



# Comportements de déplacement en milieu urbain : les modèles de choix discrets

Vers une approche désagrégée  
et multimodale



# **Comportements de déplacement en milieu urbain : les modèles de choix discrets**

**Vers une approche  
désagrégée et multimodale**

**Juin 1998**

Centre d'études sur les réseaux, les transports, l'urbanisme et  
les constructions publiques

Agence de l'environnement et de la maîtrise de l'énergie

Comité de suivi du cas d'application Lyonnais :

M Clément, Peyrton et Klaerr Blanchard (Certu); M. Bresse (Ademe); M Matossi et Saliou (Cete de Lyon); M. Gamon (Grand Lyon); M. Marty (SLTC); M. Raux (LET); M. Faivre d'Arcier (INRETS), M. Jarrige (DRE Rhône Alpes), M. Saraillon (DDE du Rhône). M. Garnier (Sytral), M. Rince (DTT)

Comité technique du cas d'application lyonnais :

M Clément et Peyrton (Certu), M. Clarke et Ducastelle (société MVA Consultants), M. Raux (Let), M. Saliou (Cete de Lyon).

Merci, pour leurs remarques à :

M. Kunkel (Certu), M. Klaerr Blanchard (Certu), M. Nalin (DTT), M. Noël (Certu), M.Ivaldi (IEI, Université des Sciences Sociales de Toulouse).

Rapport rédigé par :

Première partie : Lionel Clément (Certu) et David Peyrton (Certu/CNRS)

Deuxième partie : Michael Clarke, Eric Tardivel, Mathurin Ducastelle (MVA Consultants), sauf exemples de compréhension (Certu)

Coordination générale : Lionel Clément.

# Sommaire

<b>Introduction générale</b>	<b>6</b>
 <b>Première partie</b>	
<b>Approche théorique et méthodologique</b>	<b>7</b>
Introduction.	9
<b>Chapitre 1 : théorie des modèles de choix discret</b>	<b>11</b>
1. Approche désagrégée et lois de comportement	11
2. Procédure de choix et théorie micro-économique	21
3. Modélisation de la procédure de choix	23
<b>Chapitre 2 : méthodologie des modèles de choix discret</b>	<b>33</b>
1. La spécification : recherche de la forme fonctionnelle et de la forme mathématique des variables	33
2. Estimation d'un modèle de choix discret	36
3. Appréciation de la qualité d'un modèle de choix discret	39
4. Sorties obtenues à l'aide d'un modèle de choix discret et méthode de validation	43
5. Agrégation	44
6. Limites des modèles de choix discrets	44
Conclusion	47
 <b>Deuxième partie</b>	
<b>Approche pratique</b>	<b>49</b>
Introduction	50
1. Contexte et objectifs de l'étude lyonnaise	51
2. La chaîne de modélisation	52
3. Les données et leur traitement	58
4. Les résultats des modèles	71
5. Validation et scénarios	100
Conclusion	120
<b>Conclusion générale</b>	<b>123</b>
<b>Annexes</b>	<b>125</b>
Annexe 1 : démonstration de la formule probit	125
Annexe 2 : démonstration de la formule logit	129
<b>Bibliographie</b>	<b>131</b>



## Avertissement

Le travail présenté dans ce rapport a pour objet de répondre à deux préoccupations du ministère de l'Équipement, des Transports et du Logement, et plus particulièrement de son centre d'études; le Certu :

- Impulser un mouvement pour la diffusion de la modélisation désagrégée et multimodale des déplacements en France. Très utilisée dans les pays anglo-saxon et nordiques, elle a l'avantage d'être fiable car basée sur des techniques définies et pédagogique au niveau de la présentation des résultats.

- Participer au renouvellement des méthodes, en testant une chaîne de modélisation qui soit basée sur la structure séquentielle à quatre étapes, utilisée dans sa dimension agrégée en France.

Ce renouvellement nécessaire se rattache à deux constats : le retard pris en France dans le domaine de la modélisation, et le développement des dossiers de planification (type Plan de Déplacements Urbains ou Dossiers de Voirie d'Agglomération), pour lesquels les outils actuellement utilisés ne sont pas toujours adaptés.

# Introduction générale

Ce document a pour objectif de «faire le point» sur les modèles de choix discret. La notion de choix discret correspond à un ensemble d'alternatives en présence. On essaie d'explicitier le choix de l'une d'elle plutôt que d'une autre. Sa valeur ajoutée réside dans ce qu'il peut apporter au lecteur qui désire obtenir un résumé, que l'on espère clair, de tout ce qui concerne les aspects théoriques, méthodologiques et pratiques de ce type de modélisation appliquée aux déplacements en milieu urbain.

En ce qui concerne les aspects théoriques et méthodologiques, une analyse d'une littérature américaine et britannique, qui n'avait pas encore été mise à la portée des professionnels sensibilisés à ce sujet en France, est effectuée. Elle constitue la première partie du rapport.

En ce qui concerne les aspects pratiques, un cas d'application sur l'agglomération de Lyon est présenté. Il constitue la deuxième partie du rapport. Le travail de développement de la base de données et de l'estimation des modèles a été réalisé en 1996 et 1997, avec 1995 comme année de référence. La maîtrise d'ouvrage a été assurée par le Certu, sur financement Certu-Ademe, et la maîtrise d'œuvre effectuée par la société MVA Consultants.

Afin de pouvoir répondre aux attentes différenciées des techniciens, des chargés d'études transports et des planificateurs, la lecture du rapport peut se faire soit en continu, soit en privilégiant une partie plutôt qu'une autre.

Première partie  
**Approche théorique  
et méthodologique**





# Introduction

Deux approches de prévision de la demande des déplacements en milieu urbain, existent :

- l'approche agrégée, dans laquelle on raisonne par rapport à un individu moyen localisé dans une zone  $i$ ,
- l'approche désagrégée, dans laquelle on raisonne par rapport à un individu  $t$  ayant des caractéristiques socio-économiques et de différenciation des offres de transport qui lui sont propres, qui font qu'il est intrinsèquement différent de tout autre individu.

Dans leur usage le plus courant (quasiment 100% des études en milieu urbain en France), les modèles séquentiels traditionnels (génération-distribution-choix modal-affectation), se rapportent à la première approche <sup>2</sup>.

La modélisation désagrégée renvoie à la mesure du poids des variables intervenant dans le processus de décision des individus. Elle permet de comprendre leurs logiques de comportement de déplacement en matière de choix discret simple (mode de transport par exemple) ou combiné (choix de mode et de destination par exemple). La notion de choix discret renvoie donc à un ensemble d'alternatives dénombrables.

Pour ce faire, on associe, à chaque alternative constituant l'ensemble étudié, une fonction d'utilité, qui dépendra essentiellement des caractéristiques des individus et des caractéristiques de l'alternative elle-même. Les résultats sont présentés sous forme de probabilité. Grâce à eux, on obtient le niveau de la demande de déplacements futurs (par mode si le choix concerne les modes), compte tenu des tests de politique de transport qui ont été effectués en faisant varier les composantes des fonctions d'utilité. Les modèles de ce type sont appelés modèles de choix discret, par analogie avec le terme anglo-saxon (Discret Choice Model).

---

2. Pour une caractérisation détaillée des modèles et des dimensions qui les sous-tendent, on consultera avec profit les travaux de F. Leurent. [Leurent (F.), 1997, chapitre 1]. Les travaux sont orientés sur la phase d'affectation.

Le contenu de cette partie s'articule autour d'une présentation logique. Dans un premier temps on rappelle ce que recouvre l'approche désagrégée (I). Une fois la procédure de choix resituée dans la théorie micro-économique (II), le détail des formulations utilisées pour la modéliser est explicité (III). Ces éléments relevant du domaine théorique, d'autres, plus directement liés à la méthode de modélisation sont abordés dans un deuxième chapitre : les problèmes liés au choix de la forme fonctionnelle et à la spécification des variables sont étudiés (I), le problème de l'estimation (II), des méthodes d'appréciation de la qualité d'un modèle de choix discret (III), des sorties obtenues (IV), de l'agrégation (V). Enfin, une dernière partie traite des limites de ces modèles (VI).

## Chapitre 1

# Théorie des modèles de choix discret

### 1. Approche désagrégée et lois de comportement

Après avoir présenté l'objectif sous-tendu par le type de modélisation étudié, nous présenterons le genre de résultats auxquels on aboutit. La méthode à mettre en oeuvre pour les obtenir sera alors présentée, ainsi que la mise au point des scénarios et de la logique de prévision.

#### 1.1. Objectifs des études et modèles de choix discret

Le principal objectif des études de modélisation de choix discrets est l'analyse et la prévision des comportements de déplacement <sup>3</sup>, dans le cadre de la programmation des infrastructures de transport à court et moyen terme (de 5 à 10 ans). Deux utilisations se dégagent cependant. Elles dépendent du niveau de la politique des transports auquel on se situe.

En effet, toute politique de transport se décline en différents niveaux <sup>4</sup>:

— **les options premières** se rapportent aux grandes orientations stratégiques. Elles relèvent de la politique urbaine au sens large et se traduisent sous la forme d'objectifs tels la déviation des flux de transit, le développement de l'usage des transports collectifs...

— **le choix de schémas** correspond à la vision «achevée», et pas toujours formalisée, des systèmes en présence. C'est à ce niveau que sont déterminés «les grands traits» qui s'inscriront dans le futur. On les retrouve dans la «lettre d'intention» des candidats aux élections locales, ou dans les documents d'ordre généraux relatifs à la politique menée (exemple : Schémas Directeurs d'Aménagement et d'Urbanisme).

---

3. Comment un individu se comporte-t-il vis à vis de plusieurs alternatives en présence. Quel choix va-t-il faire si l'on modifie le niveau des variables qui interviennent dans l'élaboration de ce choix ?

4. [Bonnafeous (A.), 1992].

— **le choix de projets**, c'est-à-dire la formalisation du niveau précédent, sous la forme de projets d'investissements et de leur rythme de réalisation. C'est à ce niveau qu'est définie la cohérence des systèmes précédents, en tenant compte de la contrainte financière de la collectivité et de la dynamique interne de développement de l'agglomération. C'est également ici qu'intervient le calcul économique, pour l'évaluation a priori de la rentabilité financière, économique et sociale des projets, qui tient compte de leur adéquation à la dynamique de l'agglomération (exemple : PDU, DVA).

— **le choix de variante** enfin, correspond à la version quasi définitive du ou des projets. Après avoir traduit les options premières (les objectifs), en choix de schémas, puis décidé des projets (les moyens), leur réalisation est effectuée. Celle-ci entraîne alors des études de faisabilité détaillées.

Dans la réalité, ces quatre niveaux sont en interaction, car pas toujours définis de manière très chronologique.

#### **Les études ne concernant que le choix de schéma**

On ne cherche pas forcément à aller à un niveau de détail relativement fin (d'où un niveau de détail spatial grossier, alors que les réseaux ne sont pas décrits). Une seule composante est alors nécessaire : celle concernant les comportements de déplacement. Pour ce faire, il n'est pas nécessaire d'aller jusqu'à la phase d'affectation.

Dans ce cadre, on effectue des tests seulement sur les variables exogènes relatives à l'offre. Prenons l'exemple d'une étude visant à pré-orienter le choix relativement à la réalisation de projets de transports collectifs. Le maître d'ouvrage désire savoir de quelle façon les projets vont modifier le comportement de déplacement des individus. Pour répondre à cette attente, on pourra réaliser une enquête préférence déclarée. Les variables intervenant dans la modélisation seront modifiées dans des proportions équivalentes à celles représentant le futur niveau de service des projets en lice, pour simuler le nouveau comportement de déplacement des individus.

#### **Les études concernant le choix de schéma et le choix de projet (voire le choix de variante)**

On cherche, dans ce cas, à aller à un niveau de détail relativement fin (d'où une description fine des réseaux et un découpage fin de l'aire étudiée).

Dans ce cas, deux composantes sont alors nécessaires : l'une concerne les comportements de déplacements, l'autre les flux de trafic et/ou de voyageurs. Dans ce cadre, on a besoin de résultats sur les volumes (de trafic, de voyageurs). Une modélisation à structure séquentielle

sera donc établie. Les trois premières phases, éventuellement partiellement emboîtées (génération, distribution, choix-modal) se rattachent à l'approche désagrégée de choix individuel, alors que la dernière (affectation) se rattache à l'approche agrégée. Aux deux niveaux du travail on pourra privilégier une approche probabiliste.

Précisons que, dans ce cadre, les variables exogènes relatives à l'offre et à la demande sont modifiées lors de la phase de prévision, puisque des hypothèses d'évolution socio-démographique sont faites pour les zones du découpage de l'aire d'étude, avec l'intégration des projets en lice dans les scénarios testés.

## 1.2. Résultats issus de la modélisation

La modélisation du comportement de déplacement des individus conduit à élaborer un ou des «outils» d'analyse permettant de connaître les variables qui déterminent la procédure de choix étudiée (on modélise donc le processus de décision). C'est en ce sens que l'on parle de «lois de comportement». Ces lois sont établies à partir de fonctions d'utilité qui traduisent l'attractivité de chaque alternative. La mise au point des fonctions d'utilité permet ainsi l'évaluation de l'impact de modification de variables de type suivant sur la demande de transport :

- caractéristiques des individus et/ou des ménages (sexe, âge, profession, position dans le ménage...)
- caractéristiques de l'offre de transport (fréquence, amplitude de service, vitesse...)
- caractéristiques des parties urbaines (surface de centres commerciaux, type d'habitat...)

Voici un exemple de fonctions d'utilité intégrant seulement des caractéristiques concernant l'offre de transport :

Supposons, qu'à l'issue d'une enquête réalisée auprès de 1 200 personnes d'une même profession<sup>5</sup>, et concernant leur comportement modal vis-à-vis de la voiture et du bus on arrive à :

$$\begin{aligned} V_{\text{voiture}} &= 0,5 - 0,25 \text{ tp}_{\text{voiture}} - 0,42 \text{ ta}_{\text{voiture}} - 0,1 \text{ ct}_{\text{voiture}} \\ V_{\text{bus}} &= - 0,28 \text{ tp}_{\text{bus}} - 0,45 \text{ ta}_{\text{bus}} - 0,1 \text{ ct}_{\text{bus}} \end{aligned}$$

5. Pour une autre profession, tous les coefficients peuvent être différents.

Où :

**tp** = temps de parcours dans le mode

**ta** = temps d'accès au mode

**ct** = coût du déplacement

Nous avons là deux fonctions d'utilité concernant la voiture et le bus. Chaque variable représente un attribut de chacune des options pour les individus qui se déplacent. L'influence relative de chaque attribut, en terme de contribution à la satisfaction globale produite par l'alternative, est fournie par les coefficients. Ainsi, pour la voiture, un changement d'une unité dans le temps d'accès au véhicule (ta) aura presque deux fois plus d'impact qu'un changement d'une unité dans le temps parcours (tp) et quatre fois plus qu'un changement d'une unité dans le coût du déplacement (ct). La même analyse peut-être faite pour le bus. Le niveau de la constante (0,5 pour la voiture, il n'y en a pas pour le bus), traduit l'influence de tout ce qui n'est pas observable, et de toutes les autres caractéristiques qui ne sont pas incluses dans la loi de comportement. Du point de vue de l'interprétation, ces résultats montrent qu'une augmentation du temps d'accès à la voiture (aller chercher sa voiture qui est en stationnement) aura un impact fort sur l'attractivité de l'alternative, donc réduira son utilité pour les individus enquêtés. Dès lors, une politique de stationnement contraignante pourra entraîner un certain nombre de reports modaux de la voiture vers le bus et aura beaucoup plus d'impact qu'une politique visant à augmenter le coût d'usage de la voiture.

Du traitement des fonctions d'utilité précédentes, on peut déduire certaines lois de comportement. L'estimation des fonctions d'utilité précédentes permet d'aboutir, pour une situation moyenne constatée dans l'échantillon ou dans la population étudiée ici, à un résultat du type suivant<sup>6</sup>:

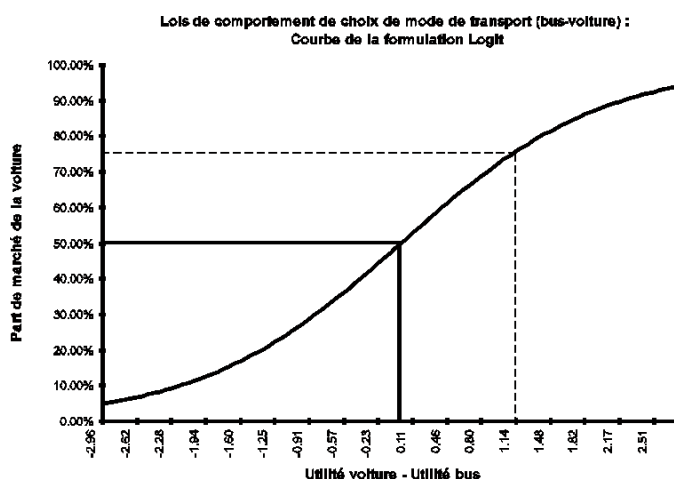
$$P_{\text{voiture}} = 0,76$$

La probabilité qu'un individu de la population enquêtée choisisse la voiture est de 0,76. Ce qui revient à dire que la part de marché de la voiture est de 76% et celle du bus de 24%.

Il s'agit donc bien d'une «loi», dans le sens où il est possible de considérer les différentes valeurs que peuvent prendre les variables, compte tenu de la situation de référence, afin d'obtenir les différentes possibilités de choix. La courbe suivante représente graphiquement cette loi (obtenue à partir d'une formulation logit, explicitée par la suite).

6. On montrera, par la suite, de quelle manière on arrive à un tel résultat.

C'est parce qu'il y a un écart de 1,14 entre l'utilité de la voiture et celle du bus, que la part de marché de la voiture est plus élevée (à une différence égale à zéro correspond une part de marché de 50%)<sup>7</sup>.



Une fois un tel résultat obtenu, la méthode utilisée consiste à le comparer, pour une situation de référence, au résultat obtenu après modification d'une ou plusieurs variables. Des politiques très complexes peuvent donc être testées, ainsi que des modifications d'ordre socio-économiques concernant la structure de la population, si celles-ci font partie des variables explicatives. Ainsi, si l'on désire tester l'impact d'une politique de stationnement contraignante, en sachant que la suppression d'une partie du stationnement entraînera un accroissement d'environ 20% du temps moyen d'accès au véhicule, la part de marché de la voiture sera, toutes choses égales par ailleurs.

$$P'_{\text{voiture}} = 0,71$$

Une telle mesure devrait entraîner un certain nombre de reports modaux, de l'ordre de 5%, de la voiture vers les transports collectifs,

7. On notera qu'une telle courbe est universelle, dans le sens où elle a toujours cette forme, avec un point d'inflexion marquant une rupture plus ou moins forte. La distinction entre des modélisations différentes s'appréciera en terme de sensibilité (positionnement sur la courbe du niveau de part de marché, plus ou moins important en fonction des différences d'utilité). En général, la courbe précédente n'est pas présentée. On préfère lui substituer des courbes réalisées en ayant fixé une variable à un certain niveau.



des individus changeant leur comportement du fait qu'il devient moins attractif pour eux de prendre leur véhicule. Ce résultat, sous forme de probabilité, peut-être transformé en terme de volumes; il suffit de multiplier par le nombre de personnes enquêtées, ou par la population de la zone d'enquête avec application d'un coefficient de redressement. Cependant, ce qui paraît être un avantage - passer de proportions à des valeurs absolues - peut être un inconvénient. En effet, les résultats auxquels on aboutit doivent être considérés sous forme «d'indicateurs», comme des résultats issus de tests de sensibilités sur le niveau des variables. Une transformation en valeur absolue devra donc être effectuée avec précaution.

### 1.3. Méthode de modélisation<sup>8</sup>

Après entretien avec Michaël Clarke de la société MVA Consultants, nous avons cherché à résumer la méthode de développement d'un modèle de choix discret. Elle passe par cinq phases, qui sont en partie valables pour les autres types de modélisations. Seule la phase de validation, telle que décrite ici, concerne uniquement les modèles de choix discret.

#### ➤ a. La préparation

Elle concerne la mise au point du fichier des données qui permettront de reconstituer le choix des individus dans la phase d'estimation. On travaille les données en prenant en compte l'ensemble des éléments qui décrivent le choix des individus. Les sources sont très diverses : enquêtes, calculs par logiciels et/ou par SIG c'est-à-dire systèmes d'information géographique (temps d'accès aux modes de transport par exemple). On décrit aussi l'ensemble des zones (surface, équipement urbains...). Enfin, on reconstitue l'ensemble des alternatives, autres que celles pour lesquelles on possède déjà des renseignements, par exemple : pour le choix de mode, on tient compte de la disponibilité, pour les individus, d'une voiture particulière ou non.

#### ➤ b. La spécification

Cette phase concerne la recherche de la meilleure spécification des fonctions d'utilité soit la sélection et la combinaison des variables

---

8. Cf les travaux de F. Leurent pour plus de précisions sur la définition des terminologies, ici adaptées en référence à tout type de modélisation. [Leurent (F.), op. cit.]

explicatives, puis la recherche de la forme mathématique la plus adaptée pour chaque variable. C'est à ce moment qu'apparaît la notion de coefficient. Ceux-ci, une fois estimés, donneront un «poids» différent aux variables prises en compte pour l'explication des comportements (cf. phase d'estimation). En effet, lors de cette phase, on réfléchit sur les variables à considérer, mais aussi sur les coefficients qui vont les pondérer. Pour ce faire, il convient de bien «dégrossir» l'ensemble des possibles. Par exemple, le nombre de motifs et de modes pouvant être important, il risque d'y avoir un grand nombre de modèles (correspondant au nombre de motifs x nombre de modes). Dans certains cas, les données ne permettront pas d'obtenir de lois de comportement; par exemple dans le cas de l'utilisation du vélo pour le motif domicile-achat.

### ► c. L'estimation

Cette phase concerne l'ajustement du modèle, c'est-à-dire la recherche des coefficients «les meilleurs», par une procédure d'optimisation. La phase 2 et la phase 3 renvoient à ce qui est dénommé le «calage», ou «calibrage».

### ► d. La validation

Trois types de validation doivent être distingués.

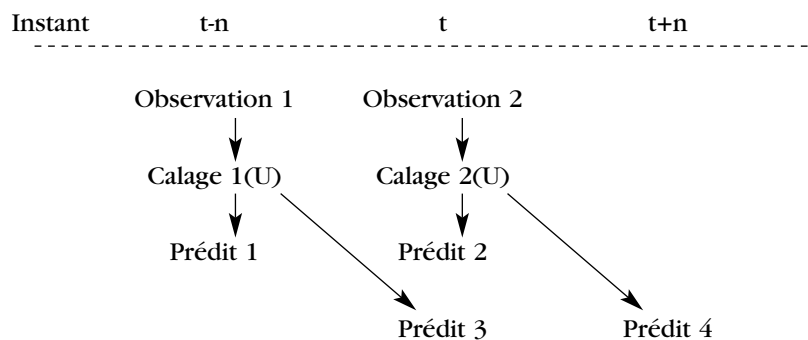
**Validation «désagrégée»:** elle est réalisée à l'instant de référence (situation de référence reconstituée). Les coefficients estimés sont appliqués sur un sous-échantillon de l'enquête de référence, en tenant compte de variables exogènes aux spécifications des modèles. Ensuite, on regarde ce que l'on peut observer par rapport au choix effectué par les individus constituant ce sous-échantillon. Si l'on constate un biais, par exemple en appliquant les modèles à des segments de population, classifiés par motorisation (si cette variable n'apparaît pas dans les modèles estimés), on essaiera d'incorporer cette variable dans leur spécification pour augmenter leur qualité. Par la suite, la prévision sera faite pour ces sous-échantillons, à partir de méthodes statistiques robustes.

**Validation «agrégée»:** elle est généralement faite à l'instant de référence. Le modèle est appliqué à toute la population d'enquête, pour l'aire de référence entière. On cherche à déterminer les changements intervenant entre une situation de référence et une situation pour laquelle les données d'offre et/ou les données socio-économiques ont été modifiées. Ces changements sont ensuite appliqués sur les matrices de référence (VP et TC) tenant compte des comptages disponibles sur les réseaux. Ainsi, la validation dite «agrégée» ne pourra

concerner que les résultats d'une application du système de prévision permettant de mesurer et d'analyser des changements résultant de tests de scénarios et de voir quelle est leur «fiabilité». La «fiabilité» correspond au fait que les modifications apportées sur la situation de référence sont telles, qu'elles permettent de vérifier que les élasticités trouvées sont conformes à celles issues d'analyses socio-économiques.

Les deux validations précédentes correspondent à deux types d'application des modèles : si les modèles sont considérés comme valides, on pourra soit les appliquer sur l'échantillon d'estimation (application désagrégée) , soit les appliquer au niveau zonal (application agrégée).

**Validation générale de la procédure de modélisation** : elle est réalisée en projection. Quel que soit le modèle, on arrive presque toujours à reconstituer une situation de référence. Dès lors, il convient de chercher à vérifier l'hypothèse de prévision suivante : y-a-t-il stabilité du modèle et des coefficients dans le temps? Pour vérifier cette hypothèse, il est nécessaire de réunir les données à deux dates différentes, ce qui pose des problèmes de mesurabilité qui peuvent être redoutables. On supprime ainsi les hypothèses faites par le modélisateur dans la phase de prévision, ce qui permet d'isoler le comportement du modèle.



où  $U$  représente les formes des fonctions d'utilité associées à chaque alternative.

L'utilisation d'un modèle en prévision implique la prise comme hypothèse d'une **stabilité dans le temps** du modèle (dans sa forme mathématique) et des coefficients (dans leurs valeurs).

À partir des résultats du calage à l'instant  $t-n$  pour les formes des fonctions d'utilité  $U$  (calage 1( $U$ )), une comparaison des résultats,

à l'instant  $t$ , de projection Predit 3 et de l'observation 2 permet de vérifier l'hypothèse de la prévision (stabilité du modèle et des coefficients). S'il y a une trop grande différence, l'hypothèse ne se vérifie pas. Il n'y a donc pas stabilité du modèle et/ou des coefficients.

Deux cas sont alors possibles :

- dans le premier, il y a stabilité du modèle, mais pas des coefficients,
- dans le second, il n'y a pas stabilité du modèle (donc a fortiori, pas stabilité des coefficients).

Pour savoir dans quel cas on se trouve, il faut vérifier si la spécification du modèle (dans sa forme),  $U$ , qui se caractérise par le choix des variables et des relations entre elles, est bonne. On refait alors un calage, calage 2( $U$ ), à partir de la même spécification des fonctions d'utilité,  $U$ , et de l'observation 2. On obtient ainsi de nouveaux coefficients à partir du même modèle. La comparaison entre le prédit 2 et l'observation 2 à l'instant  $t$  donne le taux de reconstitution à cet instant. Si ce taux de reconstitution de l'instant  $t$  est de bonne qualité, alors seuls les coefficients ne sont pas stables (cas 1). Dans le cas contraire, c'est le modèle lui-même qui est à remettre en cause (cas 2).

Dès lors, dans les deux cas, le modèle ne peut pas être raisonnablement utilisé en projection. Dans le second cas, tout le modèle est à revoir. Par contre, dans le premier cas, avant de tout revoir, on peut essayer de modéliser l'évolution des coefficients. À partir d'une série chronologique de données, et en considérant les coefficients comme étant variables, on modélise cette évolution.

Prenons l'exemple où, dans une fonction d'utilité linéaire, un des coefficients change entre deux calages à deux instants différents.

Soit  $U$  la fonction d'utilité concernée :  $U = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2$

Supposons maintenant que deux calages à deux instants différents ont donné deux valeurs différentes pour  $a_1$ . Nous pouvons alors essayer de modéliser cette évolution. On obtient :

$$U = a_0 + (b_0 + b_1 \cdot X_{j+1} + b_2 \cdot X_{j+2}) \cdot X_1 + a_2 \cdot X_2$$

Avec cette nouvelle forme de fonction d'utilité, on recommence la procédure de validation.

Cette forme de validation est très séduisante et nous semble fondamentale car elle est à même de rendre compte de la fiabilité du modèle. Un manque dans les données à deux dates de référence contraint généralement son application. Aucun exemple de ce type n'a été recensé à ce jour, certainement car la fonction d'utilité précédente n'est pas linéaire, donc difficile à caler et à interpréter.

### ► e. La prévision

C'est la dernière phase, celle où l'on effectue la projection pour tester des scénarios (cf. ci-après).

## 1.4. Scénarios et horizon de prévision

Une fois que l'ensemble des données nécessaires pour effectuer la modélisation sont réunies, une analyse de celles-ci est effectuée dans la phase de préparation. En fonction du type de scénarios que l'on désire tester<sup>9</sup>, le choix des motifs et des modes va alors être réalisé. Suivra la recherche des périodes horaires à prendre en considération, et surtout la manière dont on les regroupe. Il n'y a donc pas de modèles «types», sauf si, compte tenu des croisements retenus dans la phase de préparation entre les motifs, les modes et les périodes horaires, on pense couvrir un maximum de possibilités (on travaille alors quasiment à «l'exhaustif»). Notons, enfin, que l'orientation que l'on cherche à donner à la réalisation de la modélisation (caractère opérationnel ou test de méthodologie) aura un impact sur les modèles estimés par la suite. Par exemple, la prise en compte des variables «tarif des transports collectifs» et «coût d'usage de la voiture» pourront être omises, pour des raisons de coût dans le cadre d'une étude à caractère méthodologique et pédagogique, alors qu'elles devront retenir l'attention dans le cadre d'une étude à caractère opérationnel. En ce qui concerne l'horizon de prévision, on préférera privilégier le court ou moyen terme (5 à 10 ans) à un horizon de long terme. Comme l'on cherche à obtenir des «lois de comportement» au niveau de l'individu, basées sur des observations passées, ces modèles ont une grande sensibilité à travers le temps : les comportements ne sont donc pas stables, surtout à long terme.

Les modèles de choix discret permettant de définir des lois de comportement en matière de choix discret, il convient, maintenant, de préciser ce que recouvre la notion de procédure de choix, dans la théorie micro-économique.

---

9. Politique de croissance de l'offre en transport collectif, développement du périurbain, scénarios socio-démographiques...

## 2. Procédure de choix et théorie micro-économique<sup>10</sup>

Tout choix est réalisé à partir d'un ensemble non vide d'alternatives. L'environnement de l'individu définit l'ensemble universel d'alternatives. Prenons, à ce niveau, un ensemble universel discret. Parmi cet ensemble, l'individu ne considère qu'un sous-ensemble appelé ensemble de choix. Il contient les alternatives qu'il peut réaliser. La réalisation d'une alternative sera définie par diverses contraintes de temps, d'argent, d'information... L'attractivité d'une alternative, elle, est évaluée en fonction d'un vecteur de valeurs d'attributs. Quand les alternatives sont homogènes (exemple : biens tangibles), une alternative est juste un vecteur de quantités des produits. Cependant, lorsque les alternatives sont hétérogènes, ce qui est le cas du transport, et lorsque les individus peuvent avoir différents ensembles de choix, il est plus naturel de caractériser chaque alternative par ses attributs plutôt que seulement par les quantités qui lui sont associées.

