



**HAL**  
open science

# Décisions dans l'incertain d'un ménage et dynamique de ses comportements automobiles

Giulia Cernicchiaro

► **To cite this version:**

Giulia Cernicchiaro. Décisions dans l'incertain d'un ménage et dynamique de ses comportements automobiles. Economies et finances. Université Panthéon-Sorbonne - Paris I, 2013. Français. NNT : 2013PA010001 . tel-00984250

**HAL Id: tel-00984250**

**<https://theses.hal.science/tel-00984250>**

Submitted on 28 Apr 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PARIS I - PANTHÉON SORBONNE  
École Doctorale d'Économie Panthéon–Sorbonne (EPS)

**THÈSE DE DOCTORAT**

Présentée par

**Giulia CERNICCHIARO**

Pour obtenir le grade de  
**DOCTEUR en SCIENCES ÉCONOMIQUES**

Spécialité

**Économétrie**

Sujet de la thèse :

**DÉCISIONS DANS L'INCERTAIN D'UN MÉNAGE ET  
DYNAMIQUE DE SES COMPORTEMENTS AUTOMOBILES**

Soutenue le **10 janvier 2013**

Devant le jury composé de :

<b>Pr. François GARDES</b>	Université Paris 1	Directeur de thèse
<b>Dr. Matthieu de LAPPARENT</b>	IFSTTAR, chargé de recherche	Co-Directeur de thèse
<b>Dr. Jean–Loup MADRE</b>	IFSTTAR, directeur du DEST	Rapporteur
<b>Pr. Alain TROGNON</b>	CREST	Rapporteur
<b>Pr. Emmanuel DUGUET</b>	Université Paris–Est	Examineur
<b>Dr. Christophe STARZEC</b>	Université Paris 1	Examineur



Laboratoire de rattachement IFSTTAR  
Financé par ADEME

## Résumé en Français

Au cours des dernières quarante années un grand nombre d'études ont analysé les comportements de possession et utilisation des véhicules. Cependant, la plupart de ces études traitent le problème d'un point de vue statique, en omettant les dynamiques qui vraisemblablement guident ces décisions.

Le caractère durable des véhicules suggère l'importance de ne pas ignorer l'aspect dynamique des problèmes d'achat, possession et utilisation automobile.

Ce travail utilise une approche désagrégée structurelle basée sur des choix réels (collectés par l'enquête "Parc Auto").

Le document est composé de trois parties.

La première propose une description du problème, par l'analyse de la littérature existante. La seconde présente les opérations effectuées sur les bases de données utilisées pour l'application. La base "Parc Auto" nécessite, en effet, un traitement préliminaire si on veut l'utiliser comme panel de véhicules. La dernière partie a pour objectif d'introduire les quatre applications réalisées au cours de ma thèse de doctorat.

Une première application propose une approche hédoniste des prix des véhicules.

Le deuxième chapitre de cette dernière partie estime un modèle de Rust estimant les décisions concernant la voiture pour des ménages mono-équipés (c'est à dire possédant une seule voiture chaque année).

Un modèle dynamique discret \ continu des choix de la durée de possession et d'utilisation des véhicules pour des agents tant myopes que prévoyants.

Le dernier chapitre présente un programme dynamique tourné vers le futur pour le problème de la planification des achats.

Dans les trois applications dynamiques, un facteur d'actualisation est utilisé pour tester différentes hypothèses pour le comportement dynamique. Il n'est pas estimé mais plusieurs niveaux de préférence pour le futur sont fixés. J'ai donc estimé ces modèles à l'aide des données françaises de l'enquête panel "Parc Auto".

Mes résultats indiquent que la préférence temporelle d'un ménage est un élément crucial dans l'achat de voiture, l'utilisation et les décisions de détention.

Mots Clés : Agent tourné vers le futur ; Modèle de choix Discret\Continu ; Demande de Déplacements

## **TITLE : Decisions in the uncertain of a household and dynamics of its automobile behavior**

### **Abstract**

During the last forty years, a large number of studies have analyzed the behaviors related to car holding and use. However, the majority of these studies consider the subject from a static point of view, ignoring the household dynamics and the driver's needs, which are very likely to drive such decisions (purchase, ownership and use).

The durable character of a vehicle suggests however that such a dynamic cannot be ignored in the study of the vehicle purchase, holding and use decisions.

This work employs a disaggregated, structural approach based on real choice (collected by "Parc Auto" survey) to address these dynamics.

The document is composed of three parts.

In the first part a description of the problem is provided through an analysis of the available literature. The second part includes the description of the operations which are performed on the database used in the applications. The "Parc Auto" data needs in fact some preliminary treatment if one wants to use it as a vehicle's Panel. In the last part the four applications considered are presented.

A first application proposes an hedonic approach to the vehicles' prices.

The second chapter of this last part estimates the Rust model for mono-equipped (i.e. only one car owned each year) households' decisions concerning their car.

A dynamic discrete\continuous choice model of car holding duration and use for both myopic and forward-looking agents follow.

In the last chapter a dynamic forward-looking program for the purchase planning problem is presented.

In the three dynamic applications, a discount factor tests different effects of dynamic behavior. This is not estimated, but fixed at different future preference levels. Furthermore, I estimate these models using French panel survey data.

My findings indicate that a household's time preference is a crucial element in car purchase, use and holding decisions.

Key words : Forward-Looking Agents ; Discrete\Continuous Choice Modelling ; Transportation Demands

# Remerciements

Au terme de cette expérience je tiens à remercier sincèrement tous ceux avec qui j'ai partagé ce chemin.

Mes premières pensées et tous mes remerciements vont à Matthieu de Lapparent, co-directeur et superviseur de ma thèse, qui est venu me chercher quand je n'avais pas encore trouvé ma voie et m'a proposé un sujet si intéressant à étudier. Son soutien et ses conseils ont été précieux et m'ont accompagnée tout au long de la thèse.

Merci à François Gardes, mon directeur de thèse qui m'a aidé à faire le lien avec l'école doctorale, ainsi qu'à tout son groupe de doctorants. Je voudrais aussi remercier les autres membres du jury pour leur disponibilité malgré les courts délais auxquels ils ont été confrontés.

Un merci encore à l'IFSTTAR, et notamment à l'équipe du DEST, à son directeur Jean-Loup Madre et au responsable de la base Parc Auto Laurent Hivert, pour leur accueil et leur disponibilité. Je tiens à remercier ensuite Virginie Dejoux et Sophie Roux, qui m'ont donné confiance, Zéhir Kolli, Mariame Mbacke et Roger Collet pour les moments de partage, et tous les autres collègues du site de Marne-la-Vallée.

Une pensée riche de gratitude va à Christophe Boucher et Hélène Hubert qui m'ont permis de faire l'expérience de l'enseignement. Un grand merci aussi à Maria Kuecken et Ysaline Padieu, rencontrées à cette occasion.

J'adresse un remerciement chaleureux à ma famille, qui a su être à mes côtés malgré la distance.

Pour conclure, un merci tout spécial du fond du coeur à Jordan, qui a subi mes changements d'humeur, a participé à la correction de mes défauts en français et fait tout son possible pour me simplifier la vie pendant les moments les plus chargés.



# Glossaire

- **Agent Tourné vers le Futur** : Un agent (dans ce travail un ménage) qui anticipe les évolutions de son état et de celui de son environnement, dans le but de prendre en compte les conséquences de ses choix lorsqu'il y est confronté.
- **Approche Désagrégée Structurelle, basée sur des Choix Réels** : Approche de plus en plus utilisée dans les travaux récents visant à étudier les comportements des individus, qui a pour objectif de proposer une structure de décision représentant la réalité. Les données à utiliser pour une telle approche sont des données désagrégées, collectant par agent l'information concernant les choix qu'il prend dans son quotidien.
- **Choix d'Équipement Automobile** : Les choix portant sur la taille et la composition de la flotte de véhicules qui seront à disposition de l'agent au cours de l'année. À la différence de la définition habituelle, je suppose que ces choix ne sont pas traités comme un bloc unique par l'individu<sup>1</sup>, mais qu'ils sont plutôt obtenus par harmonisation des résultats des processus optimaux pour chaque véhicule. Ces choix peuvent alors être décomposés en choix d'achat, de possession et d'utilisation des véhicules (hérités par le parc précédent ou disponibles sur le marché).
- **Configuration de Vie** : Le cadre défini par les choix de localisation, des activités et de démographie que l'agent prévoit pour l'année. Je suppose qu'il place dans ce cadre ses décisions concernant les biens et les services qu'il devra financer pour pouvoir subvenir à ses besoins.
- **Décision Absorbante** : Un choix qui interrompt le flux futur d'utilité. Dans le cas des décisions d'utilisation et possession d'un véhicule par exemple, lorsque l'agent décide de se séparer d'un véhicule il décide en même temps de ne plus l'avoir à disposition les années suivantes<sup>2</sup> et donc de ne plus profiter des services que celui-ci peut fournir.
- **Effet de l'Habitude** : L'effet double que la permanence dans un même état peut avoir sur la satisfaction de l'individu. Dans le cas des décisions concernant les biens durables (comme par exemple les véhicules), l'habitude cumulée au cours de la possession pourrait avoir alors deux effets. D'un côté le confort produit par le fait d'utiliser un bien auquel on est habitué pourrait pousser l'agent à maximiser la durée de possession. De l'autre le besoin de nouveauté que les individus pourraient ressentir contrebalancerait ce premier effet.
- **Facteur d'Actualisation** : Paramètre indiquant la façon dont l'agent balance entre le présent et le futur (i.e. préférence pour le futur). Il varie entre 0 (i.e. aucune préférence pour le futur, seule la satisfaction obtenue dans le présent est considérée) et 1 (i.e. aucune préférence pour le présent, seul le flux des satisfactions futures est considéré).

---

1. par exemple Train (1986), nombre et types des véhicules

---

2. Même dans le cas hypothétique où l'agent rachète son véhicule des années après l'avoir remis sur le marché, cela impliquera un achat et est à mon avis considéré par l'individu comme un nouveau véhicule (pour lequel une expérience précédente existe).



- **Période Continue d’Observation** :  
Pour chaque ménage répondant au moins une fois à l’enquête, une suite d’années pendant lesquelles le ménage répond à toutes les vagues annuelles du panel.
  
- **Processus à Arrêt Optimale avec Régénération** : Processus de décision dynamique dans lequel la variable causant l’arrêt du processus n’en cause pas pour autant la fin (comme c’est le cas si la variable est absorbante). L’arrêt du processus en engendre un nouveau (i.e. Régénération). Le flux des utilités futures n’est donc pas interrompu dans ce cas.
  
- **Véhicule hérité** : Tout véhicule qui composait le parc l’année précédente et que l’agent a choisi de conserver une période de plus. Ces véhicules sont déjà à disposition en début de période et l’agent ne doit prendre, pour ce qui les concerne, que les décisions de possession et d’utilisation (et pas d’achat).

# Table des matières

<b>Introduction Générale</b>	<b>v</b>
<b>I Le problème</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Définition du Cadre . . . . .	7
<b>2 Les choix d'équipement automobile</b>	<b>15</b>
2.1 Le Véhicule . . . . .	20
2.2 Le Parc Automobile . . . . .	26
2.3 Le choix optimal . . . . .	30
<b>II Les données</b>	<b>33</b>
<b>3 Présentation des Données</b>	<b>35</b>
3.1 Les Bases des Données . . . . .	36
3.1.1 Parc Auto . . . . .	36
3.1.2 DIREM . . . . .	42
3.1.3 Base des Données Automobile . . . . .	44
3.2 Analyse des Prix du Carburant . . . . .	46
3.3 Choix de la Période. . . . .	49
<b>4 Appariement des Véhicules</b>	<b>51</b>
4.1 Traitement Préliminaire . . . . .	54
4.2 La Procédure . . . . .	55
4.2.1 Identification des Véhicules . . . . .	56
4.2.2 Sélection des Informations . . . . .	59
4.3 La Taille de l'Échantillon des Véhicules Identifiés . . . . .	62

<b>5</b>	<b>L'échantillon Retenu</b>	<b>67</b>
5.1	L'appariement au Niveau Ménage . . . . .	68
5.2	L'Analyse de l'Échantillon Retenu . . . . .	69
5.2.1	Analyse Socio-Démographique . . . . .	72
5.2.2	Analyse par Comparaison avec les Données Brutes . . . . .	75
5.3	L'Analyse Préliminaire de la Problématique . . . . .	79
5.3.1	Les Ménages Mono-équipés . . . . .	83
5.3.2	Les Véhicules . . . . .	86
<b>III</b>	<b>Les Applications</b>	<b>95</b>
<b>6</b>	<b>Les Prix Hédonistes</b>	<b>97</b>
6.1	Les Données . . . . .	98
6.2	Le Modèle . . . . .	103
6.3	Les Résultats . . . . .	109
6.4	Conclusion . . . . .	120
<b>7</b>	<b>Possession et Utilisation : Mono-Equipés</b>	<b>123</b>
7.1	Le Modèle . . . . .	127
7.2	Les Données . . . . .	137
7.3	Les Résultats . . . . .	141
7.4	Conclusion . . . . .	148
<b>8</b>	<b>Le programme dynamique discret\continu</b>	<b>151</b>
8.1	Model . . . . .	155
8.1.1	Framework . . . . .	155
8.1.2	A short discussion about Assumptions . . . . .	161
8.1.3	Estimation . . . . .	162
8.2	Data . . . . .	163
8.2.1	Statistical sources . . . . .	163
8.2.2	Descriptive statistics . . . . .	164
8.3	Results . . . . .	167
8.4	Conclusions . . . . .	175
<b>9</b>	<b>Les Décisions d'Achat</b>	<b>177</b>
9.1	Le Modèle . . . . .	180
9.2	Les Données . . . . .	190
9.3	Les Résultats . . . . .	196
9.3.1	Les transitions . . . . .	196
9.3.2	Le modèle . . . . .	202

9.4 Conclusion . . . . .	208
<b>Conclusion Générale</b>	<b>211</b>



# Introduction Générale

La mobilité occupe, de nos jours, une place centrale dans la société et, par conséquent, on le compte parmi les thèmes les plus traités dans la recherche actuelle.

Chaque être humain, même le plus sédentaire, a besoin de se déplacer pour pouvoir conduire sa vie. La possibilité de se déplacer permet en effet à l'individu d'élargir son horizon, en accédant à un nombre plus important d'activités, aussi bien professionnelles que personnelles, et de services. Le fait d'être limité dans ses déplacements pourrait alors être cause d'une insatisfaction importante pour l'individu puisque ceci réduit son accès à la vie sociale et le volume d'activités qu'il peut pratiquer.

