.40)	0.063 (8.02)	2.19 (6.28)	17.63%
70	25	87%	-0.0039 (-4.58)	-0.037 (-3.99)	0.065 (8.30)	2.04 (6.13)	18.51%
80	23	88%	-0.0045 (-4.73)	-0.031 (-3.04)	0.062 (7.15)	2.00 (5.86)	18.14%
90	23	87%	-0.0045 (-4.56)	-0.031 (-2.95)	0.063 (7.04)	1.97 (5.60)	18.19%
100	23	87%	-0.0044 (-4.41)	-0.031 (-2.88)	0.064 (6.93)	1.95 (5.38)	18.24%

(source: LET based on the Lyon household travel survey)

3.4.2. A calibration method that retains all the data and small-scale zoning

As in the case of the first method, we have tested the effect of the selected threshold for the number of trips. However, this effect is radically different from with the previous method as we are no longer eliminating part of the available information but retaining it all and creating groups of origin-destination pairs with sufficiently large numbers of surveyed trips for the estimation of the market share for each mode to be sufficiently reliable. In the same way, we have attempted to test the effect of the choice of zoning. The issue here was to know whether the higher degree of uncertainty affecting the measurement of the explanatory variables included in the utility functions can have an influence on the estimation of the coefficients. The results are set out in Table 8.

Table 8: Estimation of coefficients with the calibration method that retains all the data and small-scale zoning

	Threshold 70		Threshold 80		Threshold 90		Threshold 120		Threshold 150		Threshold 200	
	coefficient	σ	Coefficient	σ	Coefficient	σ	Coefficient	σ	Coefficient	σ	Coefficient	σ
D196												
constant	3.896	0.510	3.793	0.587	3.783	0.675	3.659	0.774	3.599	0.993	3.256	1.189
τ_{pt_m}	0.013	0.009	0.016	0.014	0.017	0.015	0.020	0.020	0.022	0.023	0.022	0.030
τ_{pc_m}	-0.040	0.012	-0.039	0.013	-0.040	0.014	-0.039	0.016	-0.038	0.022	-0.030	0.030
Density δ_m	-0.0053	0.0018	-0.0054	0.0020	-0.0054	0.0020	-0.0055	0.0024	-0.0059	0.0029	-0.0060	0.0028
R ²	62.2%	4.7%	65.7%	4.8%	67.8%	5.2%	71.4%	5.7%	75.1%	6.3%	78.0%	6.8%
number of classes	112		99		89		66		53		40	
PT's estimated %	17.0%		17.3%		17.5%		18.0%		18.5%		18.5%	
D87												
constant	3.698	0.540	3.702	0.553	3.709	0.712	3.657	0.795	3.636	0.944	3.579	1.004
τ_{pt_m}	0.015	0.011	0.015	0.010	0.017	0.014	0.019	0.016	0.020	0.018	0.021	0.022
τ_{pc_m}	-0.037	0.006	-0.038	0.008	-0.040	0.010	-0.040	0.013	-0.041	0.017	-0.040	0.020
Density δ_m	-0.0049	0.0008	-0.0048	0.0009	-0.0048	0.0012	-0.0048	0.0015	-0.0051	0.0020	-0.0052	0.0024
R ²	59.2%	3.2%	61.2%	3.3%	63.0%	3.7%	67.1%	4.5%	71.9%	4.5%	75.1%	5.3%
number of classes	108		95		87		66		52		40	
PT's estimated %	16.2%		16.3%		16.6%		17.1%		17.6%		18.0%	
D25												
constant	3.125	0.231	3.097	0.227	3.128	0.208	3.236	0.239	3.185	0.314	3.225	0.306
τ_{pt_m}	0.033	0.005	0.037	0.005	0.037	0.005	0.038	0.005	0.034	0.005	0.033	0.004
τ_{pc_m}	-0.037	0.004	-0.040	0.005	-0.041	0.004	-0.045	0.006	-0.042	0.006	-0.044	0.005
Density δ_m	-0.0047	0.0003	-0.0044	0.0003	-0.0043	0.0003	-0.0042	0.0003	-0.0041	0.0003	-0.0039	0.0004
R ²	66.5%	4.3%	70.0%	3.5%	72.3%	3.9%	75.0%	3.6%	80.1%	3.5%	83.7%	2.6%
number of classes	83		76		68		53		44		36	
PT's estimated %	16.5%		16.6%		16.8%		17.0%		17.0%		17.2%	
D7												
constant	2.388	0.299	2.337	0.239	2.245	0.294	2.096	0.000	2.419	0.035	2.198	0.162
τ_{pt_m}	0.059	0.004	0.057	0.004	0.056	0.004	0.057	0.000	0.061	0.002	0.052	0.004
τ_{pc_m}	-0.040	0.006	-0.037	0.005	-0.033	0.004	-0.031	0.000	-0.041	0.002	-0.030	0.004
Density δ_m	-0.0041	0.0002	-0.0043	0.0002	-0.0045	0.0002	-0.0045	0.0000	-0.0042	0.0002	-0.0047	0.0001
R ²	85.5%	1.3%	86.1%	2.3%	86.6%	2.1%	84.9%	0.0%	82.2%	1.0%	91.8%	1.3%
number of classes	30		29		27		26		23		21	
PT's estimated %	17.4%		17.4%		17.5%		17.9%		17.3%		17.8%	