Un choix parmi un ensemble de décisions nécessitant une règle de décision, on va supposer la commensurabilité des attributs. A partir de cela, on peut définir une fonction exprimant l'attraction d'une alternative en fonction de ses attributs. Cela signifie que l'attractivité d'une alternative (exprimée par un vecteur de valeurs), peut être réduite à un scalaire. Cette fonction, appelée utilité, est une mesure que l'individu essaiera de maximiser par son choix. Dès lors, le point de départ le plus commun pour les modèles de choix est la notion de maximisation de l'utilité. Il est supposé que chaque individu est capable d'ordonner les différentes alternatives en fonction de leurs relatives attractions. Étant rationnel, l'individu choisit l'alternative qui a l'utilité maximale (c'est-à-dire celle qui maximise les avantages), tout en respectant ses contraintes.

Rappelons le principe de base permettant de décrire le choix du consommateur entre différents produits lorsqu'il est soumis à une contrainte de budget, dans la théorie micro-économique, puis attachons nous sur les cas des alternatives hétérogènes :

— soit un ensemble de  $N$  produits, dont le consommateur peut choisir d'acquérir des quantités respectives  $x_i$ , au prix  $p_i$  ( $i=1..N$ ).

---

10. Des détails plus importants peuvent être obtenus en lisant l'ouvrage de A. Kanafani. [Kanafani (A.), 1975].

Posons  $X = (x_1, \dots, x_N)$

et  $P = (p_1, \dots, p_N)$

— soit  $U(X)$  la fonction d'utilité associée.

— soit  $B$  son budget.

Selon le principe de maximisation de l'utilité, le consommateur va choisir le vecteur  $X^*$  des quantités de produits qui maximise  $U(X)$  sous la contrainte  $P \cdot X = B$ .

Dans le cas d'un choix discret, les quantités vaudront soit 0, soit 1. En fait, seulement une quantité vaudra 1 et toutes les autres 0. Si on se base sur le principe précédent, on cherche le maximum des alternatives. Une alternative est définie par un vecteur de la forme suivante :  $(0, 0, \dots, 0, 1, 0, \dots, 0, 0)$ . Notons  $q_i$  le vecteur de quantités tel que toutes les composantes sont nulles sauf la  $i^{\text{ème}}$ . On cherche alors le maximum des  $U_i$ , où  $U_i$  est l'utilité associée à  $q_i$ . La fonction d'utilité dans le cas de choix discret n'est pas différentiable par rapport aux quantités. On a supposé que l'attractivité d'une alternative exprimée par un vecteur de valeurs d'attributs peut être réduite à un scalaire, défini comme une seule fonction objectif exprimant l'attraction d'une alternative en fonction de ses attributs. Dans notre problème, cela se traduit dans l'expression de l'utilité  $U_i$  comme n'étant plus fonction des quantités, mais des attributs de l'alternative  $i$ . Il est important de bien garder à l'esprit que ce sont les caractéristiques des biens plutôt que les biens eux-mêmes qui sont supposés générer l'utilité <sup>11</sup>. Ce concept est approprié dans l'application de la théorie au transport, pour lequel il implique, par exemple, que ce ne sont pas les modes ou les chemins spécifiques qui sont sujets au choix mais leurs caractéristiques, telles que le temps et le coût de déplacement.

Le choix de l'expression de l'utilité, fonction des caractéristiques, est conforté par la liaison entre demande de transport et activités. Toutes les analyses de la demande en services de transport précisent que cette demande est dérivée de la demande en activités. En effet, on ne se déplace pas pour le plaisir de se déplacer, mais on se déplace pour participer à des activités (travailler, faire des achats, faire du sport ...). On intégrera donc des variables différentes selon les motifs. De plus, il semble raisonnable de supposer que les individus voudront, autant que possible, minimiser leurs temps et coût de déplacement, maximiser leur confort, etc. Dans ce contexte, l'utilité représente simple-

11. Ce qui déroge un peu de la théorie néo-classique conventionnelle qui ne fait référence qu'aux biens homogènes, mais reste un élément fondamental pour un auteur comme J. Lancaster, cité par J. Tirole, in [Tirole (J.), 1993].

ment une fonction de commodité généralisée qui prend en compte les éléments «positifs» et «négatifs» impliqués dans l'action de se déplacer, et qui constitue la base pour la décision de l'individu.

### 3. Modélisation de la procédure de choix

La modélisation de la procédure de choix discret consiste à associer à chaque alternative constituant l'ensemble étudié (modes par exemple), une utilité qui dépend des caractéristiques des individus et de l'alternative elle-même. On déduira, alors, pour chaque individu, la probabilité de choisir une alternative.

Si l'on tente d'être plus complet, on peut dire que le point de départ le plus commun, pour quantifier la notion de choix à l'aide de la modélisation, est la notion de maximisation de l'utilité : on suppose que les individus sont capables d'effectuer un classement ordinal des alternatives en présence, à partir de leurs caractéristiques (attributs). Pour un individu localisé dans une zone  $i$  et qui a la possibilité, pour se rendre à sa destination, de prendre le bus, le tramway ou le métro, une préférence se dégagera en fonction des caractéristiques de chaque mode (fréquence, vitesse, confort...). Chaque individu choisira donc l'alternative qui lui apporte une utilité maximale.

La micro-économie conventionnelle fait des hypothèses très fortes à propos de l'aptitude des individus à utiliser parfaitement toute l'information disponible afin de prendre leur décision et de la rendre complètement rationnelle. Un assouplissement des hypothèses de base peut être possible par la prise en compte du concept «d'utilité aléatoire». En effet, dans la pratique, les individus ne choisissent pas toujours ce qui est objectivement le meilleur pour eux. La théorie de l'utilité aléatoire supposera toujours qu'un individu choisira l'alternative qui **apparaît** maximiser son utilité au moment où le choix est effectué.

#### L'utilité est supposée avoir deux composantes :

- **une composante déterministe** qui est identique à la fonction d'utilité conventionnelle de la micro-économie et qui représente le niveau d'utilité moyen mesurable que l'individu  $t$  attache à l'éventualité  $i$ ,
- **une composante aléatoire** représentant tout ce qui n'est pas mesurable quant à l'individu  $t$ , pour l'alternative  $i$ . Ce terme est fonction des préférences de l'individu pour l'alternative et des caractéristiques non mesurables de celle-ci. La prise en compte de cette composante aléatoire a pour origine les critiques des biologistes et des psychologues sur la rationalité des individus. On retrouva aussi ce manque de rationalité en économie expérimentale.



Si l'on suppose que la composante aléatoire entre dans la fonction d'utilité de façon additive <sup>12</sup>, alors l'utilité d'une action  $i$  pour un individu  $t$  peut être exprimée comme :

$$U_{it} = V_{it} + \varepsilon_{it} \quad [1]$$

où :

$U_{it}$  = Utilité de l'alternative  $i$  pour l'individu  $t$

$V_{it}$  = Portion "observable", déterministe de l'utilité

$\varepsilon_{it}$  = Portion aléatoire de l'utilité

La partie observable <sup>13</sup> est supposée être une fonction des attributs de l'alternative  $X_i$  et des caractéristiques socio-économiques de l'individu (de son ménage), soit  $S_t$ . En particulier, il est typiquement supposé, pour des raisons de commodité analytique et de calage <sup>14</sup>, que  $V_{it}$  est donnée par :

$$V_{it} = \theta_1 Z_{it1} + \theta_2 Z_{it2} + \dots + \theta_n Z_{itn} = \theta Z_{it} \quad [2]$$

Avec :

$\theta$  = vecteur de paramètres

$\theta_j$  pour la constante si il y en a une. Celle-ci représente les variables non incluses dans la modélisation.

$$Z_{it} = f(X_i, S_t) \quad [3]$$

Le chargé d'études ne pourra pas savoir avec certitude quelle alternative aura l'utilité maximum pour un individu (ou groupe d'individus). Ce qu'il peut estimer, c'est seulement la probabilité qu'une alternative donnée, parmi un ensemble d'alternatives disponibles, sera celle qui aura une utilité maximum pour cet individu, et sera donc choisie. Compte tenu de l'équation [1] et d'un ensemble d'alternatives  $C_t$ , la probabilité que l'alternative  $i$  soit choisie par l'individu  $t$  est donnée par  $P_{it}$  :

$$P_{it} = P(U_{it} \geq U_{jt} \quad \forall j \in C_t, j \neq i) \quad [4]$$

Ou :

$$\begin{aligned} P_{it} &= P(V_{it} + \varepsilon_{it} \geq V_{jt} + \varepsilon_{jt} \quad \forall j \in C_t, j \neq i) \\ P_{it} &= P(\varepsilon_{jt} - \varepsilon_{it} \leq V_{it} - V_{jt} \quad \forall j \in C_t, j \neq i) \end{aligned} \quad [5]$$

12. Tout simplement car c'est ce qui est le plus pratique.

13. Dans la pratique, cette partie est souvent dénommée, à tort, utilité (on ne considère plus dans le discours et les écrits la partie aléatoire, et  $V$  remplace  $U$ ).

14. Il n'existe pas de procédure de calage adaptées pour des utilités autres que linéaires.

L'équation [5] est une expression pour la fonction de distribution des variables aléatoires  $\varepsilon_{jt}$ - $\varepsilon_{it}$  évaluées aux points  $V_{it}$ - $V_{jt}$ . Ainsi, si l'on peut estimer la distribution des termes aléatoires, l'équation précédente pourra être utilisée pour calculer la probabilité qu'un individu réalise tel ou tel choix. Une des hypothèses les plus raisonnables, car plusieurs ont été testées, est de supposer que les termes aléatoires sont distribués selon une loi normale. Cette hypothèse génère ce que l'on appelle une formulation probit (3.1.). Une autre hypothèse est de supposer que les termes aléatoires sont distribués selon une loi de Weibull (cas particulier de la loi de Gumbel). Cette hypothèse génère une formulation logit (3.2.).

### 3.1. La formulation probit <sup>15</sup>

Supposons que les  $\varepsilon_{it}$  et les  $\varepsilon_{jt}$  sont distribués selon une loi normale, avec une moyenne nulle, et de variances respectives  $s_i^2$  et  $s_j^2$ . Supposons de plus que leur covariance est  $s_{ij}^2$ .

Sous ces hypothèses, le terme  $\varepsilon_{jt}-\varepsilon_{it}$  est aussi normalement distribué de moyenne nulle, mais avec une variance  $s^2 = s_i^2 + s_j^2 - 2s_{ij}^2$ .

On peut utiliser ce résultat pour déduire les probabilités de choix par :

$$P_{it} = P(\varepsilon_{jt} - \varepsilon_{it} \leq V_{it} - V_{jt}) =$$

$$\int_{-\infty}^{V_{it}-V_{jt}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2\right] d\varepsilon, \quad \sigma > 0.$$

d'où

$$P_{it} = \Phi\left(\frac{V_{it} - V_{jt}}{\sigma}\right)$$

où  $\Phi$  est la distribution normale de moyenne nulle et  $\sigma$  l'écart type (racine de  $s$ ).

La formulation probit ne peut pas être facilement utilisée dans le cas multinomial, sauf sous certaines hypothèses simplificatrices. Ainsi, si l'expression précédente a un intérêt théorique évident, on constate que, dans la pratique, l'utilisation d'une formulation logit est préférée lorsque l'univers de choix est supérieur à deux.

15. Une justification mathématique des résultats est fournie en annexe 1.

### 3.2. La formulation logit<sup>16</sup>

Une hypothèse alternative concernant la distribution des termes aléatoires, est qu'ils sont Indépendants et Identiquement Distribués (IID), selon une distribution de Weibull<sup>17</sup> dont la fonction de distribution est donnée par :

$$F(\varepsilon) = \varepsilon^{-e} \quad [6]$$

Cela veut dire que la probabilité  $P_{it}$  sera uniquement fonction des différences  $V_{it} - V_{jt}$  (d'où le graphique présenté en 1.2.). Le choix de la distribution est entièrement motivé par des considérations de convenance analytique. En intégrant  $F(\varepsilon)$ , on peut démontrer que l'expression finale pour  $P_{it}$  est un modèle logit multinomial donné par :

$$P_{it} = \frac{e^{V_{it}}}{\sum_{j \in Ct} e^{V_{jt}}} \quad [7]$$

#### Exemple de calcul de la probabilité de choix d'une alternative :

La formulation logit est utilisée pour effectuer le calcul de proportions : elle permet de déduire la probabilité de choix d'une alternative.

Si l'on reprend l'exemple étudié en 1.2. avec les fonctions d'utilité suivantes...

$$V_{voiture} = 0,5 - 0,25 tp_{voiture} - 0,42 ta_{voiture} - 0,1 ct_{voiture}$$

$$V_{bus} = -0,28 tp_{bus} - 0,45 ta_{bus} - 0,1 ct_{bus}$$

Où :

**tp** = temps de parcours dans le mode

**ta** = temps d'accès au mode

**ct** = coût du déplacement

...et que l'on suppose que les données moyennes suivantes sont connues suite à l'enquête :

	<b>tp</b>	<b>ta</b>	<b>ct (en unité mon)</b>
Voiture	30	3	50
Bus	40	6	5

16. Une justification mathématique des résultats est fournie en annexe 2.

17. Cas particulier de la distribution de Gumbel, avec comme paramètres  $m=1$  et  $b=0$ .

Le calcul de choix de l'alternative voiture sera le suivant :

$$P_{\text{voiture}} = \frac{e^{0,5 - 0,25 \cdot 30 - 0,42 \cdot 3 - 0,1 \cdot 50}}{e^{0,5 - 0,25 \cdot 30 - 0,42 \cdot 3 - 0,1 \cdot 50} + e^{-0,28 \cdot 40 - 0,45 \cdot 6 - 0,1 \cdot 50}}$$

Soit  $P_{\text{voiture}} = 0,76$

La formulation logit est pratique et commode. En particulier, elle peut être calibrée<sup>18</sup> relativement facilement et de façon efficace. Une des caractéristiques majeures de cette formulation logit est l'hypothèse d'indépendance des alternatives non associées (Independance of Irrelevant Alternatives - IIA). En effet, l'estimation des coefficients ne sera correcte que si la distribution des variables est indépendante des choix effectués par les individus.

### 3.2.1. L'hypothèse IIA

Cette propriété peut-être facilement illustrée en observant dans l'équation [7], que la probabilité relative qu'un individu  $t$  choisisse l'alternative  $i$  plutôt que l'alternative  $j$ , ne dépend que des caractéristiques des alternatives  $i$  et  $j$ .

$$\frac{P_{it}}{P_{jt}} = \frac{e^{V_{it}}}{e^{V_{jt}}} \quad [8]$$

La probabilité relative de choix (entre  $i$  et  $j$ ) est donc indépendante des autres alternatives disponibles : tant que les valeurs de  $V_{it}$  et de  $V_{jt}$  ne changent pas, la probabilité relative ne changera pas, au regard des autres alternatives qui sont ajoutées ou supprimées de l'ensemble de choix.

La propriété IIA est à la fois la force et la faiblesse de la formulation logit.

**Elle est sa force** dans le sens où le modèle peut être calibré sur un ensemble d'alternatives et utilisé pour prédire les choix pour un ensemble d'alternatives modifié. Par exemple, un modèle de choix de mode peut être calibré en se basant sur les modes disponibles et utilisé pour examiner l'impact d'un nouveau mode dans le système. Cependant, cet avantage cache un inconvénient, car l'introduction d'une alternative nouvelle (i.e. d'un mode nouveau), entraînera une «rupture» importante, puisque l'on se base sur des fonctions d'utilité qui traduisent des «lois de comportement» à un instant donné. Celles-ci

18. Rappel : le calibrage correspond à l'estimation des coefficients  $\theta_i$  de la partie déterministe  $V_{it}$  de la fonction d'utilité  $U_{it}$ .

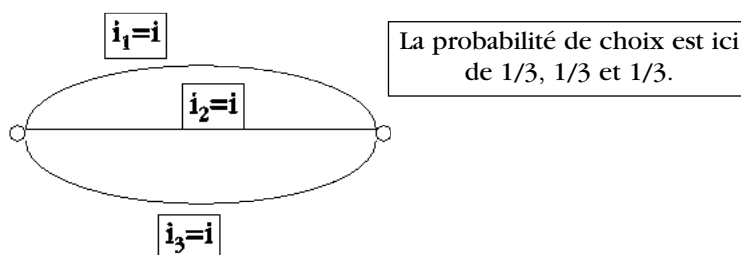
peuvent donc difficilement être utilisées pour prévoir des choix relativement à un mode qui n'existe pas, c'est-à-dire qui n'est pas intégré dans le comportement des individus <sup>19</sup>.

Cette propriété peut aussi être exploitée dans le cas où l'ensemble des choix potentiels est très large (exemple : choix de localisation résidentielle), afin d'éliminer le besoin d'inclure l'ensemble des possibilités de choix dans les procédures de calcul.

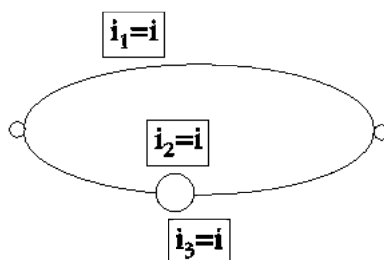
**Elle est sa faiblesse** car il faut veiller à ce que les alternatives prises en compte dans l'ensemble des choix soient bien indépendantes les unes des autres.

L'exemple suivant illustre cette faiblesse.

Soit trois itinéraires de même temps de trajet entre une origine et une destination, constituant l'ensemble des alternatives



Supposons maintenant que deux des chemins soient fortement corrélés : le troisième itinéraire, revenant à une modification à la marge du deuxième itinéraire, en créant un détour.



D'un point de vue empirique, il y a trois itinéraires possibles différents, mais le trafic ne devrait pas se répartir équitablement sur ceux-ci, comme précédemment.

19. C'est d'ailleurs l'une des conclusions à laquelle l'INRETS aboutit dans son étude de prévision sur les parts modales liées à la réalisation d'Orlyval. Cf. [INRETS, 1988].

Or, la simple application d'un modèle logit dans ce deuxième cas prévoira une répartition de  $1/3$ ,  $1/3$  et  $1/3$ , considérant les trois chemins comme étant totalement différents, alors que l'on devrait avoir  $1/2$ ,  $1/4$  et  $1/4$  du trafic. Ceci est le résultat direct de l'application d'un modèle logit à un ensemble de choix qui viole clairement la propriété IIA : les alternatives 2 et 3 ne sont pas indépendantes l'une de l'autre. Elles sont mêmes très dépendantes, et la probabilité de choisir l'une est très corrélée à la probabilité de choisir l'autre. Ainsi, la prise en compte d'une alternative qui modifie l'ensemble des possibles pourra avoir un impact aussi important que l'introduction d'une alternative modifiant fondamentalement les possibilités de choix.

Dès lors, **si l'ensemble de choix laisse présager une violation de la propriété IIA**, il sera plus prudent d'utiliser une formulation de type probit qui est capable de manipuler des corrélations entre les alternatives. Mais rappelons que l'utilisation d'une formulation probit est délicate à mettre en oeuvre. On pourra, alors, modifier ce que l'on croît être la structure du processus de choix des individus, afin de réduire ou d'éliminer la dépendance entre les alternatives. L'utilisation d'une structure de choix «hiérarchique» a cette ambition.

### 3.2.2. La structure de choix hiérarchique

Une approche, pour résoudre la violation de la propriété IIA, est de considérer le problème comme un processus de décision imbriqué à plusieurs étages. La détermination de la structure de choix à adopter dépendra bien évidemment de l'étude réalisée et des données disponibles. L'élément central reste que la supposition que l'on fera, aura pour principe de traduire le processus de comportement des individus. D'où l'importance de la réflexion en amont. En effet, deux problèmes doivent être résolus :

- le problème de l'indépendance des alternatives, certaines étant similaires entre elles, par exemple pour les transports collectifs ;
- le problème de la variation du goût des individus. Ici, on préférera un modèle avec des coefficients «marginiaux» plutôt que moyens, c'est-à-dire tels qu'obtenus à l'aide d'une forme logit.

Dans le cas d'une structure de choix hiérarchique, les décisions sont toujours prises séquentiellement, mais un haut niveau de décision pourra inclure, quant aux calculs, des attentes concernant un niveau de décision inférieur. Il y aura donc une interaction étroite entre les deux niveaux d'utilité. Pourtant une structure hiérarchique ne permettra jamais de résoudre le problème de la variation du goût des individus, qui reste irrésolu.

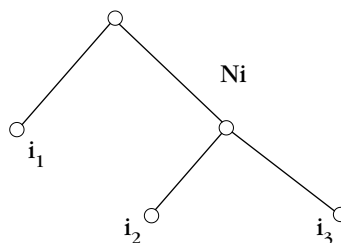
### Exemple

D'un point de vue pratique, si l'on reprend l'exemple de choix d'itinéraire préalablement présenté, on considère qu'un premier choix sera effectué entre les itinéraires  $i_1$  et l'ensemble  $\{i_2, i_3\}$  puis qu'un autre choix le sera entre  $i_2$  et  $i_3$ . On calculera tout d'abord un Logit pour les alternatives du "nid" de bas niveau ( $N_i$ ), tout en omettant les variables qui prennent la même valeur pour les deux alternatives en présence ( $i_2$  et  $i_3$ ). Celles-ci seront introduites plus tard dans le «nid» supérieur pour affecter le choix entre l'alternative composite  $N_i$  et le reste des options (ici  $i_1$ ), toujours en calculant un logit.

Dès lors, la probabilité qu'un individu  $t$  effectue un choix parmi les alternatives en présence, peut se traduire comme le produit de la probabilité marginale de choisir l'alternative composite  $N_i$  ou  $t_1$  dans le «nid» supérieur et de la probabilité conditionnelle de choisir l'option  $t_2$  ou  $t_3$  dans le «nid» inférieur.

Soit :

$$\begin{aligned} P_{t1} &= P(i_1) \\ P_{t2} &= P(i_2/N_i) P(N_i) = P(i_2/N_i) (1-P_{t1}) \\ P_{t3} &= P(i_3/N_i) P(N_i) = (1-P(i_2/N_i)) (1-P_{t1}) \end{aligned}$$



Aussi intéressante soit elle, la formulation dite logit hiérarchique peut poser des problèmes dans son utilisation, surtout si les données dont on dispose ne sont pas suffisantes pour calculer le logit de niveau inférieur. Il manquera une partie de l'information nécessaire et les erreurs passeront au(x) niveau(x) supérieur(s). Cependant, certains logiciels que l'on peut maintenant facilement trouver sur le marché, permettent de traiter des structures hiérarchiques complexes <sup>20</sup>.

20. Alogit, Hielow, par exemple.

Une autre interrogation concerne l'intérêt de modéliser à l'aide d'une structure hiérarchique plutôt qu'avec un logit multinomial «simple». Tout d'abord, il convient de préciser que la décision de réaliser une structure hiérarchique doit être issue de réflexions émanant d'enquêtes ou d'études effectuées par ailleurs. Un coefficient, proposé par les logiciels existants sur le marché permet cependant de confirmer l'orientation choisie après réflexion. Celui-ci est nommé «logsum» (ou «théta»  $\theta$ ). C'est un coefficient qui pondère le logarithme de la somme de l'exponentielle des utilités élémentaires (i.e. d'un nid supérieur). S'il est égal à un, alors le fait de réaliser une structure hiérarchique est équivalent à ne pas en réaliser une. S'il est compris entre 0 et 1, alors il est plus juste de réaliser, ou de garder, la structure hiérarchique. Enfin, s'il est inférieur à zéro ou supérieur à 1, il y a une erreur qu'il convient de chercher à comprendre. Ce coefficient intervient dans le calcul des probabilités de choix d'alternatives.

Le choix d'une alternative ne s'effectue plus de la même manière dans un modèle hiérarchique.

Dans un modèle non hiérarchique, chaque combinaison entre les alternatives possibles (par exemple : voiture particulière-VP ou transport collectif-TC type bus ou métro) représente une alternative : le choix probabiliste est effectué sur l'ensemble des alternatives.

Alors que dans le modèle hiérarchique, on estime tout d'abord l'utilité  $U$  du mode individuel (1 alternative) ou du mode collectif (2 alternatives).  $U$  s'exprime à partir des utilités de toutes les alternatives du mode correspondant et du logsum :

$$U = \text{logsum} \times \ln\left(\sum_i e^{u_i}\right)$$

où les  $U_i$  représentent les utilités des alternatives pour le type de mode considéré.





## Chapitre 2

# Méthodologie des modèles de choix discret

Les éléments relevant plutôt du domaine de la théorie des modèles de choix discret ayant été traités, d'autres relevant plus de la pratique de la construction de ces modèles doivent être abordés. Le terme de méthodologie a été retenu car on considère, ici, le rapport du chargé d'étude avec son modèle. Nous reprenons ainsi, dans le détail, les phases de spécification et d'estimation qui ont déjà été explicitées dans la première partie du rapport et qui sont fondamentales. Ensuite, nous nous intéressons au problème de l'appréciation de la qualité du modèle par le chargé d'étude, aux sorties obtenues, à la nécessaire étape d'agrégation (et à la moins systématique étape de transférabilité), et aux limites des modèles de choix discret.

### 1. La spécification : recherche de la forme fonctionnelle et de la forme mathématique des variables

Bien que la phase de spécification des modèles de choix discret ne leur soit pas spécifique, puisqu'elle concerne tous les types de modèles, elle doit être présentée.

La procédure de modélisation, dans une approche désagrégée, consiste en la recherche de lois de comportement, basées sur la notion de fonctions d'utilité. Ainsi, lorsque l'on étudie le comportement des individus vis à vis d'un ensemble de modes de transport, on obtient une fonction d'utilité par mode. Les fonctions d'utilité seront alors «introduites» dans la formulation logit (simple ou hiérarchique) ou probit, afin de traduire quelle est la procédure de choix «de base» à partir de laquelle des tests de politiques de transport pourront être effectués (Cf. exemples présentés précédemment).

Arriver à une spécification correcte des fonctions d'utilité, suppose que deux problèmes de fond aient été résolus a priori <sup>21</sup>. Ils concernent la forme fonctionnelle de la fonction d'utilité et la forme mathématique des variables.

**Le premier problème** n'est jamais vraiment traité. Dans la plupart des cas, on postule l'existence de fonctions additives <sup>22</sup>, dans lesquelles on suppose que tous les facteurs qui affectent le choix ont des effets additifs indépendants. Les formes multiplicatives <sup>23</sup>, qui postulent des effets interdépendants ne sont quasiment pas utilisées dans la pratique, pas plus que la forme exponentielle qui combine les caractéristiques précédentes d'ailleurs <sup>24</sup>. Le choix d'une forme additive est lié à la plus grande simplicité de sa mise en oeuvre. Cependant, des recherches qui visent à optimiser la forme fonctionnelle existent.

**Le deuxième problème** peut demander beaucoup de temps au chargé d'études. En effet, si une part importante de réflexion est nécessaire pour définir la représentation des variables, des procédures plus automatiques existent pour définir leur spécification.

### Représentation des variables

Prenons une procédure de choix discret entre deux modes de transport, et le cas d'une variable comme le temps de déplacement. On a  $TD_1$  pour le mode 1 et  $TD_2$  pour le mode 2. Pour affiner l'analyse on pourra créer une variable  $TD^*$ , qui sera une représentation composite de différences absolues entre les temps de déplacements (soit  $TD^* = TD_1 - TD_2$ ) ou de différence relative (soit  $TD^* = (TD_1 - TD_2)/TD_2$ ). On testera alors quelle est la représentation la meilleure pour la modélisation.

### Spécification des variables

Restons dans le cadre d'une forme fonctionnelle linéaire, avec une représentation des variables définie a priori, mais supposons que l'on ne connaît pas la spécification de ces variables : quelle forme va prendre la variable  $x$  :  $e^x$ , ou  $\ln x$  ou  $x^2$ ,...?

Plutôt que de fixer une forme a priori (ce qui se fait habituellement, on arrive alors à la meilleure forme par tâtonnement), on peut utiliser des procédures de transformation automatiques. La plus utilisée est celle de «Box-Cox» :

21. En dehors des problèmes «pratiques» liés à la récolte des données, sur lesquels nous ne nous attardons pas ici.

22. Par exemple,  $X = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3$

23. Par exemple,  $X = \theta_0 \cdot X_1^{\theta_1} \cdot X_2^{\theta_2} \cdot X_3^{\theta_3}$

24. Par exemple,  $X = e^{\theta_0 + \theta_{1X1} + \theta_{2X2} + \theta_{3X3}}$

La transformation  $x(\gamma)$  d'une variable positive  $x$  est donnée par :

$$X(\gamma) = \begin{cases} (x^\gamma - 1) / \gamma, & \text{si } \gamma \neq 0 \\ \log x & \text{si } \gamma = 0 \end{cases}$$

si il y a continuité pour toutes les valeurs de  $\gamma$  possibles.

Dès lors l'équation traditionnelle :

$$V_{it} = \sum_k \theta_{ki} X_{kit}$$

peut être réécrite sous la forme :

$$V_{it} = \sum_k \theta_{ki} X_{kit}^\gamma$$

On remarque alors que si  $\gamma_1 = \gamma_2 = \dots = \gamma_k = 1$ , l'équation précédente se réduira à une forme linéaire et que si  $\gamma_k = 0$ , on obtiendra une forme log-linéaire.

Une autre procédure, moins utilisée, est celle de «Box-Tukey». Celle-ci introduit un paramètre supplémentaire.

La transformation  $x(\gamma, \mu)$  d'une variable positive  $x$  est alors donnée par :

$$(X + \mu)^\gamma = \begin{cases} ((x + \mu)^\gamma - 1) / \gamma, & \text{si } \gamma \neq 0, x + \mu > 0 \\ \log (x + \mu) & \text{si } \gamma = 0, x + \mu > 0 \end{cases}$$

De même, on peut alors réécrire  $V_{in}$  sous la forme :

$$V_{in} = \sum_k \theta_{ji} (x + \mu_i)^{\gamma_{ji}}$$

Que l'on utilise une transformation de «Box-Cox» ou de «Box-Tukey», on obtiendra une transformation appropriée des variables. Elle n'aura donc pas besoin d'être explicitée de façon exhaustive et figée a priori pour chacune d'entre elles. Si l'on utilise pas une procédure «assistée», le travail devra être réalisé «manuellement».

## 2. Estimation d'un modèle de choix discret

La méthode présentée ici est la plus utilisée, la plus courante. D'autres existent mais sont très peu, voire pas du tout, utilisées dans la pratique.