L'évolution économique et urbaine, que notre société connaît à un rythme de plus en plus pressant depuis plusieurs décennies, multiplie les solutions de déplacement et offre à chacun de ses membres la possibilité de considérer un ensemble de possibilités de plus en plus grand lorsqu'il prend les décisions définissant sa vie. Cela permet alors, par exemple, une distance entre le lieu d'habitation et le lieu d'activité principale sensiblement plus importante que celle envisageable il y a encore 30 ans.

D'un autre côté, une dépense découle des décisions de mobilité et l'individu doit faire face à cette dépense sous la contrainte que son budget disponible lui impose. Comme le montrent les chiffres concernant la structure de consommation publiés par l'[INSEE](#), cette dépense constitue l'une des sorties (avec celles concernant le logement et la consommation alimentaire) les plus importantes pour l'individu.

La hausse des prix, contrepartie de la rapidité d'évolution de l'environnement, les restrictions budgétaires auxquelles les individus sont confrontés dans la conjoncture actuelle, ainsi que l'avancement d'Internet dans notre quotidien (permettant de plus en plus souvent de satisfaire une part importante des besoins de l'individu), pourraient alors pousser l'individu à souhaiter réduire

son volume de déplacement.

Les études récentes portant sur la mobilité en France ([Massot and Orfeuill \(2005\)](#), [Guidez \(2007\)](#), [Birchen and Richard \(2008\)](#), [Hivert et al. \(2010\)](#)) mettent en lumière une hétérogénéité dans les comportements de déplacement qui se fait de plus en plus importante, en réponse aux évolutions du rythme de vie des individus. La mobilité individuelle (i.e. le nombre de déplacements que chaque individu effectue au quotidien), qui n'avait pas cessé d'augmenter depuis 1960, semble s'être stabilisée dans les années 2000 et connaît depuis 2005 une légère régression. Cependant, l'analyse de ce volume en fonction de certaines caractéristiques individuelles permet de remarquer la présence de tendances bien distinctes selon l'état de l'individu étudié. La mobilité urbaine baisse en fait alors que celle rurale n'a pas cessé d'augmenter. Les jeunes adultes réduisent leurs déplacements, contrairement aux personnes âgées qui présentent des nouveaux besoins de plus en plus importants.

Le fait que la mobilité de la population ait une multiplicité d'enjeux (économiques, d'organisation urbaine, environnementaux et écologiques) sur l'organisation économique et territoriale de la société, suggère l'importance d'être en mesure d'anticiper et répondre aux évolutions de ces comportements. L'hétérogénéité mise en lumière laisse alors à penser que, en descendant dans l'analyse au niveau individu, on pourrait mieux comprendre les dynamiques qui caractérisent les comportements de mobilité.

L'offre de moyen de transport est vaste et les comportements de déplacement, ainsi que leurs évolutions, varient selon le type de transport considéré.

Je vais m'intéresser ici à un moyen de transport en particulier : la voiture. [Hivert et al. \(2010\)](#) présentent une analyse complète des évolutions de comportement vis-à-vis de ce moyen de transport.

Pour ce moyen de transport, des nouvelles tendances ont été mises en lumière. La circulation automobile s'est stabilisée au niveau national, mais l'usage de la voiture semble toutefois reculer dans les grandes agglomérations et ceci ne s'était jamais vérifié auparavant. De plus, si l'utilisation automobile baisse, cela n'est pas le cas de la possession. Le volume du parc automobile continue de croître, mais le rythme est ralenti.

Les changements d'utilisation automobile peuvent être reliés aux évolutions connues par les prix des carburants. Ces prix connaissent en effet une hausse continue depuis 2002 pour les deux types principaux (Diesel et Essence).

Ces deux prix n'ont cessé de se rapprocher, jusqu'à ce que le premier rejoigne le deuxième en 2008.

Comme le montrent [Hivert et al. \(2010\)](#), la variation de la consommation totale de carburant est seulement en partie imputable à la variation de son prix. Les auteurs avancent qu'au moins une partie de la variation d'utilisation moyenne est imputable à la variation de la circulation totale, mais la baisse de la circulation n'est pas traduite par une baisse du volume du parc. Les évolutions de l'offre ne causent donc pas des variations majeures de la demande d'utilisation automobile mais plutôt un changement de comportement pour les choix de possession. Les auteurs mettent en lumière une plus forte préférence pour les véhicules Diesel et un vieillissement du Parc.

Le caractère durable de ce type de bien (le véhicule) ne permet pas de séparer ces deux dimensions. En effet, le déclin de la valeur d'un bien durable à l'augmenter de l'utilisation que l'agent en fait implique le besoin de les considérer conjointement. Les véhicules sont donc caractérisés par une dualité entre utilisation et possession qui rend délicat l'étude des décisions les concernant.

Les besoins de mobilité des agents répondent aux préférences de l'agent et sont conditionnés par son planning et son environnement. En effet, bien que l'agent ait la liberté de choisir aussi bien de quels moyens de transport se servir que des trajets à parcourir, son besoin de kilométrage est tout même conditionné conjointement par son habitation, ses activités et la composition du ménage dans lequel il vit.

Dans ce travail je présente une approche structurelle pour l'étude des comportements de possession et utilisation automobile des ménages français. Le choix de traiter le ménage comme un seul individu dérive du fait que, lorsque ses membres se confrontent au choix de déplacement, ils sont conditionnés par un même budget. Pour ce qui concerne l'utilisation automobile, de plus, les membres du ménage ont un parc unique à disposition, qu'ils doivent se partager.

Je postule que traiter les individus séparément dans ce type de problème impliquera le besoin de modéliser aussi les interactions qui ont lieu entre eux et qui contribuent à la formation de leurs choix de transport. Je préfère alors étudier un agent unique.



## Présentation du Document

Ce travail utilise une approche structurale pour l'étude des comportements automobiles. Cette étude est appliquée aux données concernant les ménages français fournies par la base "Parc Auto"<sup>1</sup>.

Le document s'articule sur trois parties.

- La **première partie** a pour objectif la définition d'un cadre complet pour l'estimation des comportements d'équipement automobile des ménages. Cette première partie se développe sur deux chapitres.

Au cours du chapitre 1 j'essaie de relier les différentes suggestions offertes par la littérature micro-économique, avec une attention particulière pour les travaux traitant des comportements de décision, afin de suggérer un cadre idéal d'estimation du problème complet de décision, sans pourtant proposer une méthode d'estimation.

En raison des limites dictées par le cerveau humain (nombre d'informations traitables, permanence des chemins de raisonnement construits) je suggère l'existence d'un *processus de décision globale de long terme* visant à définir ce cadre

On pourrait décomposer le problème en trois sous problèmes : l'estimation de la combinaison optimale de "choix de vie" par maximisation d'une fonction équilibrant les utilités attachées à chaque choix, l'estimation des volumes de consommation<sup>2</sup> par minimisation de la "fonction de coût complète", visant à équilibrer la satisfaction et la dépense annuelle, et l'estimation des décisions d'achat, possession et utilisation pour chaque bien (durable ou courant).

L'étude des décisions portant sur l'organisation du parc automobile faisant l'objet de mon travail, je ne vais présenter de façon détaillée que le processus concernant cette sous-classe de choix.

Le chapitre 2 fournit une revue de la littérature et de la méthodologie concernant le sujet ici traité, ainsi que la présentation d'un processus

---

1. Le lecteur peut se référer à la section 3.1.1 pour une présentation de cette base.

2. Comme il sera dit par la suite, pour la possession d'un bien durable on peut supposer que l'agent considère le volume de satisfaction que ceci lui apporte (continu) et non pas le nombre (discret). L'utilisation d'un tel bien est par contre représentable par le volume de biens de consommation courant à payer à cette fin (ex. carburant pour les véhicules).

de décision pour l'équipement automobile, à placer dans le cadre défini au premier chapitre.

Le caractère durable des véhicules suggère que l'achat d'un tel bien est à traiter comme un investissement et fait des choix d'équipement automobile un problème de long terme.

De plus, je suggère que l'utilisation d'un bien durable engendre un effet de l'habitude qui joue un rôle non négligeable dans le processus de décision associée. En d'autres termes, l'utilité de continuer à posséder un bien durable acheté dans le passé n'est pas la même que celle obtenue par la possession d'un bien à acheter.

Pour ce modèle une méthode d'estimation séquentielle des sous-problèmes composant cette classe de choix est suggérée. Les hypothèses formulées pour l'estimation sont très restrictive et devraient être étudiées avec beaucoup plus d'attention pour que le modèle soit applicable, cependant l'objectif de ce chapitre est de permettre de mieux comprendre l'idée développée dans ce travail.

- Les données sont introduites au cours de la **deuxième partie**.

Le premier chapitre de cette partie (chapitre 3) a pour objectif de présenter les données utilisées pour l'application qui conclut le document.

Les bases de données exploitées dans le cadre de ma thèse sont présentées au cours de la section 3.1. La première base (**Parc Auto**, 3.1.1) fournit une description représentative du parc automobile des ménages français au cours des dernières trois décennies. Les caractéristiques techniques des véhicules composant le marché automobile mondial sont récoltées dans la deuxième base (**Base des Données Automobile**, 3.1.2). La dernière base concerne les prix des carburants en France depuis 1999 (**Prix DIREM**, 3.1.3). L'utilisation jointe de ces 3 bases permet donc une étude complète de la problématique de la mobilité automobile, qui fait l'objet de mon travail.

Une analyse des évolutions des prix du carburant est effectuée au cours de la section 3.2, dans le but de définir la période d'observation la plus adaptée à l'étude du problème (comme expliqué en section 3.3).

Les données Parc Auto nécessitent un traitement préliminaire lorsque

l'on souhaite utiliser la base comme un panel de véhicules<sup>3</sup>.

En effet, si la base Parc Auto inclut une “clé d'identification” pour chaque ménage enquêté, aucun identifiant unique pour les véhicules n'est disponible et doit donc être créé si l'on veut suivre un même véhicule tout au long de sa présence dans le parc automobile.

Cela étant les cas de ma thèse, j'ai consacré une partie considérable de mon temps à la rédaction d'un code SAS permettant d'apparier les véhicules décrits par les ménages participant à l'enquête Parc Auto (Chapitre 4).

Après une brève présentation des opérations préliminaires de préparation des données au niveau véhicule (section 4.1), cette “procédure d'appariement” est décrite au cours de la section 4.2. En permettant d'identifier la presque totalité des véhicules déclarés entre 1996 et 2008, cette procédure laisse apparaître des erreurs de déclaration/saisie qui ne sont pas visibles lorsque l'observation est faite au niveau ménage(4.2.1). La section 4.2.2 présente alors une procédure, permettant de compléter et affiner les descriptions techniques des véhicules à l'aide de la Base de Données Automobiles.

La taille de l'échantillon identifié, obtenu en appliquant une telle procédure d'appariement des véhicules, est discutée en conclusion de ce chapitre (section 4.3).

L'“échantillon retenu” est analysé au cours du chapitre 5.

En premier lieu, les manipulations à effectuer au niveau ménage pour la construction de cet échantillon sont présentées au cours de la section 5.1.

La section 5.2 présente la construction de l'échantillon. Comme il sera expliqué par la suite, celui-ci est composé des véhicules possédés par les ménages répondant à 5 conditions d'inclusion découlant des procédures d'appariement. Il peut être traité aussi bien au niveau ménage qu'au niveau véhicule. Une analyse succincte socio-démographique de la “population des ménages retenus” est menée et présentée au cours du chapitre avec un double objectif. Les tendances démographiques de l'échantillon sont avant tout tracées dans les grand lignes, en utili-

---

3. La base Parc Auto est complète et simplement exploitable lorsque l'unité d'observation est le ménage.

sant une variable définissant le “type de chef” (5.2.1), dans le but d’en évaluer la représentativité par comparaison avec les résultats publiés par l’INED et l’INSEE (INSEE (2010), INSEE (2004)). Je compare alors l’échantillon obtenu après traitement avec les données brutes (i.e. tels que fourni par L. Hivert, responsable des vagues Parc Auto conservées au sein de l’IFSTTAR) dans le but d’observer l’importance des biais de sélection engendrés par la procédure d’appariement (5.2.2).

Pour conclure, le panel retenu est décrit au cours de la section 5.3 par une première analyse statistique au “niveau ménage” et une au “niveau véhicule”.

- La **troisième partie** propose une spécification économétrique pour quatre des sous problèmes dans le quel le processus de décision d’équipement et utilisation automobile peut être décomposé.

Les chapitres qui composent cette dernière partie sont caractérisés par une même structure que je vais présenter de suite. Seulement le chapitre 6 présente une différence dans la présentation du travail. En effet, dans ce cas les premières deux sections sont inversées.

Chaque chapitre s’ouvre avec l’introduction du sous-problème que l’application présentée vise à étudier. Ces introductions reprennent et développent pour le problème étudié, la revue de la littérature présentée au cours de la première partie. De cette façon, il est possible de lire ces chapitres indépendamment des autres parties du document.

- La première section<sup>4</sup> décrit le cadre économétrique, défini par les hypothèses formulées pour le sujet d’étude, dans le quel se place la spécification du modèle.

Les 4 problématiques sont présentées par la suite. Les cadres de travail choisis, ainsi que les spécifications définies et les méthodes d’estimation, sont plutôt simples et méritent d’être discutés. Cette section présente alors une critique aux simplifications adoptées<sup>5</sup>, choix dûs principalement au temps disponible.

- La deuxième section<sup>6</sup> présente les données.

Les informations utilisées sont extraites des bases décrites au cours

---

4. Deuxième au chapitre 6

5. Une critique plus complète est en fin présentée en Conclusion du document

6. Première au chapitre 6

de la deuxième partie de ce document. Le traitement des données présenté plus haut ayant demandé un temps considérable pour sa réalisation, les données utilisées au cours de ces chapitres ne sont pas toutes issues de l'échantillon retenu<sup>7</sup>, décrit au chapitre 5. Cette section a alors pour objectif de présenter l'échantillon utilisé. Là aussi le temps disponible en est pour beaucoup dans le choix de ne pas répéter les études avec cet échantillon final. De plus, deux de ces chapitres correspondent respectivement à une publication (de Lapparent and Cernicchiaro (2012)) et une communication (Cernicchiaro and de Lapparent (2012)), je préfère alors présenter ici les résultats comme fait à ces occasions.

- La troisième section présente les résultats de l'application. Ces résultats y sont commentés de façon assez rapide. Le choix de ne pas rentrer dans le détail de ces commentaires vient des limites d'interprétation causées par les hypothèses et les méthodes d'estimation choisies. L'objectif de ce travail n'est pas réellement celui d'étudier les comportements des ménages, mais plutôt celui de reprendre les travaux précédemment réalisés à ce sujet, dans le but de suggérer un point de vue plus large et plus proche de la réalité pour le traitement du problème.