Before commenting on the results, a few explanations about this table are necessary. An iterative process is employed in order to estimate the coefficients (see Section 2). We therefore need to analyze the convergence of this process. We have conducted a large number of tests which show that good convergence is achieved after the first iterations, but after a few hundred iterations we arrive at an oscillating result with no convergence whatsoever (We come back on this question later in the paper). In order to avoid the effect which the number of iterations has on the results we have estimated each coefficient by averaging the results of the last iterations. The data in Table 8 was obtained after 2,500 iterations with the coefficient computed for the last 300 iterations. Furthermore, in order to analyze convergence the table shows, next to the value of the coefficient, the standard deviation for the estimation of the coefficient for the last 300 iterations. The same calculation was performed for the R^2 value. We have not produced Student's t values for the coefficients, but these are all sufficiently high for each of the studied configurations.

Whatever the division and threshold that are selected, the estimated market share of public transport is close to that observed, even though it is systematically slightly lower for the lowest thresholds. The second positive factor is that for a given division the value of the coefficients change little as the threshold is varied. Estimation of the coefficients is therefore only slightly sensitive to the selected threshold, as long as this is large enough (for the lowest thresholds the stability is lower, but so is the accuracy of the market share estimation). However, differences are apparent when different divisions are selected. Although the results for the two smallest-scale divisions are very close, this is not the case for the other zonings, in particular the 7-zone division. This may be due to the uncertainty generated by the size of the zones. However, the data does not allow us to reach a conclusion on this point. It is nevertheless the case that the aim of this method is to retain the smallest-scale zoning possible. This method is therefore of much less value when it is used with large-scale zoning.

For the 7-zone division, the results for this method are very close to those obtained with the "classical" calibration method, as the number of trips in many matrix cells is greater than the threshold value. The regression classes are therefore fairly similar for the two methods.

The last criterion we have used for diagnosis is the standard deviations of the coefficients that are estimated in the course of the iterative process. These provide information about the method's convergence. Logically, these standard deviations will be higher the smaller the zones. However, we observe that the values for the 87 and 196-zone divisions are almost of the same order of magnitude as the coefficients. This result is rather disappointing as it indicates that the convergence of the iterative process is poor. Therefore, we have tried to analyse this convergence repeating a great number of iterations. We can observe a very rapid average convergence of the coefficients. This convergence is more rapid as the number of zones or the threshold are smaller. We furnish the results for the 196-zone division with a threshold of 150 observed trips (Figure 3). After 10 iterations, the evolutions are very small (Figure 4). Similar results are obtained for the 3 variables and the constant.

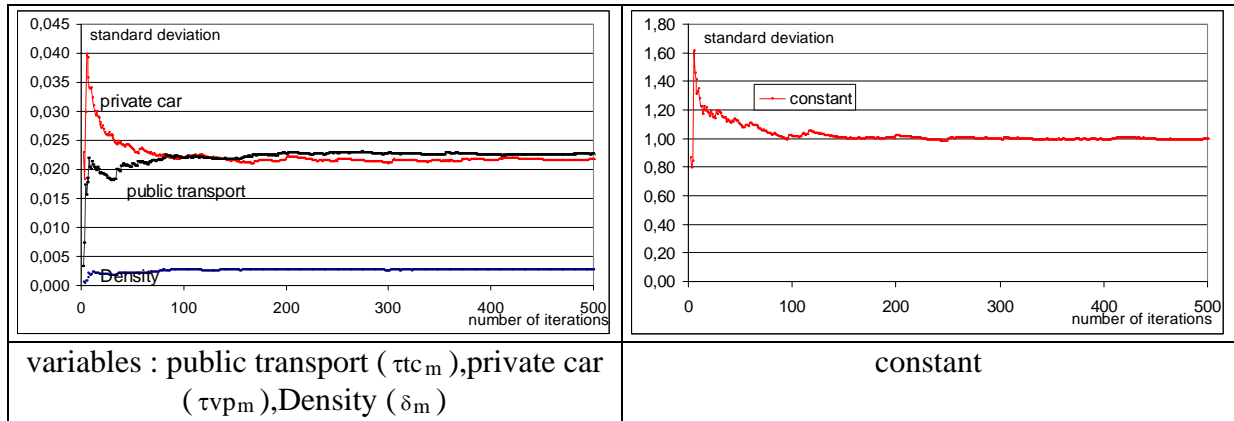


Figure 3: Evolution of the standard deviation on all iterations, 196-zone division, threshold 150 trips