### 2.1. La méthode du maximum de vraisemblance : explications

La procédure d'estimation d'un modèle de choix discret se fait habituellement en utilisant la méthode du maximum de vraisemblance, introduite dans les années 1920 par Ronald Fischer. Elle consiste à définir, à l'aide d'un échantillon d'individus et de probabilités données par le modèle, une «fonction de vraisemblance», notée  $L(\theta)$  et à chercher la valeur des  $\theta_i$  qui rendent cette fonction maximum. Le plus souvent, on a comme informations pour chaque échantillon d'individus :

- le choix effectué,
- les valeurs des caractéristiques des individus,
- les valeurs des caractéristiques des différentes alternatives

Un modèle de choix discret est estimé sur un échantillon. Le maximum de vraisemblance correspond à la recherche de la valeur des paramètres qui permettent de mesurer la probabilité (que l'on doit maximiser) qui est la plus proche d'avoir généré l'échantillon observé.

#### Explications

Supposons que l'on réalise une enquête, ou que l'on en possède une, pour étudier le comportement d'une population en terme de choix de mode (alternatives : bus ou voiture ?). L'échantillon que l'on possède de notre population a une fonction de probabilité  $p(X/\theta)$ , avec  $X$  représentant le choix de chaque individu dans l'échantillon (il varie de 1 à  $n$ ).  $\theta$  est l'ensemble des paramètres du modèle déterminant le choix que l'on souhaite estimer. La fonction de probabilité simultanée pour l'ensemble de la population correspondra à la multiplication des probabilités de choix de l'alternative pour chaque individu  $t$  (pondérée par son poids dans l'échantillon), soit :

$$P(X_1, X_2, \dots, X_t, \dots, X_n / \theta) = P(X_1 / \theta) P(X_2 / \theta) \dots P(X_t / \theta) \dots P(X_n / \theta)$$

La vraisemblance de l'échantillon sera donc :

$$L = \prod_t P_t(\text{choix de } t)$$

Où,  $P_t(i)$  représente la probabilité pour l'individu  $t$  de choisir l'alternative  $i$ .  
Et (choix de  $t$ ) représente le choix effectivement réalisé par l'individu  $t$ .

Définissons la variable suivante :

$$g_{jt} = \begin{cases} 1 & \text{si l'individu } t \text{ a choisi l'alternative } j \\ 0 & \text{sinon} \end{cases}$$

On peut alors réécrire la fonction de vraisemblance par :

$$L = \prod_{t=0}^N \prod_{j \in \text{choix possibles pour } t} (P_t(j))^{g_{jt}}$$

Soit  $V_{it} = f(\theta_j, x_{itj})$  la partie déterministe de la fonction d'utilité  $U_{it}$ . On cherche à estimer au mieux les coefficients  $\theta_j$ , à partir de certaines données :

- les caractéristiques  $x_{itj}$  de l'individu  $t$  et de l'alternative  $i$
- le choix effectué par cet individu,

Ceci pour un certain échantillon d'individus.

Cette fonction correspond au produit pour tous les individus de l'échantillon, des probabilités de choisir le mode qu'ils ont effectivement choisi. L'ajustement sera le meilleur lorsque la probabilité de choisir le mode réellement choisi sera proche de 1. Dans le cadre des études, on maximise le logarithme de la fonction de vraisemblance, afin de se ramener à une somme plutôt qu'à un produit  $L$  :

$$\ln(L) = \sum_{t=1 \text{ à } N} \sum_{j=1 \text{ à } I} g_{jt} \ln(P_t(j))$$

On va donc maximiser  $\ln(L)$  en fonction des coefficients  $\theta_j$  de la fonction d'utilité. Cependant, il convient de faire attention à la définition du modèle, certaines erreurs dans sa spécification pouvant entraîner des problèmes dans la recherche du maximum de vraisemblance.

## 2.2. Les problèmes survenant dans la recherche du maximum de vraisemblance

Les causes les plus courantes des problèmes sont au nombre quatre :

### 1. Un trop grand nombre de termes constants dans la fonction d'utilité

Dans la plupart des applications, il est habituel d'introduire des termes constants. Par exemple, on peut définir :

$$x_{it1} = \begin{cases} 1 & \text{si } i = 1 \\ 0 & \text{sinon} \end{cases}$$

Cela aura pour effet de rendre  $\theta_1 x_{it}$  terme constant dans l'utilité de la première alternative.

En pratique, la restriction dans l'utilisation de termes constants est la suivante : pour K alternatives, on peut définir au plus K-1 constantes. Les constantes supplémentaires sont redondantes, et feront échouer l'estimation des coefficients. Le choix de la constante à éliminer est entièrement arbitraire, n'a pas d'effet sur les termes non constants, et ne modifie pas la probabilité de choix prédite.

## 2. Une spécification incorrecte des variables socio-économiques

Lors de l'utilisation d'une variable qui est indépendante des différentes alternatives (une variable socio-économique par exemple), il faut qu'un coefficient différent soit attribué à chaque alternative ( $\theta_i$ ). Dans la pratique, ce type de spécification n'est pas toujours fait, ce qui est incorrect.

Par exemple, dans le cas d'une fonction d'utilité linéaire à deux alternatives faisant intervenir le revenu, si celui-ci a le même coefficient  $\theta_1$  pour chaque alternative, on obtient :

$$Pt(i) = \frac{e^{\theta_1 \text{revenu}_t + \theta_2 x_{it}}}{e^{\theta_1 \text{revenu}_t + \theta_2 x_{it}} + e^{\theta_1 \text{revenu}_t + \theta_2 x_{it}}}$$

où  $\text{revenu}_t$  est le revenu de l'individu t. Si on réarrange le numérateur et le dénominateur en factorisant le terme  $e^{\theta_1 \text{revenu}_t}$ , il résulte :

$$Pt(i) = \frac{e^{\theta_1 \text{revenu}_t} e^{\theta_2 x_{it}}}{e^{\theta_1 \text{revenu}_t} (e^{\theta_2 x_{it}} + e^{\theta_2 x_{it}})} = \frac{e^{\theta_2 x_{it}}}{(e^{\theta_2 x_{it}} + e^{\theta_2 x_{it}})}$$

Ce paramètre n'a donc aucun effet sur la probabilité de choix s'il a le même coefficient dans toutes les alternatives.

## 3. Une spécification du modèle avec un coefficient non borné

Cette erreur arrive lorsque l'on inclut une variable dont le coefficient peut prendre une valeur très grande lors de l'estimation. Par exemple, supposons que l'on définisse un attribut dans  $V_{it}$  qui prend toujours une valeur positive pour les individus choisissant l'alternative i, et qui est nul pour les autres. Alors, en donnant au coefficient de ce paramètre une valeur très grande, on explique parfaitement les décisions des individus qui choisissent i (car à la limite, lorsque le coefficient croît,  $P_{it}$  approche 1). Or, de tels coefficients vont entraîner des dépassements de capacité lors des calculs nécessaires pour leur estimation.

#### 4. Une colinéarité des paramètres.

Il est impossible d'obtenir le maximum de vraisemblance estimé pour un modèle de choix à partir d'un échantillon dans lequel certaines des variables sont parfaitement corrélées. Par exemple, si le coût d'utilisation d'un mode est toujours une fonction linéaire du temps de déplacement, alors on ne peut pas inclure ce temps et ce coût comme variables distinctes. On retrouve ce problème dans tous les travaux économétriques, et il faut lui apporter une certaine attention.

### 3. Appréciation de la qualité d'un modèle de choix discret

Six critères sont conseillés pour apprécier la qualité d'un modèle de choix discret. Ils sont présentés par ordre d'importance :

#### 3.1. La significativité statistique des paramètres

Comme cela a été précisé, l'estimation des coefficients  $\theta_i$  se fait par la technique du maximum de vraisemblance. L'ensemble des paramètres estimés  $\theta^*$  qui est obtenu est distribué normalement  $[N(\theta, S^2)]$  où :

$$S^2 = \frac{1}{E\left(\frac{\partial^2 L(\theta)}{\partial \theta^2}\right)}$$

C'est cette expression explicite de la matrice de covariance  $S^2$ , déterminant les paramètres  $\theta^*$  qui impliquera un processus itératif. Dans le cas d'une forme linéaire de la fonction d'utilité, le processus convergera facilement et toujours à un maximum unique, ce qui explique que beaucoup de logiciels permettant d'estimer de tels modèles sont disponibles sur le marché. Cette remarque rejoint celle qui a été faite sur les formes fonctionnelles.

Dès lors, si  $\theta_k^*$  a une variance estimée  $s_{kk}^2$ , ou  $S^2 = \{s_{kk}^2\}$ , qui est calculée durant l'estimation, alors, si sa moyenne  $\theta_k = 0$  ; le coefficient  $st$  sera égal à :

$$st = \theta_k^* / s_{kk}$$

a une distribution normale standard  $N(0,1)$ . C'est pour cela que l'on peut tester si  $\theta_k^*$  est significativement différent de 0.

Un paramètre ne devrait donc être retenu que s'il est statistiquement significatif, c'est-à-dire si l'on est sûr, à hauteur d'un certain seuil, qu'il



est significativement différent de 0. Les niveaux de significativité sont :

à 1%	+/- 2,58
à 5%	+/- 1,96
à 10%	+/- 1,65

Cependant, dans la pratique, il se peut que le chargé d'études décide qu'un modèle soit de bonne qualité alors que tous les paramètres ne sont pas statistiquement significatifs ! Ceci est surtout vrai pour les variables dites «cruciales», c'est-à-dire pour celles qui doivent logiquement se trouver dans la modélisation, car la pratique de la modélisation et les analyses empiriques montrent qu'elles ont un pouvoir explicatif déterminant (exemple pour le choix de mode : la disposition d'un véhicule, le coût...). Un problème de qualité des données entrées pourra conduire à réaliser une telle observation.

### 3.2. Le signe des paramètres

Parallèlement à la significativité statistique des paramètres, le chargé d'études doit à tout prix regarder le signe attribué aux paramètres. Ces signes doivent être étudiés un par un, et les variables dont le coefficient a un signe contraire à ce que l'on attend doivent normalement être rejetées. Ainsi, tout comme la significativité des paramètres, la cohérence de leur signe sert à évaluer le modèle et sa méthode d'estimation. Au cas où un paramètre n'a pas le bon signe, il faut normalement revoir le modèle. Cependant dans la pratique, dans le cas des variables dites «cruciales», la découverte d'un signe contraire à ce que l'on attend doit attirer toute l'attention. Son rejet pourra porter préjudice à la procédure de modélisation. Par exemple, un signe positif pour le paramètre de coût des transports collectifs, alors que l'on modélise le choix de mode sera très gênant et entraînera une étude attentive dont l'objectif sera de comprendre le dysfonctionnement auquel on aboutit.

### 3.3. La valeur du $\rho^2$

Le  $\rho^2$  va permettre d'estimer quel est le gain d'information apporté par un ensemble de variables exogènes (explicatives), relativement à une variable endogène (à expliquer), à l'instar du coefficient de détermination dans les analyses de régression ( $r^2$ ). Celui-ci s'obtient, pour chaque modèle, en faisant le rapport de la log-vraisemblance obtenue pour les probabilités de choix calculées à partir des coefficients estimés  $\theta_0^*, \theta_1^*, \dots, \theta_h^*$  à la valeur de la log-vraisemblance pour les

probabilités de choix calculées à partir de la répartition enregistrée dans l'échantillon (modèle équiprobable, à coefficients nuls).

### Exemple

Prenons l'exemple d'une modélisation de choix de mode de transport (bus, voiture), à partir d'un échantillon de 500 personnes. Supposons que 100 personnes ont choisi l'autobus et 400 la voiture. La probabilité de choix associée à la première alternative est : 100/500 et 400/500 pour la deuxième. En supposant que les tirages sont indépendants, la vraisemblance du partage modal s'écrit :

$$L = \prod x P = \frac{100^{100}}{500} \times \frac{400^{400}}{500}$$

d'où la log-vraisemblance :

$$\begin{aligned} \ln(L) &= 100 \ln 100/500 + 400 \ln 400/500 \\ &= -160,94 - 89,25 = -250,20 \end{aligned}$$

Si la log-vraisemblance  $L_0$  calculée à partir des coefficients estimés est :  $\ln L = -170,30$

On aura alors :

$$\rho^2 = 1 - \frac{\ln(L)}{\ln(L^0)} \quad \text{d'où } \rho^2 = 1 - [-170,30 / -250,20] = 0,32$$

Les valeurs du  $\rho^2$  sont toujours plus faibles que celles du  $r^2$ . Une valeur de 0,4 correspond à un ajustement excellent. Cependant, il faut se méfier de cet indicateur, car il croît automatiquement avec le nombre de variables. C'est pour cela qu'on lui préférera le  $\rho^2$  (rho barre carré), qui introduit une correction au niveau du nombre de coefficients estimés. Soit :

$$\rho^2 = 1 - \frac{\ln(L) - n}{\ln(L^0)} \quad (\text{avec } n = \text{nombre de coefficients})$$

Dès lors, il est intéressant de tenir compte de cet indicateur d'un point de vue relatif, pour des modèles qui ont un nombre de variables équivalent et compte tenu des autres critères qui permettent de juger de la qualité d'un modèle. Nous les présentons dans ce paragraphe.

## 3.4. La valeur du temps

La valeur du temps correspond au prix que chaque individu (ou groupe d'individus) serait prêt à payer pour économiser une unité de temps marginale, pour un motif donné et sous contrainte de revenu. C'est

un coût d'opportunité. Cette valeur est obtenue par le rapport entre deux paramètres de deux variables parfois prises en compte dans la procédure de modélisation de choix discret en matière de transport : le temps et le coût <sup>25</sup>.

Le résultat du calcul de la valeur du temps permettra de vérifier la qualité du modèle (par confrontation avec d'autres résultats connus), mais aussi d'enrichir d'autres études de transport (exemple : évaluation a priori des projets), tout comme le résultat des calculs d'élasticités simples et croisées effectués dans le cadre de la modélisation.

### 3.5. Les coefficients d'élasticité

Ce concept d'élasticité est très présent dans les modèles de choix discret. Pour un individu  $t$ , l'élasticité directe de la probabilité de l'alternative  $i$  se rapportera à une caractéristique  $x_{itk}$  de cette alternative, alors que l'élasticité croisée sera par rapportera à une caractéristique  $x_{jtk}$  d'une autre alternative. Dans le cas d'un modèle logit, on obtient : comme élasticité directe :

$$E_{x_{itk}}^{P_i(i)} = \frac{\frac{\Delta P_i(i)}{P_i(i)}}{\frac{\Delta x_{itk}}{x_{itk}}} = \frac{\Delta \ln P_i(i)}{\Delta \ln x_{itk}} = (1 - P_i(i)) \theta_k$$

où  $\theta_k$  est le coefficient de  $x_{itk}$  dans la fonction d'utilité de l'alternative  $i$ .  
et comme élasticité croisée :

$$E_{x_{jtk}}^{P_i(i)} = \frac{\frac{\Delta P_i(i)}{P_i(i)}}{\frac{\Delta x_{jtk}}{x_{jtk}}} = \frac{\Delta \ln P_i(i)}{\Delta \ln x_{jtk}} = -P_n(j) \theta_k$$

On remarque que l'élasticité croisée est uniforme, c'est à dire que l'élasticité croisée de toutes les alternatives par rapport au changement d'un attribut affectent seulement l'utilité de l'alternative  $j$  et est égale pour toutes les alternatives  $i \neq j$ .

Une manière d'apprécier la qualité d'un modèle de choix discret est de vérifier la pertinence des coefficients d'élasticité obtenus à l'aide de la modélisation. On vérifiera que l'on se situe dans des fourchettes

25. Sous réserve qu'elles ne soient pas corrélées.

«acceptables», c'est à dire que l'on ne s'éloigne pas des résultats obtenus avec ce genre de modèles et qu'il n'y a pas d'incohérence (sensibilités trop marquées).

Ainsi, les modèles de choix discret permettent de fournir toute une batterie de coefficients d'élasticité, qui peuvent avoir une certaine utilité pour les études socio-économiques.

### 3.6. La propension du modèle à reproduire les comportements observés

Une des manières de procéder pour vérifier si le modèle obtenu permet de reproduire correctement les comportements observés, est de construire un tableau qui croise les choix prédits et les choix observés. Cette méthode est très utilisée.

## 4. Sorties obtenues à l'aide d'un modèle de choix discret et méthode de validation

Le type de sorties obtenues à l'aide d'un modèle de choix discret est regroupé dans le tableau suivant.

En général, on obtient :

$\theta_i$	La valeur des coefficients associés aux variables
st	L'indicateur pour la significativité statistique des coefficients
Log(L)	Le logarithme du maximum trouvé pour la fonction de vraisemblance
Log(L°)	La valeur du logarithme de la fonction de vraisemblance avant évaluation des paramètres
$2 \times [\text{Log(L)} - \text{Log(L}^\circ)]$	La statistique du taux de vraisemblance
$1 - [\text{Log L} / \text{Log L}^\circ]$	La statistique rho-carré
$(N_{ii} + N_{jj}) / n$	Le pourcentage de prédictions correctes à l'instant de référence.
$P_t(i)$ $E_{x_{jtk}}$	Les élasticités simples et croisées

## 5. Agrégation

Pour un grand nombre de personnes le terme de dimension «désagrégé», auquel se rapportent les modèles de choix discret, est impropre. L'INRETS préfère utiliser le terme de «niveau de désagrégation». En effet, bien que l'on parte d'observations «désagrégées» (i.e. relatives aux individus), et que les fonctions d'utilité soient estimées pour chaque individu, les résultats sont présentés sous une forme agrégée pour des segments de population ou de ménages. La procédure «d'agrégation» porte aussi sur l'aspect spatial. En effet, les contraintes que fait peser la récolte des données <sup>26</sup> poussent à découper l'agglomération en zones. Logiquement, l'aire d'étude ne devrait pas être découpée, et l'on devrait travailler au niveau le plus fin (l'adresse). L'impossibilité pratique de le faire dans la plupart des cas, fait que l'on recherche le niveau de finesse spatial le plus adéquat du point de vue de l'analyse, compte tenu des contraintes précédentes. Dès lors, cela réduit l'ensemble des possibles du point de vue des destinations, lorsque l'on étudie les comportements de choix de mode/destination combinés par exemple. Mais cette réduction n'est souvent pas suffisante pour l'analyse et l'on regroupe alors les zones en fonction de certaines caractéristiques pour la définition des fonctions d'utilité. Ainsi, l'agrégation peut porter sur deux dimensions (individus/ménages et types de zones de l'aire d'étude).

## 6. Limites des modèles de choix discrets

Comme tous les modèles, les modèles de choix discret sont soumis à un certain nombre de limites, cinq seront citées.

### 6.1. Les limites liées à l'opérationnalité

La première des limites concerne la performance de ces modèles. A. Bonnaïous <sup>27</sup>, a ainsi montré que la performance d'un modèle était fonction de «son aptitude à réaliser les objectifs qui lui étaient assignés»,

---

26. Problème pour utiliser des données très fines à cause de la Convention Nationale Informatique et Liberté (CNIL), problème du coût d'obtention et de l'existence même de certaines d'entre elles (données fines sur l'offre de transports collectifs par exemple).

27. Cf [Bonnaïous (A.), 1989].

le niveau de la performance dépendant de l'opérationnalité du modèle. Cette dernière est soumise à trois conditions nécessaires et suffisantes, que l'on peut résumer ainsi :

- Un modèle sera pertinent si la structure mathématique sur laquelle il est fondé est conforme à la réalité : la qualité de l'approximation de ce que l'on croit être la réalité va dépendre de la pertinence du modèle.
- Un modèle sera cohérent si son objectif est réalisable au plan théorique. Un modèle ne peut présenter des contradictions tant d'un point de vue interne que par rapport à ses objectifs.
- Un modèle sera mesurable lorsque les variables et les paramètres pourront être estimés à partir d'échantillons statistiques accessibles.

Or, ces trois règles sont contradictoires. Par exemple, en voulant améliorer la cohérence du modèle par l'introduction de nouvelles hypothèses, on risque de remettre en cause sa pertinence. Ou encore, pour rendre plus pertinent un modèle on pourra introduire de nouvelles variables, plus nombreuses, mais cela risque d'agir sur sa cohérence. De même, la recherche d'une meilleure mesurabilité peut entraîner l'abandon de certaines variables explicatives, donc affecter la cohérence du modèle et réduire sa pertinence. Ainsi, pour les modèles de choix discret, l'ambiguïté tient surtout à ce dernier point.

## 6.2. Les limites liées aux données

Sans considérer le cadre d'interaction permanente précédent, la qualité des modèles de choix discret est très étroitement liée aux données nécessaires à leur fonctionnement, comme nous avons déjà eu l'occasion d'y faire allusion. Les problèmes de récolte de données peuvent être conséquents, d'autant plus que, dans certains cas, ces données n'existent même pas. Ainsi, il est souvent nécessaire de les produire (à l'aide d'enquêtes spécifiques), et/ou de faire appel à des outils informatiques performants ; ce type d'outils permettant, par exemple, de redresser des matrices de temps existantes, ou de calculer des temps d'accès au réseau de transports collectifs.

## 6.3. Les limites liées aux problèmes de stabilité des comportements

Parce que ces modèles traduisent des «lois de comportement», ils sont sensibles aux variations concernant les variables explicatives qui les caractérisent. Il est donc délicat de les utiliser sur le long terme, et d'effectuer des tests qui entraînent une «rupture» trop importante.

Par exemple, le test d'une offre de transport collectif nouvelle qui n'existe pas dans l'agglomération. On testera ce genre de mesure sur le court ou moyen terme, en faisant varier les variables de base concernant l'offre de transports collectifs pour les liaisons concernées.

#### 6.4. Les limites économétriques

Ces limites concernent le fait que l'on raisonne toujours «toutes choses égales par ailleurs». Le fait qu'un certain nombre de variables explicatives soient présentes au sein de ces modèles ne fait pas sortir de ce type de raisonnement. Ainsi, on réfléchira toujours de la même manière, même si le nombre de variables est conséquent, certaines variables n'ayant de toute façon pas été intégrées dans la modélisation. La procédure de validation désagrégée permettra de vérifier le comportement du modèle au regard d'autres variables non introduites dans la modélisation.

#### 6.5. Les limites des procédures décalage

D'un point de vue pratique, on ne peut jamais dire que les résultats issus de la procédure de calage du modèle sont les meilleurs, et ceci qu'elle que soit la procédure de calage. Ainsi, si l'on considère la notion de formulation d'un modèle, telle que :

$$M=F(\theta,X)$$

avec :

M : modèle

F : mécanisme causal

$\theta$  : paramètres

X : entrées

Dans le cas d'un modèle désagrégé, les paramètres  $\theta$  sont calés à partir d'un échantillon de  $n$  individus dont nous connaissons les caractéristiques X ainsi que le choix. La procédure de calage consiste à résoudre un problème d'optimisation :

$$\underset{\theta \in \mathfrak{R}}{\text{Minimiser}} (Obs - F(\theta))^2$$

où Obs représente l'observation, et  $F(\theta)$  la prévision du modèle pour les paramètres  $\theta$ .

Ceci peut correspondre à un calage à partir du principe des moindres carrés (l'écart entre l'observé et le prédit est minimisé) ou à partir du maximum de vraisemblance. La procédure de calage se résolvant par

une méthode d'optimisation, un problème intervient. En effet, rien ne permet d'affirmer que son résultat est le meilleur (on obtient un optimum local et non un optimum global). Pour s'en convaincre, il suffit de partir d'un jeu de paramètre  $\theta_1$ , et de générer un échantillon, ainsi que le choix de cet échantillon en se servant du modèle et des paramètres. Ensuite, à partir de l'échantillon généré, nous pouvons essayer de retrouver nos paramètres par une procédure de calage. Nous avons de fortes chances de tomber sur un jeu de paramètre  $\theta_2$ , qui d'une part sera différent de  $\theta_1$ , et d'autre part ne reproduira pas notre situation de départ (connue), alors que les paramètres  $\theta_1$  existent bien et qu'ils reproduisent exactement la situation. Ceci est dû à la non convexité du modèle.

## Conclusion

Le principal objet des modèles de choix discret consiste en l'analyse et en la prévision des comportements de déplacements. Ce type de modèle peut satisfaire un grand nombre d'attentes : toute procédure de choix peut ainsi être modélisée, de façon très poussée, ou plus «générale». On pourra aller jusqu'à réaliser une modélisation qui reprend les trois étapes de la modélisation séquentielle (en combinant quelques unes éventuellement), et chercher par la suite à effectuer des affectations robustes allant à un niveau très fin. Au contraire, on pourra s'en tenir à la recherche de fonctions d'utilité à l'aide d'enquêtes de type préférence révélées ou déclarées. En ce qui concerne la prise en compte des fonctions d'utilité et leur exploitation du point de vue de la procédure de choix, on préférera, pour des raisons «pratiques», utiliser une formulation logit.

Dans tous les cas, la maîtrise du processus de modélisation par le chargé d'études reste un élément central. En dehors de la récolte et de la préparation des données, il est nécessaire de porter une attention soutenue aux phases de spécification (sélection et combinaison des variables explicatives, forme mathématique des variables) et d'estimation des modèles<sup>28</sup>. Dans ce dernier cas, la prise en considération des problèmes pouvant intervenir dans la modélisation n'est pas toujours évidente. Par ailleurs, il convient de faire attention à ne pas s'en tenir aux résultats fournis par les logiciels de modélisation. Il faut chercher à apprécier la qualité des modèles obtenus, avant toute interprétation.

---

28. Il convient de rappeler le fait que la définition des scénarios doit se faire ex ante dans ce type de modélisation.





## Deuxième partie

# Approche pratique

## Introduction

Cette partie présente la méthodologie et les résultats de l'étude de modélisation des comportements de déplacement en milieu urbain, entreprise sous la maîtrise d'ouvrage du Certu et de l'Ademe, sous la direction technique du Certu, et avec maîtrise d'oeuvre de la société MVA Consultants. Cette étude a eu pour site d'application l'agglomération de Lyon et s'est déroulée de 1995 à 1997.

Dans un premier temps, une description précise des modèles réalisés est présentée. Elle s'attache à expliquer la structure de la chaîne de modélisation retenue, ainsi que les différentes étapes la constituant.

Dans un deuxième temps, une description du fichier de calibrage est fournie. Ce fichier est la base de données à partir de laquelle les modèles sont construits. Sa mise au point représente un travail important : mise en adéquation des données concernant la voirie, les transports collectifs, les données socio-économiques (caractéristiques des ménages, emplois, population...) et des caractéristiques de zones (équipements publics, surfaces commerciales, superficies...).

Dans un dernier temps, les résultats des modèles et plusieurs scénarios sont présentés.

## 1. Contexte et objectifs de l'étude lyonnaise

### 1.1. Les modèles désagrégés

Plusieurs modèles désagrégés de choix existent en France. Parmi ces modèles, ceux ayant été appliqués sont :

- Impact, pour la RATP : modèle de choix de mode et de destination pour l'Île-de-France, version 3 ;
- pour le District de Rennes, un modèle de choix modal ;
- pour la DDE 67, des modèles de motorisation, génération et choix de mode et de destination pour la région Alsace ;
- à Metz, un modèle de choix modal dans le contexte du PDU (plan de développement urbain).

D'autres modèles de ce type sont actuellement en développement dans le cadre de diverses études de projet. Ils ont été lancés après l'étude commanditée par le Certu et présentée dans ce rapport (PDU de Rouen et de Saint-Etienne...).

Les principales raisons de l'utilisation accrue de cette modélisation sont la qualité statistique des estimations, la flexibilité dans l'utilisation des données et la variété des domaines d'application. Notons également que ces modèles sont considérés comme «comportementaux» dans la mesure où les segmentations, variables et coefficients utilisés sont directement issus du processus d'estimation des modèles et non pré-spécifiés, et que l'unité statistique de base est l'individu.

De plus, les modèles désagrégés offrent l'avantage de pouvoir intégrer des sources de données différentes, comme les données externes provenant d'enquêtes de type «préférences déclarées». Ces données peuvent être intégrées à celles des enquêtes ménages et utilisées conjointement pour estimer des modèles faisant intervenir des dimensions plus qualitatives. On peut aussi y inclure des estimations de paramètres qui ne sont pas directement observables ou trop corrélés dans les données observées.

Enfin, l'approche peut être utilisée pour incorporer de nombreuses variables caractéristiques des alternatives de choix et donc déboucher sur des applications permettant de tester une large gamme de politiques de transport.

## 1.2. Les objectifs de l'étude

La présente étude se place dans une perspective d'amélioration des outils de prévision existants, destinés à évaluer des politiques de transport et d'aménagement dans les grandes agglomérations. Dans cette optique, le choix s'est porté sur le développement d'un modèle désagrégé de choix de mode et de destination combiné. Pour rendre ce modèle applicable, il faut l'intégrer dans une structure séquentielle. Pour étoffer ses possibilités d'utilisation, des modèles de motorisation et de génération ont été également développés : la prise en compte de la motorisation des ménages permet en effet, par exemple, d'évaluer des politiques liées à l'environnement.

La souplesse du modèle, c'est-à-dire ses possibilités d'évolution mais aussi sa transférabilité, représente un point clé de l'étude. En effet, si l'agglomération de Lyon a été choisie, les modèles développés ici doivent pouvoir s'appliquer à d'autres agglomérations de taille similaire, de manière relativement aisée.

Pour assurer cette transférabilité, deux dimensions de base doivent être prises en compte :

- l'utilisation d'un cadre de réflexion qui permette de couvrir les différentes préoccupations en matière de planification des transports et des problématiques associées (aménagement, environnement...),
- l'emploi de données aussi « courantes » que possible, afin de s'assurer qu'elles puissent être récoltées dans les autres agglomérations.

## 2. La chaîne de modélisation

### 2.1. Structure de la chaîne de modélisation

Le modèle de choix de mode et de destination, qui constitue le cœur de l'étude, fait en réalité partie d'une chaîne de modélisation plus globale. Cette chaîne comprend un modèle de motorisation, un modèle de génération puis le modèle de choix de mode et de destination proprement dit. Ce dernier est suivi de l'affectation des matrices origine-destination sur les réseaux. Cette structure est résumée dans la figure de la page suivante.

**Trois classes de modèles constituent cette chaîne :**

- Un modèle de motorisation : ce modèle permet d'estimer par type de ménage le nombre de voitures possédées par le ménage ; la moto-

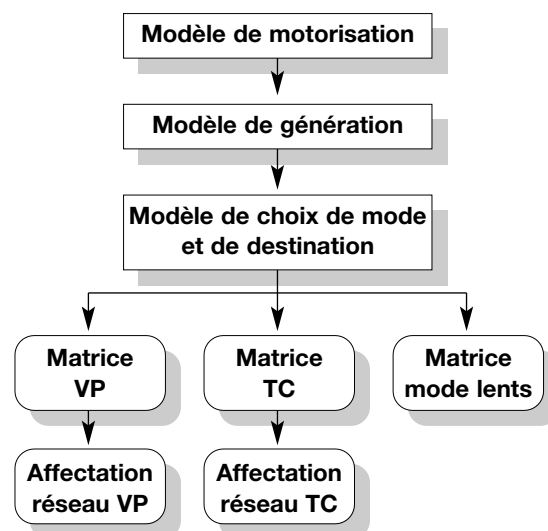
risation des ménages joue en effet un rôle important en terme de déplacements générés ainsi qu'en terme de choix de mode. Le nombre de voiture par ménage ainsi modélisé, sera donc utilisé comme variable explicative dans les étapes suivante.