Les conclusions sont tirées au cours de la dernière section.

Une étude hédoniste des prix des véhicules est présentée dans un premier temps au chapitre 6.

Les véhicules sont des "biens composés". En raison des multiples composantes qui caractérisent un tel bien, on ne peut traiter un véhicule comme une unité lorsqu'on veut étudier son prix. La méthode adoptée dans ce chapitre permet d'étudier le prix d'un bien composé comme fonction de ses caractéristiques. Bien que cette méthode soit assez répandue dans la littérature, elle est rarement appliquée pour l'étude des prix des véhicules.

Une hypothèse importante est à commenter. La consommation en carburant me semble être une des caractéristiques que l'agent traiterait en priorité (avec par exemple, la puissance du moteur, les dimensions et la carrosserie du véhicule) en évaluant un véhicule. Or, pour les lois

---

7. Comme je l'explique en introduction du chapitre 4, ces échantillons sont issus des procédures temporaire, composantes les différents étapes du travail préliminaire d'analyse réalisé sur les données.

de la physique, la consommation est fonction de ces caractéristiques (Les effets des autres caractéristiques sont vraisemblablement aussi liés. Cependant je postule que l'on puisse isoler ces interactions par la simple introduction d'effets croisés dans la spécification<sup>8</sup>). Un modèle d'équations est alors proposé pour l'estimation simultanée de la consommation et le prix du véhicule.

Le modèle est traité à l'équilibre en raison de l'absence de données du marché automobile suffisamment détaillées<sup>9</sup>. Pour cette raisons certaines caractéristiques du ménage sont utilisées dans la spécification.

La base utilisée pour cette application est issue de l'échantillon construit au niveau véhicule par la procédure en 4.2. Pour pouvoir disposer de l'information concernant le ménage, seulement les véhicules déclarés l'année de leur achat sont utilisés. Puisque il est rare d'observer plusieurs achats pour un même ménage au cours de leur participation à l'enquête (moins de 6,4%), la structure panel des données disponibles n'est pas exploitée ici.

La méthode des prix hédonistes est appliquée aux prix neufs des véhicules. Les prix d'occasion sont supposés être fonction géométrique du prix neuf, décroissant selon l'age et l'utilisation du véhicule.

Deux modèles de choix de possession et utilisation automobile, sont traités par la suite.

Un premier modèle est construit pour les ménages mono-équipé dans un premier temps (présenté au chapitre 7).

Le chapitre suivant (8) présente une évolution de ce premier modèle, permettant de considérer tout ménage, quelque soit le nombre de véhicules qui composent son parc.

Pour conclure, un programme dynamique de planification d'achat est introduit au chapitre 9.

Dans les trois cas l'hypothèse d'agent "tourné vers le futur", évaluant les conséquences de ses décisions dans le cadre défini par les prévision des évolution qu'il formule pour son état, est adopté. Le cadre utilisé est celui proposé par Rust (1987). Le choix d'utiliser de ne relaxer aucune

---

8. Par exemple, les dimensions du véhicule sont liées par des lois géométriques connues, qui peuvent être utilisés directement dans la régression

9. Si l'on veut connaître les prix de marché détaillés par type et caractéristiques principales du véhicule (ou même seulement marque et modèle), la seule table Argus disponible est en format papier et ils n'ont pas la possibilité de le numériser.

des hypothèses restrictives formulées pour ce cadre d'estimation vient du fait que ce cadre est le plus maniable disponible. Deux programmes d'estimation SAS ont été écrits au cours de la thèse pour estimer le programme dynamique à arrêt optimal de Rust. Le premier, utilisé pour les applications concernant le problème de la possession et de l'utilisation du véhicule (chapitres 7 et 8), estime le problème à horizon fini et nécessite de données de panel au niveau véhicule. Le problème de planification présenté au chapitre 9 est estimé à horizon infini par un deuxième programme d'estimation. Puisque une partie du temps a été consacrée à l'écriture de ces programmes (notamment celui à horizon infini) je me suis limitée au cas le plus simple à traiter.

La conclusion du document présente et critique de façon générale le travail réalisé.

Quelque évolution possible est proposée pour clore le document.

# Première partie

## Le problème.

Une présentation de l'état de l'art et  
du point de vue adopté





# Chapitre 1

## Introduction

Le développement récent de l'économie comportementale (voir par exemple [Patalano and Rizzello \(2002\)](#), [Marsh \(2012\)](#)) a mis l'accent sur l'importance de prendre en compte les caractéristiques propres au raisonnement de l'individu dans l'étude de ses comportements économiques. Simon et Hayek (dans [Marsh \(2012\)](#)) suggèrent l'existence pour l'individu du besoin de normes fixées qui définissent le cadre dans lequel il agit. Ce besoin serait, selon les auteurs, dicté par l'incapacité de traiter un trop grand nombre d'informations qui caractérise le cerveau humain. Dans la littérature d'économie comportementale ce besoin est traduit, comme suggéré par les auteurs, par le besoin de pouvoir se référer aux institutions ([Patalano and Rizzello \(2002\)](#)).

Je suggère que le raisonnement proposé par Simon et Hayek peut être traduit, lorsque l'on étudie les comportements de choix de l'individu, par le besoin pour celui-ci de connaître le cadre de vie dans lequel il agira avant d'affronter ses décisions au quotidien.

En même temps, la dimension temporelle de notre vie implique que chaque choix effectué aujourd'hui aura un impact, plus ou moins important, sur notre quotidien futur.

La dualité de ce problème a été mise en lumière en premier dans les travaux macroéconomiques de [Hall \(1978\)](#) et [Hansen and Singleton \(1982\)](#), [Hansen and Singleton \(1984\)](#) entre autres. Motivés par les critiques aux hypothèses d'agrégations au niveau macro formulées par [Mankiw et al. \(1985\)](#) et [Deaton \(1988\)](#), les premiers travaux microéconomiques ([Altonji \(1986\)](#), [Browning et al. \(1985\)](#)) traitent le problème complet dynamique visant à optimiser pour le long terme la consommation, la force-travail fournie et les dépenses de l'agent.

On peut alors définir deux types de choix en fonction de l'importance de cet impact :

- les choix dont le résultat a un effet à **long terme** ;
- les choix dont l'effet n'a un impact que à **court terme**.

Je suppose que ces choix sont le résultat de deux processus de décision distincts mais intrinsèquement liés. En d'autres termes, on pourrait supposer que l'individu prenne les décisions de long terme, comme par exemple le lieu d'habitation, les activités principales ou dans le cas de la mobilité, l'achat d'un véhicule, dans un premier temps, en anticipant la configuration de vie que l'ensemble de ces choix lui permet d'avoir. Il ajuste dans un deuxième temps son quotidien par le biais des choix de court terme.

Dans un sens alors les décisions prises pour le long terme définissent l'ensemble de choix que l'agent considérera face aux décisions du deuxième type.

Pour mieux comprendre, imaginons les exemples suivants.

Lorsqu'un agent choisit sa location, la présence de commerces dans l'environnement a sûrement un impact mais c'est seulement dans l'organisation quotidienne de son planning qu'il décide exactement auprès duquel se fournir. Ou pour revenir au cas de la mobilité, il pourrait prendre en compte la distance domicile-travail lorsqu'il décide de s'équiper d'un véhicule, mais décider ponctuellement si utiliser ce véhicule disponible ou un moyen de transport alternatif au quotidien.

On pourrait alors voir le processus de décision complet comme deux problèmes imbriqués. Considérons  $t = 0$  le moment initial, quand l'individu commence à décider pour sa vie. À partir de ce moment l'individu prend toutes les décisions pour sa vie, jusqu'au moment  $t = T$  où il n'est plus en mesure de décider pour lui. Un individu entre dans le processus donc lorsqu'il atteint un certain âge (par exemple la sortie du lycée) et il peut en sortir par la maladie ou le décès.

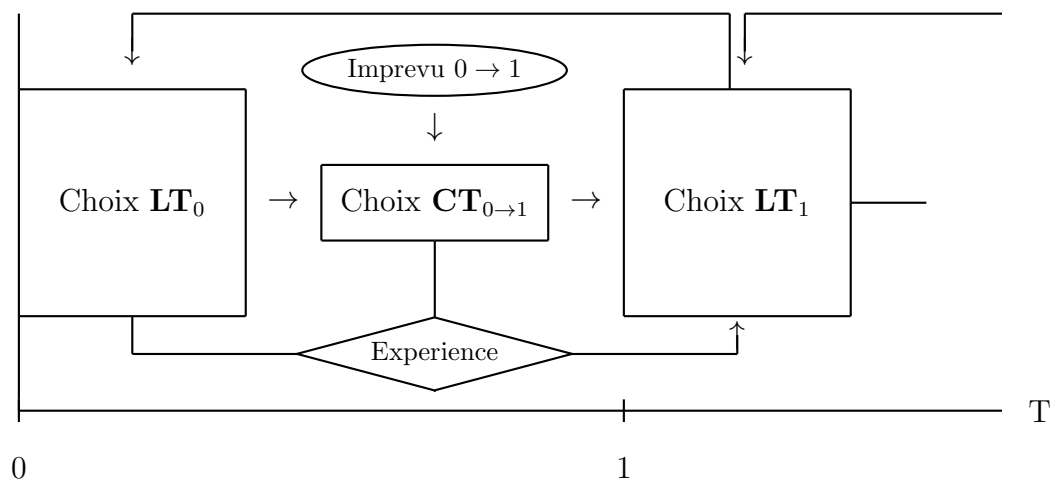
Dans ce travail l'agent est un individu, un couple ou une agrégation d'individus en mesure de décider. Par simplicité le processus de décision "interne à l'agent", c'est à dire la façon dans laquelle les individus le composant balancent entre leurs besoins respectifs, n'est pas considérée ici. Une satisfaction est donc estimée pour chaque agent.

Un agent peut être aussi responsable d'un ou plusieurs individus hors-processus.

Dans ce cas les besoins de ces individus sont considérés dans le processus de décision de l'agent.

Chaque année  $t$  l'agent estime dans un premier temps simultanément les décisions de long terme optimales, en fonction de l'expérience qu'il a sur ses préférences tout en anticipant les conséquences que ces décisions ont sur les choix "de court terme" auxquelles il sera confronté au quotidien et sur les périodes successives. Il détermine alors les choix optimaux au quotidien entre  $t$  et  $t + 1$  dans le cadre découlant des décisions prises au moment initial.

**Fig. 1.0** : PROCESSUS COMPLET



Au moment  $t + 1$  l'expérience de l'agent est régénérée en incluant le vécu depuis  $t$  et il peut réévaluer l'ensemble de ses décisions de long terme de la même façon qu'au moment précédent. Les paramètres attachés aux facteurs jouant sur le processus de décision de long terme à chaque moment  $t$  pourraient être considérés comme fonction de cette expérience. En supposant pouvoir connaître, ou au moins approximer, une fonction de distribution pour l'expérience de l'agent on pourrait alors supposer que celles attachées aux paramètres sont une mixture normale de cette distribution.

Comme dit plus haut, le problème ici considéré concerne des choix de long terme. Je vais donc m'intéresser seulement au processus guidant ce type de

choix. Le processus interne (concernant les décisions de court terme) n'étant pas traité, ni l'expérience, ni le caractère mixte de la distribution des paramètres ne sont considérés dans cette proposition.

Lorsque l'agent définit les décisions optimales de long terme pour l'année, il doit prendre en compte un grand nombre de facteurs. Or, le cerveau humain est capable de traiter en même temps un nombre limité d'informations. Il semble alors vraisemblable que le processus de décision de long terme soit à son tour décomposable en plusieurs sous-processus. Je suggère que l'on puisse définir un groupe de classes de décisions, qui définissent sa "configuration de vie" annuelle, traité en priorité par l'agent.

Chaque choix apporte une satisfaction et comporte une dépense pour l'agent. Cette dépense est composée d'une partie directe et une partie due aux besoins d'achat engendrés, que l'agent veut équilibrer.

Les besoins portent sur deux types de biens : les biens durables, pour lesquels l'agent doit déterminer en début de période les décisions de possession et d'utilisation optimales, et les autres biens de consommation pour lesquels il veut estimer un volume annuel. L'agent a donc pour objectif, dans une configuration fixée, de définir les demandes optimales pour chaque bien.

Or, un bien durable produit des services sur une longue durée mais son état se dégrade en fonction du temps et de l'utilisation que l'on en fait. Je suppose alors l'agent capable de définir un "cycle de vie" optimal pour chaque bien qu'il possède, qui dépend de ses préférences et des caractéristiques du bien. Par cycle de vie j'entends la suite de décisions de possession et utilisation, entre l'achat et l'abandon.

Les décisions d'achat, utilisation et abandon de ce type de bien doivent alors être telles que l'équipement choisi subvient aux besoins annuels de l'agent tout en respectant le cycle de vie optimal de chaque bien le composant.

Le sujet d'étude dans ce travail est l'équipement automobile. Je vais ici présenter une modélisation idéale du processus qui guide les décisions portant sur cette classe de biens durables : les véhicules.

Au cours du premier chapitre je vais présenter de façon plus détaillée le cadre dans lequel se place le modèle idéal de décisions d'équipement automobile. Ce deuxième est présenté au deuxième chapitre.

## 1.1 Définition du Cadre

Chaque année l’agent doit prendre toute une série de décisions portant sur différents aspects de sa vie, qui comme dit plus haut vont définir son cadre de vie pour la période. On pourrait par exemple décomposer l’ensemble des décisions auxquelles il est confronté en 3 classes : localisation, activités et démographie. Ce que j’appellerai par la suite “configuration de vie” est alors l’une des combinaisons possibles des résultats pour ces trois décisions. Les choix pour les achats et les déplacements, découlant de ces classes, sont traités dans un deuxième temps.

Les choix de localisation de long terme incluent la gestion du ou des biens immobiliers possédés ou loués. Les choix de planning portent sur la ou les activités principales de l’agent. On peut supposer qu’en début d’année l’agent décide aussi bien de son activité professionnelle que des activités qu’il pratiquera régulièrement (une passion<sup>1</sup> sportive ou artistique, par exemple, pourrait ne pas déboucher sur le choix d’en faire son activité principale. Cependant, l’agent pourrait décider de lui consacrer un temps et des ressources considérables). Les choix démographiques sont les choix d’organisation familiale, c’est-à-dire aussi bien dans quel type de ménage vivre que toutes les décisions qui en découlent (par exemple celles de santé, d’alimentation ou d’assurance).