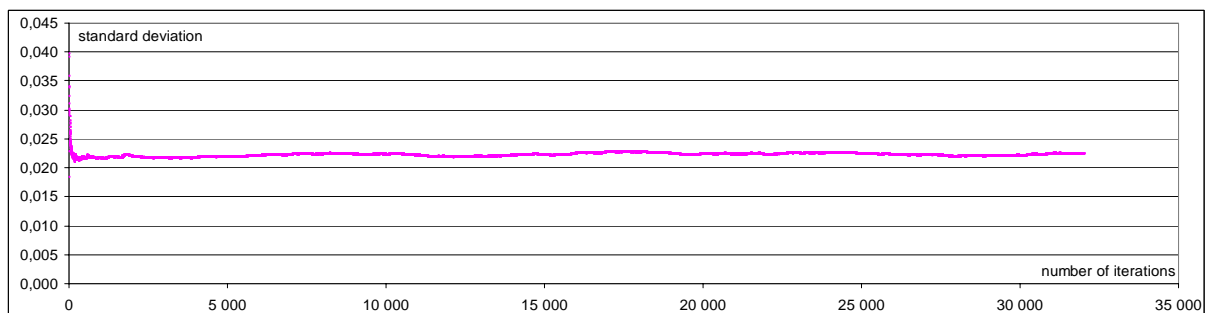


Figure 4: Evolution of the standard deviation on the last 5.000 iterations, variable private car (τ_{vp_m}), 196-zone division, threshold 150 trips

From these results, we have tried to look for periodicity in the results. A great number of iterations are necessary, especially for the 196-zone division. As computation calculation time are quite long (about 20 seconds per iteration for the 196-zone division), the analysis has been limited to the two most significant thresholds (120 and 150 observed trips) in terms of simulated public transport share). For each configuration, we can observe a periodicity (table 9, the table can be read as follow: for the 196-zone division and 120 trips threshold, from iteration 27.188 we have cyclic results, i.e. iteration 37.680 (27.188 + 10.492) have the same results that iteration 27.188 and so on). The identification of the cycle is interesting as it confirms that we can only observe an average convergence and not an absolute convergence. It indicates also the number of iterations on which it's possible to calculate the average coefficients of the utility function which are independent of the number of iterations. In fact results are very closed when obtained from the last 300 iterations (as for table 8) or obtained from the number of iterations included in the cycle (10.492 for example for the 196-zone division and the 120 trips threshold). In almost all cases, differences are less than 1%. Therefore it's probably better to estimate the coefficient on the greater number of iterations than in table 8 to obtain results which are independent of the number of iterations which are taken into account (Nevertheless we should keep in mind that this precision is illusory as this variation is much less than the uncertainty on coefficients estimation. Therefore this variation has little incidence on final results).

Table 9: Results periodicity in the coefficient estimation iterative process

87-zone division	Iteration number for which begin results periodicity	Number of iterations within the cycle
Threshold 70	2.491	1.303
Threshold 80	2.219	5.448
Threshold 90	1.706	3.024
Threshold 120	8.863	5.297
Threshold 150	1.657	4.496
Threshold 200	330	5.515
196-zone division		
Threshold 120	27.188	10.492
Threshold 150	22.983	17.187

We have carried out the same analysis for the other trip purposes. This led to very similar conclusions to those which we reached for the work trips (Ferey, 2002). We shall therefore now state the principal lessons we have drawn from this research in the conclusion.

4. Conclusion

The methods that are usually used to estimate an aggregate modal choice model must comply with two contradictory objectives. On the one hand, they must be based on the smallest-scale zoning possible in order to produce accurate data for the explanatory variables that are included in the utility functions. This is because the estimation of generalized times is more reliable and homogeneous for a given origin-destination pair if the origin and destination zones are quite small. In addition, the number of surveyed trips should be sufficient for each origin-destination pair in order for the market share observed for the origin-destination pair to be sufficiently accurate. In practical terms, in view of the number of trips generally available, this implies a high degree of zonal aggregation.