— Un modèle de génération : ce modèle estime le nombre de déplacements effectués par les ménages suivant leur type (âge de la personne de référence, nombre d'actifs, motorisation, etc.) et pour différents motifs. Pour chaque motif, ce modèle estime donc le nombre de déplacements générés par chacune des zones du domaine de l'étude,

— Un modèle de choix de mode et destination : ce modèle est utilisé pour distribuer l'ensemble des déplacements générés lors de l'étape précédente, dans l'univers des modes disponibles et des destinations possibles; il constitue la dernière étape de la chaîne de modélisation.

Le modèle de génération distingue sept motifs pour la période de pointe du soir : domicile-travail, domicile-école, domicile-loisirs, domicile-achats, domicile-autres, déplacements secondaires et affaires professionnelles. Le modèle de choix de mode et de destination distingue ces mêmes motifs en période de pointe du soir et en période creuse, sauf pour le motif affaires professionnelles, qui est estimé sur l'ensemble des deux périodes.

**Figure 1 : Structure de la chaîne de modélisation**



## 2.2. Le modèle de motorisation

L'objectif du modèle de motorisation est d'estimer le nombre de voitures (ou niveau de motorisation) que les ménages possèdent suivant leur type (taille du ménage, revenu...) et suivant leur lieu d'habitation (centre ville, proximité des TC...). Le niveau de motorisation conditionne en effet fortement le nombre de déplacements effectués ainsi que le mode et la destination choisis.

On comprend donc, par exemple, que le modèle de motorisation est particulièrement indispensable pour tester certaines politiques liées à l'environnement :

- il permet, d'une part, de tenir compte de la typologie des ménages et donc de son évolution supposée;
- il permet, d'autre part, de prendre en compte l'impact de la qualité de l'offre TC (essentiellement l'accessibilité) sur la motorisation des ménages et donc sur leur utilisation de la VP.

Il est donc important d'avoir estimé auparavant le niveau de motorisation d'une population. Cela se fait à l'aide d'un modèle de motorisation de type désagrégé dit «stop and go». Ce modèle est construit à partir de l'échantillon constitué de l'ensemble des ménages recensés dans l'enquête ménage de Lyon 1995.

D'autre part, les raisons amenant un ménage à posséder au moins une voiture de celles l'amenant à posséder au moins deux voitures ont été différenciées. Cette méthodologie entraîne l'estimation de deux «sous-modèles» :

- un modèle estimant le nombre de ménages possédant au moins une voiture,
- un deuxième modèle estimant parmi ces ménages, le nombre de ceux possédant au moins une deuxième voiture.

Cette différenciation à l'estimation entre l'utilité de posséder une ou deux voitures et plus est dite de type «stop and go».

Chacune des alternatives possibles (posséder 0, 1, 2 voitures ou plus) est représentée mathématiquement par une fonction d'utilité. Cette fonction d'utilité est de la forme :

$$U = \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_n X_n$$

Les  $x_i$  sont les variables explicatives choisies telles que le revenu, le nombre d'actifs dans le ménage, l'accessibilité au TC, etc.

Les  $\theta_i$  sont les paramètres associés à ces variables. Leur valeur est obtenue à l'estimation de manière à expliquer le maximum d'observations de l'échantillon.

La connaissance de ces utilités permet ensuite d'estimer la probabilité pour un ménage de posséder un nombre de voitures  $i$ . Cette probabilité  $P_i$  s'exprime, par exemple ainsi pour le modèle estimant l'opportunité de posséder 0 ou 1 voiture et plus :

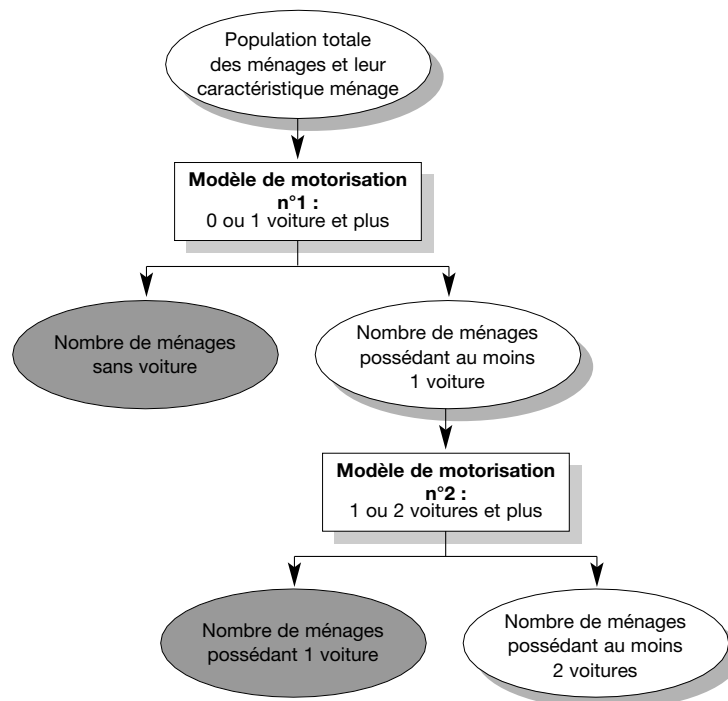
$$P_i = \frac{e^{U_i}}{(e^{U_0} + e^{U_1})}$$

Ici,  $i$  est égal à 0 ou 1 (pour 0 voiture et 1 voiture et plus).

$U_0$  est l'utilité de ne posséder aucune voiture,  $U_1$  est celle d'en posséder au moins une.

La structure globale du modèle de motorisation est donnée dans la figure suivante.

**Figure 2 : Structure du modèle de motorisation**





### 2.3. Le modèle de génération

Le but de ce modèle est d'estimer, par motif, le nombre de déplacements générés par chacune des zones. Il constitue la deuxième étape de la chaîne de modélisation et s'appuie notamment sur les données de motorisation estimées auparavant. Du fait du manque de certaines données sur Lyon, la génération a été estimée à l'aide de coefficients issus des enquêtes ménages.

### 2.4. Le modèle de choix de mode et de destination

La troisième et dernière étape de la modélisation est assurée par le choix de mode et de destination: il s'agit en fait de distribuer les déplacements générés dans l'univers des choix de mode et de destination possibles. L'intérêt de la démarche entreprise ici est de lier le choix de mode et le choix de la destination, en estimant la distribution des déplacements à l'aide d'un modèle désagrégé.

Trois types de modes ont été distingués après analyse de l'enquête ménages. Le mode voiture particulière (VP), qui regroupe les déplacements effectués en VP conducteurs et passagers, ainsi que les 2 roues de plus de 50 cm<sup>3</sup>. Le mode transports collectifs (TC), qui comprend les déplacements effectués en bus seul, métro seul et autres TC. Le mode lent (ou modes alternatifs) qui regroupe les déplacements effectués en marche, bicyclette et 2 roues inférieurs à 50 cm<sup>3</sup>.

Ce modèle s'appuie sur un échantillon d'observations constitué de l'ensemble des déplacements recensés dans l'enquête ménage pour la tranche horaire considérée (période de pointe ou période creuse). Chaque observation de cet échantillon représente donc le choix effectué par un individu en termes de mode et de destination pour un déplacement dont le motif est connu. Ce choix représente une alternative qui est confrontée, lors de l'estimation, à un ensemble d'alternatives différentes. La construction de cet univers d'alternatives possibles est décrite dans le chapitre suivant: elle intervient dans l'élaboration du fichier de calibrage.

De même que pour le modèle de motorisation, une alternative est représentée mathématiquement par une fonction d'utilité. Ces fonctions d'utilité ont la forme suivante:

$$U = A + t.B$$

Avec :

$$- A = \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_n X_n$$

$$- B = \ln(X_{n+1} + e^{\theta_{n+2} X_{n+2}} + \dots + e^{\theta_n X_n})$$

— t est un coefficient de taille.

**A** comprend les variables liées aux choix de mode (temps de parcours, rabattement, etc.) ; ce terme est avant tout une mesure de l'offre de transport. Mais il prend également en compte la motorisation du ménage; en effet, le taux de motorisation a un effet direct sur l'utilisation préférentielle de la VP.

Les variables comprises dans le terme **B** sont des variables dites «de taille» et caractérisent la zone de destination : ce sont par exemple le nombre d'emplois, la population, etc. La nature de ces variables varie avec le motif de déplacement ; **B** est donc une mesure de l'attractivité de la zone de destination.

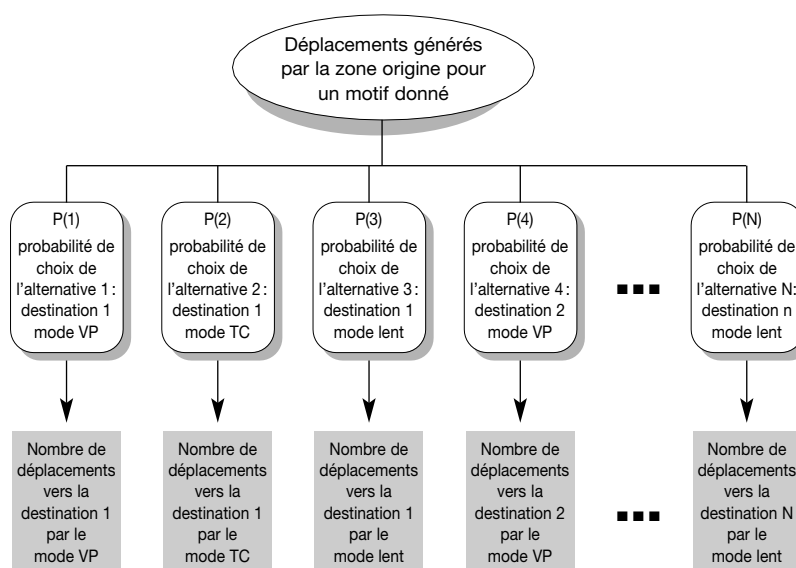
L'estimation conduit à l'obtention de valeurs pour les paramètres  $\theta_i$ . Les fonctions d'utilité estimées pour chacune des alternatives permettent de calculer la probabilité  $P_{d1m}$ , de choisir la destination d avec le mode m. Cette probabilité s'exprime sous la forme :

$$P_{d1m} = \frac{e^{U_{d1m}}}{\sum_{ij} e^{U_{ij}}}$$

où (i,j) décrit toutes les alternatives mode x destinations possibles

La structure du modèle de choix de mode et de destination est représentée dans la figure suivante :

**Figure 3 : Structure du modèle de choix de mode et de destination**



### 3. Les données et leur traitement

#### 3.1. Le caractère déterminant du traitement des données

La qualité des résultats issus des modèles est toujours tributaire de la qualité des données utilisées. Cette réalité est particulièrement sensible pour le modèle de choix de mode et de destination. En effet, celui-ci ne s'appuie pas uniquement sur les données d'enquête ménage, puisqu'il nécessite une description de l'offre de transport (tous modes) ainsi qu'une description de l'attractivité des zones. Dans ce chapitre, le recueil des données pour le modèle joint de choix de mode et de destination ainsi que le traitement de ces données est présenté.

### 3.2. Éléments de cadrage généraux

L'estimation et l'application de la chaîne de modélisation ont été effectués sur l'agglomération de Lyon. Il convient de rappeler que l'objectif n'était pas de réaliser un modèle propre à Lyon mais pouvant se transposer à d'autres villes de taille identique. Un souci de simplicité dans l'utilisation des données était donc omniprésent : une prise en compte de données encore plus spécifiques que celles utilisées permettrait certainement d'élaborer un modèle plus précis, pour Lyon, mais difficilement transférable.

La première étape dans l'élaboration d'un modèle de transport est d'établir un zonage géographique de l'espace sur lequel sont recensés les déplacements. Dans notre cas, le zonage de base utilisé est le zonage le plus fin de l'enquête ménage de Lyon en 1995 <sup>29</sup>.

Ce zonage donne en effet satisfaction pour quatre raisons principales :

- il représente un découpage très fin de l'agglomération et permet d'obtenir des informations précises sur l'offre de transport (accessibilité aux TC par exemple),
- il est composé de zones a priori homogènes en termes de caractéristiques socio-économiques,
- il permet d'utiliser les données de l'enquête ménage à l'état brut et, par conséquent, évite d'introduire des biais inhérents à un changement de zonage,
- il provient d'une enquête ménage ; ces enquêtes sont maintenant très répandues dans les agglomérations françaises de taille moyenne, l'utilisation de ce zonage participe donc à une transférabilité plus aisée.

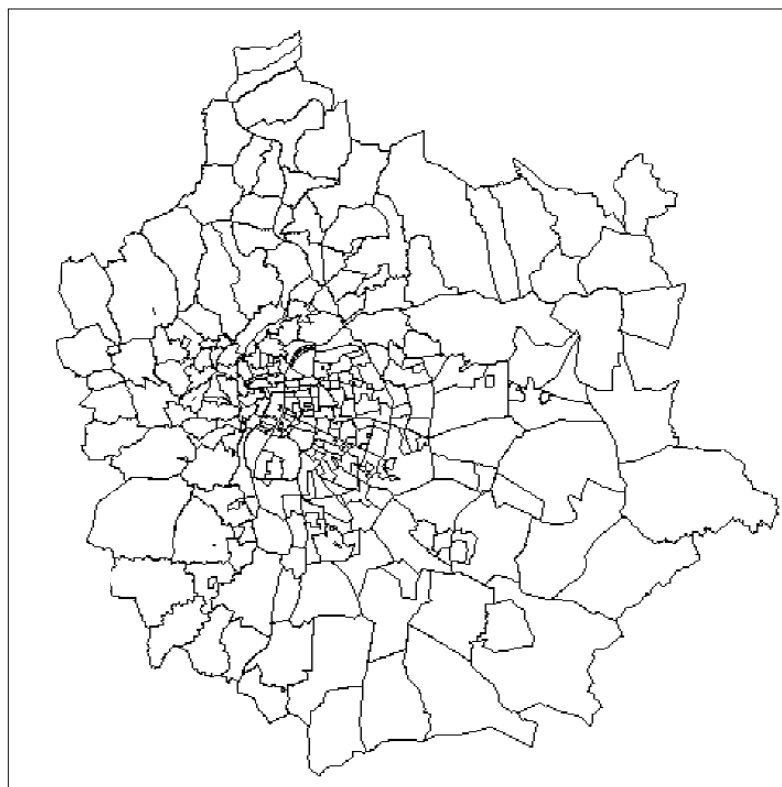
Pour autant, seule une partie de ce zonage a pu être utilisée. En effet, le réseau de voirie fourni par le Cete de Lyon ne recouvre pas totalement le zonage de l'enquête ménage, composé de 357 zones : quelques zones de grande banlieue ont donc été mises à l'écart. Le zonage utilisé ne comprend lui que 310 zones. Le tableau et la figure, page suivante, résument ses caractéristiques.

29. Soit 444 zones, modifié en 357 zones pour être en conformité avec les directives édictées par la CNIL.

**Tableau 1 : Données sur le zonage utilisé**

Zonage	Nombre de zones	Nombre de déplacements pondérés recensés en période de pointe du soir
Enquête ménage	357	1 136 033
Zonage utilisé	310	1 025 370
Différence	47	110 663

**Figure 4 : Carte du zonage de l'agglomération lyonnaise utilisé pour l'estimation des modèles**



### 3.3. L'utilisation du logiciel Trips pour l'estimation de l'offre de transport

Le logiciel Trips a été utilisé pour caractériser l'offre de transport en terme de temps de parcours pour chaque mode. Ceci a nécessité l'importation du réseau de voirie fourni par le Cete de Lyon et le recodage des zones à considérer (les 310 zones de l'enquête ménage recouvertes par ce réseau).

Pour décrire l'offre de transport en mode TC, le réseau des Transports Collectifs de Lyon a été codé sur le réseau du Cete de Lyon et les lignes en site propre ont été rajoutées (métro, funiculaire). La Société Lyonnaise des Transports en Commun (SLTC), avait aimablement mis à notre disposition les données concernant son réseau.

Ce travail préliminaire effectué, l'estimation des temps de parcours zone à zone pour chaque mode était alors possible. Des images du réseau VP et TC tels qu'ils apparaissent dans Trips sont présentées par la suite.

#### Le mode VP

Pour le mode VP, le travail a consisté à affecter la matrice de déplacements heure de pointe du soir (HPS) fournie par le CETE de Lyon sur le réseau. Cette affectation a été réalisée de manière itérative avec contrainte de capacité jusqu'à convergence.

À la première itération, les déplacements sont donc affectés sur le réseau de sorte que, pour chacun d'entre eux, ce coût soit minimal. Les vitesses considérées alors sont les vitesses à vide. Cette affectation conduit à estimer des volumes de déplacements sur le réseau de voirie.

Ces volumes, comparés aux capacités des voies, permettent à leur tour le calcul de nouvelles vitesses. Au cours de la deuxième itération, la matrice de déplacements initiale est donc réaffectée sur le réseau en tenant compte de ces nouvelles vitesses, et toujours en minimisant le coût VP défini précédemment. Cette itération conduit à d'autres volumes, etc. Au bout d'une vingtaine d'itérations de ce type, un équilibre est atteint, de sorte que chacun des déplacements utilise le chemin minimisant son coût. Les vitesses finales de chacune des voies permettent alors le calcul des temps de parcours pour l'heure de pointe du soir zone à zone, ainsi que de la distance des chemins parcourus.

Le travail a été effectué de la même façon en période creuse.

### Le mode TC

Pour l'estimation de l'offre TC, nous avons tenu compte des temps de parcours effectivement réalisés par les bus et métros sur chacune de leur ligne en période de pointe du soir (16h00 - 18h30) et en période creuse (9h00-16h00). Les fréquences réelles ont également été prises en compte: celles-ci permettent notamment de calculer les temps d'attente.

Pour ce mode, le coût d'un trajet tel qu'il est ressenti lors d'un déplacement a été estimé en pénalisant, en temps, le temps de marche à pied, le temps d'attente ainsi que le nombre de correspondances. Les pénalités considérées sont résumées dans le tableau suivant :

**Tableau 2 : Pénalités pour le mode TC** (source : MVA Consultants)

Segment du déplacement	Pénalités
Temps dans le véhicule	Non pénalisé
Temps d'attente	Multiplié par 2
Temps de marche à pied (liaison)	Multiplié par 1.5
Correspondance	Pénalité de 5 minutes

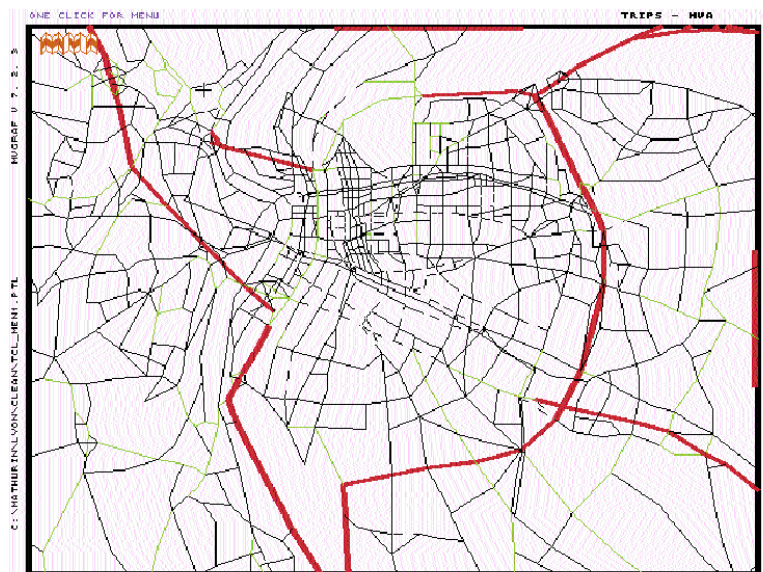
Le calcul du chemin pour joindre deux zones en mode TC a donc été établi de sorte qu'il minimise le temps pénalisé (cf. tableau 2) de parcours total. Une fois le chemin connu, nous retenons alors uniquement les temps réels (non pénalisés). Ces temps permettent une description du niveau de service existant en mode TC pour se rendre d'une zone à une autre.

Notons que l'accessibilité au réseau n'est pas prise en compte à cette étape : le paragraphe 3.4. décrit la méthodologie employée à cet effet.

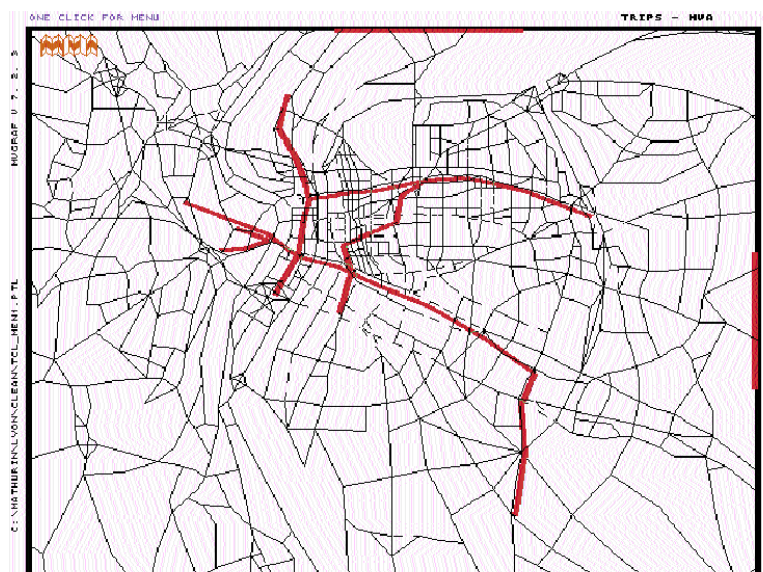
### Les modes alternatifs

La représentation de l'offre mode alternatif a été estimée très simplement en considérant le chemin le plus court en distance entre deux zones : l'utilisation des modes alternatifs (marche à pied, vélo) est en effet presque uniquement liée à la distance de parcours.

**Figure 5 : Réseau VP dans Trips (réseau autoroutier et périphériques en rouge)**



**Figure 6 : Réseau TC dans Trips (réseau de métro et funiculaire en rouge)**





### 3.4. L'accessibilité aux TC : utilisation de MapInfo

#### Pourquoi utiliser un Système d'Information Géographique ?

L'estimation de l'accessibilité au réseau TC a fait l'objet d'un travail particulier. En effet, il apparaît que cette accessibilité est d'un poids important dans le choix d'utiliser ou non les transports en commun. De manière évidente, le mode TC ne sera choisi pour un déplacement que s'il est accessible. On comprend donc que, de manière moins brutale, la facilité d'accès à un réseau TC est une donnée primordiale en terme de politique de transport. Pour cette raison, une attention particulière a été portée à ce point pour représenter au mieux la réalité des choses.

De manière courante, cette accessibilité est comprise dans le temps de rabattement estimé lors du calcul des plus courts chemins (en temps pénalisés) effectué à l'aide de logiciels tel que Trips (cf. paragraphe précédent). Cette estimation du temps de rabattement a ses limites puisqu'elle dépend très fortement de la manière dont les zones sont connectées au réseau. D'autre part, la zone étant représentée physiquement par un point situé souvent en son centre (appelé centroïde), le temps d'accessibilité est un temps moyen ne tenant pas compte de la distribution interne de la population.

Pour pallier ces lacunes, le choix a été fait d'estimer l'accessibilité en utilisant un Système d'Information Géographique (SIG), ici le logiciel MapInfo.

#### Principe d'utilisation d'un SIG

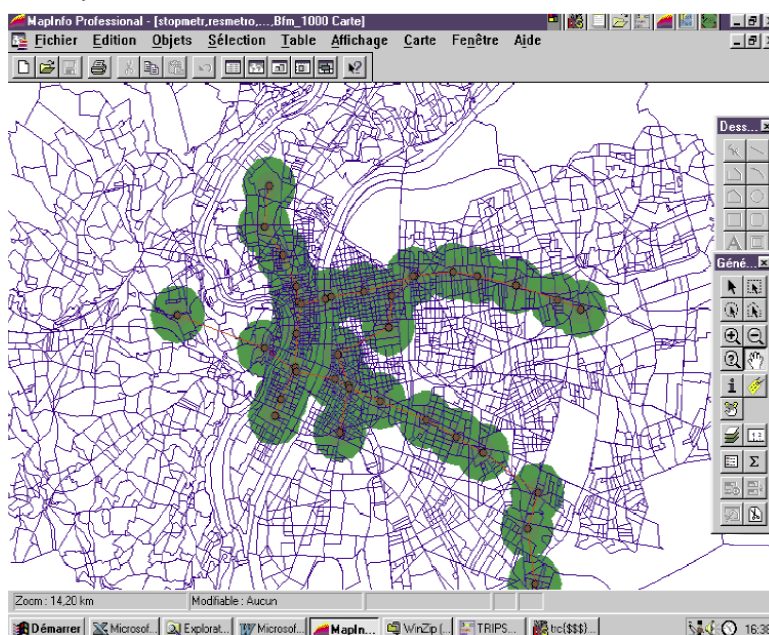
Le principe est le suivant : la représentation du réseau TC de la communauté urbaine de Lyon est superposée à la carte des îlots Insee correspondant aux 310 zones sur lesquelles le modèle est estimé. Chacun des 5 175 îlots est renseigné par sa population et son nombre d'emplois. Il est ensuite possible d'estimer la population (ou le nombre d'emplois) située à moins d'une certaine distance  $X$  du réseau TC :

- si un îlot est totalement compris dans la région située à moins de  $X$  mètres du réseau, sa population l'est évidemment aussi,
- si un îlot est totalement en dehors de cette région, la totalité de sa population est éloignée de plus de  $X$  mètres du réseau,
- si seule une partie d'un îlot est située à moins de  $X$  mètres du réseau, la population considérée à moins de  $X$  mètres est calculée au prorata de la surface de l'îlot recouverte : ceci signifie en fait que l'on considère la densité de population comme étant uniforme à l'intérieur de l'îlot. Cette hypothèse n'est pas très pénalisante du fait de la très petite taille d'un îlot.

Ces données recueillies à l'îlot pour plusieurs distances sont ensuite agrégées au niveau zonal (310 zones): la population (ou le nombre d'emplois) située à moins d'une certaine distance du réseau TC est donc connue avec une précision appréciable. Il est alors possible de calculer un temps moyen de rabattement par zone de manière correcte. Il est aussi possible de garder ces données brutes dans le modèle comme des données zonales à part entière: population à moins de 100 mètres, 250 mètres, 500 mètres..., du réseau TC. Les paramètres associés à ces variables seront alors calculés lors de l'estimation.

La figure suivante donne l'exemple de l'utilisation de MapInfo pour calculer l'accessibilité aux stations de métro à Lyon; des «filaires» ont été utilisés pour les lignes de bus c'est-à-dire que l'on calcule l'accessibilité le long de la ligne et non pas à chaque arrêt.

**Figure 7 : Accessibilité aux stations de métro (les cercles représentent la population à moins de 100 m, 200 m, 300 m... d'une station)**



### 3.5. Les données zonales

Un modèle de choix de mode et de destination nécessite, d'une part, la description de l'offre de transport pour chacun des modes retenus et, d'autre part, une description de l'offre en terme de destinations possibles. Pour cela, les zones sont caractérisées en utilisant des données socio-économiques susceptibles d'expliquer les comportements observés.

Ainsi, par exemple, la donnée zonale «surface commerciale» permet de bien représenter le caractère attractif d'une zone pour le motif domicile-achat. Malheureusement, pour un certain nombre de motifs qui représentent parfois beaucoup de déplacements, il s'avère difficile d'obtenir des données zonales explicatives en terme de choix de destination. Ces raisons sont diverses :

— Les données zonales peuvent exister à un niveau zonal plus grossier. Par exemple, sur Lyon, la majorité des données n'étaient disponibles qu'au niveau du zonage 87 zones de l'enquête ménage (hormis la population et l'emploi total), alors que nous travaillions au niveau 357. Il était donc nécessaire de désagréger ces données.

— Le zonage 357 de l'enquête ménage comprend des zones dites «générateurs ponctuels» correspondant à des centres commerciaux, des hôpitaux, des parcs, des lycées, etc. La désagrégation de données en devient donc plus compliquée.

— Dans des cas plus extrêmes, les données utilisables n'existent pas : le motif domicile-loisirs nécessite une donnée zonale de type équipement. Or, les données disponibles dans ce cas sont de mauvaise qualité. De même, les données sur le stationnement n'existent pas en terme de nombre de places payantes et non payantes de surface.

— Enfin, certains motifs représentant en volume beaucoup de déplacements ne peuvent être mis en relation à une donnée zonale de façon simple : comment caractériser la destination d'un déplacement pour motif secondaire ?

Pour ces raisons, le choix s'est porté sur l'utilisation de données zonales simples et de bonne qualité (population et emploi surtout car ces données étaient disponibles au niveau de l'îlot). Ainsi, les modèles obtenus peuvent être considérés comme solides même s'il ne peuvent tout expliquer. D'autre part, ils seront transposables plus aisément : une donnée zonale de mauvaise qualité sur Lyon risque d'être de mauvaise qualité ailleurs.

Le tableau suivant présente les données ayant été utilisées pour caractériser les zones en fonction des motifs considérés.

**Tableau 3 : Données zonales utilisées comme variables de taille**

Motifs	Variables de taille
Domicile-travail	Population, emplois
Domicile-école	Population, nombre de lycéens et de collégiens
Domicile-achats	Population, emplois et surface commerciale
Domicile-loisirs	Population, emplois
Domicile-autres	Population, emplois
Déplacements secondaires	Population, emplois
Affaires professionnelles	Population, emplois

### 3.6. La constitution du fichier de calibrage

Toute l'acquisition de ces données et de leur traitement constitue donc le travail préparatoire à la constitution d'une base de données, sur laquelle le modèle de choix de mode et de destination peut être estimé. La structure de cette base de données est aussi très importante et sera décrite ci-après. Le principe, pour construire cette base de données, ou fichier de calibrage, est le suivant : chaque observation recensée (chaque déplacement de l'enquête ménage) doit être confrontée à un certain nombre d'alternatives en termes de choix de mode et de destination, chacune de ces alternatives, dont celle que la personne a effectivement choisie, étant décrite aussi précisément que possible avec les données explicitées dans les chapitres précédents.