Je définis l’ensemble de choix  $\mathbb{C}_t^i$  pour l’agent  $i$  au cours de la période  $t$ , où les définitions d’agent et de période ont été données en introduction de cette partie, toutes les “configurations de vie”  $c_{i,t} = \{l_{i,t}, a_{i,t}, d_{i,t}\}$ , c’est-à-dire les combinaisons de lieu d’habitation ( $l_{i,t} \in \mathbb{L}$ ), activités principales ( $a_{i,t} \in \mathbb{A}$ ) et situation démographique ( $d_{i,t} \in \mathbb{D}$ ).

Je suppose que chacun de ces processus est lui-même composé de deux sous-processus : un programme “externe” estimant s’il est optimal de se confronter à la décision ou si l’état hérité (dicté par les décisions à la période précédente et les ajustements de court terme face aux imprévus) est toujours optimal. Un deuxième processus “interne” estime le choix optimal, dans le cas où l’agent décide de changer d’état. Je ne considère alors pas crédible l’hypothèse que l’agent se confronte en repartant “à zéro” à ces décisions chaque année. En effet des décisions comme déménager, changer de profession ou avoir un enfant ont une importance primordiale sur la définition de la

---

1. Je ne parle pas ici des hobbies, pratiqués occasionnellement, qui peuvent être traités comme choix de court terme.

satisfaction d'un agent et impliquent une rupture dans son habitude. L'agent n'a donc pas pour objectif d'optimiser chaque choix annuel mais la suite entière des décisions. Pour pouvoir définir l'optimalité d'un tel choix l'agent doit donc pouvoir prévoir les effets demain de ses décisions aujourd'hui, ainsi que tous les autres changements d'état possibles.

On pourrait alors supposer l'existence de processus dynamiques<sup>2</sup> de décision optimale distincts qui estiment, pour chaque classe, la part de satisfaction obtenue par les décisions la composant en prenant en compte les évolutions d'état que celles-ci impliquent.

À chaque choix  $j$  d'une classe  $J \in \{\mathbb{L}, \mathbb{A}, \mathbb{D}\}$  peut être attachée une utilité inter-temporelle  $U^J(j, z_{i,t})$ , qui est définie en fonction des caractéristiques de l'agent (y compris ses préférences et son expérience)  $z_{i,t}$  et des caractéristiques propres au choix considéré. L'agent prévoit alors les scénarii possibles qui découlent de sa décision actuelle et choisi  $c$  qui maximise le flux d'utilités obtenu.

Or, pour une classe fixée, les décisions prises pour les autres classes ont vraisemblablement un impact sur le processus de choix la définissant. En effet, les ensembles de choix possibles de ces classes se définissent mutuellement (Ex. Choisir d'habiter en Bretagne exclura à priori les Alpes comme lieu d'activité principale possible), et aussi bien les caractéristiques de l'agent (ex. classe socio-professionnelle, nombre d'individus à charge) que celles propres aux choix mais dépendant de son état (ex. distance domicile-travail) peuvent varier en fonction de la configuration.

La première observation implique que l'utilité de choisir  $j$  pour une classe  $J$  est nulle lorsqu'elle n'est pas compatible avec les décisions prises pour les autres classes. On peut alors supposer que l'agent ne peut pas accéder à l'ensemble complet mais seulement à un sous-ensemble  $J_t^i \in J$ , conditionné par les décisions optimales  $k^*$  pour les autres classes.

En deuxième lieu, comme souligné plus haut, l'état de l'agent est aussi fonction des  $k^*$  ( $z^j = z_{i,t|k^*}$ ). Les transitions d'état découlant du choix  $j$  sont par conséquent représentés par la fonction de probabilité  $f_x(x_{i,t+1}|j, k^*x_{i,t})$ . Les trois classes n'interagissent donc pas seulement annuellement mais aussi

---

2. La distinction entre les processus interne et externe mentionnée plus haut n'est pas faite ici. Cependant, une hypothèse semblable serait développée pour le processus de décision d'achat.

dans les transitions d'un état à l'autre au fil du temps.

Il me semble alors plus crédible que l'agent n'ait pas pour objectif d'optimiser chaque problème séparément mais plutôt la configuration de vie qui en découle. L'ensemble de choix  $\mathbb{C}_t^i$  que l'agent prend en considération n'est composé que des combinaisons possibles de localisation, emploi du temps et composition familiale.<sup>3</sup>

Je suppose par simplicité que l'ensemble des configurations  $c_{i,t}$  possibles peut être défini de façon objective aussi bien par l'agent que par le chercheur. On pourrait par exemple définir une distance maximum entre le lieu d'habitation et le/les lieux d'activité principale, ou l'école la plus proche en cas de présence d'enfants à charge de l'agent. L'agent évalue ces distances, comme toute autre caractéristique qui pourrait être considérée comme contraignante pour la compatibilité des choix, au moment de la décision mais il est conscient de ne pas pouvoir dépasser les "limites objectives de faisabilité".

Il me semble alors crédible de penser que l'agent évalue tous les choix possibles pour une classe, sur la base des cadres définis par chaque ensemble de choix découlant des processus de décision propres aux autres classe. Pour chaque classe  $J$  il connaît donc l'utilité inter-temporelle  $U^J$  et la valeur escomptée  $V^J$  pour la période  $t+1$  attachées à chaque choix  $j$ . L'agent optimise alors simultanément les décisions (pour les trois classes) ainsi déterminées, en assignant des poids différents à chaque classe de décision. Un agent pourrait donc accepter par exemple de restreindre l'ensemble de choix d'une classe qui a un poids plus faible à ses yeux en faveur de ceux d'une plus importante pour sa satisfaction, afin d'obtenir une "configuration de vie" optimale.

Les paramètres  $\gamma = \{\gamma^L, \gamma^A, \gamma^D\}$  attachés aux utilités des classes  $J$  peuvent être considérés comme les poids avec lesquels chaque classe de choix rentre dans le processus complet<sup>4</sup>.

---

3. La définition de cet ensemble mériterait d'être étudié de plus près. En effet, la définition de compatibilité entre ces décisions est vraisemblablement entaché par les préférences et l'expérience de l'agent et est donc le résultat d'une évaluation subjective du monde qui l'entoure. Or, cela demanderait des années de travail et n'est donc pas considéré ici.

4. Ces paramètres aussi mériteraient beaucoup d'attention. En effet il est crédible que l'agent les définisse selon ses préférences et son expérience. On pourrait donc supposer que ces paramètres suivent une distribution propre. Cependant, je ne me penche pas ici sur la question.



Plusieurs auteurs se sont intéressés à la spécification des  $U^J$ . Il serait intéressant de s'interroger sur comment intégrer ces différents travaux, cependant je ne vais pas traiter cette question ici.

La configuration de vie, jointe aux caractéristiques et préférences de l'agent, définit ses besoins annuels.

On peut identifier deux types de besoins : les besoins de consommation et de temps libre.

Le besoin de consommation est le volume total de “biens à consommer”. Les biens auxquels je fais référence, dans les deux cas, peuvent être aussi bien réels (ex. alimentaire, équipement, parc automobile, ...) que des services (ex. assurances, soins médicaux, ...).

Le besoin de temps libre est représentable par le volume de temps à consacrer aux activités de tous les jours, comme le sommeil, la nutrition, la vie sociale, les loisirs, les vacances, etc. Ce besoin dépend de l'agent et de ses préférences mais aussi de la configuration de vie. On peut vraisemblablement penser en effet que les activités que l'on pratique au quotidien aussi bien que le type de ménage dans lequel on vit ont des influences différentes sur le besoin de temps libre que l'on ressent. Cette classe n'est pas traitée ici.

Les dépenses engendrées par ces besoins doivent être considérées par l'agent lorsqu'il évalue l'utilité de la configuration.

Le besoin de consommation peut être décomposé en un volume total de biens et services à consommer au quotidien (ex. alimentations, soins médicaux) et le nombre de biens durables à posséder et utiliser au cours de l'année (ex. électroménager, véhicule, assurance).

- Les *biens durables* sont achetés à l'unité et utilisés pendant plusieurs périodes. Le volume de biens durables est donc une quantité discrète. De plus, ces biens comportent un important nombre de caractéristiques qui ne permettent pas de considérer chaque bien comme une unité équivalente (à moins de discrétiser de façon extrême la définition de “bien”). Dans le cas d'un véhicule, par exemple, il me semble peut crédible que le fait de posséder une Renault Espace soit équivalent à posséder une Fiat 500. On pourrait alors supposer que l'agent définit des “niveaux de qualité” pour chaque classe de biens en fonction des caractéristiques du bien (par exemple, pour un véhicule, on pourrait définir différents niveaux selon les performances du moteur).

- Pour ce qui concerne les *biens quotidiens*, puisque l'on parle ici de prévisions annuelles et non pas de choix ponctuels de consommation il est plus plausible que l'agent ne considère pas à cette étape le volume exacte à consommer mais plutôt un "niveau de consommation". L'information est donc à considérer, à mon avis, comme discrète.

En supposant que l'agent attache une quantité d'utilité à chaque "niveau de qualité" d'un bien durable et à chaque "niveau de consommation" d'un bien quotidien, on peut alors représenter les besoins de consommation par le volume (continu) de "satisfaction (ou qualité) à financer", pour chaque classe de biens, plutôt que le nombre de biens à consommer (discret).

Le cas auquel je m'intéresse dans ce travail concerne une classe de biens durables : les véhicules. Je vais donc traiter ici seulement ce type de bien.

Pour les biens durables un volume d'utilisation est aussi à déterminer et ce volume répond à son tour aux besoins de l'agent. Pour ce qui concerne les véhicules, par exemple, aussi bien le niveau de qualité que la quantité des besoins de déplacements de l'agent que ceux-ci peuvent satisfaire sont à considérer. Le besoin de déplacement est représentable par le kilométrage que l'agent doit parcourir au cours de l'année pour pouvoir accomplir ses activités et subvenir à ses autres besoins<sup>5</sup>.

Comme dit plus haut aussi bien la possession d'un bien durable que son utilisation engendre des dépenses qui vont gréver l'utilité d'un choix (ex. on ne peut pas décider d'habiter à 100km de son lieu de travail sans considérer les dépenses dues à l'importance du kilométrage qui en découlent.). On pourrait alors supposer que les caractéristiques liées à un choix  $j$ , évaluées au moment de la définition de la part d'utilité apportée par chaque classe  $J$ , incluent les prix unitaires de ces volumes.

Pour ce qui concerne les volumes d'utilisation ces prix sont ceux des composants permettant de faire marcher le bien (ex. le prix du carburant pour les véhicules). La définition du prix de la satisfaction est plus délicate. En effet seulement un prix total pour un bien est observable mais ceci ne correspond pas à une unité de qualité. Cependant, en supposant pouvoir définir ces unités de qualité de façon objective, on pourrait définir un prix unitaire

---

5. Je parle ici aussi bien des besoins de l'agent que ceux d'un "individu dépendant", c'est-à-dire un individu dont l'agent est responsable et qui n'est pas en mesure de se confronter seul au processus de planification de vie annuelle (par exemple un enfant). Ses décisions sont donc intégrées aux processus de l'agent.

universel comme le rapport entre le prix moyen des biens ayant un même volume de qualité et l'utilité moyenne escomptée par les agents<sup>6</sup>. Bien qu'il ne soit pas crédible que l'agent soit capable de connaître un tel prix, même s'il est calculable, je suppose que le cerveau humain est capable de l'approximer. C'est à dire que par observation de son environnement, un individu est capable de reconnaître la valeur moyenne attachée à un bien et d'évaluer combien une variation dans sa qualité influencera cette valeur. Je suppose donc que l'agent est capable de faire la différence entre la valeur objective d'un bien et celle subjective qui inclut son ressenti et ses préférences par rapport au bien considéré.

Or, ces prix dépendent des choix précis que l'agent fait pour chaque classe de biens, les processus de décision de la configuration de vie et des choix visant à satisfaire les besoins engendrés devraient alors être traités simultanément. Puisque je crois plus plausible que l'homme hiérarchise ses décisions, plutôt que de traiter toutes les informations disponibles, et vu qu'une telle hypothèse permet de définir un cadre assez simple à présenter, je vais supposer ici que ce ne sont pas les prix auxquels il va réellement être confronté que l'agent considère pour définir l'utilité obtenue par une configuration de vie. Il estime plutôt un vecteur de "prévisions optimistes" (i.e. les prix minimum qu'il pourrait payer) et un de "prévisions pessimistes" (i.e. les prix maximum) pour ces prix. Ces estimations se basent aussi bien sur les préférences de l'agent et son expérience (ex. On ne prendra en compte que l'offre disponible à l'entour de son domicile, dans le cas des biens électroménagers par exemple, si l'on ne préfère pas se déplacer en voiture pour ce genre de tâches ou se faire livrer et l'on sait par expérience qu'une machine à laver ne peut pas être transportée en bus.) que sur les observations qu'il peut faire de son "marché disponible", et ce marché dépend de la configuration de vie (ex. On ne sera pas confrontés aux mêmes prix des carburants ou des transports urbains selon la zone d'habitation). Ces deux prix sont évalués simultanément par l'agent en définissant l'utilité de la configuration.

On peut supposer que, bien que l'utilité choisie doive être celle qui maximise le flux d'utilités futures, l'agent a aussi l'objectif, à plus court terme, d'équilibrer la satisfaction qu'il espère tirer de son année et les ressources (économiques, physiques, intellectuelles, etc.) qu'il doit dépenser. L'agent peut alors optimiser ses dépenses annuelles par le biais des volumes de besoins

---

6. Cela impliquerait un travail de discrétisation fine du marché (automobile, par exemple, dans le cas des véhicules) et la connaissance des processus de formation du prix de marché.

à satisfaire, en minimisant la fonction de dépense sous contrainte de l'utilité obtenue. En supposant que l'utilité totale de consommation  $U(D_1, \dots, D_b, \dots, D_B)$  soit exprimable comme une fonction Cobb-Douglas, l'objectif de l'agent est de déterminer le vecteur des demandes  $D$  tel que

$$C(v(c, x_{it}), p_{i,t}) = \min_D \left\{ \sum_{b \in \mathbb{B}_i^t} D_b p_b \mid a \prod_b D_b^{\alpha_b} \geq v(c, x_{it}) \right\} \quad (1.1)$$

où  $a, \alpha_b > 0$  et  $\sum_b \alpha_b = 1$ .  $p_{i,t}$  est le vecteur complet des prix auxquels l'agent  $i$  est confronté au moment  $t$ , pour tout bien à consommer afin de satisfaire ses besoins. Puisque je suppose que l'agent n'a pas comme seul objectif l'équilibre économique mais souhaite aussi un équilibre de satisfaction, je suppose qu'il considère ici des prix virtuels, définis en fonction du marché réel et de ses préférences.