This contradiction is responsible for a number of problems that affect the estimation of the coefficients:

- the selected threshold for the number of surveyed trips must be as high as possible to guarantee an acceptable level of precision for the market share matrix for each mode. This leads to the elimination of all the origin-destination pairs on which the number of trips is lower than this threshold. If zoning is conducted on an excessively small scale, a large proportion of the available information is therefore eliminated (Tables 2 to 5);
- the division must contain as many zones as possible to guarantee an acceptable level of precision for the measurement of the explanatory variables at origin-destination pair level;
- the selected number of regression classes must be as high as possible in order to increase the number of degrees of freedom in the regression and very probably the initial variance of the data set. This can be achieved in two ways; the threshold of surveyed trips can be reduced, but this also reduced the accuracy of the observed market share matrix, or alternatively zonal aggregation can be performed, but this means that the explanatory variables are measured less accurately;
- the number of trips made for the selected origin-destination pairs must be as high as possible in order to make the best use of the available data. But this means we have to face the same contradiction as above.

Our application to the Lyon conurbation has revealed the consequences of this contradiction (see Tables 2 to 7):

- the number of origin-destination pairs for which the surveyed number of trips is sufficient is far too small for the 87 and 196-zone divisions. The number of trips is still small for the 25-zone division;
- the calibration coefficients are highly dependent on the selected zoning. They are also highly dependent on the selected threshold for the number of surveyed trips, except in the case of the most aggregated 7-zone division.

These results lead us to recommend the use of highly aggregated zoning when the “classical” modal choice calibration method is employed. However, the reliability of the computed values for the explanatory variables is open to question in the case of the 7-zone division. In particular, does the generalized time between two zones in this division accurately represent the diversity of the surveyed situations? It was with a view to dealing with this problem that we have proposed a new calibration method which allows us both to retain the totality of the available information from the surveys and also to work with the smallest-scale zoning (the scale of zoning is nevertheless dependent on the reliability of the zoning data, which is, however, considerably more reliable than origin-destination data). The application of this method has led to the following conclusions:

- for the two divisions with the largest number of zones, the value of the coefficients is little affected by the number of surveyed trips threshold that is adopted for each origin-destination class. In addition, the estimated market share is close to that which is observed. However, in the case of the larger-scale zonings, the results diverge;
- convergence of the method can only be observed on an average. It is therefore necessary to estimate the calibration coefficients by averaging the last iterations. However, convergence on an average is achieved very rapid. A few hundred iterations are generally sufficient;
- although the convergence on an average is good, there is a high degree of variability between successive iterations.

These findings do not allow us to conclude with absolute certainty that our calibration method is superior from the empirical point of view. We have therefore analyzed the results at a more disaggregated level using the matrix obtained with the 7-zone division. Once again, we observed that the quality of the results with the new calibration method for the 87-zone and 196-zone divisions is similar to that of the 7-zone division with the more classical method. We therefore feel that additional investigation for different situations would be necessary for us to be able to state categorically that our calibration method is superior. However, from empirical point of view, we can identify the following important advantages

- coefficient estimation is not depending on the zone division system when the zone division is enough disaggregated;
- coefficient estimation is not depending on the threshold choice for the number of observed trips necessary for each origin-destination.

Furthermore, from the theoretical point of view its advantages are obvious:

- it provides the possibility of retaining much smaller-scale zoning which gives much more accurate data for explanatory variables;
- all the information which is available from travel surveys can be retained to calibrate the model.

Nevertheless the main limit of the method is that it's not possible to produce statistical test on the coefficient estimation to identify the quality of the estimation. We can only calculate an average R^2 , and control that t student are acceptable for the coefficient estimation of each iteration. But we think that this limit is overcompensated by empirical advantages we have listed above. The most important advantage is probably the stability of results over zone divisions and thresholds of number of observed trips.

Acknowledgement

The author would like to thank Jean-Baptiste Ferey for use of the test of calibration methods he performed while working on a post-graduate dissertation (Ferey, 2002).

References

Abraham, C., Coquand, R., 1961. La répartition du trafic entre itinéraires concurrents, réflexions sur le comportement des usagers – application au calcul des péages, *Revue Générale des Routes et Aérodrômes*, N°357, pp. 57-76.

Ben-Akiva, M., Lerman, S.R., 1985. *Discrete choice analysis, theory and application to travel demand*, MIT Press, Cambridge, Massachusetts, United states, 4th Edition, 390p.

Bhat, C., 1997. Recent methodological advances relevant to activity and travel behavior analysis, in *Recent developments in travel behavior research*, Pergamon, Oxford.