#### Contenu de chaque observation

Chaque observation comprend :

- les renseignements concernant la personne et son ménage : sexe, âge, profession, nombre de voitures à disposition, revenu du ménage, etc. Ces renseignements sont directement issus de l'enquête ménage,
- les informations liées à la nature du déplacement : motif, mode utilisé, période de la journée, zone d'origine, zone de destination, etc. Ces informations sont aussi issues de l'enquête ménage,
- la description de 30 alternatives, ces 30 alternatives correspondant à 10 destinations différentes combinées aux 3 modes considérés (VP, TC, mode alternatif). L'une de ces alternatives correspond au choix effectif, les autres aux choix qu'aurait pu faire chaque individu. Le fait de construire un échantillon d'alternatives de la sorte n'a pas d'impact

sur la qualité du modèle<sup>30</sup>. Nous cherchons ici à établir des lois de comportements qui seront ensuite appliquées à toutes les alternatives possibles (3 modes x 310 zones).

### Les alternatives de destination

Les 10 alternatives de destination doivent représenter le mieux possible l'univers de choix offert à l'individu pour son déplacement. A priori, il existe pour le zonage considéré ici, 310 alternatives possibles en terme de destination ; cet univers de choix étant beaucoup trop lourd pour pouvoir estimer un modèle, nous considérons donc pour chaque observation un échantillon de 10 d'entre elles, en sachant que la zone effectivement choisie appartient à cet échantillon. Cet échantillon est construit sur la base de quatre strates :

- la zone d'origine correspond à la strate 1 ; elle est toujours prise en compte comme alternative,
- 3 zones sont tirées au sort à l'intérieur de la strate 2 qui comprend les 10 zones voisines de la zone d'origine ; ce tirage au sort est réitéré pour chaque observation,
- 3 zones sont tirées au sort dans la strate 3 ; cette troisième strate est constituée des zones du centre non comprises dans les deux premières strates ; là encore, le tirage aléatoire est répété pour chaque observation,
- 3 zones sont finalement choisies au hasard parmi les zones de la strate 4 qui rassemble le reste des zones, ce tirage étant lui aussi renouvelé à chaque observation.

Le tableau suivant résume la stratification des zones et leur représentation dans l'échantillon des 10 alternatives :

**Tableau 4 : Échantillonnage du choix de destination**

Type de zone	Numéro de strate	Taille de la strate	Représentation dans l'échantillon
Zone origine	1	1 zone	1 zone
Zone voisine	2	10 zones	3 zones au hasard
Zone du centre	3	50 zones	3 zones au hasard
Reste des zones	4	250 zones	3 zones au hasard

30. [Watanatada, (T.) et Ben-Akiva (M.); 1979].

### Le choix de la stratification des zones

Le choix de la précédente stratification des zones est lié à la spécificité des déplacements en ville : un déplacement est souvent local ou dirigé vers le centre. Ces deux types de déplacement sont donc bien représentés ici (7 zones de l'échantillon sur 10). Il convient de noter également que, statistiquement, le tirage aléatoire de 3 zones pour chaque observation dans les strates 2, 3 et 4 s'avère suffisant pour estimer le modèle, au vu du nombre d'observations présentes dans le fichier de calibrage.

Le tableau suivant donne le nombre d'observations par période horaire et par motif :

**Tableau 5 : Nombre d'observations dans le fichier de calibrage**

Motifs	Période horaire	Nombre d'observations utilisées
Domicile-travail	PPS (16h00 – 18h30)	1 452
Domicile-école	PPS	1 495
Domicile-loisirs	PPS	1 147
Domicile-achats	PPS	1 585
Domicile-autres	PPS	3 596
Déplacements secondaires	PPS	2 394
Affaires professionnelles	PPS	284
Domicile-travail	PCR (9h00 – 16h00)	2 515
Domicile-école	PCR	2 313
Domicile-loisirs	PCR	1 385
Domicile-achats	PCR	3 733
Domicile-autres	PCR	5 081
Déplacements secondaires	PCR	4 616
Affaires professionnelles	PCR	865

### Le «redressement» des fonctions d'utilité

Une fois l'estimation du modèle effectuée à partir du fichier de calibrage ainsi construit, il convient tout de même de «redresser» les fonctions d'utilité obtenues par un terme correctif prenant en compte la sous-estimation ou la surestimation du choix d'une strate de zones.

Cette correction est indispensable lorsque l'on applique le modèle : en effet, cette application s'effectue sur l'ensemble des zones (310 au total) et les strates retrouvent donc leur taille réelle. Les fonctions d'utilité sont corrigées comme suit :

$$U_i(\text{corrigée}) = U_i(\text{estimée}) - \ln \left( \frac{S_i}{s_i} \right)$$

$U_i$  est l'utilité correspondant à une zone de destination appartenant à la strate  $i$ ,

$S_i$  est le nombre d'alternatives total de la strate  $i$ ,

et  $s_i$  est le nombre d'alternatives tirées dans la strate  $i$  pour construire l'échantillon.

Le terme correctif  $\ln (S_i/s_i)$  prend les valeurs suivantes :

**Tableau 6 : Termes de correction dus à la stratification de l'échantillon**

Type de zone de destination	Terme correctif dans les fonctions d'utilité
Zone origine	0
Zone voisine	-1,204
Zone du centre	-2,813
Autre zone	-4,423

Les modèles de choix de mode et de destination estimés lors de cette étude prennent en compte des constantes de destination pour les strates «zone origine», «zones voisines» et «zones centre» ; la constante associée à la strate «autres zones» est nulle. Pour cette raison, les fonctions d'utilités ont été corrigées de manière à garder un terme constant nul pour le choix de destination «autres zones». Les termes de correction prennent alors les valeurs suivantes :

**Tableau 7 : Termes de correction utilisés**

Type de zone de destination	Terme correctif dans les fonctions d'utilité
Zone origine	4,423
Zone voisine	3,219
Zone du centre	1,609
Autre zone	0

## 4. Les résultats des modèles

### 4.1. Modèle de motorisation

#### Premier modèle de motorisation

Pour le premier modèle de motorisation estimant l'opportunité des ménages à posséder 0 ou 1 voiture et plus, les variables explicatives suivantes ont été utilisées :

- le revenu du ménage (revenu mensuel en milliers de francs),
- l'accessibilité au TC (temps d'accès au réseau TC en minutes),
- le nombre d'actifs.

L'utilité de ne posséder aucune voiture a été prise égale à 1. Le modèle estime donc l'utilité de posséder au moins une voiture par rapport à cette valeur.

Les résultats obtenus pour l'estimation de l'utilité à posséder au moins 1 voiture sont donnés dans le tableau suivant :

**Tableau 8 : Modèle de motorisation n°1 <sup>31</sup>**

Paramètres	Estimations	T ratio
Constante ( $\theta_0$ )	-0.8559	-7.7
Paramètre lié au revenu du ménage ( $\theta_1$ )	0.3274	15.6
Terme correctif sur $\theta_1$ si la zone d'habitation du ménage a une densité d'emplois par km <sup>2</sup> supérieure à 2000	-0.0609	-5.2
Terme correctif sur $\theta_1$ si le ménage ne comprend aucun actif	-0.0431	-1.9
Paramètre lié au temps d'accessibilité aux TC ( $\theta_2$ )	0.02527	4.4
Paramètre lié au nombre d'actifs dans le ménage ( $\theta_3$ )	0.7362	4.5

31. Nombre d'observations : 5 165,  $\rho^2 \text{ bar} = 0.52$



Les utilités s'expriment donc :

$$U(0 \text{ voiture}) = 1$$

Dans le cas général :

$$\begin{aligned} U(1 \text{ voiture et plus}) = & - 0.8559 + (0.3274 \times \text{revenu}) \\ & + (0.02527 \times \text{temps accès TC}) \\ & + (0.7362 \times \text{nombre actifs}) \end{aligned}$$

Dans le cas où le ménage habite dans une zone à forte densité d'emplois :

$$\begin{aligned} U(1 \text{ voiture et plus}) = & - 0.8559 + (0.2665 \times \text{revenu}) \\ & + (0.02527 \times \text{temps accès TC}) \\ & + (0.7362 \times \text{nombre actifs}) \end{aligned}$$

Dans le cas où le ménage ne comprend aucun actif :

$$\begin{aligned} U(1 \text{ voiture et plus}) = & - 0.8559 + (0.2843 \times \text{revenu}) \\ & + (0.02527 \times \text{temps accès TC}) \end{aligned}$$

Dans le cas où le ménage habite dans une zone à forte densité d'emplois et ne comprend aucun actif :

$$\begin{aligned} U(1 \text{ voiture et plus}) = & - 0.8559 + (0.2234 \times \text{revenu}) \\ & + (0.02527 \times \text{temps accès TC}) \end{aligned}$$

Le revenu pris en compte ici est le revenu moyen mensuel du ménage en milliers de francs. Le paramètre multiplicatif de ce revenu est positif, ce qui semble naturel : plus un ménage a un revenu élevé, plus il a de chance de posséder au moins une voiture.

Deux termes correctifs ont été pris en compte sur ce paramètre : ces termes sont intéressants car chacun d'eux fait baisser la valeur du paramètre lié au revenu. Ceci signifie qu'à revenu équivalent, un ménage aura moins tendance à posséder une voiture s'il habite dans une zone à forte densité d'emploi (au centre ville en particulier), ou s'il ne comprend aucun actif (cas des ménages étant à la retraite par exemple).

Le paramètre lié au temps d'accessibilité est positif (le temps est pris en minutes). Ceci montre que les ménages éloignés du réseau TC, et dont le temps d'accès à ce réseau est donc plus important, ont plus de chance d'être motorisés.

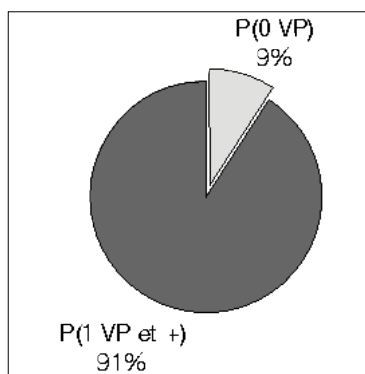
Enfin le paramètre lié au nombre d'actifs est fortement positif : la probabilité qu'un ménage possède au moins une voiture augmente avec son nombre d'actifs.

### Exemple de compréhension

Dans le cas général, nous pouvons estimer, à l'aide de ce type de modèle, la probabilité d'avoir une voiture particulière au moins, pour les individus ayant un ménage composé par 2 actifs, avec un temps d'accès au TC de 10 mn, et ayant un revenu mensuel moyen de 30000F :

On a donc :  $U(1 \text{ voiture et plus}) = 10.69$

Une simple transcription sous Excel permet d'effectuer très rapidement les tests :



Il y a 91% de chances qu'un individu appartenant à un ménage du type précédent (30 KF de revenus mensuels, 10 mn d'accès aux TC et 2 actifs), soit équipé d'au moins 1 véhicule.

### Deuxième modèle de motorisation

Le deuxième modèle de motorisation estime, parmi les ménages possédant au moins une voiture, combien vont en posséder au moins deux. Ce modèle s'appuie sur les mêmes variables que le modèle n°1, c'est-à-dire :

- le revenu du ménage (revenu mensuel en milliers de francs),
- l'accessibilité au TC (temps d'accès au réseau TC en minutes),
- le nombre d'actifs.

La seule différence réside dans le fait qu'il n'existe plus de termes correctifs pour le paramètre lié au revenu du ménage suivant son lieu d'habitation ou son nombre d'actifs.

L'utilité de ne posséder qu'une seule voiture a été prise égale à 1. Le modèle estime donc l'utilité de posséder au moins deux voitures par rapport à cette valeur. Les résultats obtenus pour l'estimation de l'utilité à posséder au moins 2 voitures sont donnés dans le tableau suivant :

**Tableau 9 : Modèle de motorisation n°2**<sup>32</sup>

Paramètres	Estimations	T ratio
Constante ( $\theta_0$ )	-5.255	-23.1
Paramètre lié au revenu du ménage ( $\theta_1$ )	1.822	21.9
Paramètre lié au temps d'accessibilité aux TC ( $\theta_2$ )	0.04297	11.6
Paramètre lié aux nombre d'actifs dans le ménage ( $\theta_3$ )	0.7811	14.2

Les utilités s'expriment donc ainsi :

$$U(1 \text{ voiture}) = 1$$

$$U(2 \text{ voitures et plus}) = -5.255 + (1.822 \times \text{revenu}) \\ + (0.04297 \times \text{temps accès TC}) \\ + (0.7811 \times \text{nombre actifs})$$

La comparaison de l'estimation des paramètres de ce deuxième modèle avec ceux du premier est assez intéressante.

On remarque que le paramètre lié au revenu est beaucoup plus fort dans le choix de posséder une deuxième voiture : le niveau de revenu influe donc beaucoup sur le fait qu'un ménage possède deux voitures ou plus.

De plus, le paramètre lié au temps d'accès a doublé : cette variable a donc un poids plus fort dans le choix de posséder au moins deux voitures.

D'une manière générale, il n'est pas étonnant de voir la valeur des paramètres augmenter dans le deuxième modèle : l'élasticité aux variables correspondantes y est donc plus grande. En effet, l'achat d'une deuxième voiture n'est pas aussi «automatique» que l'achat de la première et ces variables ont donc une plus grande signification.

32. Nombre d'observations = 4 051,  $\rho^2 \text{ bar} = 0.26$ .

## 4.2. Modèle de génération

De la même manière que pour les modèles de motorisation, il est possible d'estimer des modèles de génération désagrégés de type «stop and go». A partir des observations de l'enquête ménage, on estime alors l'utilité d'un ménage à effectuer 1, 2, 3 déplacements ou plus pour un motif donné.

Les trois tableaux suivants présentent l'estimation d'un modèle de ce type, basé sur l'enquête ménage de Lyon pour le motif domicile-achats en période de pointe du soir :

Le tableau 10 donne l'estimation des coefficients intervenant dans l'utilité pour le ménage d'effectuer au moins un déplacement pour le motif achats en période de pointe du soir (déplacement domicile-achats ou achats-domicile) ; l'utilité de ne pas effectuer de déplacement est égale à 1.

Le tableau 11 présente l'estimation des coefficients intervenant dans l'utilité d'effectuer au moins deux déplacements pour les ménages effectuant déjà au moins un déplacement pour le motif achats en période de pointe du soir ; l'utilité d'effectuer uniquement un déplacement est égale à 1.

Le tableau 12 présente l'estimation des coefficients intervenant dans l'utilité d'effectuer au moins trois déplacements pour les ménages effectuant déjà au moins deux déplacements pour le motif achats en période de pointe du soir ; l'utilité d'effectuer uniquement deux déplacements est égale à 1.

**Tableau 10 : Modèle désagrégé de génération «0 – 1 déplacement et plus» pour le motif domicile-achats en période de pointe du soir**

Paramètres	Estimations	T ratio
Constante ( $\theta_0$ )	-0.2133	-3.1
Paramètre lié au nombre d'actifs dans le ménage ( $\theta_1$ )	-0.6952	-16.6
Paramètre lié à la taille du ménage ( $\theta_2$ )	0.7469	18.3
Paramètre lié au nombre de scolaires dans le ménage ( $\theta_3$ )	-0.7676	-15.1

**Tableau 11 : Modèle désagrégé de génération «1 – 2 déplacements et plus» pour le motif domicile-achats en période de pointe du soir**

Paramètres	Estimations	T ratio
Constante ( $\theta_0$ )	-1.184	-8.5
Paramètre lié au nombre d'actifs dans le ménage ( $\theta_1$ )	-0.1759	1.9
Paramètre lié au nombre de scolaires dans le ménage ( $\theta_2$ )	0.2177	4.0
Paramètre lié au nombre de voitures possédées ( $\theta_3$ )	0.2389	3.6
Paramètre lié au nombre de retraités dans le ménage ( $\theta_4$ )	0.4546	4.8

**Tableau 12 : Modèle désagrégé de génération «2 – 3 déplacements et plus» pour le motif domicile-achats en période de pointe du soir**

Paramètres	Estimations	T ratio
Constante ( $\theta_0$ )	-1.342	-4.8
Paramètre lié au nombre de scolaires dans le ménage ( $\theta_2$ )	0.4133	3.7
Paramètre lié au nombre de voitures possédées ( $\theta_3$ )	0.3090	2.2

L'estimation de ce type de modèle est donc réalisable en se basant sur les observations de l'enquête ménage, mais l'application pose problème : les variables intervenant dans le calcul des utilités ne sont pas toutes disponibles au niveau zonal, à Lyon; c'est le cas notamment pour le nombre d'enfants scolarisés par ménage. Pour cette raison majeure, les modèles de génération ont été estimés de manière agrégée en utilisant une segmentation de la population compatible avec les données zonales disponibles.

Les modèles de génération ont été directement estimés à partir de l'enquête ménage. Il existe sept modèles de génération qui correspondent aux sept motifs considérés pour la période de pointe du soir (16h00-18h30). Les modèles des cinq motifs de déplacement liés au domicile (domicile-travail, domicile-école, domicile-achats, domicile-loisirs, domicile-autre) ont été estimés en calculant des taux de déplacements sur différentes segmentations de la population. Ces taux de déplacements ont été obtenus en considérant les deux sens de déplacement possibles : par exemple, pour le motif domicile-travail, les taux estimés représentent pour différentes classes de la population, le nombre moyen de déplacements domicile-travail ou travail-domicile effectués en période de pointe du soir.

Les modèles des deux motifs restants (affaires professionnelles et déplacements secondaires) ont, quant à eux, été estimés à l'aide de régressions linéaires.

#### 4.4. Modèle de choix de mode et de destination

Nous présentons dans ce paragraphe les résultats obtenus lors de l'estimation des 13 modèles de choix de mode et de destination. Ces modèles correspondent aux 7 motifs de déplacement pris en compte pour la période de pointe du soir et pour la période creuse (2 périodes), le motif affaires professionnelles étant, quant à lui, estimé sur l'ensemble des deux périodes pour obtenir un nombre d'observations consistant.

Le tableau suivant donne la liste de ces modèles.

**Tableau 13 : Modèles de choix de mode et de destination estimés**

Motifs de déplacement	Période de pointe du soir (16h00 – 18h30)	Période creuse (9h00 – 16h00)
Domicile-travail	oui	oui
Domicile-école	oui	oui
Domicile-loisirs	oui	oui
Domicile-achats	oui	oui
Domicile-autre	oui	oui
Déplacements secondaires	oui	oui
Affaires professionnelles	Estimé sur l'ensemble des deux périodes	

Il convient de noter que la dénomination de ces motifs ne tient pas compte du sens du trajet. Ainsi, lorsque nous parlerons du motif domicile-travail, il s'agit en fait de la réunion des deux motifs : domicile-travail et travail-domicile. Pour tous les motifs que nous prenons en compte ici, l'estimation des modèles a en effet été basée sur l'ensemble des observations d'une période, qu'elles soient dans un sens ou dans un autre.

Pour chacun de ces motifs, les résultats sont présentés sous forme de tableaux où apparaissent les valeurs estimées des paramètres des différentes fonctions d'utilité ainsi que les coefficients de student correspondants ( $t$ ). Les constantes liées au choix de la destination sont présentées telles qu'obtenues directement après l'estimation et une fois le coefficient de correction appliqué.

Les coefficients estimés liés au temps (temps de rabattement, temps dans le véhicule, etc.) correspondent à des temps exprimés en minutes.

Le temps de rabattement pour le mode VP est le temps de marche à pied nécessaire pour se rendre à sa voiture, plus le temps nécessaire pour se garer une fois arrivé à destination.

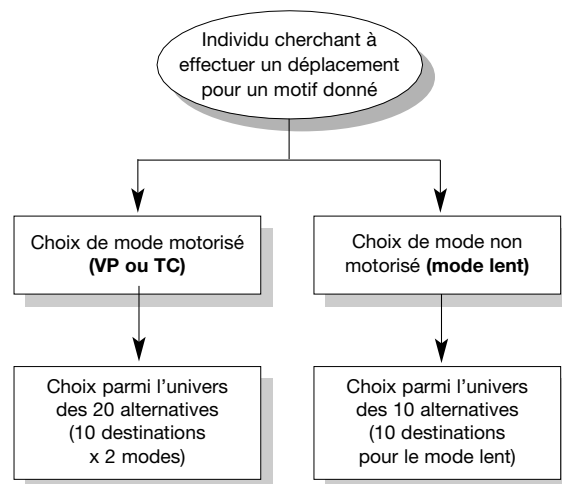
Le temps de rabattement pour le mode TC prend en compte le temps d'accessibilité au réseau TC à l'origine et à la destination, ainsi que le temps d'attente et le temps de correspondance.

Certains modèles sont renseignés par un terme appelé «logsum» dont la valeur est estimée avec le modèle et comprise entre 0 et 1. Ce terme mesure la hiérarchie existant entre le mode alternatif et les modes motorisés (VP et TC ; cf. première partie du rapport pour plus de renseignements).

La connaissance des deux utilités  $U$  (mode motorisé, mode non motorisé) permet de calculer la probabilité de choix  $P1$  de l'un et l'autre des deux types de mode. Ensuite, de la même manière que pour un modèle à structure non hiérarchique, on estime la probabilité de choix  $P2$  d'une alternative parmi l'ensemble des alternatives correspondant au même type de mode (10 pour le mode alternatif, 20 pour le mode motorisé). La probabilité de choix d'une alternative sur les 30 possibles est donc obtenue par le produit des deux probabilités  $P1$  et  $P2$ .

La structure des modèles hiérarchiques estimés ici est présentée en figure 8.

**Figure 8 : Structure hiérarchique des modèles de choix de mode et de destination**



Pour le premier modèle (motif domicile-travail), le tableau est accompagné, pour plus de clarté, de résultats sous forme d'équations. Pour les autres modèles, les résultats apparaissent sous forme de tableau uniquement, leur structure étant identique.



### Motif domicile-travail

**Tableau 14 : Modèle de choix de mode et de destination pour le motif domicile-travail en période de pointe du soir (PPS) et en période creuse (PCR)**

Description des variables	Paramètres estimés en PPS	T ratio PPS	Paramètres estimés en PCR	T ratio PCR
<b>Constante modale</b>				
Constante TC	-1,233	-5,0	-1,746	-8,5
<b>Constante zone de destination</b>				
Constante zone centre*	-1,951 -0,321	-20,6	-1,714 -0,084	-22,9
Constante zone origine*	-3,403 1,020	-17,4	-2,816 1,607	-21,1
Constante zone voisine*	-2,838 0,381	-24,1	-2,373 0,846	-28,6
<b>Mode VP</b>				
Constante si le ménage possède au moins 2 VP	1,541	9,0	1,728	11,0
Temps de rabattement mode VP	-0,07099	-4,0	-0,1151	-9,0
Temps du trajet VP inférieur à 10 mn	-0,1528	-7,2	-0,1391	-9,3
Temps du trajet VP supérieur à 10 mn	-0,08447	-14,3	-0,1285	-22,5
<b>Mode TC</b>				
Temps de rabattement mode TC	-0,05264	-4,8	-0,06793	-6,4
Temps dans le véhicule mode TC	-0,03028	-6,0	-0,03361	-5,0
<b>Mode alternatif (marche à pied)</b>				
Temps de marche à pied	-0,07407	-11,0	-0,07322	-16,8
<b>Variables de taille de la zone de destination</b>				
Coefficient de taille	0,9085	23,7	0,7253	27,1
Population de la zone	1	Fixé	1	Fixé
Nombre d'emplois dans la zone**	3,385	7,2	3,218	9,0
<b>Hiérarchie entre mode alternatif et mode motorisé</b>				
Logsum	0,7654	13,5	0,7016	17,5

\* Pour ces constantes, la valeur en italique est la valeur estimée, l'autre correspond à la valeur corrigée tenant compte du phénomène d'échantillonnage.

\*\* C'est l'exponentielle du paramètre estimé qui est prise en compte comme coefficient.

**Le modèle**

Il est basé sur 1 412 observations pour la période de pointe ( $\rho^2 \text{ bar} = 0.31$ ) et sur 2 465 observations pour la période creuse ( $\rho^2 \text{ bar} = 0.29$ ).

**Construction d'une fonction d'utilité**

L'expression de l'utilité du mode TC vers une zone du centre en période de pointe du soir s'exprime donc de la manière suivante :

$$U = -1,233 - 0,321 - (0,05264 \times \text{temps de rabattement TC}) \\ - (0,03028 \times \text{temps dans le véhicule TC}) \\ + 0,9085 \times \ln [\text{population} + [\exp(3,385) \times \text{emplois}]]$$

Pour la même période, l'expression de l'utilité du mode VP vers une zone quelconque (ni origine, ni voisine, ni centre) pour un ménage possédant 2 voitures et pour un temps supérieur à 10 mn s'exprime ainsi :

$$U = 1,541 - (0,07099 \times \text{temps de rabattement VP}) \\ - (0,1528 \times 10) \\ - (0,08447 \times (\text{temps de trajet VP} - 10)) \\ + 0,9085 \times \ln [\text{population} + [\exp(3,385) \times \text{emplois}]]$$

Si l'on veut spécifier l'expression de l'utilité du mode VP vers une zone du centre, il suffit de rajouter dans la fonction d'utilité le coefficient 0,321.

Pour le mode VP, le temps du trajet dans le véhicule est donc décomposé de la manière suivante :

- si le temps VP est inférieur à 10 minutes, il correspond à la variable «temps du trajet VP inférieur à 10 minutes» exprimée en minutes,
- si le temps VP est supérieur à 10 minutes, la variable «temps du trajet VP inférieur à 10 minutes» est égale à 10 et la variable «temps du trajet VP supérieur à 10 minutes» correspond au reste du temps exprimé en minutes.

Pour le motif domicile-travail, parler de choix de destination est un abus de langage ; il n'existe pas réellement de choix de destination puisque le lieu de travail est la destination obligée. Il existe par contre bien un choix de mode. Le modèle a tout de même été estimé comme un modèle joint de choix de mode et de destination. Il permet une bonne description des tendances actuelles sur la ville de Lyon : où les gens travaillent-ils ? Suivant leur lieu de travail, quel mode de transport utilisent-ils ?

### Qualité de l'estimation des coefficients

Tous les coefficients estimés dans les deux modèles (un modèle par période) le sont avec une bonne précision (les T ratios sont significativement supérieur à 2). Le modèle comprend une constante modale liée au TC et trois constantes de destination (zone origine, zones voisines et zones du centre), à la période de pointe du soir (PPS) comme à la période creuse (PCR).

Pour le mode VP, il existe une constante de correction pour les ménages possédant au moins deux voitures. Cette constante est fortement positive, ce qui indique qu'un actif appartenant à un ménage de ce type a une beaucoup plus forte probabilité de choisir le mode VP qu'un actif faisant partie d'un ménage ne possédant qu'une voiture (si le ménage ne possède aucune voiture, l'alternative mode VP est impossible).

### Expliquer le niveau des coefficients

En période de pointe comme en période creuse, le coefficient du temps de rabattement en mode VP est bien sûr négatif, mais moins grand en valeur absolu que le temps de trajet dans le véhicule.

Pour les deux périodes, mais de façon beaucoup plus sensible en PPS, le coefficient du temps de trajet en VP est plus grand en valeur absolue pour les 10 premières minutes du trajet. Ceci semble indiquer que les premières minutes d'un trajet sont ressenties comme plus contraignantes que les autres. On peut l'expliquer par le fait qu'il existe une pénibilité fixe à prendre sa voiture (recherche de stationnement par exemple) qui ne dépend pas de la longueur du trajet.

En ce qui concerne le mode TC, on remarque qu'en PPS comme en PCR, les coefficients liés au temps sont négatifs (comme attendu) et le temps de rabattement est plus pénalisé que le temps dans le véhicule, donc ressenti plus péniblement.

### Le danger de l'optique «comparative»

La comparaison entre les coefficients obtenus pour les différents modes (mode alternatif, VP et TC) ne peut être effectuée directement. Ainsi, le fait d'obtenir un coefficient lié au temps de trajet VP de l'ordre de -0,15 pour les 10 premières minutes à la PPS, ne permet pas de conclure à une plus grande pénibilité du temps passé en voiture en comparaison avec le temps passé dans un transport en commun dont le coefficient est de l'ordre de -0,03. On peut seulement dire que le temps de parcours en voiture, surtout les 10 premières minutes, est très mal perçu par les utilisateurs de la VP (-0,15), en comparaison avec le temps au delà de 10 minutes (-0,08) et le temps de rabattement (-0,07). En ce qui concerne les usagers des TC, le temps dans

le véhicule est mieux perçu que le temps de rabattement (-0,03 contre -0,05).

Par ailleurs, dans chacune des utilités intervient le terme lié à la destination. Ce terme, invariant suivant le choix de mode, influe sur les coefficients liés au choix de mode, suivant son poids relatif dans chacune des fonctions d'utilité.

En résumé, pour deux modes différents, la comparaison entre les coefficients liés au choix de mode n'est pas possible directement. On remarquera que les ordres de grandeur obtenus ici sont ceux que l'on retrouve habituellement dans ce type de modèle. Cette remarque vaut également pour les modèles suivants.

#### **Le rôle important de la variable de taille**

La variable de taille «nombre d'emplois» est très explicative des choix observés. Elle a un poids 29,5 fois plus important que celui de la population [ $\exp(3,385)$ ]. D'autre part, le coefficient de taille est relativement élevé (proche de 1) ce qui souligne le caractère significatif de ces variables (population et emplois).

#### **L'existence d'une hiérarchie en terme de choix**

Le logsum est significativement différent de 1 en PPS (0,7654) comme en PCR (0,7016). Par conséquent, la hiérarchie en terme de choix entre les modes motorisés et le mode alternatif existe bien pour le motif domicile-travail.