On peut supposer que l'agent est capable de quantifier l'intérêt qu'il porte à une classe de biens et de définir donc un "prix ressenti" unitaire de possession et un par unité d'utilisation, inversement proportionnel à cet intérêt. Cela revient à dire que moins un bien est intéressant pour l'agent, plus la possession et l'utilisation de ce bien lui demandera un effort important. Le prix de chaque bien  $b$ , élément du vecteur  $p_{i,t}$ , peut être exprimé en fonction des prix minimum et maximum observés et de ce prix ressenti. Les prix directement liés aux "volumes de satisfaction" choisis pour les trois classes principales<sup>7</sup> sont dictés par la configuration et l'agent ne cherche plus à les optimiser (i.e. les mêmes prix sont utilisés pour les utilités min. et max.). Ils constituent donc un coût fixe qui est déjà évalué en définissant l'utilité de la configuration.

En dérivant la fonction de dépense par rapport au prix de possession et au prix d'utilisation pour un bien  $b$ , on obtient respectivement la satisfaction et le volume d'utilisation total optimal.

Or, les décisions d'achat suivent vraisemblablement un processus propre à chaque classe de biens. Ces processus sont contraints alors par ces valeurs optimales.

Je vais m'intéresser ici à une seule de ces classes de biens : les véhicules.

---

7. Rappel : localisation planning et démographie

Des processus semblables à celui que je vais présenter maintenant pourraient être imaginés pour les autres classes de biens durables.

## Chapitre 2

# Les choix d'équipement automobile

En raison du caractère discret du problème, je vais dans un premier temps présenter ce type de modélisation.

Les méthodes permettant d'estimer un modèle dont la variable expliquée est discrète ont vu le jour dans le champ de la biométrie (par exemple [Berkson \(1951\)](#) et [Finney \(1971\)](#)). [Amemyia \(1981\)](#) et [Ben-Akiva and Lerman \(1985\)](#) fournissent une bonne présentation de cette classe de modèles.

### LE PROBLÈME DISCRET

Soit  $J$  le nombre de choix possibles et  $y_j = (y_{1j}, \dots, y_{Nj})$  les  $J$  variables binaires définies par

$$y_{ij} = \begin{cases} 1 & \Leftrightarrow V_{ij} = \max_{h \in (1, 2, \dots, J)} V_{ih} \\ 0 & \Leftrightarrow V_{ij} \neq \max_{h \in (1, 2, \dots, J)} V_{ih} \end{cases} \quad (2.1)$$

où  $V_{ih}$  est la valeur que prend l'utilité indirecte lorsque le ménage  $i$  fait le choix  $h$ . On a donc  $V_j > V_h \forall k \neq j$ .

Rappelons que l'utilité indirecte  $V$  est définie comme

$$V(p, R) = U[q^*(p, R)] \quad (2.2)$$

où  $U$  est l'utilité directe et  $q^*$  la demande qui la maximise.

Si l'on s'inspire des travaux menés en psychologie (Voir de Palma et Thisse (1987) pour une présentation historique et épistémologique des modèles de

choix discrets.) on voit que le niveau d'utilité peut être considéré comme le niveau de stimulus. Ce dernier a un caractère aléatoire. L'utilité  $V$  se décompose donc en  $V = \bar{V} + \varepsilon$  où  $\bar{V}$  est la composante connue de l'utilité et  $\varepsilon$  est l'effet aléatoire dû aux informations non observables.

Le choix de se servir de l'utilité indirecte  $V$  se justifie en observant que la forme fonctionnelle la plus simple pour  $\bar{V}$  est

$$\bar{V}_j = z_j \gamma + w \delta_j \quad (2.3)$$

où  $z_j$  sont des caractéristiques du choix  $j$  et  $w$  des caractéristiques individuelles.

Cette forme fonctionnelle, linéaire en ses paramètres, dépend en fait des prix  $p$  (qui font partie des  $z_j$ ) et du revenu  $R$  (une des caractéristiques  $w$ ).

Une fois la forme fonctionnelle de  $\bar{V}$  définie, il nous reste à spécifier la distribution des  $\varepsilon_j$ . En se basant sur les travaux de McFadden (1968, 1973), on peut supposer les  $\varepsilon_j$  indépendants et identiquement distribués selon une loi de Gumbel dont la fonction de répartition et la fonction de densité sont respectivement

$$F(a) = \exp[-e^{-a}] \quad \text{et} \quad f(u) = \exp[-u - e^{-u}] \quad (2.4)$$

Si l'on fait l'hypothèse d'indépendance des caractéristiques  $x$  pour tous les  $\varepsilon_j$ , la probabilité qu'un ménage choisisse le kilométrage  $j$  compte tenu de ses caractéristiques  $x$  s'écrit

$$P(y_i = 1 | x) = P[(V_1 < V_j, \dots, V_J < V_j) | x] \quad (2.5)$$

Pour traduire cette probabilité en fonction de répartition il faut la réécrire en décomposant  $V$  et conditionner par  $\varepsilon_j$ .

L'indépendance supposée entre les  $\varepsilon_j$  nous permet d'écrire la probabilité sous la forme

$$P(y_i = 1 | x) = \int \prod_{h \neq j} P[\varepsilon_h < \varepsilon_j + \bar{V}_j(x) - \bar{V}_h(x) | x, \varepsilon_j] \cdot f(\varepsilon_j) d\varepsilon_j \quad (2.6)$$

Rappelons la propriété suivante : soit  $u$  une variable aléatoire de fonction de répartition  $F$  et  $z$  un vecteur de variables aléatoires ; si  $u$  est indépendante de  $z$ , alors  $p[u < g(z) | z] = F[g(z)]$ .

Puisque nous avons supposé les  $\varepsilon_j$  indépendantes des  $x$  dont dépend la fonction  $\bar{V}$  on a

$$P(y_j = 1 | x) = \int \prod_{h \neq j} P[\varepsilon_j + \bar{V}_j(x) - \bar{V}_h(x)] \cdot f(\varepsilon_j) d\varepsilon_j \quad (2.7)$$

Après substitution de  $F$  et  $f$  par leurs expressions on obtient, pour  $j = 1, \dots, J$

$$P(y_j = 1 | x) = \frac{\exp \bar{V}_j(x)}{\sum_{h=1}^J \exp \bar{V}_h(x)} \quad (2.8)$$

En remplaçant la forme fonctionnelle que nous avons définie en 2.3 pour  $\bar{V}$  dans la fonction de répartition des  $\varepsilon_j$ , le modèle s'écrit

$$P(y_j = 1 | x) = \frac{\exp [z_j \gamma + w \delta_j]}{\sum_{h=1}^J \exp [z_h \gamma + w \delta_h]} \quad (2.9)$$

La dépendance du choix  $j$ , des paramètres attachés aux variables individuelles, est nécessaire si l'on veut se servir de ces variables dans l'explication. On peut aisément voir que pour  $\delta_j = \delta \forall j$  les caractéristiques individuelles disparaissent du modèle. En revanche, l'indépendance des paramètres attachés aux caractéristiques des choix, permet l'interprétation en termes d'élasticité des choix aux variations de prix.

Arrêtons-nous un instant pour définir deux modèles polytomiques non ordonnés.

Soit  $x_i$  le vecteur décrivant l'individu  $i$ . La probabilité que le choix de l'individu  $i$  soit  $j$  est

$$P(j | x_i) = \frac{\exp x_i \beta_j}{\sum_{h=1}^J \exp x_i \beta_h} \quad (2.10)$$

Il s'agit d'un modèle logit multinomial. Supposons maintenant que la décision de choisir  $j$  ne dépende pas des caractéristiques individuelles mais des caractéristiques  $z_j$  propres au choix  $j$ . Le modèle s'écrit alors

$$P(j | x_i) = \frac{\exp x_j \beta}{\sum_{h=1}^J \exp x_h \beta} \quad (2.11)$$

et il est dit logit conditionnel.

Comme on peut le voir, ces deux modèles sont des cas particuliers du modèle de choix discret que l'on a décrit. Lorsque  $\gamma = 0$  on est dans le cas de logit multinomial et pour  $\delta_j = 0$  on retrouve le logit conditionnel. On dénomme alors ce modèle logit multinomial/conditionnel.

Revenons maintenant à la définition du programme. Nous avons dit que l'individu  $i$  choisit  $j$  lorsque  $V_i j = \max_{h \in (1, 2, \dots, J)} V_i h$ . Si l'on multiplie  $V$  par une constante  $\sigma$  strictement positive et si on lui ajoute une deuxième constante  $\mu$ , le programme s'écriera

$$y_{ij} = 1(0 = \max_{h \in (1, 2, \dots, J)} \sigma(V_i h - V_i j)) \quad (2.12)$$



Cela signifie que l'ordre de préférences des agents est invariant pour toute transformation monotone croissante de la fonction d'utilité indirecte.

Le caractère durable du véhicule suggère que le comportement optimal vis-à-vis des décisions portant sur son utilisation soit un comportement “tourné vers le futur”. En effet, si l'on considère l'achat d'un bien durable comme un investissement qui est remboursé à la fois par les services qu'il fournit et la valeur finale que l'on peut en tirer par sa revente, il semble correct de supposer que les choix de possession et utilisation optimales d'un véhicule au fil du temps soient le résultat d'un programme dynamique tourné vers le futur, prenant en compte le déclin de la valeur du véhicule en fonction du temps et de l'utilisation. En d'autres termes, on peut supposer que l'agent prend en compte les conséquences que l'utilisation d'un véhicule aujourd'hui a sur son efficacité demain (comme suggéré par [Rust \(1987\)](#)), ainsi que les évolutions futures de son état, lorsqu'il décide de la possession et l'utilisation d'un véhicule.

Les décisions d'achat me semblent aussi issues d'un processus dynamique tourné vers le futur. En effet, l'importance de l'investissement économique que l'achat automobile comporte (“[...] probablement le plus important après l'achat immobilier”, cite), ainsi que la continue évolution du marché automobile, pourraient pousser l'agent à planifier ces décisions pour le long terme. Je suppose en plus qu'un tel processus de décision soit affecté par l'effet de l'habitude. Un agent pourrait alors aussi vouloir, par la planification de ses achats automobiles, satisfaire ses besoins de nouveauté ou habitude (selon ses préférences).

Puisqu'un des choix qui s'offrent à l'agent est d'être multi-équipé (i.e. de s'équiper de plusieurs véhicules), et les caractéristiques de son parc automobile définissent l'état dans lequel il se trouve au moment de la décision, l'hypothèse d'un tel comportement implique que l'agent est capable d'anticiper les décisions de possession et d'utilisation des autres véhicules qui composaient le parc à la date précédente. De plus, l'agent est libre d'acheter un nouveau véhicule à tout moment. Il devrait donc être en mesure d'anticiper aussi ses décisions d'achat futures.

Dans l'idéal il faudrait alors résoudre un système de  $n_i$  problèmes dynamiques, déterminant simultanément les décisions optimales de possession et

utilisation, pour les  $n_i - 1 \geq 0$  véhicules composant le parc de l'agent  $i$  à la date  $t - 1$ , et d'achat pour la période  $t$ .

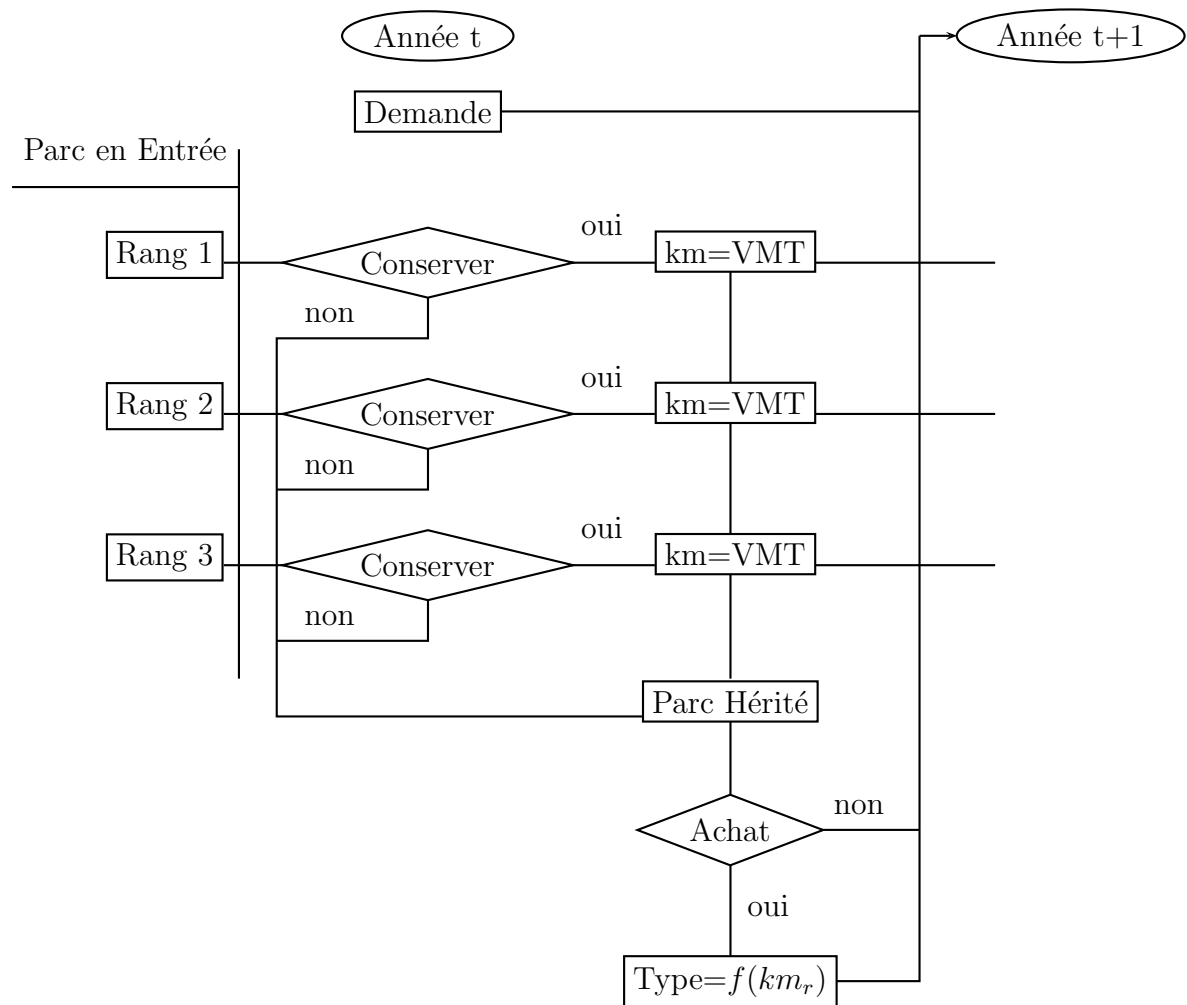
Comme dit plus haut, ces décisions visent à servir aux besoins de l'agent et doivent donc satisfaire la fonction de dépense liée à l'utilité complète de la configuration de vie. Ce système de problèmes dynamiques devrait alors être intégré à la solution du problème complet de l'agent.