Bhat, C., 2000. Flexible model structures for discrete choice analysis, in *Handbook of transport modelling*, D A Hensher, K J Button (eds.), Pergamon, Elsevier, pp. 71-90.

Bonnel, P., 2002. *Prévision de la demande de transport*, Habilitation à diriger les recherches, Université Lumière Lyon 2, Lyon, 409p.

Bonnel, P., Bechar, M., Julien, H., Odent, P., Saint-Marc, L., 2002. *Didacticiel de formation à la modélisation transport*, produit sous CD-Rom, Laboratoire d'Economie des Transports, ENTPE, IMTRANS, MVA pour le compte de l'ADEME, Lyon.

Bonnel, P., Cabanne, I., 2000. 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, *Proceedings of the European Transport Conference*, Cambridge (UK), 11-13 September 2000, Seminar K, pp. 205-226.

Bouyaux, P., 1988. Modélisation désagrégée des transports urbains : une application à la ville de Rennes, *Revue d'Economie Régionale et Urbaine*, N°5, pp. 783-809.

CERTU, 1998a. *L'enquête ménages déplacements méthode standard*, Collections du CERTU, éditions du CERTU, Lyon, 295p.

CERTU, 1998b. *Comportements de déplacement en milieu urbain: les modèles de choix discrets, vers une approche désagrégée et multimodale*, Dossier du CERTU mobilité-Transport, éditions du CERTU, Lyon, 133p.

CETE de Lyon, CETE de l'Ouest, CETUR, MELATT, 1986. *Modèles désagrégés, principes généraux, méthodologie, applications* Grenoble, Nantes. *Journées de rencontre sur les modèles désagrégés*, Paris, 10-11 juin 1986, 76p.

CETE de Lyon, INSEE, SYTRAL, 1995. Enquête déplacements auprès des ménages de l'agglomération lyonnaise, document technique, Lyon, 119p.

CETUR, 1985. Les déplacements domicile-travail et domicile-école, modèles désagrégés de choix modal, application au cas de l'agglomération grenobloise, Bagneux, 71p.

Daly, A., 1985. A study of transferability of disaggregate mode choice models from Grenoble to Nantes, Cambridge Systematics, 19p.

Domencich, T.A., McFadden, D., 1975. Urban travel demand: a behavioural analysis, North-Holland, Elsevier, Amsterdam.

Ferey, J-B., 2002. Analyse des méthodes d'estimation des modèles de répartition modale, mémoire de DEA, ENTPE, Université Lumière Lyon 2, Lyon, 137p.

Henderson, J.M., Quandt, R.E., 1980. Microeconomic theory a mathematical approach, McGraw-Hill International editions, economics series, 3rd edition, 420p.

Hivert, L., Orfeuil, J-P., Troulay, P., 1988. Modèles désagrégés de choix modal : réflexions méthodologiques autour d'une prévision de trafic, INRETS, rapport N°67, Arcueil, 65p.

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

Lichère, V., Raux, C., 1997b. Développement d'un modèle stratégique de simulation des déplacements, guide de l'utilisateur, SEMALY, LET, Lyon, 33p.

Luce, R., 1959. Individual choice behaviour: a theoretical analysis, Wiley, New York

Luce, R., Suppes, P., 1965. Preference, utility and subjective probability, in Handbook of Mathematical psychology, vol. 3, Luce R, Bush R, Galanter E (eds.), Wiley, New York.

McFadden, D., 2000. Disaggregate behavioral travel demand's RUM side, a 30-year retrospective, in pre-print 9th International Association for Travel Behaviour research Conference, Gold Coast, Queensland, Australia, 2-7 July 2000, vol. 1, 38p.

Ortuzar, J. de D., Willumsen, L.G., 2001. Modelling transport, 3rd edition, John Wiley & Sons, UK.

PTV-Isis, 2001. Manuel de l'utilisateur Davisum (30/07/01), version 7.0, PTV-Isis, Lyon, France.

RATP, Cambridge Systematics, 1982. Etudes de politiques de transports en région Ile-de-France, mise au point et utilisation de modèles désagrégés de choix modal, Paris, 91p.

Richardson, A.J., Ampt, E.S., Meyburg, A.H., 1995. Survey methods for transport planning, Eucalyptus press, University of Melbourne, Australia, 459p.

Rousseau, J., Saut, C., 1997. Un outil de simulation de politiques de transport : impact 3, Revue générale des chemins de fer, pp. 77-83.

SEMALY, 2000. TERESE, modèle d'affectation de voyageurs dans les études de transport collectif, document pédagogique, cours analyse et prévision de la demande de transport du DESS Transports Urbains et Régionaux de Personnes, ENTPE, Université Lumière Lyon 2, Lyon.