#### **Exemple de compréhension**

Considérons le modèle à la PPS, à destination d'une zone quelconque et prenons l'exemple d'une personne (ou d'un groupe de personnes) dont les caractéristiques des alternatives sont les suivantes pour effectuer le déplacement :

### VP

Tps rabattmt :	5 mn
Temps < 10 mn :	10 mn
Temps > 10 mn :	5 mn
Temps total :	15 mn

### TC

Tps rabattmt :	10 mn
Temps dans le véhicule :	30 mn

### Mode alternatif (lent)

Temps de MAP :	50 mn
----------------	-------

**Les caractéristiques de la zone de destination sont les suivantes :**

Taille :	10 000 habitants
Emplois :	3 200

Le calcul des différentes utilités conduit à :

$$U(VP) = 8.141$$

$$U(TC) = 7.778$$

$$U(ML) = 6.743$$

L'utilité des modes motorisés sera :

$$U(MOT) = LOGSUM * LN[U(TC)+U(VP)]$$

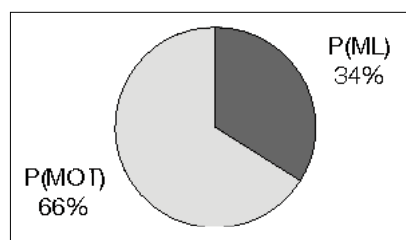
$$U(MOT) = 2.118$$

L'utilité des modes non motorisés sera :

$$U(NMOT) = LOGSUM * LN [U(ML)]$$

$$U(NMOT) = 1.461$$

La formulation logit conduit dans ce cas, à :



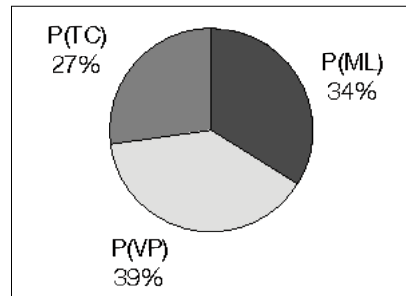
Si l'on désire «descendre» à un niveau inférieur, on aura :

$$P(VP) = [\exp(U(VP)) / (\exp(U(VP)) + \exp(U(TC)))] * P(MOT)$$

$$P(TC) = [\exp(U(TC)) / (\exp(U(VP)) + \exp(U(TC)))] * P(MOT)$$

$$P(ML) = [\exp(U(ML)) / (\exp(U(ML)) + \exp(U(MOT)))]$$

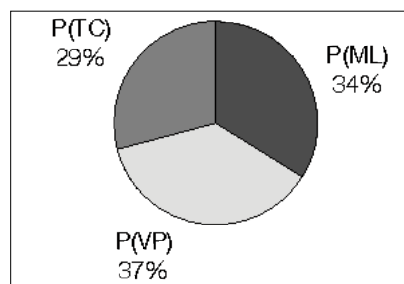
Soit :



En résumé : pour se rendre à la précédente destination, 27% des individus prendront les transports collectifs, 34% un mode alternatif (marche, vélo), et 39% la voiture.

A l'aide d'une telle modélisation on pourra effectuer des tests. Par exemple, on pourra évaluer l'impact d'une politique des transports dont l'objectif est d'améliorer la performance du réseau des transports collectifs, en développant des systèmes en sites propres de surface performants, largement subventionnés par l'Etat (à hauteur de 40%). Après étude, on constate que leur réalisation devrait entraîner une pénalisation d'environ 5 % du temps moyen de parcours en voiture (les TC accaparant une partie de la voirie), et une amélioration de l'ordre de 20% du temps de parcours en TC et de 8% du temps de rabattement (accès) aux services de TC. On se positionne ainsi le plus près possible du client et on accroît l'aire de chalandise des TC car ils deviennent plus performants.

#### Résultat du test



Le résultat d'une politique aussi audacieuse (les variations de temps annoncées restent très importantes), permettraient de faire varier à la hausse la part de marché des transports collectifs (plus 2%).

**Motif domicile - école****Tableau 15 : Modèle de choix de mode et de destination pour le motif domicile-école en PPS et en PCR**

Description des variables	Paramètres estimés en PPS	T ratio PPS	Paramètres estimés en PCR	T ratio PCR
<b>Constante modale</b>				
Constante mode alternatif	1,868	13,2	1,895	16,3
<b>Constante zone de destination</b>				
Constante zone centre*	-1,288 0,342	-10,7	-1,205 0,425	-10,3
Constante zone origine*	-2,064 2,359	-11,8	-2,028 2,395	-13,7
Constante zone voisine*	-1,934 1,285	-15,6	-1,657 1,562	-15,5
<b>Mode VP</b>				
Constante si le ménage possède au moins 2 VP	1,291	9,6	1,070	9,2
Temps de rabattement mode VP	-0,1221	-5,9	-0,1553	-8,0
Temps du trajet VP inférieur à 10 mn	-0,2998	-11,9	-0,3171	-14,1
Temps du trajet VP supérieur à 10 mn	-0,1253	-8,6	-0,1483	-8,7
<b>Mode TC</b>				
Temps de rabattement mode TC	-0,06057	-9,3	-0,06424	-10,5
Temps dans le véhicule mode TC	-0,04138	-8,0	-0,06626	-11,9
<b>Mode alternatif (marche à pied)</b>				
Temps de marche à pied	-0,07017	-16,3	-0,06976	-20,5
<b>Variables de taille de la zone de destination</b>				
Coefficient de taille	0,6318	11,0	0,6046	13,7
Population de la zone	1	Fixé	1	Fixé
Nombre d'élèves**	3,893	6,5	3,646	8,4
Nombre de lycéens**	4,171	6,7	3,842	8,6

\* Pour ces constantes, la valeur en italique est la valeur estimée, l'autre correspond à la valeur corrigée tenant compte du phénomène d'échantillonnage.

\*\* C'est l'exponentielle du paramètre estimé qui est prise en compte comme coefficient.

Il existe, pour le motif domicile-école une constante modale pour le mode alternatif (marche à pied). Cette constante est positive et joue donc en faveur du choix de la marche à pied, ce qui n'est guère étonnant sur une population jeune.

La constante de destination liée à la zone d'origine est également fortement positive : on retrouve ici le caractère local de ce type de déplacements.

### **Le modèle**

Il est basé sur 1 457 observations pour la période de pointe ( $\rho^2 \text{ bar} = 0.34$ ) et sur 2 271 observations pour la période creuse ( $\rho^2 \text{ bar} = 0.39$ ).

### **Expliquer le niveau des coefficients**

Le coefficient du temps de rabattement TC est une fois et demi plus grand en valeur absolue que celui du temps dans le véhicule en PPS. Par contre, ces deux coefficients sont pratiquement identiques (-0,065) en PCR.

### **Le rôle des variables de taille**

Les variables de taille «nombre d'élèves» et «nombre de lycéens» sont significatives et estimées avec une bonne précision. Par contre, les coefficients de taille ont des valeurs assez faibles (0,6).

### **La non existence d'une hiérarchie dans les choix**

Ces modèles ne sont pas de type hiérarchique. Cela signifie que les observations des choix réels ne font pas apparaître de différences significatives au sein de l'univers des destinations choisies, selon que l'on utilise un mode motorisé ou non. On peut comprendre ce phénomène : les déplacements pour le motif domicile-école étant très localisés, la marche à pied n'est pas un mode discriminant en terme de choix de destination.



**Motif domicile - loisirs****Tableau 16 : Modèle de choix de mode et de destination pour le motif domicile-loisirs en PPS et en PCR**

Description des variables	Paramètres estimés en PPS	T ratio PPS	Paramètres estimés en PCR	T ratio PCR
<b>Constante modale</b>				
Constante TC	-1,859	-5,5	-1,378	-4,5
Constante mode alternatif	1,126	-4,8	1,245	-5,8
<b>Constante zone de destination</b>				
Constante zone centre*	-1,066 -0,564	-8,6	-0,6343 -0,9957	-6,3
Constante zone origine*	-2,168 2,255	-10,5	-1,924 2,499	-10,4
Constante zone voisine*	-2,004 1,215	-14,2	-2,034 1,185	-16,3
<b>Mode VP</b>				
Constante si le ménage possède au moins 2 VP	0,7902	4,3	1,424	7,5
Temps de rabattement mode VP	-0,1365	-5,1	-0,1632	-6,5
Temps du trajet VP inférieur à 10 mn	-0,2073	-7,8	-0,1509	-6,4
Temps du trajet VP supérieur à 10 mn	-0,1041	-9,3	-0,1217	-11,3
<b>Mode TC</b>				
Temps de rabattement mode TC	-0,04033	-3,7	-0,04832	-4,5
Temps dans le véhicule mode TC	-0,04282	-5,3	-0,03903	-5,3
<b>Mode alternatif (marche à pied)</b>				
Temps de marche à pied	-0,08669	-13,8	-0,07255	-14,5
<b>Variables de taille de la zone de destination</b>				
Coefficient de taille	0,6045	14,1	0,2912	9,0
Population de la zone	1	Fixé	1	Fixé
Nombre d'emplois dans la zone**	0,8082	2,8	1,284	3,1
<b>Hiérarchie entre mode alternatif et mode motorisé</b>				
Logsum	0,5863	8,2	0,5425	9,0

\* Pour ces constantes, la valeur en italique est la valeur estimée, l'autre correspond à la valeur corrigée tenant compte du phénomène d'échantillonnage.

\*\* C'est l'exponentielle du paramètre estimé qui est prise en compte comme coefficient.

Pour le motif domicile-loisirs, deux constantes modales sont prises en compte (TC et mode alternatif). La constante TC est négative, ce mode est donc pénalisé par rapport à la voiture. En revanche, la constante liée à la marche à pied est positive, et donc avantagée par rapport à la voiture. Cette estimation conduit à penser que les lieux de loisirs choisis sont souvent proches du lieu de domicile.

Les constantes liées à l'origine et aux zones voisines sont elles aussi positives de façon très marquée, ce qui confirme la remarque précédente.

### **Le modèle**

Il est basé sur 1 106 observations pour la période de pointe ( $\rho^2 \text{ bar} = 0.25$ ) et sur 1 332 observations pour la période creuse ( $\rho^2 \text{ bar} = 0.21$ ).

### **Expliquer le niveau des coefficients**

Les coefficients du mode VP connaissent de très forts écarts entre la période de pointe du soir et la période creuse. On constate, en effet, que la constante liée aux ménages possédant au moins une voiture double en période creuse. En période de pointe du soir, le début du trajet est deux fois plus pénalisé que la suite du trajet (paramètre temps à moins et plus de 10 minutes), alors qu'en période creuse, la différence est bien moins sensible. La pénibilité d'un trajet en voiture en PPS, même s'il est court, peut être une des explications à cette observation.

On remarque que pour le mode TC, le temps de rabattement et le temps dans le véhicule ont des coefficients à peu près semblables en PPS. En PCR, le temps de rabattement est, par contre, plus pénalisé que le temps dans le véhicule; sans doute à cause de la plus grande longueur des temps d'attente.

### **Le rôle de la variable de taille**

Il convient de remarquer que les coefficients de taille sont faibles, surtout pour la période creuse, où le coefficient se rapproche beaucoup de 0 (0,2912). Ceci s'explique par le fait que l'emploi et la population d'une zone ne sont évidemment pas des variables suffisantes pour expliquer des déplacements domicile-loisirs. On mesure ici l'importance de données liées à l'équipement, les services, les parcs, etc.

### **L'existence d'une hiérarchie**

Le «logsum» est quant à lui très différent de 1: la hiérarchie entre modes motorisés et non motorisés est donc très marquée pour ce motif.

**Motif domicile - achat****Tableau 17 : Modèle de choix de mode et de destination pour le motif domicile-achats en PPS et en PCR**

Description des variables	Paramètres estimés en PPS	T ratio PPS	Paramètres estimés en PCR	T ratio PCR
<b>Constante modale</b>				
Constante TC	-2,388	-8,3	-1,343	-6,0
Constante mode alternatif	0,9736	5,4	1,360	11,2
<b>Constante zone de destination</b>				
Constante zone centre*	-1,819 -0,189	-13,6	-1,186 -0,444	-13,7
Constante zone origine*	-2,229 2,194	-12,9	-1,781 2,642	-15,3
Constante zone voisine*	-2,080 1,139	-17,5	-1,787 1,432	-21,9
<b>Mode VP</b>				
Constante si le ménage possède au moins 2 VP	0,9849	6,3	0,9984	9,1
Temps de rabattement mode VP	-0,1020	-5,0	-0,1469	-9,5
Temps du trajet VP inférieur à 10 mn	-0,2225	-10,5	-0,1900	-12,9
Temps du trajet VP supérieur à 10 mn	-0,1484	-14,1	-0,1487	-19,4
<b>Mode TC</b>				
Temps de rabattement mode TC	-0,04808	-3,8	-0,1245	-8,5
Temps dans le véhicule mode TC	-0,05067	-5,2	-0,04435	-5,5
<b>Mode alternatif (marche à pied)</b>				
Temps de marche à pied	-0,1039	-17,4	-0,1017	-26,8
<b>Variables de taille de la zone de destination</b>				
Coefficient de taille	0,7796	19,8	0,8367	29,6
Population de la zone	1	Fixé	1	Fixé
Nombre d'emplois dans la zone**	1,782	6,1	1,359	7,6
Surface commerciale >1000m <sup>2</sup> (en m <sup>2</sup> )	1,641	6,1	1,537	9,7
<b>Hiérarchie entre mode alternatif et mode motorisé</b>				
Logsum	0,9204	13,3	0,8080	20,1

\* Pour ces constantes, la valeur en italique est la valeur estimée, l'autre correspond à la valeur corrigée tenant compte du phénomène d'échantillonnage.

\*\* C'est l'exponentielle du paramètre estimé qui est prise en compte comme coefficient.

Le motif domicile-achats présente deux constantes modales : la constante TC qui est négative (de façon marquée en PPS) et la constante mode alternatif qui est positive. Il présente également trois constantes de destination.

### Le modèle

Il est basé sur 1 541 observations pour la période de pointe ( $\rho^2 \text{ bar} = 0.38$ ) et sur 3 666 observations pour la période creuse ( $\rho^2 \text{ bar} = 0.41$ ).

### Expliquer le niveau des coefficients

En ce qui concerne le mode VP, on remarque comme dans les modèles précédents, que la pénibilité du temps de trajet est plus forte dans les 10 premières minutes, en PPS comme en PCR. Il existe donc une pénibilité à effectuer un trajet en voiture même si ce dernier est relativement court.

Ici, on note que le coefficient du temps de trajet supérieur à 10 minutes est le même (-0,148) pour les deux périodes. Par contre, pour le temps de trajet inférieur à 10 minutes, le coefficient du modèle pour la PPS est plus grand en valeur absolue que celui du modèle pour la PCR : il est plus pénible de prendre sa voiture en heure de pointe qu'en heure creuse.

Pour le mode TC, en PPS, le coefficient du temps de rabattement est identique à celui du temps dans le véhicule (-0,050). Mais pour la PCR, ce même coefficient est pratiquement trois fois plus grand que celui du temps dans le véhicule (temps de rabattement : -0,125; temps dans le véhicule : -0,044). On peut expliquer cette pénalisation du temps hors véhicule en mode TC par le fait que le type d'achats effectués en PCR n'est probablement pas le même qu'en PPS. En PPS, les achats sont souvent des achats quotidiens, habituels (courses) et sont peu volumineux. En PCR, par contre, il peut s'agir d'achats moins courants et plus volumineux.

### Le rôle des variables de taille

Les coefficients de taille sont assez grands (proches de 0,8) : les variables de taille expliquent correctement les choix observés en terme de destination. En plus de la population et du nombre d'emplois utilisés ici comme variables de taille, le modèle prend également en compte pour chaque zone le total des surfaces commerciales supérieures à 1000 m<sup>2</sup>. Cette dernière variable de taille est estimée avec précision (T ratio de 6,1 en PPS et de 9,7 en PCR) et son poids est comparable à celui du nombre d'emplois de la zone.

### L'existence d'une faible hiérarchie

Le logsum est assez proche de 1, surtout en PPS. La hiérarchie entre les modes motorisés et non motorisés est donc peu marquée. Ceci peut être dû au fait que pour ce motif le choix de mode (VP ou mode alternatif surtout) est fonction du type d'achats effectués et pas uniquement de la destination.

### Exemple de compréhension

Un exemple de compréhension plus complet que pour le modèle concernant le motif domicile-travail peut être réalisé ici. En effet, la modélisation désagrégée de choix de mode-destination, se prête très bien à la prise en compte du motif domicile-achat, pour lequel les caractéristiques de la zone de destination peuvent fortement influencer le choix de mode. Cet exemple ne sera pas très lisible dans la logique de présentation par «camembert» : plusieurs destinations possibles seront prises en compte, au lieu d'une seule. Cette logique n'est à considérer qu'à travers sa valeur illustrative.

Considérons un individu ou un groupe d'individus **ayant des caractéristiques d'alternatives de modes équivalentes à l'exemple précédent (cf. motif domicile-travail)**, et quatre zones de destination possibles ayant les caractéristiques suivantes (à la PPS) :

**Zone centrale (Z1)**

Taille (population)	17 000
Nombre d'emplois	20 000
Surface commerciale > 1000 m <sup>2</sup>	1 000

**Zone origine (Z2)**

Taille (population)	3 000
Nombre d'emplois	1 500
Surface commerciale > 1000 m <sup>2</sup>	9 000

**Zone voisine (Z3)**

Taille (population)	6 500
Nombre d'emplois	5 000
Surface commerciale > 1000 m <sup>2</sup>	10 000

**Autre zone (Z4)**

Taille (population)	4 000
Nombre d'emplois	3 000
Surface commerciale > 1000 m <sup>2</sup>	30 000

On calcule l'ensemble des fonctions d'utilités composites concernées, soit :

U (VP, Z1)	5.577	U (TC, Z1)	4.665	U (ML, Z1)	4.833
U (VP, Z2)	7.272	U (TC, Z2)	6.361	U (ML, Z2)	6.528
U (VP, Z3)	6.536	U (TC, Z3)	5.624	U (ML, Z3)	5.792
U (VP, Z4)	5.942	U (TC, Z4)	5.030	U (ML, Z4)	5.198

Chaque utilité composite correspond à la somme de deux utilités : celle concernant chaque mode plus celle concernant chaque zone de destination.

Les probabilités de choix sont alors les suivantes :

P (VP, Z1)	0.051	P (TC, Z1)	0.020	P (ML, Z1)	0.024
P (VP, Z2)	0.277	P (TC, Z2)	0.111	P (ML, Z2)	0.131
P (VP, Z3)	0.132	P (TC, Z3)	0.053	P (ML, Z3)	0.063
P (VP, Z4)	0.073	P (TC, Z4)	0.029	P (ML, Z4)	0.035

**Motif domicile-autres****Tableau 18 : Modèle de choix de mode et de destination pour le motif domicile-autres en PPS et en PCR**

Description des variables	Paramètres estimés en PPS	T ratio PPS	Paramètres estimés en PCR	T ratio PCR
<b>Constante modale</b>				
Constante TC	-2,755	-13,8	-2,128	-13,8
Constante mode alternatif	0,1825	1,4	0,1978	1,8
<b>Constante zone de destination</b>				
Constante zone centre*	-1,530 0,100	-19,7	-1,147 0,483	-20,5
Constante zone origine*	-2,174 2,249	-19,0	-2,185 2,238	-23,3
Constante zone voisine*	-2,063 1,156	-26,8	-2,019 1,200	-32,5
<b>Mode VP</b>				
Constante si le ménage possède au moins 2 VP	0,8292	7,8	0,9076	10,0
Temps de rabattement mode VP	-0,1122	-8,4	-0,1223	-11,0
Temps du trajet VP inférieur à 10 mn	-0,2353	-16,7	-0,2105	-18,3
Temps du trajet VP supérieur à 10 mn	-0,1192	-20,1	-0,1282	-24,9
<b>Mode TC</b>				
Temps de rabattement mode TC	-0,05756	-7,6	-0,05683	-9,6
Temps dans le véhicule mode TC	-0,01844	-3,6	-0,03673	-8,9
<b>Mode alternatif (marche à pied)</b>				
Temps de marche à pied	-0,09297	-23,0	-0,08309	-28,2
<b>Variables de taille de la zone de destination</b>				
Coefficient de taille	0,6349	25,4	0,5066	27,4
Population de la zone	1	Fixé	1	Fixé
Nombre d'emplois dans la zone**	0,8609	5,8	1,587	10,4
<b>Hiérarchie entre mode alternatif et mode motorisé</b>				
Logsum	0,7111	17,2	0,6891	19,9

\* Pour ces constantes, la valeur en italique est la valeur estimée, l'autre correspond à la valeur corrigée tenant compte du phénomène d'échantillonnage.

\*\* C'est l'exponentielle du paramètre estimé qui est prise en compte comme coefficient.

Les modèles pour le motif domicile-autres présentent deux constantes modales et trois constantes de destination. Ces cinq constantes sont estimées avec une très bonne précision hormis le cas de la constante mode alternatif dont les «T ratio» sont inférieurs à 2. Ces constantes ont à peu près les mêmes valeurs en période de pointe du soir (PPS) et en heure creuse (PCR).

#### **Le modèle**

Il est basé sur 3 459 observations pour la période de pointe ( $\rho^2 \text{ bar} = 0.28$ ) et sur 4903 observations pour la période creuse ( $\rho^2 \text{ bar} = 0.23$ ).

#### **Expliquer le niveau des coefficients**

Pour le mode VP, le temps de trajet compris entre 0 et 10 minutes est deux fois plus pénalisé que le reste du temps (après 10 minutes) en PPS comme en PCR. Les coefficients des deux périodes sont relativement semblables.

Concernant le mode TC, l'estimation du paramètre lié au temps de rabattement est identique en PPS et en PCR (-0,057). Par contre, le temps dans le véhicule est deux fois plus pénalisé en PCR qu'en PPS (-0,036 pour -0,018) : en période de pointe du soir, le temps d'attente est donc ressenti comme étant trois fois plus pénible que le temps dans le véhicule ; en période creuse, il est seulement une fois et demi plus pénible.

#### **Le rôle des variables de taille**

Les coefficients de taille sont assez moyens puisqu'ils sont proches de 0,5 : le motif de déplacement « autres » est difficilement caractérisé par une variable de taille. La population et le nombre d'emplois ne suffisent pas pour expliquer l'ensemble des observations de l'échantillon sur lequel ont été estimés les modèles.

#### **L'existence d'une hiérarchie**

Le logsum est sensiblement différent de 1 (proche de 0,7) : la structure hiérarchique des modèles pour ce motif est donc convenable.



**Motif déplacement secondaires****Tableau 19 : Modèle de choix de mode et de destination pour le motif déplacements secondaires en PPS et en PCR**

Description des variables	Paramètres estimés en PPS	T ratio PPS	Paramètres estimés en PCR	T ratio PCR
<b>Constante modale</b>				
Constante TC	-2,872	-12,2	-1,875	-10,3
Constante mode alternatif	0,9885	-4,4	0,2289	2,0
<b>Constante zone de destination</b>				
Constante zone centre*	-1,927 -0,297	-22,7	-1,663 -0,033	-27,6
Constante zone origine*	-2,105 2,318	-15,7	-2,086 2,337	-22,0
Constante zone voisine*	-2,314 0,905	-24,7	-2,153 1,066	-31,8
<b>Mode VP</b>				
Constante si le ménage possède au moins 2 VP	0,5054	3,7	0,4604	5,4
Temps de rabattement mode VP	-0,1811	-11,5	-0,1675	-14,4
Temps du trajet VP inférieur à 10 mn	-0,1431	-9,0	-0,1425	-12,4
Temps du trajet VP supérieur à 10 mn	-0,09842	-16,6	-0,1289	-22,5
<b>Mode TC</b>				
Temps de rabattement mode TC	-0,04275	-4,3	-0,1061	-4,8
Temps dans le véhicule mode TC	-0,03716	-5,1	-0,03142	-8,9
<b>Mode alternatif (marche à pied)</b>				
Temps de marche à pied	-0,08592	-16,1	-0,08462	-26,6
<b>Variables de taille de la zone de destination</b>				
Coefficient de taille	0,4494	17,8	0,4709	27,8
Population de la zone	1	Fixé	1	Fixé
Nombre d'emplois dans la zone**	0,1805	0,9	1,894	10,1
<b>Hiérarchie entre mode alternatif et mode motorisé</b>				
Logsum	0,6221	12,4	0,7465	19,3

\* Pour ces constantes, la valeur en italique est la valeur estimée, l'autre correspond à la valeur corrigée tenant compte du phénomène d'échantillonnage.

\*\* C'est l'exponentielle du paramètre estimé qui est prise en compte comme coefficient.

Pour le motif déplacements secondaires, cinq constantes ont été utilisées pour le choix de mode et de destination (constantes TC, mode alternatif, zone origine, zone voisine et zone centre). Les constantes de destination ont des estimations très proches d'une période à l'autre. Par contre, concernant les constantes modales, leur estimation est bien plus pénalisante pour le TC et le mode alternatif en période de pointe du soir (PPS).

### **Le modèle**

Il est basé sur 2 306 observations pour la période de pointe ( $\rho^2 \text{ bar} = 0.24$ ) et sur 4 469 observations pour la période creuse ( $\rho^2 \text{ bar} = 0.24$ ).

### **Expliquer le niveau des coefficients**

L'estimation des coefficients du mode VP fait apparaître un temps de rabattement légèrement plus pénalisé que la part du temps de trajet inférieur à 10 minutes, elle même un peu plus pénalisée que la part du temps de trajet supérieure à 10 minutes. L'estimation de ces coefficients est assez proche pour les deux périodes considérées.

Pour le mode TC, le temps de rabattement est plus pénalisé que le temps dans le véhicule, et ceci pour les deux périodes. Mais, alors que l'estimation du coefficient lié au temps dans le véhicule est du même ordre (-0,03) en période de pointe du soir et en période creuse (PCR), le coefficient lié au temps de rabattement en PCR est deux fois et demi plus grand en valeur absolue qu'en PPS.

L'estimation des coefficients du temps de marche à pied sont pratiquement identiques (-0,084) et estimés avec une très bonne précision.

### **Le rôle des variables de taille**

Les coefficients de taille sont assez faibles (proche de 0,5) car, là encore, il est difficile de lier une variable de taille au motif «déplacements secondaires», ces déplacements rassemblant en fait beaucoup de motifs différents.

D'ailleurs, le paramètre de la variable de taille «nombre d'emplois» est estimé avec une précision assez médiocre (T ratio de 0,9) pour la PPS. Sa comparaison avec la valeur estimée en PCR n'est donc pas réalisable.

### **L'existence d'une hiérarchie**

Les valeurs des «logsum» pour les deux périodes soulignent la solidité du modèle hiérarchique.

**Motif affaires personnelles****Tableau 20 : Modèle de choix de mode et de destination pour le motif affaires professionnelles pour la journée**

Description des variables	Paramètres estimés en PPS+PCR	T ratio PPS+PCR
<b>Constante modale</b>		
Constante TC	-1,720	-4,1
<b>Constante zone de destination</b>		
Constante zone centre*	-1,532 -0,098	-13,5
Constante zone origine*	-2,448 1,975	-11,1
Constante zone voisine*	-2,396 0,823	-15,8
<b>Mode VP</b>		
Constante si le ménage possède au moins 2 VP	1,073	4,2
Temps de rabattement mode VP	-0,1037	-5,2
Temps du trajet VP inférieur à 10 mn	-0,06921	-2,9
Temps du trajet VP supérieur à 10 mn	-0,08729	-10,3
<b>Mode TC</b>		
Temps de rabattement mode TC	-0,07889	-2,5
Temps dans le véhicule mode TC	-0,04469	-2,3
<b>Mode alternatif (marche à pied)</b>		
Temps de marche à pied	-0,08927	-8,5
<b>Variables de taille de la zone de destination</b>		
Coefficient de taille	0,6427	15,6
Population de la zone	1	Fixé
Nombre d'emplois dans la zone**	0,2315	5,9
<b>Hiérarchie entre mode alternatif et mode motorisé</b>		
Logsum	0,8742	9,6

\* Pour ces constantes, la valeur en italique est la valeur estimée, l'autre correspond à la valeur corrigée tenant compte du phénomène d'échantillonnage.

\*\* C'est l'exponentielle du paramètre estimé qui est prise en compte comme coefficient.

Le modèle de choix de mode et de destination pour le motif affaires professionnelles a été estimé sur l'échantillon des déplacements observés sur l'ensemble des deux périodes (période creuse et période de pointe du soir). En effet, le nombre d'observations de l'enquête ménage pour ce motif était trop faible (284 enregistrements) pour pouvoir estimer un modèle en période de pointe du soir.

Une seule constante modale a été utilisée : la constante liée au mode TC, qui est négative. Les trois constantes liées aux types de destination présentent des estimations assez proches des autres motifs.

#### **Le modèle**

Il est basé sur 844 observations ( $\rho^2 \text{ bar} = 0.30$ ).

#### **Expliquer le niveau des coefficients**

Pour le mode VP, les paramètres, estimés avec une bonne précision, présentent une particularité : l'estimation du paramètre lié à la variable «temps du trajet VP supérieur à 10 minutes» est plus grande en valeur absolue que celle du paramètre lié à la variable «temps du trajet VP inférieur à 10 minutes». Dans le cas de ce motif, les premiers instants passés en voiture sont considérés comme peu pénalisants, alors que les instants suivants le sont beaucoup plus. Ceci est peut-être lié au fait que le temps de transport empiète trop sur le temps de travail.

Concernant le mode TC, le temps de rabattement est presque deux fois plus pénalisé que le temps dans le véhicule et les ordres de grandeur des coefficients sont les mêmes que pour les autres motifs.

#### **Le rôle de la variable de taille**

Le coefficient de taille n'est pas trop mauvais (0,64) pour un motif comme «affaires professionnelles» où il est difficile de représenter l'attractivité d'une zone. On peut remarquer qu'en ce qui concerne les variables de taille, le poids du nombre d'emplois de la zone n'est pas beaucoup plus important que le poids de sa population [ $\exp(0,2315) = 1,26$  pour 1].

#### **L'existence d'une hiérarchie peu marquée**

Le logsum a une valeur estimée à 0,87 ; la hiérarchie entre le choix de modes motorisés ou non motorisés n'est donc pas très marquée, mais elle existe.

## 5. Validation et scénarios

### 5.1. Application désagrégée sur l'échantillon

Cette partie traite plus spécifiquement des modèles de motorisation et de choix de mode et destination de type désagrégé, estimés en utilisant le logiciel ALOGIT.

#### Validation des modèles

Lors du travail d'estimation de ces modèles, certains tests ont été effectués régulièrement pour s'assurer qu'ils représentaient au mieux les observations réelles; c'est la phase de validation. Ces tests permettent notamment de modifier la structure du modèle si nécessaire, en prenant en compte de nouvelles variables plus explicatives ou en en rejetant d'autres.

Le principe de ces tests est d'appliquer le modèle estimé sur l'échantillon d'observations ayant permis son estimation : il s'agit d'une application désagrégée. L'idée est de s'assurer que le modèle est capable de reproduire les comportements sur lesquels il est basé. Ces applications sont, d'autre part, segmentées suivant des variables non incluses dans l'estimation des modèles, ce qui renforce la qualité de ces tests.

Un deuxième type de tests est de «faire réagir le modèle» : là encore, on applique le modèle sur l'échantillon ayant permis son estimation, mais on modifie l'une des variables explicatives. Ce sont des tests d'élasticité: il s'agit ici de déterminer quels sont les changements de comportements estimés par le modèle suivant l'évolution d'une variable.

Le premier type de tests peut être réalisé sur une segmentation de la population des individus de l'échantillon : il est possible de comparer les résultats effectivement observés et estimés suivant l'âge de la personne, le revenu de son ménage, etc.

Dans le cas des modèles de choix de mode et de destination, ces tests comparent donc pour une segmentation donnée le nombre de déplacements prévus par mode et aussi par type de destination.

Quelques exemples de ces tests, pour le modèle final de choix de mode et de destination pour le motif domicile-achats en période creuse sont donnés dans les tableaux suivants.