Par souci de simplicité et puisque je crois plus crédible que l'agent hiérarchise ses choix, je ne traite pas la simultanéité entre le modèle complet et les décisions d'achat.

Les demandes optimales de satisfaction tirées du parc et de kilométrage que celui-ci peut fournir sont donc considérées ici comme contraintes externes, obtenues par solution du modèle de "configuration de vie".

L'objectif de l'agent est donc de choisir le parc automobile qui satisfait le mieux ses besoins, tout en respectant les contraintes dictées par le caractère durable de ce type de bien. Je vais proposer ici une décomposition en trois étapes séquentielles du processus de décision d'équipement automobile :

1. **Définition de l'utilité de possession et d'utilisation d'un véhicule** comme solution d'un problème discret/continu dynamique tourné vers le futur, fonction des caractéristiques du véhicules et de l'état de l'agent, à "composition du parc fixée" ;
2. **Définition de l'utilité de possession d'un parc automobile** comme solution d'un problème emboité discret dynamique tourné vers le futur, fonction des caractéristique de l'agent et de l'effet que l'habitude a sur lui, ainsi que de l'utilité tirée de la possession et l'utilisation des véhicules le composant ;
3. **Optimisation des décisions dans la "configuration de vie" optimale** comme solution d'un modèle discret multinomial contraint.



Flow-Chart : structure des décisions du ménage

## 2.1 Le Véhicule

Les biens durables sont caractérisés par un volume de satisfaction apportée et un volume d'utilisation produite. Les premiers à prendre en compte la dualité de ce type de biens dans la littérature sont [Train and Lohrer \(1982\)](#). Ils s'intéressent aux décisions d'usage de l'automobile et proposent l'utilisation de la méthode des variables instrumentales, en considérant certaines caractéristiques du véhicule comme endogène à la fonction de demande de kilométrage. En se basant sur cette première idée, [Dubin and McFadden \(1984\)](#) apportent une contribution significative au traitement de ce type de biens en proposant une méthode de correction des attentes conditionnelles. Plusieurs

auteurs ont apporté leur contribution à l'évolution de ce type de modèle depuis, comme par exemple Hanemann (1984), Hensher and Milthorpe (1987), Hensher et al. (1990), Ben-Akiva and Bolduc (1991), Bernard et al. (1996) et plus récemment Kim (2002) et Train (2003).

Hensher and Milthorpe (1987) comparent plusieurs méthodes de correction et montrent que, bien que les résultats obtenus ne soient pas différents de ceux obtenus avec la méthode des variables instrumentales proposée à l'origine, l'hypothèse d'absence de biais de sélection ne peut pas être acceptée. Les travaux de Berkowitz et al. (1990) et Hensher et al. (1990), appliqués aux décisions de possession et utilisation des véhicules, confirment ces résultats. Hensher et al. (1990) suggèrent alors que le choix de ne pas utiliser une méthode de correction n'est pas justifié et l'absence apparente de biais est une évidence empirique qui montre que la relation entre l'utilité de possession et la demande d'utilisation d'un bien durable est adéquatement estimée par l'inclusion des caractéristiques de celui-ci.

Une contribution méthodologique est ensuite apportée par les travaux de Kim (2002) dans lesquels une approche bayésienne est proposée.

L'importance des impacts que l'utilisation automobile a sur l'environnement ont motivé un grand nombre d'études sur la demande automobile des ménages. Une revue du problème est fournie par de Jong et al. (2004). Pour ce qui concerne l'application du modèle prenant en compte le caractère mixte, discret/continu, des véhicules dans l'estimation de la demande de kilométrage des agents on peut citer par exemple Hensher and Le Plastrier (1985), Mannering and Winston (1985), Berkovec and Rust (1985), Train (1986), West (2004).

Une autre observation importante pour l'étude de ce problème est mise en lumière par Hess (1977) et Goodwin and Mogridge (1981). Les deux montrent l'amélioration non négligeable apportée par un traitement dynamique du processus de choix de possession des véhicules. Manski (1980), Manski and Sherman (1980), Hocherman et al. (1983), Mannering and Winston (1985), Berkovec (1985b), Hensher and Le Plastrier (1985), Kitamura (1987), Smith et al. (1991), Gilbert (1992), van Wissen and Golob (1992) et Golob and Brownstone (1997), entre autres, prennent en compte dans leurs travaux le caractère dynamique du problème. Certains de ces auteurs (ex. Berkovec (1985b)) traitent aussi l'équilibre entre offre et demande.

Deux types d'approches dynamiques sont utilisés dans ces travaux : le modèle de durée (ex. Hensher and Le Plastrier (1985)) et le modèle de tran-

sition (ex. [Hoeherman et al. \(1983\)](#)). Bien que l'introduction de ces méthodes comporte un pas en avant considérable dans l'étude de la possession automobile, les dynamiques sont considérées comme tournées vers le passé. En d'autres termes l'agent ajuste ses décisions à posteriori.

[Rust \(1987\)](#), en accord avec les remarques faites par [Lucas \(1976\)](#), propose de considérer que la suite de décisions optimales de possession et utilisation pour un véhicule soit solution d'un problème dynamique discret tourné vers le futur. Il développe un cadre d'estimation restrictif mais maniable, et propose l'utilisation de l'algorithme du point fixe pour l'approximation de la fonction valeur du programme dynamique ([Rust \(1988\)](#), [Rust \(1994\)](#)).

### LE PROBLÈME DYNAMIQUE

Le temps  $t$  dans les modèles dynamiques de choix discret à un seul agent est discret  $t = \{0, \dots, T\}$  et son horizon  $T$  peut être fini ou infini. Chaque individu  $i$  au moment  $t$  est caractérisé par un vecteur  $s_{it}$  dit d'état et il effectue le choix  $a_{it}$ , sur une liste discrète  $A = \{0, 1, \dots, J\}$ , qui maximise l'utilité espérée

$$E\left(\sum_{j=0}^{T-t} \beta^j U(a_{i,t+j}, s_{i,t+j}) | a_{it}, s_{it}\right) \quad (2.13)$$

où  $\beta \in (0, 1)$  est le coefficient d'actualisation, supposé indépendant de l'individu  $i$  observé, et  $U(a_{it}, s_{it})$  la fonction d'utilité au moment  $t$ . On suppose que la décision prise au moment  $t$  par l'individu  $i$  conditionne les évolutions futures selon une distribution de transition de Markov  $F(s_{i,t+1} | a_{it}, s_{it})$ . Le programme dynamique est ainsi défini.

La fonction valeur  $V(s_{it})$  est définie comme solution de l'équation de Belleman

$$V(s_{it}) = \max_{a \in A} \left\{ U(a, s_{it}) + \beta \int V(s_{i,t+1}) dF(s_{i,t+1} | a, s_{it}) \right\} \quad (2.14)$$

et la règle de choix optimal associé est  $\alpha(s_{it}) = \arg \max_{a \in A} \{v(a, s_{it})\}$ .

On appelle  $\theta$  le vecteur des paramètres à estimer et l'on choisit un critère d'estimation  $g_N(\theta)$ . Si l'échantillon est aléatoire et l'estimation est faite à l'aide de la méthode du maximum de log-vraisemblance, on a  $g_N(\theta) = \sum_{i=1}^N l_i(\theta)$  où la contribution à la fonction de vraisemblance de l'individu  $i$  est

$$l_i(\theta) = \log Pr(\alpha(s_{it}, \theta) = a_{it}, x_{it} : t = 1, 2, \dots, T_i | \theta) \quad (2.15)$$

Cela montre la difficulté de ce type de modèles. Pour pouvoir estimer les paramètres  $\theta$  il faut résoudre le problème dynamique afin d'obtenir la règle optimale  $\alpha$ , qui dépend aussi de  $\theta$ . La résolution du problème requiert donc un processus emboîté.

Le vecteur des variables d'état  $s_{it}$  contient à la fois les variables  $x_{it}$  observées par le chercheur et les informations non-observables  $\varepsilon_{it}$ . Comme je l'ai annoncé précédemment, les hypothèses faites sur le comportement de l'information non-observable ont des conséquences importantes sur l'estimation du modèle et c'est sur cette base que l'on distingue les deux principales méthodes.

### Le modèle de Rust

Il s'agit du plus restrictif cadre d'estimation pour ce type de programme. Rust (1987) pose, en fait, cinq conditions sur les relations existant entre l'information observée et celle non-observable, qui seront en partie relaxées dans le modèle de Eckstein and Wolpin (1989). Les hypothèses sont les suivantes :

- *AS (Additive separability)* : La fonction d'utilité au moment  $t$  peut être écrite comme somme d'une composant observable et d'une non-observable de la façon suivante :

$$U(a, x_{it}, \varepsilon_{it}) = u(a, x_{it}) + \varepsilon_{it}(a) \quad (2.16)$$

où  $\varepsilon_{it}(a)$  est une variable aléatoire de moyenne nulle définie sur  $\mathfrak{R}$  et elle a la même dimension  $(J + 1) \times 1$  de l'espace de choix  $A$ . Le vecteur des paramètres de la fonction d'utilité  $u$  est appelé  $\theta_u$ .

- *IID (iid unobservables)* : Les variables d'état non-observables  $\varepsilon_{it}$  sont indépendantes et identiquement distribuées sur les individus et le temps avec une fonction de répartition  $G_\varepsilon(\varepsilon_{it})$  de moyenne finie et deux fois différentiable en  $\varepsilon_{it}$ .
- *CI-X (Conditional independence of futur  $x$ )* : L'information non-observable  $\varepsilon_{it}$  n'a pas d'influence sur la transition de l'état présent  $x_{it}$  au suivant  $x_{i,t+1}$  et la fonction de répartition est donc équivalente à la fonction de transition des  $x_{it}$

$$CDF(x_{i,t+1}|a_{it}, x_{it}, \varepsilon_{it}) = F_x(x_{i,t+1}|a_{it}, x_{it}) \quad (2.17)$$

Le vecteur des paramètres de  $F_x$  est appelé  $\theta_f$ .

- *CLOGIT* : L'information non-observable  $\{\varepsilon_{it}(a) : a = 0, 1, \dots, J\}$  est indépendante sur les choix et suit une distribution de Gumbel.
- *DIS (Discrete support of  $x$ )* : Le support des  $x_{it}$  est discret et fini.

Une conséquence importante des hypothèses IID et CI-X est l'hypothèse d'indépendance conditionnelle de Rust (1987) :

$$F(x_{i,t+1}, \varepsilon_{i,t+1}|a_{it}, x_{it}, \varepsilon_{it}) = G_\varepsilon(\varepsilon_{i,t+1})F_x(x_{i,t+1}|a_{it}, x_{it}) \quad (2.18)$$

L'information non-observable ne constitue donc pas un obstacle à l'estimation des probabilités de transition puisque  $x_{it}$  est une statistique exhaustive pour le choix  $a_{it}$ . En outre, on peut intégrer la fonction valeur (2.14) sur la distribution de l'information non-observable en obtenant la fonction *Emax*

$$\begin{aligned} \bar{V}(x_{it}) &\equiv \int V(x_{it}, \varepsilon_{it}) dG_\varepsilon(\varepsilon_{it}) \\ &= \int \max_{a \in A} \left\{ u(a, x_{it}) + \varepsilon_{it}(a) + \beta \sum_{x_{i,t+1}} \bar{V}(x_{i,t+1}) f_x(x_{i,t+1} | a, x_{it}) \right\} dG_\varepsilon(\varepsilon_{it}) \end{aligned} \quad (2.19)$$

qui caractérise complètement le programme. En plus, la dimension de l'espace  $X$  est moins importante que celle de l'espace  $S$ , le programme peut être alors résolu.

Pour terminer, on définit la probabilité conditionnelle du choix  $a$  (CCP) comme :

$$P(a|x, \theta) = \int I\{v(a, x_{it}) + \varepsilon_{it}(a) > v(a', x_{it}) + \varepsilon_{it}(a') \text{ for all } a'\} dG_\varepsilon(\varepsilon_{it}) \quad (2.20)$$

Rust montre que les hypothèses IID et CLOGIT pour  $\{\varepsilon_{it}(a)\}$  permettent d'obtenir une solution analytique pour l'intégral de Bellman et les CCP. On a, en fait

$$\bar{V}(x_{it}) = \log \left( \sum_{a=0}^J \exp \left\{ u(a, x_{it}) + \beta \sum_{x_{i,t+1}} \bar{V}(x_{i,t+1}) f_x(x_{i,t+1} | a, x_{it}) \right\} \right) \quad (2.21)$$

et

$$P(a|x_{it}, \theta) = \frac{\exp\{v(a, x_{it})\}}{\sum_{j=0}^J \exp\{v(j, x_{it})\}} \quad (2.22)$$

**Nested fixed point** L'algorithme d'estimation proposé consiste en l'emboîtement d'un nested fixed point algorithm dans un BHHH. Ce deuxième estime le vecteur  $\theta$  des paramètres structuraux de la règle de décision optimale  $\alpha(x_{it}, \varepsilon_{it}, \theta)$  par itérations successives définies comme

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \left( \sum_{i=1}^N \nabla l_i(\hat{\theta}_k) \nabla l_i(\hat{\theta}_k)' \right) \left( \sum_{i=1}^N \nabla l_i(\hat{\theta}_k) \right) \quad (2.23)$$

où  $\nabla l_i(\hat{\theta})$  est le gradient en  $\hat{\theta}$  de la fonction du maximum de vraisemblance pour l'individu  $i$ .

Dans le cas d'un problème à horizon infini l'on choisit un vecteur de paramètres arbitraires que nous nommons  $\hat{\theta}_0$  et on l'utilise comme vecteur initial d'estimation de l'algorithme du nested fixed point emboîté pour construire le vecteur

$\bar{V}(\hat{\theta}_0)$  par itérations successives, jusqu'à convergence, de l'équation de Bellman  $\bar{V}_{h+1} = \log \left( \sum_{a=0}^J \exp \left\{ u \left( a, \hat{\theta}_0 \right) + \beta F_x \left( a \right) \bar{V}_h \right\} \right)$ , en partant d'une valeur supposée  $\bar{V}_0$ . On utilise alors les  $\hat{\theta}_0$  et  $\bar{V}(\hat{\theta}_0)$  trouvés pour construire les probabilités  $P(a|x, \hat{\theta}_0)$ , la matrice  $\frac{\partial \bar{V}(\hat{\theta}_0)}{\partial \theta'}$  et le gradient  $\nabla l_i(\hat{\theta}_0)$ .

Lorsque le problème est à horizon fini nous pouvons résoudre l'équation de Bellman, son gradient et ses probabilités de choix en utilisant la backward induction. Pour la dernière période  $T$  la fonction de Bellman s'écrit

$$\bar{V}_T(\hat{\theta}_0) = \log \left( \sum_{a=0}^J \exp \left\{ u_T \left( a, \hat{\theta}_0 \right) \right\} \right) \quad (2.24)$$

et chaque période  $t \leq T - 1$  est définie par récurrence comme

$$\bar{V}_t(\hat{\theta}_0) = \log \left( \sum_{a=0}^J \exp \left\{ u_T \left( a, \hat{\theta}_0 \right) + \beta F_{x,t} \left( a \right) \bar{V}_{t+1} \left( \hat{\theta}_0 \right) \right\} \right) \quad (2.25)$$

Les probabilités sont définies comme

$$P_t \left( a \mid \hat{\theta}_0 \right) = \frac{\exp \left\{ u_T \left( a, \hat{\theta}_0 \right) + \beta F_{x,t} \left( a \right) \bar{V}_{t+1} \left( \hat{\theta}_0 \right) \right\}}{\left( \sum_{j=0}^J \exp \left\{ u_T \left( j, \hat{\theta}_0 \right) + \beta F_{x,t} \left( j \right) \bar{V}_{t+1} \left( \hat{\theta}_0 \right) \right\} \right)} \quad (2.26)$$

et les gradients ont la forme suivante

$$\begin{cases} \frac{\partial \bar{V}_T(\hat{\theta})}{\partial \theta'_u} = \sum_{a=0}^J P_t \left( a \mid \hat{\theta} \right) * \left\{ \frac{\partial u_t(a, \hat{\theta})}{\partial \theta'_u} + \beta F_{x,t} \left( a \right) \frac{\partial \bar{V}_{t+1}(\hat{\theta})}{\partial \theta'_u} \right\} \\ \frac{\partial \bar{V}_T(\hat{\theta})}{\partial \theta'_f} = \beta \sum_{a=0}^J P_t \left( a \mid \hat{\theta} \right) * \left\{ \frac{\partial F_{x,t}(a)}{\partial \theta'_f} \bar{V}_{t+1} \left( \hat{\theta} \right) + \beta \hat{F}_{x,t} \left( a \right) \frac{\partial \bar{V}_{t+1}(\hat{\theta})}{\partial \theta'_f} \right\} \end{cases} \quad (2.27)$$

Une approche plus simple pour l'estimation du nested fixed point, proposée par Rust, est de décomposer la log-vraisemblance comme suit

$$l_i(\theta) = \sum_{t=1}^{T_i} \log P(a_{it} | x_{it}, \theta) + \sum_{t=1}^{T_i-1} \log f_x(x_{i,t+1} | a_{it}, x_{it}, \theta_f) \quad (2.28)$$

et de l'estimer en deux étapes. Cette décomposition est possible sous les hypothèses IID et CI-X. Un terme apportant la contribution des conditions initiales  $Pr(x_{i1} | \theta)$  devrait apparaître dans la décomposition, mais puisque l'on utilise la vraisemblance conditionnelle ce terme est négligeable.

La maximisation du deuxième terme permet l'estimation du vecteur des



paramètres  $\theta_f$  sans besoin de résoudre le programme. En utilisant le vecteur ainsi estimé on maximise ensuite la probabilité conditionnelle de choix (CCP) pour obtenir le vecteur complet des paramètres  $\theta$ . Cette approche baisse le cout total de l'estimation en réduisant le nombre de paramètres à estimer dans le BHHH et la dimension de l'espace des paramètres.

Le modèle que je propose s'inspire du modèle de [Rust \(1987\)](#). Il se différencie de celui-ci pour le choix de définir le problème inter-temporel comme discret/continu. L'utilisation de telles probabilités permet de traiter la demande du kilométrage comme endogène au modèle.

En disposant de données de panel un tel modèle peut être estimé à horizon fini. Puisque la valeur escomptée future est calculable dans ce cas, il est possible de traiter les autres décisions ayant un impact sur la composition du parc automobile de l'agent comme exogène au problème. Cela revient à faire l'hypothèse que la probabilité de prendre toute les décisions qui définissent la composition du parc hérité, avant la décision pour le véhicule considéré, est connue par l'agent.

Un tel modèle fournit alors une estimation de la fonction définissant l'utilité de posséder et utiliser un véhicule  $j$  pour l'agent  $i$ , lorsque la composition du parc auquel ce véhicule pourrait appartenir est connue, en fonction de ses caractéristiques et des caractéristiques de l'utilisateur principal. Les décisions d'appariement véhicule-conducteur principal sont considérées aussi comme exogènes.

La spécification de ce modèle est présentée au chapitre 6 de la deuxième partie de ce travail.

## 2.2 Le Parc Automobile

La multiplicité de décisions d'équipement auxquelles l'agent est confronté et la simultanéité qui les caractérise, fait de la composition du parc automobile un problème délicat à traiter.

[Train \(1986\)](#) propose un modèle logit emboîté qui permet de prendre en compte l'interaction existante entre les différentes composantes de ce

type de décision. Plusieurs auteurs se sont intéressés à ce problème en utilisant ce type de modèle. [Brownstone et al. \(2000\)](#), [Page et al. \(2000\)](#) et [Birkeland and Jordal-Jorgensen \(2001\)](#), par exemple, étudient ainsi les comportements d'achat automobile. [Choo and Mokhtarian \(2004\)](#) s'intéressent, de leur côté, aux effets des habitudes automobiles (pour le véhicule le plus utilisé par l'agent l'année précédente). Une revue complète de ce type de modèles peut être trouvée dans [Bhat and Sen \(2006\)](#).

[Bierlaire \(1995\)](#) propose une méthode d'estimation simultanée pour ce modèle. [Bhat and Sen \(2006\)](#) présentent un modèle plus sophistiqué qui permet de traiter le cas d'agents multi-équipés. Il s'agit du modèle multinomiale discret/continu Extreme Value (qui ne sera pas décrit ici).

### LE PROBLÈME DISCRET EMBOITÉ

Un modèle logit emboité permet de mieux modéliser une suite de choix ainsi structurée. Par simplicité je me limiterai ici à la présentation du cas d'un arbre à deux niveaux.

Soit  $L$  le nombre de groupes dans lesquels l'espace des choix se répartit et  $J_l$  le nombre de choix dans chaque groupe  $l$ . On indicera les  $J = J_1 + \dots + J_L$  choix par  $j(l)$

On peut alors supposer que l'utilité indirecte  $V_j = \bar{V}_j + \varepsilon_j$  se décompose de la façon suivante :

$$V_j = \bar{V}_l + \bar{V}_{j(l)} + \varepsilon_j \quad (2.29)$$

où  $\bar{V}_l$  est l'utilité associée au groupe  $l$ ,  $\bar{V}_{j(l)}$  est propre au choix  $j(l)$  et  $\varepsilon_j$  est la partie stochastique de l'utilité.

Nous allons spécifier, dans un premier temps, la distribution des  $\varepsilon_j$ . McFadden (1978) a démontré qu'en considérant que  $(\varepsilon_1, \dots, \varepsilon_J)$  est distribué selon une loi ayant comme fonction de répartition

$$F(\varepsilon_1, \dots, \varepsilon_J) = \exp \left( - \sum_{l=1}^L \left( \sum_{j=1}^{J_l} e^{-\varepsilon_j/\lambda_l} \right)^{\lambda_l} \right) \quad (2.30)$$

On peut écrire un modèle *RUM* (*Random Utility Maximisation*) de choix discret, si et seulement si  $0 < \lambda_l \leq 1$ , qui associe au choix  $j$  la probabilité

$$P(j | x) = \frac{\exp(\bar{V}_j/\lambda_l) \left( \sum_{j=1}^{J_l} \exp(\bar{V}_{j(l)}/\lambda_l) \right)^{\lambda_l - 1}}{\sum_{l=1}^L \left( \sum_{j=1}^{J_l} \exp(\bar{V}_{j(l)}/\lambda_l) \right)^{\lambda_l}} \quad (2.31)$$

C'est l'expression du logit emboîté à deux niveaux.

Dans ce modèle les résidus  $\varepsilon_j$  correspondant aux choix d'un même groupe  $l$  sont corrélés et leur degré d'indépendance est mesuré par le paramètre  $\lambda_l$ .

Il reste à définir une forme fonctionnelle pour  $\bar{V}_j$ . Nous pouvons répartir les variables explicatives  $x$  en deux groupes, les caractéristiques  $x_l$  communes à tous les choix d'un même groupe et les caractéristiques  $x_{j(l)}$  propres à chaque choix  $j(l)$ . En supposant que  $\bar{V}_l$  et  $\bar{V}_{j(l)}$  sont des combinaisons linéaires des variables explicatives, l'utilité indirecte s'écrit

$$\bar{V}_j = x_l \beta + \lambda_l x_{j(l)} \beta_l \quad (2.32)$$

L'introduction du paramètre  $\lambda_l$  revient à normer  $\beta_l$ .

Revenons maintenant sur une propriété des modèles logit polytomiques, dit propriété *IIA* (*Independence from Irrelevant Alternatives*).

Cette propriété affirme que le rapport des probabilités de deux choix distincts ne dépend pas des autres choix offerts. En rappelant la probabilité 2.8 on voit en effet que le rapport des probabilités associées à deux choix  $j_1$  et  $j_2$  est

$$\frac{P(j_1 | x)}{P(j_2 | x)} = \frac{\exp[\bar{V}_{j_1}(x)]}{\exp[\bar{V}_{j_2}(x)]} \quad (2.33)$$

Retournons à notre modèle logit emboîté et observons le rapport des probabilités associées à deux choix  $j_1$  et  $j_2$  appartenant au même groupe  $l$ . Nous obtenons

$$\frac{P(j_1 | x)}{P(j_2 | x)} = \frac{\exp(\bar{V}_{j_1}/\lambda_l)}{\exp(\bar{V}_{j_2}/\lambda_l)} \quad (2.34)$$

Le modèle de choix dans un groupe  $l$  est donc un logit polytomique. Supposons maintenant  $\lambda_l = 1 \quad \forall l$ . La probabilité 2.31 s'écrit alors

$$P(j | x) = \frac{\exp(\bar{V}_j)}{\sum_{l=1}^L \sum_{j=1}^{J_l} \exp(\bar{V}_{j(l)})} = \frac{\exp \bar{V}_j(x)}{\sum_{h=1}^J \exp \bar{V}_h(x)} \quad (2.35)$$

Le modèle logit polytomique est donc un cas particulier du logit emboîté.

Comme dit plus haut, je suppose que les décisions d'achat automobile sont issues d'un programme dynamique de planification de long terme. En effet, je pense qu'effectuer un achat pourrait constituer pour le ménage un point de rupture dans l'habitude. Je suggère alors de représenter l'utilité de choisir un parc par l'arbre présenté en figure 2.2.

Le niveau supérieur capture les effets propres à l'acte d'acheter. Les utilités de possession et utilisation attachées à chaque composition possible, définie comme l'intersection de l'ensemble des parcs que l'agent peut hériter (i.e. l'ensemble de décisions qu'il peut prendre pour les véhicules éventuellement possédés à la période précédente) et l'ensemble de types de véhicules que l'agent peut acheter sont représentés aux niveaux inférieurs. L'abandon d'un véhicule impliquant aussi un changement dans la routine de l'agent, je suggère que l'effet de cette rupture devrait aussi être isolé au deuxième niveau de ce modèle. L'utilité découlant de chaque composition possible peut être définie comme fonction des utilités estimées à l'étape précédente pour chaque véhicule inclus dans le parc considéré.

Le problème inter-temporel du programme de planification est donc défini comme un modèle logit emboîté à trois niveaux. Les probabilités de possession et utilisation conditionnées à la composition du reste du parc ayant été obtenues à l'étape précédente, ce modèle est estimable à horizon infini.

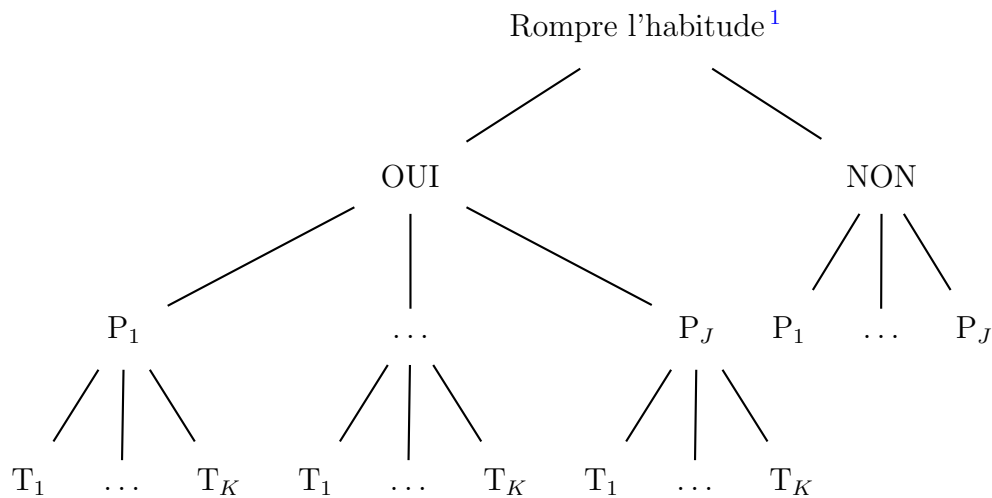


FIGURE 2.1 – Choix de Configuration du Parc

Une spécification simplifiée pour ce modèle, visant à prouver l'importance de considérer l'effet lié au fait même d'acheter est présentée de façon détaillée au chapitre 7 de la deuxième partie de ce travail.

## 2.3 Le choix optimal

Pour conclure cette première partie je vais maintenant écrire le programme complet jusqu'ici décrit.

La détermination de la “configuration de vie” optimale pour l’année, qui a été présentée au chapitre précédent, définit les besoins automobiles annuels de l’agent. Son objectif inter-temporel est donc de choisir le parc lui apportant le volume de satisfaction et de services qui s’approchent le plus, respectivement, de la demande d’utilité  $D_u$  et la demande de kilométrage  $D_{km}$  qui minimisent sa fonction de dépense annuelle.