**Explications pour le premier type de test****Tableau 21 : Tests du choix de destination pour le motif domicile-achats suivant l'âge de la personne en période creuse**

Type de zone de destination	Choix réel / estimation	0 – 20 ans	21 – 50 ans	51 ans et plus	Total
Zone d'origine	Observés	76	481	956	1513
	Estimés	63	527	929	1519
Zone voisine	Observés	46	457	622	1125
	Estimés	55	441	643	1139
Zone du centre	Observés	20	118	151	289
	Estimés	17	109	169	295
Autre zone	Observés	41	336	364	741
	Estimés	49	334	385	768
Total	Observés	183	1392	2093	3668
	Estimés	184	1411	2126	3721

Le tableau précédent donne la répartition des déplacements observés dans l'échantillon et estimés par le modèle sur ce même échantillon, en fonction du type de destination choisi et de l'âge de la personne.

Le fait que le total de déplacements observés et le total de déplacements estimés soient différents s'explique par le fait que, lors de l'estimation, certains enregistrements ne sont pas pris en compte, parce qu'ils sont erronés au niveau de leur description du choix de mode ou de destination. Par contre, à l'application, ces mêmes enregistrements ne sont pas rejetés puisqu'on ne tient alors plus compte du choix effectif en terme de mode et de destination effectué par la personne.

On remarque que le modèle représente donc relativement bien les comportements observés, même s'il existe toujours certains écarts. Il est important de noter que, comme décrit dans le chapitre précédent, le modèle de choix de mode et de destination pour le motif domicile-achats ne comprend pas de variables explicatives liées à l'âge de la personne. Cette segmentation montre que cela n'est pas nécessaire, puisque le modèle représente bien les phénomènes observés sans cette variable.

Le même type de commentaire est possible pour le tableau suivant :

**Tableau 22 : Tests du choix de mode en période creuse pour le motif domicile-achats suivant le sexe de la personne**

Type de mode	Choix réel / estimation	Hommes	Femmes	Total
VP	Observés	729	954	1683
	Estimés	699	1003	1702
TC	Observés	80	154	234
	Estimés	95	145	240
Mode alternatif	Observés	660	1091	1751
	Estimés	697	1082	1779
Total	Observés	1469	2199	3668
	Estimés	1491	2230	3721

#### Explications pour le deuxième type de tests

Le deuxième type de test est un test d'élasticité. Nous présentons dans le tableau suivant deux tests sur les variations de choix de mode ayant été effectués pour le motif domicile-achats en période creuse quand :

- le temps passé dans le bus est diminué de 10%, c'est-à-dire que la vitesse est augmentée de 10% ;
- le temps d'attente est diminué de 20% car on augmente la fréquence de passage des TC.

**Tableau 23 : Tests d'élasticité du choix de mode pour le motif domicile-achats en période creuse**

Mode	Evolution du nombre de déplacements pour un temps de bus diminué de 10%	Evolution du nombre de déplacements pour un temps d'attente diminué de 20%
VP	- 0,2%	- 0,5%
TC	+ 3,8%	+ 9,5%
Mode alternatif	- 0,3%	- 0,8%

On remarque donc que l'élasticité au temps dans le bus en période creuse est un peu moins forte que l'élasticité au temps d'attente. On observe en effet pour le mode TC une élasticité de -0,38 au temps

dans le bus, alors que l'élasticité au temps d'attente est de -0,47. En effet, le temps d'attente en période creuse étant relativement long (fréquences de passage moins importantes), il est naturel de constater que son poids est déterminant dans le choix d'utiliser le mode TC. Nous sommes, ici, dans des fourchettes «habituellement» constatées. Les sensibilités sont bonnes.

### **Calculer des élasticités pour «alimenter» les études de transport**

Si les élasticités sont intéressantes pour tester les modèles, elles sont aussi généralement très utilisées dans les études concernant les transports en milieu urbain. Afin de mettre en évidence un des intérêts des modèles désagrégés, des calculs d'élasticités pouvant être utilisés dans le cadre d'études de transports ont été réalisés <sup>33</sup>.

Les calculs d'élasticités ont été effectués sur l'ensemble des motifs pour la période de pointe du soir, c'est-à-dire le moment où les moyens mis en oeuvre pour les transports collectifs sont les plus importants et le moment où les externalités négatives liées à la voiture sont les plus fortes (hormis pour le motif affaires professionnelles où il s'agit de la journée entière). Pour chacun des motifs, quatre tests ont été réalisés en prenant en compte pour chacun d'eux l'une des variations suivantes :

- diminution du temps dans le véhicule pour le mode TC de 10% ; cette variation correspond par exemple à une plus grande vitesse commerciale des bus,
- diminution du temps hors du véhicule pour le mode TC de 10% (appelé temps de rabattement) ; l'accessibilité au réseau TC est plus facile, les temps d'attentes sont moins longs,
- augmentation du temps dans le véhicule pour le mode VP de 10% ; la vitesse des voitures est diminuée du fait de ralentissements sur la route ou d'une diminution des vitesses autorisées,
- augmentation du temps de rabattement pour le mode VP de 10% ; les parkings sont plus éloignés du lieu de destination ou ils sont plus rares.

---

33. Pour des études sur l'agglomération de Lyon. Pour des études dans d'autres grandes agglomérations, les prendre avec précaution, les valeurs pouvant varier quelque peu (nous considérons que les résultats s'inscrivent dans des fourchettes). Un programme de recherche, piloté par le Certu, est en cours dans le domaine de la transposabilité de ce type de modèle. Il permettra de vérifier la «stabilité» ou non de ces coefficients d'élasticité entre les grandes agglomérations.



Ces quatre tests d'élasticité correspondent, en fait, à diverses actions possibles pour augmenter la part de marché des TC : il s'agit d'améliorer le service TC ou de rendre plus pénible l'usage de la voiture.

Les deux tableaux suivants présentent les résultats obtenus pour ces différents tests d'élasticité.

Le premier tableau donne l'évolution en pourcentage des volumes par rapport à la situation de base : ce tableau décrit donc en fait les élasticités aux différents temps. Il se lit de la manière suivante : une amélioration de 10% du temps de parcours en transports collectifs entraînera une diminution de 0,7% du trafic, pour le motif domicile-travail.

Le deuxième tableau donne l'évolution de la part de marché du TC dans l'univers des modes motorisés (VP + TC) et dans l'univers de tous les modes (VP + TC + mode alternatif) en point de pourcentage.

**Tableau 24 : Évolution du nombre de déplacements par mode et par motif pour quatre tests d'élasticités en PPS (sauf affaires professionnelles : période journée)**

Motifs	Modes	-10% temps TC	-10% temps rabatement TC	+10% temps rabatement VP	+10% temps VP
Domicile-travail	VP	-0,7%	-1,3%	-2,1%	-0,7%
	TC	4,1%	7,0%	6,4%	1,9%
	LT	-1,2%	-2,0%	4,6%	1,5%
Domicile-école	VP	-1,6%	-3,0%	-7,9%	-3,2%
	TC	5,3%	9,3%	2,6%	1,0%
	LT	-1,6%	-2,7%	2,3%	1,0%
Domicile-achats	VP	-0,4%	-0,5%	-3,2%	-1,3%
	TC	6,1%	7,1%	4,8%	1,8%
	LT	-0,7%	-0,8%	4,0%	1,7%
Domicile-loisirs	VP	-1,0%	-1,0%	-3,9%	-2,3%
	TC	5,0%	5,2%	4,0%	2,5%
	LT	-0,7%	-0,7%	2,2%	1,2%
Domicile-autre	VP	-0,3%	-0,9%	-2,7%	-1,4%
	TC	2,8%	9,0%	5,0%	2,5%
	LT	-0,3%	-1,0%	3,7%	1,8%
Secondaires	VP	-0,4%	-0,5%	-1,8%	-2,0%
	TC	4,0%	5,0%	4,5%	5,4%
	LT	-0,5%	-0,6%	3,3%	3,4%
Aff. pro.	VP	-0,2%	-0,4%	-0,9%	-0,8%
	TC	4,5%	8,9%	4,9%	4,2%
	LT	-0,5%	-0,9%	4,8%	3,9%

Ce tableau amène plusieurs commentaires.

— Pour le mode TC, l'élasticité au temps de rabatement TC est toujours plus importante que l'élasticité au temps dans le véhicule TC : le temps en dehors du véhicule (temps d'attente, temps de correspondance, accessibilité) est donc le plus pénalisant. Cette différence d'élasticité varie suivant les motifs de déplacement : elle est par exemple très sensible pour le motif domicile-autre et presque inexistante pour le motif domicile-loisirs. D'une manière générale, l'élasti-

cité au temps (temps TC dans le véhicule ou hors du véhicule) pour le mode TC est comprise entre -0,4 et -0,9.

— Une augmentation de 10% du temps passé dans la VP ou du temps de rabattement VP entraîne des augmentations de volumes de déplacements pour le mode TC et le mode alternatif relativement comparables. Les élasticités du mode TC et du mode alternatif aux temps VP sont du même ordre, l'élasticité du mode TC restant tout de même toujours plus forte.

— Il apparaît également que pour le mode TC, quelque soit le motif, l'élasticité au temps de rabattement TC est plus forte que l'élasticité au temps VP ou au temps de rabattement VP. Parallèlement, pour le mode VP, quel que soit le motif, l'élasticité au temps VP est plus forte que l'élasticité au temps TC ou au temps de rabattement TC. Ces résultats tendraient donc à montrer qu'une politique pro-transport collectif est plus efficace si elle s'appuie sur une amélioration de l'offre TC. Par contre, une politique environnementale, destinée à diminuer l'utilisation de la VP, serait plus efficace en s'appuyant sur une politique visant à contraindre l'offre VP (politique de stationnement par exemple).

**Tableau 25 : Évolution des parts de marché du mode TC pour quatre tests d'élasticités pour tous les motifs en PPS (sauf affaires professionnelles : période journée)**

Motifs	Evolution part de marché du TC	-10% temps TC	-10% rabattement temps TC	+10% temps VP	+10% rabattement temps VP
Domicile-travail	Modes motorisés	0,7%	1,2%	1,3%	0,4%
	Tous modes	0,7%	1,2%	1,0%	0,3%
Domicile-école	Modes motorisés	1,7%	3,0%	2,7%	1,0%
	Tous modes	1,2%	2,2%	0,6%	0,2%
Domicile-achats	Modes motorisés	0,7%	0,8%	0,8%	0,3%
	Tous modes	0,5%	0,5%	0,4%	0,1%
Domicile-loisirs	Modes motorisés	1,1%	1,2%	1,5%	0,9%
	Tous modes	0,7%	0,7%	0,6%	0,4%
Domicile-autre	Modes motorisés	0,4%	1,1%	0,9%	0,5%
	Tous modes	0,3%	0,8%	0,5%	0,2%
Second.	Modes motorisés	0,5%	0,6%	0,7%	0,8%
	Tous modes	0,4%	0,5%	0,4%	0,5%
Aff. pro.	Modes motorisés	0,2%	0,5%	0,3%	0,3%
	Tous modes	0,2%	0,4%	0,2%	0,2%

Ce tableau donne les évolutions de part de marché du TC en point de pourcentage dans l'univers des modes motorisés (VP et TC) et en considérant tous les modes (VP, TC et mode alternatif). Ainsi par exemple, ce tableau nous indique que pour le motif domicile-achats, en période de pointe du soir, la diminution de 10 % du temps dans le véhicule pour le mode TC entraîne une augmentation de 0,7 point de la part de marché du TC, en considérant uniquement les modes motorisés.

Il apparaît que le motif domicile-école présente, d'une manière générale, les plus grandes évolutions en terme de gain de points de part de marché pour les différents tests effectués. Pour ce motif, le gain est le plus important dans le cas d'une diminution du temps de rabattement TC : le TC concurrence alors notamment plus le mode alternatif.

Mis à part le cas des déplacements secondaires, on note que pour tous les motifs le gain en terme de part de marché pour le mode TC est plus prononcé dans le cas d'une diminution du temps de rabattement TC.

## 5.2. Application globale de la chaîne de modélisation

L'application globale de la chaîne de modélisation ne constituait pas le but premier de cette étude. Elle est présentée ici pour donner une idée des possibilités offertes par ce type de modélisation : application pour tester des politiques de transport, d'urbanisme, d'environnement. Elle permet également de vérifier la solidité du modèle. Mais il serait nécessaire de consacrer plus de temps à cette dernière partie pour obtenir une représentation plus fine des comportements de déplacements.

### Une application agrégée

L'application globale de la chaîne de modélisation s'effectue à un niveau agrégé. En effet, contrairement à l'application des modèles dans ALOGIT sur leurs échantillons d'estimation composés d'observations de déplacements, les modèles sont ici appliqués à un niveau zonal (le zonage de l'enquête ménage).

Pour chaque zone, la population est donc segmentée en fonction des caractéristiques, du ménage ou de la personne nécessaires à l'application des modèles. Par exemple, le modèle de génération pour le motif domicile-travail estime des taux de déplacements par âge et par activité (cf. chapitre 4.2), son application nécessite donc de connaître dans chaque zone la population correspondant à cette segmentation.

### Description de l'application

La chaîne de modélisation comprend trois grandes étapes :

- l'application des modèles de motorisation,
- l'application des modèles de génération,
- l'application des modèles de choix de mode et de destination.

A la suite de l'application de ces trois types de modèles, on obtient pour chaque motif et pour chaque mode une matrice de déplacement pour la période considérée. Dans ce paragraphe, nous traitons de l'application pour la période de pointe du soir (PPS). Ces matrices peuvent ensuite être affectées sur les réseaux (VP ou TC) afin d'estimer les trafics.

Une fois l'application globale effectuée, il est nécessaire de pouvoir comparer les résultats globaux obtenus et les comportements réellement observés. Cette comparaison s'effectue à partir des matrices de déplacements de base. Elles sont de deux types :

- la matrice de déplacements tous motifs du Cete de Lyon pour le mode VP : elle décrit les déplacements UVP en heure de pointe du soir et date de 1995. Le zonage utilisé pour cette matrice est un zonage de 382 zones ;
- la matrice de déplacements tous motifs fournie par la SLTC pour le mode TC : elle donne les déplacements de personnes en heure de pointe du soir pour le zonage 87 zones de l'enquête ménage. Elle date de 1996.

### Les difficultés existantes

Quelques difficultés existent concernant la comparaison des matrices estimées et observées, au niveau du mode VP notamment.

D'une part, le zonage de l'enquête ménage en 357 zones ne correspond pas du tout au zonage utilisé par le Cete qui comprend 382 zones. D'autre part, il s'avère que l'équivalence entre ces deux zonages est extrêmement difficile à obtenir car le zonage de l'enquête ménage comprend des zones dites «générateurs ponctuels». Or, nous ne possédons pas d'informations suffisantes sur le nombre de déplacements les concernant.

L'utilisation de la matrice des déplacements issus de l'enquête ménage (en considérant les poids des observations) peut cependant permettre d'obtenir une matrice de base pour le zonage 357 zones (D357). Mais cette matrice n'aura qu'une qualité médiocre car le nombre d'observations n'est pas suffisant pour un zonage si fin.

### Résoudre les difficultés

Pour effectuer les comparaisons des modes VP et TC, les moyens suivants ont donc été utilisés.

Pour le mode VP, la matrice estimée a été comparée à la matrice issue de l'enquête ménage sur un zonage bien moins fin que le zonage D357, afin de prendre en compte des volumes de déplacements zone à zone suffisamment importants pour que le biais dû à la pondération basée sur la population des ménages diminue fortement. Ce zonage est le zonage D33 (33 zones) de l'enquête ménage.

Cette comparaison matricielle a été accompagnée d'une comparaison graphique : la matrice estimée sur 357 zones et la matrice du Cete observée sur 382 zones ont toutes deux été affectées sur le réseau VP pour la période de pointe du soir. L'affectation a été effectuée en tout ou rien, c'est-à-dire que les chemins sont définis de manière à minimiser le coût du déplacement en considérant les vitesses à vide sur les voies. Il est ainsi possible de comparer les structures des deux matrices par l'intermédiaire des structures des réseaux chargés. Ces affectations n'ont donc aucun rapport avec les volumes de trafic effectifs sur le réseau.

Pour le mode TC, la comparaison est effectuée plus facilement sur le zonage D33 (zonage de l'enquête ménage) en agrégeant les résultats de la matrice TCL fourni au D87. Ce zonage assez grossier permet en effet de bien faire ressortir les structures de différentes matrices.

### Comparaisons générales

Avant de passer au détail des comparaisons, sont donnés, dans le tableau suivant, les différents volumes de déplacements par motif puis par mode obtenus après application de la chaîne de modélisation. Ces volumes sont confrontés aux volumes issus de l'enquête ménage, volumes obtenus par prise en compte du poids des ménages.

**Tableau 26 : Volumes de déplacements estimés par mode pour tous motifs en période de pointe du soir**

Mode	Nombre de déplacements dans l'enquête ménage	Nombre de déplacements estimés	Différence
VP	541 618	583 168	+ 7,7 %
TC	142 317	137 574	- 3,3 %
Mode alternatif	341 349	341 579	+ 0,1 %
Total	1 025 284	1 062 321	+ 3,6 %

Le mode VP semble surestimé alors que le mode TC est quelque peu sous-estimé. Le mode alternatif est quant à lui bien représenté en volume. L'ensemble de ces chiffres montre tout de même que les modèles de motorisation et de génération restent d'une manière globale cohérents avec les comportements observés. En effet, il convient de souligner le fait que nous comparons des nombres de déplacements estimés ici à partir de données zonales (type Insee) aux nombres de déplacements obtenus par pondération des observations de l'enquête ménage. Or, cette pondération est basée sur la population des ménages et ne correspond donc pas tout à fait à la réalité des choses en terme de déplacement.

#### Comparaisons détaillées

**Tableau 27 : Volumes de déplacements estimés par motif pour tous modes en période de pointe du soir**

Motif	Nombre de déplacements dans l'enquête ménage	Nombre de déplacements estimés
Domicile-travail	126 803	147 063
Domicile-école	118 828	128 149
Domicile-loisirs	100 277	73 576
Domicile-achats	139 980	120 272
Domicile-autre	301 981	333 197
Secondaires	209 051	221 571
Affaires professionnelles	28 364	38 493
Total	1 025 284	1 062 321

Le tableau 28 permet de s'assurer du bon ordre de grandeur des chiffres estimés. Il convient tout de même de noter que certains écarts entre les volumes estimés et observés dans l'enquête ménage sont relativement importants. Mais il est difficile de tenir compte de manière trop précise des volumes de l'enquête ménage puisque la pondération ne prend pas en compte les caractéristiques du type: âge des personnes dans le ménage, nombre de voitures possédées, etc. Or, ces caractéristiques jouent un rôle important et sont prises en compte dans l'application de la chaîne de modélisation.

De manière plus synthétique, on peut dire que les modèles prennent bien en compte la structure de la population, puisqu'ils s'appuient sur les données zonales. Par contre, l'utilisation des poids de l'enquête ménage ne permet pas une aussi bonne représentation de la population.

La différence en pourcentage des parts modales de la matrice estimée par rapport aux parts modales de la matrice provenant de l'enquête ménage, varie entre 1 et 3%. Les résultats montrent que la chaîne de modélisation représente bien la distribution des déplacements, au niveau global ainsi qu'au niveau intrazonal, par mode pour chacun des motifs.

Pour le motif affaires professionnelles, la distribution des déplacements intrazonaux suivant les modes est très différente de celle observée dans l'enquête ménage, mais les chiffres sont trop faibles pour qu'ils puissent être solides : on observe 1 253 déplacements en voiture pour l'intrazonal ce qui représente à peu près 10 observations dans l'enquête ménage, le poids moyen étant de l'ordre de 100.

#### **Analyse des résultats de la matrice de déplacements en mode VP tous motifs**

Il convient tout d'abord de rappeler qu'il n'est pas attendu une précision remarquable des résultats obtenus par la chaîne de modélisation. Ceci pour plusieurs raisons :

- les données à notre disposition, pour décrire le choix de destination pour certains motifs (domicile-loisirs par exemple), n'étaient pas suffisantes (cf. chapitre 2),
- certaines données importantes, tel que le nombre de places de stationnement payantes ou gratuites par zone, n'existaient pas; l'absence de telles données n'a pas permis de prendre en compte les coûts des différents modes, et notamment de la voiture.

L'objectif est donc plus de bien retranscrire la structure des matrices de déplacements, c'est-à-dire de vérifier que les modèles représentent bien les grandes tendances comportementales (au niveau choix de mode et au niveau choix de destination). Un travail plus précis peut



ensuite être envisagé sur la base des modèles estimés ici, mais il nécessite d'obtenir des données actuellement non disponibles.

Pour analyser la structure de la matrice de déplacements tous motifs pour le mode VP, nous la comparons avec :

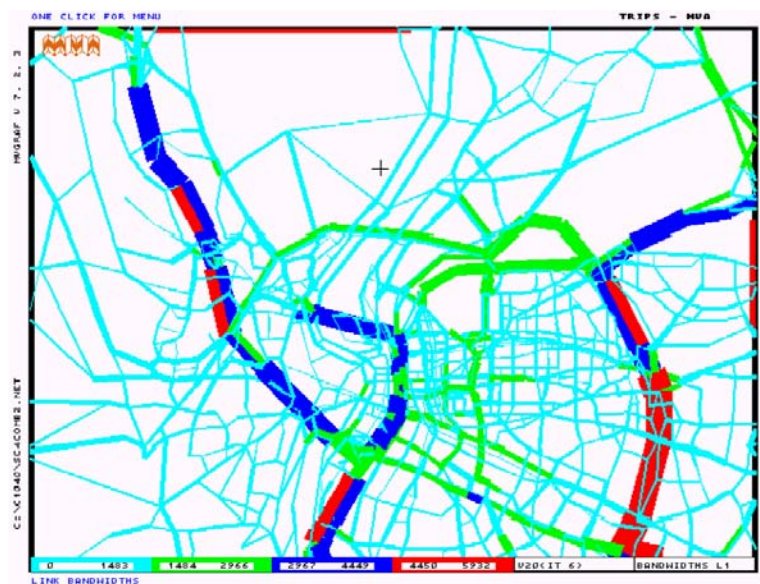
- la structure de la matrice issue de l'enquête ménage pour le zonage D33,
- la structure de la matrice du CETE en utilisant une affectation du réseau.

La comparaison des matrices obtenues à partir de l'enquête ménage et en utilisant la chaîne de modélisation permet les commentaires suivants :

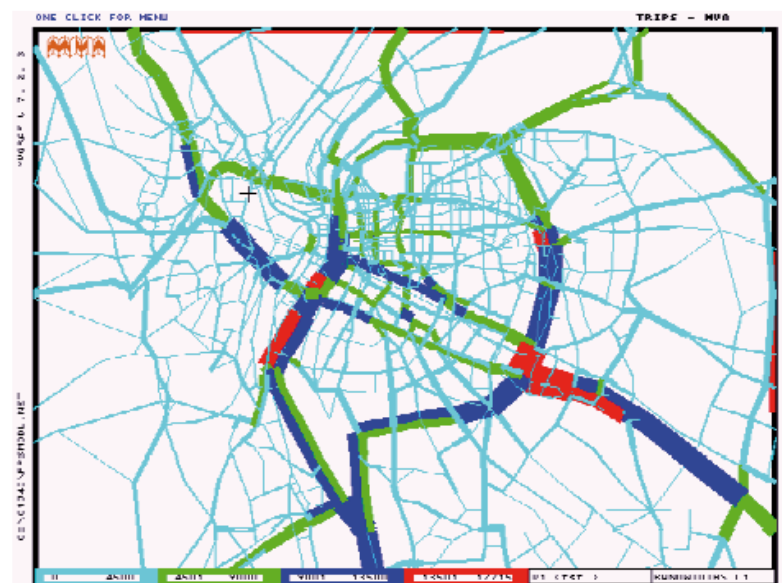
- d'une manière générale, l'estimation représente relativement bien les ordres de grandeur du nombre de déplacements entre chacune des zones d'origine et de destination ;
- l'intrazonal est, lui aussi, d'un bon ordre de grandeur dans la matrice estimée ;
- en hypercentre (zone 1,2 et 3), la voiture est surestimée. Ceci s'explique par l'absence de données sur le stationnement dans les modèles de choix de mode et de destination.

Les figures suivantes représentent l'affectation des matrices du Cete (382 zones) et de la matrice estimée (357 zones) pour le mode VP en période de pointe du soir. Cette affectation est réalisée comme décrit précédemment, elle permet donc une comparaison de la structure des déplacements estimés et observés.

**Figure 9 : Affectation sur le réseau de Lyon de la matrice du Cete (affectation tout ou rien, 1 itération)**



**Figure 10 : Affectation sur le réseau de Lyon de la matrice estimée (affectation tout ou rien, 1 itération)**



### **Analyse des résultats de la matrice de déplacements en mode TC tous motifs**

L'analyse des résultats obtenus pour le mode TC, en comparaison avec les observations réelles disponibles grâce à la matrice de la SLTC, se dégagent différents points :

Tout d'abord, il s'avère que, d'une façon globale, le mode TC est sous-estimé sur l'ensemble de la communauté urbaine.

Ceci est d'autant plus vrai pour les zones appartenant à l'hypercentre (zones 1, 2 et 3). En fait, on retrouve ici une conséquence de la non utilisation de données concernant le stationnement pour le mode VP. Ce mode était surestimé en hypercentre car il apparaît beaucoup trop concurrentiel en l'absence de coût de stationnement. Inévitablement, on sous-estime le TC pour ces mêmes zones et donc également au niveau global vu les volumes de déplacements qui sont concernés ici.

Mis à part ce point important, la structure globale de la matrice estimée est très correcte : les déplacements de courtes distances, c'est-à-dire les déplacements intrazonaux du zonage D33, représentent les volumes les plus importants, dans une matrice comme dans l'autre.

### **Conclusion**

L'analyse des résultats obtenus en appliquant l'ensemble de la chaîne de modélisation permet de souligner plusieurs points.

Les modèles ne permettent pas de connaître précisément quel nombre de déplacements est actuellement observé pour un mode déterminé entre telle et telle zone dans la communauté urbaine de Lyon. Résultat prévisible, lié aux données trop peu complètes et précises.

Les modèles permettent par contre de bien retranscrire la structure des déplacements, et c'était là un des objectifs principaux : les ordres de grandeurs sont bien respectés en termes de part modale et de choix de destination. Ils sont donc susceptibles de bien réagir à certaines variantes, en terme d'offre de transport par exemple. Par conséquent, il peut être intéressant d'utiliser ces modèles pour tester quelques scénarios de politique de transport ou d'urbanisme. Quatre scénarios testés sont présentés dans ce paragraphe.

## **5.3. Élaborations et tests de scénarios**

Ce chapitre décrit quatre scénarios de politique de transport ou d'urbanisme. Ils sont volontairement simples et n'ont pour but que de faire «réagir» les modèles estimés lors de cette étude, afin d'en montrer l'intérêt.

### Scénario 1 : Construction d'un grand centre commercial au Port Édouard Herriot

Dans l'estimation du modèle de base, la zone correspondant au Port Édouard Herriot affiche une population de 520 habitants et 2 312 emplois. On ajoute donc à ces données zonales, 500 emplois et une surface commerciale de 20 000 m<sup>2</sup>.

Les résultats de l'application de la chaîne de modélisation sont résumés dans le tableau suivant. Ce tableau donne les évolutions du nombre de déplacements concernant la zone du Port Édouard Herriot (zone 1 354 du zonage D357 de l'enquête ménage) pour chaque mode.

**Tableau 28 : Évolution pour le scénario 1**

Mode	Évolution du nombre de déplacements générés par la zone 1354 en PPS	Evolution du nombre de déplacements dirigés vers la zone 1354 en PPS
VP	+ 15,5 %	+ 7,9 %
TC	+ 15,5 %	+ 2,3 %
Mode alternatif	+ 10,5 %	+ 4,8 %
Tous modes	+ 15,3 %	+ 6,7 %

Le nombre de déplacements générés par la zone du Port Édouard Herriot évolue donc beaucoup plus que le nombre de déplacements dirigés vers cette zone.

En effet, en période de pointe du soir, les déplacements générés par les emplois d'une zone sont importants (motif travail-domicile). Par conséquent, les 500 emplois rajoutés à cette zone augmentent considérablement les déplacements qu'elle génère. D'autre part, les déplacements dirigés vers cette zone pour le motif achats (déplacements du type achats-domicile) génèrent également des déplacements ayant pour origine le Port Édouard Herriot.

L'augmentation du nombre de déplacements ayant pour destination le Port Édouard Herriot est elle aussi sensible, mais tout de même moins marquée.

Il faut également noter que les estimations de la chaîne de modélisation ne montrent pratiquement pas de changement pour les autres grandes surfaces de la communauté urbaine de Lyon. L'influence de ce nouveau centre commercial semble plutôt locale.

### Scénario 2 : Délocalisation de 10 % de la surface commerciale de la banlieue vers le centre de Lyon

Ce scénario consiste à faire basculer 10 % de la surface commerciale, actuellement située en banlieue et principalement dans les grands centres commerciaux, vers le centre ville de Lyon.

Pratiquement, les surfaces commerciales de chaque zone n'appartenant pas au centre de Lyon ont donc été diminuées de 10% ; la surface commerciale totale ainsi retranchée a ensuite été redistribuée de manière uniforme sur chaque zone du centre : un peu plus de 2 000 m<sup>2</sup> de surface commerciale ont donc été ajoutés à chacune des zones du centre (49 au total).

Dans le même temps, le nombre d'emplois de chaque zone a été modifié en considérant une équivalence de 100 emplois pour 10 000 m<sup>2</sup> de surface commerciale. Le nombre d'emplois a donc diminué en banlieue et augmenté au centre de Lyon.

Les résultats globaux de l'évolution des déplacements obtenus à l'estimation de ce scénario sont résumés dans le tableau suivant.

**Tableau 29 : Évolution pour le scénario 2**

Mode	Évolution globale du nombre de déplacements par mode en PPS	Évolution du nombre de déplacements générés par une zone du centre en PPS	Évolution du nombre de déplacements dirigés vers une zone du centre en PPS
VP	- 0,07 %	+ 1,32 %	+ 0,61 %
TC	+ 0,05 %	+ 1,11 %	+ 0,10 %
Mode alternatif	+ 0,10%	+ 1,46 %	+ 1,24 %
Tous modes	0 %	+ 1,33 %	+ 0,80 %

On observe tout d'abord que la part modale des TC augmente très légèrement, ainsi que la part modale du mode alternatif. On peut expliquer cette évolution par l'augmentation des déplacements internes dans le centre ville due au fait que les commerces y sont plus nombreux : il est moins nécessaire d'effectuer ses achats en périphérie. Les déplacements sont donc plus courts d'où une évolution positive de la part du mode marche à pied. D'autre part, pour ces déplacements, l'offre TC est relativement bonne (centre ville) et ce mode de transport est donc plus utilisé.

Pour tous les modes, le nombre de déplacements ayant pour destination ou pour origine une zone du centre augmente dans un ordre de

grandeur de 1%. Notons que cette évolution est plus forte pour les déplacements générés par le centre : en effet, en période de pointe du soir, plus de 60 % des déplacements entre le domicile et le lieu d'achats s'effectuent dans le sens achats-domicile, et l'augmentation des emplois au centre ville entraîne également une plus forte génération de déplacements secondaires ou de type travail-domicile.

Les évolutions pour le mode TC sont moins conséquentes que celles du mode VP. Il faut ici tenir compte du fait que le mode voiture est surévalué au centre ville puisqu'il n'est pas tenu compte des problèmes de stationnement.

### Scénario 3 : Multiplication par deux des fréquences de bus sur tout le réseau TCL

Le troisième scénario aborde une problématique plus directement liée au transport : il s'agit de tester quels peuvent être les effets d'une amélioration de l'offre TC. Cette amélioration est ici concrétisée par une plus forte fréquence des bus sur le réseau, puisque cette fréquence est doublée. Globalement, le temps d'attente d'un bus est donc divisé par deux.

Les résultats de l'application sont résumés dans le tableau suivant.

**Tableau 30 : Évolution pour le scénario 3**

Mode	Évolution globale du nombre de déplacements par mode en PPS	Évolution pour le centre ville du nombre de déplacements par mode en PPS
VP	- 0,60 %	- 0,25 %
TC	+ 4,69 %	+ 1,30 %
Mode alternatif	- 0,86 %	- 0,08 %
Tous modes	0%	+ 0,04 %

On constate alors que le nombre de déplacements s'effectuant en transport en commun augmente globalement de 5 % sur l'ensemble de la communauté urbaine de Lyon. L'élasticité au temps d'attente est donc de l'ordre de - 0,1. Ce chiffre est relativement faible ; deux raisons peuvent être avancées :

- la période sur laquelle est testée ce scénario correspond à la période de pointe du soir : les fréquences des TC sont donc déjà très bonnes et la diminution du temps d'attente n'est donc pas fortement ressentie,
- l'absence de contrainte de stationnement en centre ville avantage énormément le mode VP ; une diminution par deux du temps d'attente des bus n'a donc pas le même impact sur le comportement des personnes, leur choix continuant à se porter sur la voiture.

L'augmentation du nombre de déplacements par le mode TC est beaucoup moins sensible au centre ville que sur l'ensemble de l'agglomération. On remarque aussi que l'offre TC est meilleure en centre ville que dans le reste de la communauté urbaine. Par conséquent, les temps d'attente évoluent moins en grandeur absolue et l'impact sur le choix de mode est donc moindre.

**Scénario 4 : mise en place d'une partie du tronçon nord du périphérique de Lyon sans tarification et sans changement sur le reste du réseau de voirie.**

Ce dernier test est consacré à la voirie, et plus précisément à la voirie : il s'agit d'intégrer une partie du tronçon nord périphérique de Lyon (TNP) au réseau de Lyon. Aucun changement n'est effectué sur le reste du réseau (aucune évolution de capacité notamment) et cette portion du TNP est considérée comme étant gratuite.

Les résultats globaux obtenus après l'application de la chaîne de modélisation sur ce scénario sont donnés dans le tableau suivant. Nous joignons également les représentations des réseaux de voirie chargés avec, et sans, la portion du TNP (l'affectation de la matrice du CETE est effectuée comme indiqué au chapitre 3.3). Leur comparaison permet de rendre compte de l'influence «naturelle» de cette voirie sur la distribution des flux à Lyon. On utilise le terme «naturel» car la tarification des péages n'est pas prise en compte, les restrictions de capacités sur les autres voies du réseau non plus.

**Tableau 31 : Évolution pour le scénario 4**

Mode	Évolution globale du nombre de déplacements par mode en PPS	Évolution pour le centre ville du nombre de déplacements par mode en PPS
VP	+ 0,19 %	+ 0,20 %
TC	- 0,51 %	- 0,38 %
Mode alternatif	- 0,12 %	- 0,10 %
Tous modes	0%	-0,01 %

On remarque que la part modale du mode VP augmente, ce qui est normal, puisque la prise en compte de la voirie représente une amélioration de l'offre pour ce mode. Rappelons que cette portion du TNP est ici considérée comme gratuite et que le reste du réseau n'a pas évolué. Il s'agit donc d'un «plus».

L'utilisation du TC baisse proportionnellement plus en banlieue. Ceci est dû essentiellement au fait que le modèle estime que beaucoup de

personnes habitant au nord-est de Lyon délaissent à présent le TC au profit de la voiture (influence directe du TNP).

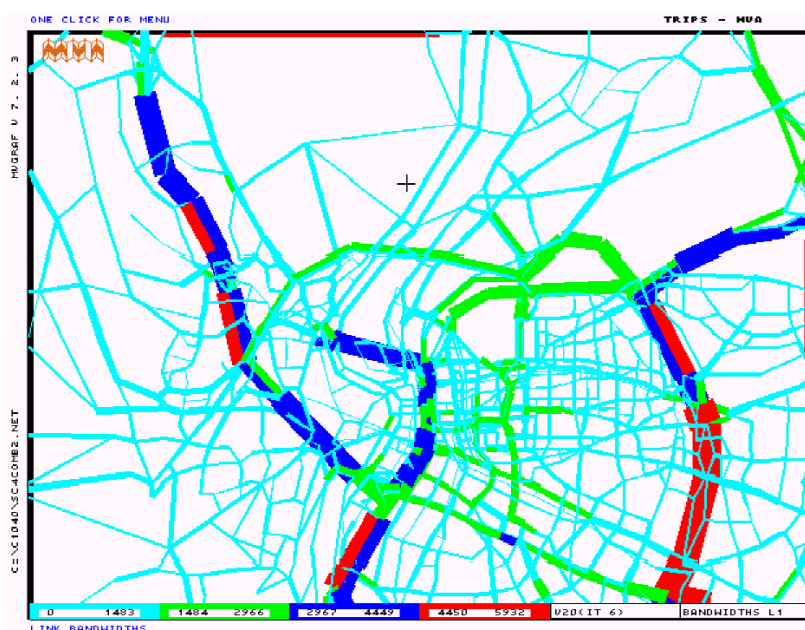
Les deux pages suivantes montrent donc les réseaux chargés sans, puis avec, le tronçon du TNP, et permettent de visualiser les changements de répartition des flux.

**Figure 11 : Réseau chargé avec la matrice du Cete (uvp tous motifs) en heure de pointe du soir sans prendre en compte le tronçon du TNP**





**Figure 12 : Réseau chargé avec la matrice du Cete (uyp tous motifs) en heure de pointe du soir en prenant en compte le tronçon du TNP**



## Conclusion

### Analyse des résultats au regard des objectifs fixés

L'élaboration de la chaîne de modélisation effectuée lors de cette étude avait un double objectif :

- permettre une amélioration des outils existants de prévision destinés à évaluer des politiques de transport et d'aménagement dans les grandes agglomérations en testant un type de modèle encore très peu utilisé : un modèle désagrégé de choix de mode et de destination,
- construire un outil de prévision suffisamment solide et suffisamment souple pour pouvoir être transféré à d'autres agglomérations et pour pouvoir servir de base à l'élaboration d'outils aux objectifs plus spécifiques.

Les résultats obtenus lors de l'étude concernant l'estimation de cette chaîne de modélisation sur l'agglomération lyonnaise répondent à ces objectifs pour les raisons suivantes :

— l'estimation des modèles, et notamment des modèles de choix de mode et de destination, a été effectuée en utilisant des données accessibles et sûres ; les résultats montrent que ces modèles sont estimés avec une très bonne précision. La structure des modèles est donc solide car elle explique bien les grandes tendances comportementales ; elle est aussi souple car des variables explicatives peuvent facilement y être intégrées,

— les résultats de l'application de la chaîne de modélisation montrent que les répartitions modales sont bien respectées ; la structure de choix de destination est elle aussi bien représentée. Certaines faiblesses demeurent pourtant ; elles sont principalement dues aux manques de certaines données importantes (sur le stationnement par exemple). Ces données pourront toujours être ajoutées dans les modèles le cas échéant.

### **Transposabilité des modèles**

La transposabilité des modèles développés ici pourra s'effectuer de plusieurs manières suivant les objectifs fixés :

— il peut s'agir d'appliquer directement les modèles estimés dans Lyon en corrigeant seulement certains paramètres pour tenir compte d'une répartition modale différente. Le transfert du modèle nécessitera alors tout de même l'utilisation de l'enquête ménage pour pouvoir mener à bien cette correction,

— il est également possible de transférer uniquement la structure des modèles en y ajoutant même certaines autres variables ; l'estimation des modèles est alors complètement reprise : cette opération nécessite toutes les données nécessaires à l'estimation des modèles développés ici, c'est-à-dire une enquête ménage, un réseau de voirie pour le mode VP et une matrice de déplacements pour ce mode, un réseau TC, une carte des îlots renseignés par le nombre d'emplois et la population pour les calculs d'accessibilité... Cette deuxième solution prend donc un peu plus de temps puisqu'il est nécessaire de construire un nouveau fichier de calibrage, mais les résultats obtenus seront certainement de meilleure qualité.



## Conclusion générale

Les modèles de choix discret dans la forme où ils sont présentés dans ce rapport, sont très intéressants pour tester des politiques de transport, (voire d'urbanisme si des variables adaptées sont considérées), dans le cadre d'un Plan de Déplacements Urbains ou d'un Dossier de Voirie d'Agglomération. En effet, ils permettent de tenir compte des différents modes que les individus utilisent pour leur déplacements : voiture, transports collectifs, modes alternatifs (marche, vélo).

L'approche désagrégée renforce leur puissance : on peut ainsi voir l'impact d'une mesure sur un segment particulier de la population (par exemple les individus à haut ou à faible revenu). Un autre intérêt est qu'ils peuvent être utilisés à un niveau plus ou moins fin de la politique des transports (choix de projet et/ou choix de variantes), si l'on se donne la peine de développer la phase d'affectation de la demande sur les réseaux. Un autre avantage est qu'ils s'inscrivent dans un cadre théorique très cohérent et qu'ils s'estiment avec des méthodes robustes. En ce sens, ils sont moins sujet à polémique, ou à caution quant à la reproduction des comportements à la date de référence choisie : un modèle sera soit conservé, soit rejeté, si on le considère de mauvaise qualité. Dans bien d'autres types de modélisation, cette logique n'est pas présente : comme on estime «manuellement» les paramètres, sans procédure statistique ni test de contrôle, on trouve toujours «le modèle» qui reproduit bien la situation de référence, même si l'on n'arrive pas à justifier les paramètres utilisés...

En dehors des limites recensées, qui concernent d'ailleurs également les autres types de modèles, on notera que les modèles de choix discrets présentés pèchent sur deux aspects, qui sont pourtant des «leviers» de plus en plus importants de la politique des transports :

- la prise en compte du stationnement.
- la prise en compte des tarifs de TC et des coûts d'usage de la VP.

D'un point de vue technique, rien n'empêche de prendre en compte ces variables. Dans le premier cas, le manque de données est flagrant, pour Lyon comme certainement pour la majorité des agglomérations françaises. Ils concernent surtout le stationnement de surface non payant non délimité et le stationnement privé des résidents, qui représentent une masse considérable de places. Dans le deuxième cas, des études appropriées doivent être menées pour intégrer ces variables au fichier de calibration (enquêtes tarifaires synchroniques, enquêtes préférences déclarées), car, sinon, le prix ne sera qu'une constante.

L'enrichissement de ces modèles sur ces deux points peut donc se faire très facilement, si les deux difficultés précédentes sont levées.

Le dernier élément important concerne le coût de développement de tels modèles. Pour ce qui est du cas pratique traité dans ce document, il faudra compter environ 900 KF, répartis de la façon suivante : 70% de maîtrise d'oeuvre de bureau d'études spécialisé, 20% de suivi technique et de maîtrise d'ouvrage et 10% de données (acquisition et mise en forme). Un tel coût peut paraître élevé, mais il est le gage de la mise à disposition d'un outil performant et sûr.

## Annexes

### Annexe 1 : Démonstration de la formulation probit

Distinguons deux cas :

#### a - le cas de la formulation probit binomiale (2 choix)

Supposons que les  $e_{it}$  et les  $e_{jt}$  sont distribués selon une loi normale, avec une moyenne nulle, et de variances respectives  $s_i^2$  et  $s_j^2$ . Supposons de plus que leur covariance est  $s_{ij}^2$ . Sous ces hypothèses, le terme  $e_{jt} - e_{it}$  est aussi normalement distribué de moyenne nulle, mais avec une variance  $s^2 = s_i^2 + s_j^2 - 2s_{ij}^2$ .

On peut utiliser ce résultat pour déduire les probabilités de choix par :

$$P_{it} = P(e_{jt} - e_{it} \leq V_{it} - V_{jt}) =$$

$$\int_{-\infty}^{V_{it} - V_{jt}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2\right] d\varepsilon, \quad \sigma > 0.$$

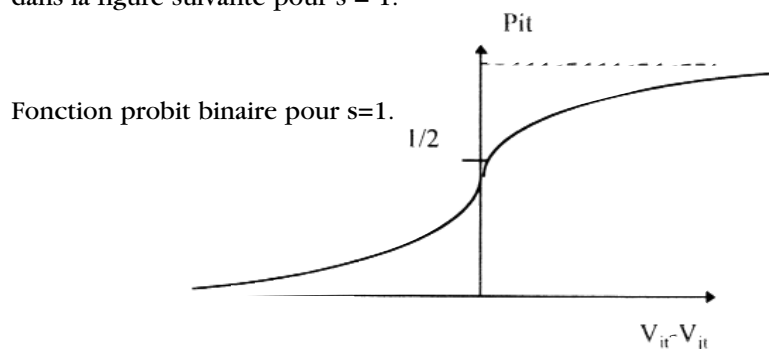
d'où

$$P_{it} = \Phi\left(\frac{V_{it} - V_{jt}}{\sigma}\right)$$

où  $\Phi$  est la distribution normale de moyenne nulle.

La probabilité de choix probit dépend seulement de  $s$ , et non de  $s_i^2$ ,  $s_j^2$  et  $s_{ij}^2$ . Aussi, les variances et covariances des deux perturbations sont sans rapport avec les probabilités de choix.

Le graphe d'une fonction probit binomiale (2 alternatives) est donné dans la figure suivante pour  $s = 1$ .



Remarque : l'hypothèse de distribution des  $\varepsilon_{it}$  et des  $\varepsilon_{jt}$  selon une loi normale, avec une moyenne nulle, et de variances respectives  $s_i^2$  et  $s_j^2$ , est équivalente à supposer que les  $U_{it}$  et les  $U_{jt}$  sont distribués selon une loi normale, avec une moyenne les  $V_{it}$  et les  $V_{jt}$ , et de variances respectives  $s_i^2$  et  $s_j^2$ .

### b - le cas de la formulation probit multinomiale (>2 choix)

Le problème d'une formulation probit est qu'elle n'est pas aisée à manipuler lorsque l'univers de choix est supérieur à deux alternatives (calibrage<sup>33</sup> délicat, peut être coûteux en temps à la mise en œuvre,...). Dans ce dernier cas, on étend en effet la formulation Probit binomiale présentée ci-avant en supposant que le vecteur des erreurs  $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Jt}]$ , est multinomial distribué normalement (MVN), avec un vecteur de moyennes égal à 0, et une matrice variance-covariance de dimension  $J \times J$ ,  $M$

$$M = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1t} \\ \sigma_{21} & \sigma_2^2 & \\ \sigma_{t1} & & \sigma_t^2 \end{pmatrix}$$

On veut calculer :

$P_{it}$  = Probabilité ( $\varepsilon_{it} \leq V_{it} - V_{jt} + \varepsilon_{jt}$ ,  $\forall j \in C_t, j \neq i$ )

Soit  $U$  le vecteur des utilités

$$U = \begin{pmatrix} U_{1t} \\ U_{2t} \\ U_{it} \\ U_{jt} \\ \vdots \\ U_{Jt} \end{pmatrix}$$

Nous supposons donc que  $U$  suit une loi normale :  $N(V, M)$  où  $V$  est le vecteur moyenne des utilités déterministes, et une matrice variance-covariance de dimension  $J \times J$ ,  $M$ .

33. Le calibrage correspond à l'estimation des coefficients  $\theta_i$  de la partie déterministe  $V_{it}$  de la fonction d'utilité  $U_{it}$ .

$$V = \begin{pmatrix} V_{1i} \\ V_{2i} \\ V_{ii} \\ V_{mi} \end{pmatrix} \quad M = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1i} \\ \sigma_{21} & \sigma_2^2 & \\ \sigma_{i1} & & \sigma_i^2 \end{pmatrix}$$

Concentrons nous tout d'abord sur la probabilité de faire le choix 1 :  $P_{1i}$ .

Posons  $W$  comme étant le vecteur des différences entre  $U_1$  et  $U_j$  pour tout  $j$ .

$$W = \begin{pmatrix} U_{1i} - U_{2i} \\ U_{1i} - U_{ii} \\ U_{1i} - U_{mi} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & & -1 \\ 1 & 0 & -1 \\ 1 & & & -1 \end{pmatrix} \cdot U = A \cdot U$$

Donc  $W$  est gaussien de moyenne  $AV$ , et de matrice de covariance.

En effet, si nous reprenons le calcul de l'espérance de  $W$ , nous obtenons :

$$E[(W - E(W))' (W - E(W))]$$

$$E[W'W] = E[A \cdot U' U' A] = A \cdot M' A = M_1$$

D'où, pour la probabilité du choix 1, son calcul nécessite la résolution d'une intégrale de dimension  $J$  :

$$P_{1i} = P(W \geq 0) = \int_0^{+\infty} \dots \int_0^{+\infty} \exp\left(-\frac{1}{2} (W - E(W))' M_1^{-1} (W - E(W))\right) dW$$





## Annexe 2 : Démonstration de la formulation du modèle logit

La distribution de Gumbel de paramètre  $m$  et  $b$  est définie par : une variable aléatoire  $e$  suit une distribution de Gumbel de paramètre  $m$  et  $b$  si celle-ci s'exprime par la formule

$F(\varepsilon) = e^{-e^{-\mu(\varepsilon-\eta)}}$ , avec  $m > 0$ , et la fonction de densité de Gumbel est

$$f(\varepsilon) = \mu \cdot e^{-\mu(\varepsilon-\eta)} \cdot e^{-e^{-\mu(\varepsilon-\eta)}}.$$

Cette distribution est similaire à la normale et donne des résultats équivalents du modèle probit sous l'hypothèse d'indépendance des  $e_{in}$ .

On a :

$P_n(i)$  = Probabilité ( $e_{jn} \leq V_{in} - V_{jn} + e_{in}$ ,  $\forall j \in C_n, j \neq i$ )

avec  $e_{in}$  et  $e_{jn}$  variables aléatoires indépendantes.

d'où,

$$P_n(i) = \int_{\alpha \in \text{domaine}(e_{in})} P(\varepsilon_{jn} \leq V_{in} - V_{jn} + \alpha, \forall j \in C_n, j \neq i / \varepsilon_{in} = \alpha) * P(\varepsilon_{in} = \alpha)$$

Si  $F(\cdot)$  est la fonction de distribution pour les termes  $e_{jn}$  de toutes les alternatives, alors

$$P(\varepsilon_{jn} \leq V_{in} - V_{jn} + \alpha, \forall j \in C_n, j \neq i / \varepsilon_{in} = \alpha) = \prod_{j \in C_n, j \neq i} F(V_{in} - V_{jn} + \alpha)$$

Si  $e_{in}$  est une variable aléatoire continue de fonction de densité marginale  $f_i(\cdot)$ . On doit intégrer sur le domaine de valeur de  $e_{in}$  :  $]-\infty, +\infty[$ .

$$P_n(i) = \int_{-\infty}^{+\infty} F(V_{in} - V_{jn} + \varepsilon_{in}, \forall j \in C_n, j \neq i) \cdot f_i(\varepsilon_{in}) \cdot d\varepsilon_{in}$$

Dans le cas où la distribution est de Gumbel, on obtient :

$$\begin{aligned} P_n(i) &= \int_{-\infty}^{+\infty} \prod_{j \in C_n, j \neq i} \exp\left[-e^{-\mu(V_{in} - V_{jn} + x - \eta)}\right] \mu \cdot e^{-\mu(x-\eta)} \cdot e^{-e^{-\mu(x-\eta)}} dx \\ &= \int_{-\infty}^{+\infty} \prod_{j \in C_n, j \neq i} \exp\left[-e^{-\mu(V_{in} - V_{jn} + x - \eta)}\right] \mu \cdot e^{-\mu(x-\eta)} dx \end{aligned}$$

$$\begin{aligned}
P_n(i) &= \int_{-\infty}^{+\infty} \exp \left[ -\sum_j e^{-\mu(V_{in} - V_{jn} + x - \eta)} \right] \mu \cdot e^{-\mu(x - \eta)} dx \\
&= \int_{-\infty}^{+\infty} \exp \left[ -\sum_j (e^{-\mu(V_{in} - V_{jn})} \cdot e^{-\mu(x - \eta)}) \right] \mu \cdot e^{-\mu(x - \eta)} dx \\
&= \int_{-\infty}^{+\infty} \exp \left[ -e^{-\mu(x - \eta)} \cdot \sum_j e^{-\mu(V_{in} - V_{jn})} \right] \mu \cdot e^{-\mu(x - \eta)} dx \\
P_n(i) &= \left[ \sum_j e^{-\mu(V_{in} - V_{jn})} \cdot \exp \left[ -e^{-\mu(x - \eta)} \cdot \sum_j e^{-\mu(V_{in} - V_{jn})} \right] \right]_{-\infty}^{+\infty} \\
&= \sum_j e^{-\mu(V_{in} - V_{jn})} \left[ \exp \left[ -e^{-\mu(x - \eta)} \cdot \sum_j e^{-\mu(V_{in} - V_{jn})} \right] \right]_{-\infty}^{+\infty} \\
&= \frac{1}{\sum_j e^{-\mu(V_{in} - V_{jn})}} \left[ \lim_{x \rightarrow +\infty} \left( \exp \left[ -e^{-\mu(x - \eta)} \cdot \sum_j e^{-\mu(V_{in} - V_{jn})} \right] \right) - \lim_{x \rightarrow -\infty} \left( \exp \left[ -e^{-\mu(x - \eta)} \cdot \sum_j e^{-\mu(V_{in} - V_{jn})} \right] \right) \right]
\end{aligned}$$

On peut voir que :

$$\lim_{x \rightarrow +\infty} (-e^{-\mu(x - \eta)}) = 0 \text{ car } \mu > 0$$

$$\text{d'où } \lim_{x \rightarrow +\infty} \left( \exp \left[ -e^{-\mu(x - \eta)} \cdot \sum_j e^{-\mu(V_{in} - V_{jn})} \right] \right) = 1$$

et

$$\lim_{x \rightarrow -\infty} (-e^{-\mu(x - \eta)}) = -\infty \text{ car } \mu > 0$$

$$\text{d'où } \lim_{x \rightarrow -\infty} \left( \exp \left[ -e^{-\mu(x - \eta)} \cdot \sum_j e^{-\mu(V_{in} - V_{jn})} \right] \right) = 0$$

D'où, on obtient :

$$\begin{aligned}
P_n(i) &= \frac{1}{\sum_j e^{-\mu(V_{in} - V_{jn})}} \\
&= \frac{e^{\mu V_{in}}}{\sum_j e^{\mu V_{jn}}}
\end{aligned}$$

## Bibliographie

Ben-Akiva (M.), Lerman (S.R.),  
***Discret choice analysis, Theory and application to travel demand,***  
 MIT press, 1985, 389 p.

Ben-Akiva, (M.),  
***Structure of passengers, travel demand models,***  
 PHD dissertation. Department of Civil Engineering, MIT, Cambridge, Mass USA., 1973.

Bonnafeous (A.),  
***Le siècle des ténèbres de l'économie,***  
 Economica, Paris, 1989.

Bonnafeous (A.),  
***«Systèmes de transport et systèmes de solidarité, un choc de priorités sociales»,***  
 Economie et Humanisme, n°323, Octobre-Décembre 1992, p. 37-42.

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

Cetur,  
***Les déplacements domicile-travail et domicile-école, Modèles désagrégés de choix modal, application au cas de l'agglomération grenobloise,***  
 1985, 71 p.

Cetur, LET, Inrets,  
***Déplacements, Modélisation,***  
 Recherches sur la mobilité quotidienne et renouvellement des méthodes de modélisation, Cetur, 1990, 88 p.

Clarke, (M.), Ryan, (J.) et Dejong (G.),  
***«Choix du lieu de résidence exemple d'application à New York et au Pays-Bas»,***  
 présentation à l'École Nationale des Ponts et Chaussées, 1989.

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

Domenich (T.), Mc Fadden (D.),  
***Urban travel demand : a Behaviourial Analysis***,  
 North-Holland, Amsterdam, 1975.

Inrets,  
***Modèles désagrégés de choix modal. Réflexions méthodologiques autour d'une prévision de trafic***,  
 Rapport Inrets n°67, Juin 1988, 65 p.+ annexes.

Kanafani (A.),  
***Transportation demand analysis***,  
 Mc Graw Hill Book Company, 1975, 320 p.

Leurent (F.),  
***«Pour certifier un modèle - application à l'affectation bicritère»***,  
 Thèse de l'ENPC, Janvier 1997, INRETS, Arceuil.

Ortuzar (J. de D.), Willumsen (L.G.),  
***Modelling transport***,  
 2d edition, Willey, 1985, 439 p.

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

Raux (C.),  
***Modèles et prévision des comportements de mobilité quotidienne***,  
 Thèse de Docteur-Ingénieur, Université Lyon II, ENTPE, 1983, 255 p.

Richards, (M.) et Ben-Akiva (M.),  
***A disaggregate travel demand model***,  
 D.C. Heath, Lexington, Mass. USA., 1975.

Rousseau (J.), Saut (C.),  
***«Tests de politiques de transport»***,  
 RATP, 1991, pp. 12-19

Thurston (L.),  
***A law of comparative judgement***,  
 Psychological review, n°34, p. 273-286, 1927.

Tirole (J.),  
***Théorie de l'organisation industrielle***,  
 Tome I, Economica, Février 1993, 417 p.

*Bibliographie*

Warner (S.L.),  
***Stochastic choice of mode in urban travel : a study in binary choice***,  
Northwestern University Press, Evanston, USA., 1962.

Wonnacott (T.H.), Wonnacott (R.J.),  
***Statistique***,  
Economica, 4<sup>ème</sup> édition, 1995, 920 p.

Watanatada, (T.) et Ben-Akiva (M.) ;  
Forecasting urban travel demand for quick policy analysis with  
disaggregate choice models : a Monte Carlo simulation approach,  
***Transport Research A13*** : p. 241-248, 1979.

© Ministère de l'Équipement, des Transports et du Logement  
Centre d'études sur les réseaux, les transports, l'urbanisme  
et les constructions publiques  
Coordination: Patrick Marchand

Achevé d'imprimer: juin 1998  
Dépôt légal: 2ème trimestre 1998  
ISSN: 0247 1159

CERTU  
Bureau de vente:  
9, rue Juliette-Récamier  
69456 Lyon Cedex 06  
☎ 04 72 74 59 59

Toute reproduction intégrale ou partielle, faite sans le consentement du CERTU est illicite (loi du 11 mars 1957). Cette reproduction par quelque procédé que ce soit, constituerait une contrefaçon sanctionnée par les articles 425 et suivants du code pénal.







## MOBILITÉ ET TRANSPORT

**L**

e travail présenté dans ce rapport a pour objet de répondre à deux préoccupations du ministère de l'Équipement, des transports et du logement, et plus particulièrement de son centre d'études (le Certu) :

— impulser un mouvement pour la diffusion de la modélisation désagrégée et multimodale des comportements de déplacements en France. Très utilisée dans les pays anglo-saxons et nordiques, elle a l'avantage d'être robuste en termes d'estimation et pédagogique au niveau de la présentation des résultats ;

— participer au renouvellement des méthodes, en testant une chaîne de modélisation qui soit basée sur la structure séquentielle à quatre étapes – génération, distribution, choix modal, affectation –, utilisée dans sa dimension agrégée en France. Ce renouvellement nécessaire se rattache à deux constats :

- le retard pris en France dans le domaine de la modélisation de la demande,
- le développement des dossiers de planification (type plan de déplacements urbains ou dossiers de voirie d'agglomération), pour lesquels les outils actuellement utilisés ne sont pas toujours adaptés.

Ce document s'adresse aux responsables du domaine des transports des collectivités locales, aux ingénieurs et économistes des transports et aux étudiants.

**The Discrete-choice models: Towards the development of disaggregate and multimodal models of trip-making patterns in urban areas**

The work presented in this report is intended to satisfy two concerns expressed by the French Ministry of Public Works, Transportation and Housing, and in particular by its research center (the CERTU):

— to instill momentum for introducing in France a disaggregate and multimodal approach to modeling trip-making behaviour. Heavily employed in both Anglo-Saxon and Scandinavian countries, this modeling approach offers the dual advantage of being robust in terms of estimation capabilities, while being instructive in terms of the way results are presented;

— to contribute to the updating of existing methods, by means of testing an entire series of modeling approaches, based on the four-stage sequential system - trip generation, trip distribution, modal choice and trip assignment - which is currently applied in its aggregated form in France. The motivation behind such an important updating effort has been drawn from the two following observations:

- the need for France to catch up in the area of demand-based modeling,
- the more widespread production of specific planning documents (such as urban transportation plans or metropolitan roadway plans), for which the tools in use are simply ill-adapted.

This document has been designed for management staff in the field of local public transit services, for transportation engineers and economists, and for students.

**Los modelos de elección discreta: hacia una modelización desagregada y multimodal de los comportamientos de desplazamiento en medio urbano**

El trabajo que se presenta en este informe tiene como objeto responder a dos preocupaciones del Ministerio de la Infraestructura, de los Transportes y de la Vivienda, y en particular de su centro de estudios (el Certu):

— impulsar un movimiento para la difusión de la modelización desagregada y multimodal de los comportamientos de desplazamientos en Francia. Dicha modelización, muy utilizada en los países anglosajones y nórdicos, presenta la ventaja de ser sólida en términos de estimación, y pedagógica a nivel de la presentación de los resultados;

— participar en la renovación de los métodos, al mismo tiempo que se prueba una cadena de modelización que se base en la estructura secuencial de cuatro etapas – generación, distribución, elección modal, afectación –

utilizada en su dimensión agregada en Francia. Esta renovación necesaria se encuentra vinculada a dos constataciones:

- el retraso que sufre Francia en el ámbito de la modelización de la demanda,
- el desarrollo de los dossiers de planificación (tipo plan de desplazamientos urbanos o dossiers de viabilidad de área urbana), para los que las herramientas que se utilizan actualmente no se encuentran siempre adaptadas.