Soit  $i$  un agent confronté aux choix d’équipement automobile pour l’année  $t$ . L’utilité indirecte de choisir un des parc automobiles  $p_{it}$  composant l’ensemble de choix disponibles  $\mathbb{P}_t^i$ , défini en fonction de l’état d’équipement choisi la période précédant  $p_{i,t-1}$  peut être écrite comme

$$\bar{V}_p = -\beta_u \|u_p - D_u\| - \beta_{km} \|km_p - D_{km}\| \quad (2.36)$$

Les paramètres  $\beta = \{\beta_u, \beta_{km}\}$ , représentant respectivement l’effet de la satisfaction produite et des services fournis par le parc sur le choix de composition finale, peuvent être estimés par la méthode du maximum de vraisemblance, où la probabilité attachée à chaque choix est définie comme un logit multinomial :

$$P(\bar{p} | x) = \frac{\exp \bar{V}_{\bar{p}}}{\sum_{p \in \mathbb{P}} \exp \bar{V}_p} \quad (2.37)$$

La satisfaction produite  $u_p$  et les services fournis  $km_p$  sont le résultat du programme dynamique décrit au cours de la section précédente. Un tel modèle est estimable par la méthode du maximum de vraisemblance en imposant le cadre proposé par [Rust \(1987\)](#). Sous ces conditions il prouve que l’on obtient des estimations consistantes des paramètres  $\theta$ , modélisant la probabilité non contrainte de choisir le parc  $\bar{p}$ , même en ne maximisant que la fonction de vraisemblance partielle due à la probabilité conditionnelle de choix.

$$P(\bar{p} | x_t, \theta) = \frac{\exp \{v(\bar{p}, x_t)\}}{\sum_p \exp \{v(p, x_t)\}} \quad (2.38)$$

où

$$\begin{cases} v(p, x_t) = U(p, x_t) + \gamma \int V(x_{t+1}) dF(x_t + 1 | p, x_t) \\ U(p, x_t) = \left( \lambda_a v(a, x_t) + \frac{\lambda_a}{\mu_a} \log \sum_p \exp(\mu_a v_{|a}(p, x_t)) \right) + \epsilon \end{cases} \quad (2.39)$$

$v(a, x_t)$  est la part de satisfaction due à l'effet de l'habitude et  $v_{|a}(p, x_t) = f(u, \psi)$ <sup>2</sup> où  $u$  est le vecteur des utilités, résultantes du programme dynamique décrit au cours de la première section de ce chapitre, attachés à chaque véhicule  $j$  composant le parc  $p$  et  $\psi$  sont des poids tels que  $\sum_j \psi_j = 1$ . En raison de la forme supposée pour ce problème, on peut écrire cette probabilité comme  $P(\bar{p} | x_t, \theta) = P(\bar{a} | x_t, \theta)P(\bar{p} | a, x_t, \theta)$  et le problème est estimable de façon séquentielle.

---

2. Je ne propose pas de spécification pour  $f(u, \psi)$  ici mais une idée pourrait être développée en partant des travaux de [Bhat and Sen \(2006\)](#).



## Deuxième partie

### Les données

#### Une Contribution à la Base Parc Auto





# Chapitre 3

## Présentation des Données

L'objectif de cette deuxième partie est de fournir une description détaillée des données utilisées pour les applications qui seront décrites dans la troisième partie du document. Ces applications sont dérivées de la problématique présentée au cours de la première partie. La présentation qui suit a aussi pour but d'expliquer une partie des limites qui seront rencontrées au cours de cette étude.

Ce travail a pour objectif d'étudier les dynamiques des comportements automobile : Possession et utilisation des véhicules et Renouvellement du parc automobiles<sup>1</sup>.

Étant donné le caractère durable de ce type de bien, l'agent est ici supposé avoir un comportement "*tourné vers le futur*" vis-à-vis des décisions concernant son parc automobile. Cela signifie que l'on suppose l'agent capable d'anticiper les évolutions futures de son environnement. Sur la base de ces prévisions, il peut alors évaluer les conséquences à long terme de ses décisions présentes.

Il faut alors choisir, pour l'application, une période qui permet l'hypothèse d'"évolutions prévisibles" de l'environnement, ici représentée par les prix du carburant.

Ce choix est commenté en conclusion de chapitre.

---

1. Le lecteur peut se référer à Part III pour une explication détaillée.

## 3.1 Les Bases des Données

La source de données principale est le panel **Parc Auto TNS-SOFRES**, le meilleur outil disponible lorsque l'on souhaite étudier la dynamique des comportements individuels relatifs à l'équipement et à l'usage de l'automobile en France. L'objectif d'un tel dispositif est de fournir une source importante d'informations sur le volume et la structure du marché (neuf et occasion) et du parc automobile français. Les comportements d'usage (kilométrages et consommations), les attitudes et les opinions des ménages sont aussi capturés par l'enquête.

Cette enquête permet donc de :

- **Décrire la structure du parc automobile des ménages** (âge, état à l'achat, durée de détention, modèles, gammes, etc)
- **Étudier le profil des automobilistes** (profil sociodémographique des utilisateurs principaux et occasionnels)
- **Analyser les comportements d'utilisation** (kilométrage annuel parcouru, fréquence et raisons d'utilisation...)
- **Étudier les attitudes à l'égard de l'automobile**
- **Suivre les tendances lourdes et identifier les évolutions récentes** (taux de motorisation, part des ménages mono-bi-tri motorisés, impact des nouvelles immatriculations sur la taille et la structure du parc)
- **Suivre les déformations structurelles** liées à l'évolution des comportements et aux changements de mains successifs des véhicules
- **Suivre des questions d'actualité**

Pour autant, les informations disponibles sont insuffisantes. La description des caractéristiques des véhicules est à compléter par d'autres variables. À ce titre est utilisée une **Base des Données Automobile** complémentaire développée par Philippe [Boursin](#) et contenant une description fine des caractéristiques d'environ 7200 types de véhicule, comme par exemple le poids ou la puissance maximale du moteur.

La description et l'appariement de cette base de données au panel Parc Auto sont présentés ci-après.

Enfin, les prix moyens annuels du carburant en France, contenus dans la base **Prix DIREM**, sont inclus aux informations du panel.

### 3.1.1 Parc Auto

Il s'agit d'une enquête annuelle auto-administrée conçue par l'institut d'études marketing et d'opinion international (TNS SOFRES) depuis 1983,

avec le conseil scientifique de l’Institut français des sciences et technologies des transports, de l’aménagement et des réseaux (IFSTTAR, ex INRETS).

L’échantillon observé repose sur un panel de ménages résidant en France. Il est constitué de 10000 ménages représentatifs de la France entière. Le taux de réponse annuel se situe entre 65% et 70%. Chaque vague est pondérée après réception des questionnaires afin de conserver la représentativité de l’échantillon. Il est renouvelé au taux d’environ 1/3 par an afin de gérer au mieux les problèmes d’attrition. Pour autant, il est tout à fait possible d’observer certains ménages sur une période de plus de trois ans, possibilité dont on profitera ici.

L’IFSTTAR se charge de traiter les fichiers et conserve, sur un système Unix, les tables des vagues annuelles seulement depuis 1993. Les fichiers des données précédentes ne sont donc pas disponibles.

La table 3.1.1 reporte le nombre de ménages répondant au cours de la période 1993 – 2008.

TABLE 3.1 – Nombre de Ménages Répondants

<b>1993</b>	<b>1994</b>	<b>1995</b>	<b>1996</b>	<b>1997</b>	<b>1998</b>	<b>1999</b>	<b>2000</b>
7120	7577	7050	7093	6762	6582	6872	6368
<b>2001</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>
6438	6466	6523	6291	6257	5600	6846	7270

Chaque début d’année, les 10000 ménages composant l’échantillon annuel sont invités à répondre par voie postale à des questions portant sur leur parc automobile « à disposition » (i.e. Véhicules dont le ménage dispose au cours de l’année. Ces véhicules constituent son parc et l’agent peut s’en servir, qu’il en soit le propriétaire ou pas. Cependant par la suite lorsque je parlerai de “véhicule possédé” cela reviendra à dire “véhicule à disposition”) et son usage au cours des 12 mois écoulés de l’année civile précédente.

Des questions visant à collecter les informations démographiques et socio-économiques susceptibles de changement annuel, comme le revenu du ménage, sont incluses dans le questionnaire annuel.

Depuis 2001 certaines questions portant sur l’utilisation des transports en

commun ont aussi été jointes au questionnaire.

À noter qu’au moment de l’entrée d’un ménage dans le panel certaines informations concernant sa structure démographique et socio-économique sont collectées. Ces informations sont mises à jour lorsque nécessaire au cours des années suivantes. Les ménages s’engagent à communiquer tout changement de situation ayant lieu au cours de la période d’observation. Des “ravalements” sont effectués par TNS-SOFRES afin de garantir des données d’une ancienneté inférieure ou égale à 6 mois.

La structure de la base de données est hiérarchique. Elle est composée d’une table “ménages” et d’une table “véhicules”.

La première contient des informations concernant le ménage et ses membres<sup>2</sup>, telles que le niveau socio-économique, les caractéristiques démographiques, la localisation résidentielle, l’équipement automobile et la motorisation.

Une description plus détaillée des véhicules possédés par le ménage, comprenant pour chacun ses caractéristiques techniques, ses paramètres d’usage et son utilisateur principal, est conservée dans une deuxième table.

La totalité de la flotte automobile est décrite lorsque celle-ci est caractérisée par une taille inférieure ou égale à 3. Le cas échéant, seulement trois véhicules, choisis par le ménage, sont décrits. Les enquêtés sont en outre invités à hiérarchiser les véhicules décrits, en les ordonnant par degré d’importance.

La quantité d’information est considérable. Le lecteur est invité à visiter ma page personnelle ([Cernicchiaro](#)) si il souhaite visionner le questionnaire pour l’année 2008.

Bien entendu, ce questionnaire a fait l’objet de modifications progressives depuis le début de cette enquête<sup>3</sup>. D’une part le champ couvert par l’information collectée s’est élargi progressivement, en incluant notamment une description succincte de la demande d’utilisation en transport en commun au niveau ménage. De l’autre, la définition choisie pour coder certaines variables collectant l’information au niveau véhicule, a connu des changements au fil du temps.

Bien que la présence d’une clé unique identifiant chaque ménage rend simplement réalisable la construction d’une table longitudinale à ce niveau,

---

2. Depuis 2004 ces dernières sont récoltées dans une troisième table “individus”.

3. Pour une description détaillée de ces changements le lecteur peut visiter la page personnelle de [Cernicchiaro](#)

des changements touchant le questionnaire découle le besoin d'un traitement préliminaire visant à harmoniser l'information collectée au cours du temps. De plus, si les évolutions connues par les prix des carburants jouent un rôle dans le choix de la période d'observation à utiliser pour l'application, comme annoncé en introduction et discuté par la suite, le besoin d'homogénéité dans l'information exploitée ne permet pas d'ignorer les changements de définition lorsque cette période est définie. L'objectif est donc, en d'autres termes, de choisir une période dont un volume d'informations couvrant un même champ, suffisamment étendu, soit collecté chaque année (autre que satisfaire les besoins liés à l'utilisation des prix du carburant).

À la lumière de ces observations, deux procédures de préparation des données sont aussi à construire. Je conclus donc la présentation de cette base (pilier de ce travail) par une analyse visant à signaler les informations importantes, concernant le volume d'informations exploitables, qui ne peuvent pas, à mon sens, être ignorées en effectuant le choix de la période. La procédure d'harmonisation des définitions pour ce qui concerne les déclarations au niveau véhicule sera présentée en introduction du chapitre 4.

### Traitement Préliminaire

Comme annoncé précédemment, les informations sont collectées par vagues annuelles et enregistrées dans des tables différentes. Lorsque l'on veut, comme dans ce cas, étudier la dynamique du ménage, plusieurs manipulations sont alors à effectuer sur les données disponibles.

En effet, comme annoncé en introduction de ce chapitre, l'étude ici présentée vise à s'interroger sur la pertinence de l'hypothèse d'agent "tourné vers le futur". Or lorsque l'on souhaite, comme ce sera le cas dans deux des applications présentées dans la dernière partie (Chapitres 7 et 8), estimer ce type de problème à horizon fini il est important de pouvoir utiliser des ménages observés continuellement au cours d'une période suffisamment longue<sup>4</sup>. De plus, comme on le verra au cours du prochain chapitre, le traitement de la base Parc Auto en qualité de Panel de Véhicule implique de devoir renoncer à une partie des observations en raison des incohérences rencontrées au cours de l'analyse longitudinale à ce niveau. La construction de la procédure présentée au prochain chapitre vise à réduire le plus possible cette perte, sans

---

4. Le lecteur peut se référer à la première partie de ce document pour une explication plus approfondie.

pourtant arriver à l'éliminer<sup>5</sup>.

Je vais donc fournir ici une rapide analyse de l'évolution de l'information collectée entre 1999 et 2008. Le choix de se réduire à cette sous-période, sur celle disponible (1993 – 2008) est fait principalement en raison du grand nombre de variables à traiter. En effet, bien qu'une documentation assez complète est fournie par L. Hivert, celle-ci n'est pas simple à consulter. De plus, il n'existe pas à ma connaissance parmi ces documents un dictionnaire complet pour les variables collectées au cours du temps. Les données d'après 2008, bien que disponibles, ne sont pas considérées dans ce travail. En premier lieu à cette date le prix du gazole a rejoint celui de l'essence (section 3.2). De plus les dernières années sont visiblement marquées de façon de plus en plus importante par les changements comportementaux, économiques et démographiques que la littérature de la mobilité met en lumière depuis 2005<sup>6</sup>.

L'analyse qui suit est alors faite sur la base d'un dictionnaire construit en joignant les différents documents fournis par L. Hivert, qui concernent la construction des variables incluses dans les différentes tables annuelles. L'information ainsi reconstituée peut alors être complétée, détaillée et rectifiée (si besoin) à l'aide des méthodes classiques de traitement et analyse des bases de données (aussi bien à l'aide des outils fournis par SAS que du langage SQL)<sup>7</sup>.

Sont reportés en table 3.1 les volumes d'informations (i.e. le nombre de variables rencontrées dans au moins une table annuelle<sup>8</sup>) par type de destin connu, i.e. :

- Variable observée tout au long de la période 1999 – 2008 (< – >)
- Variable exclue avant 2005 (– > |)

---

5. Plusieurs évolutions possibles seront traitées en conclusion.

6. De plus, ma thèse ayant débuté en fin 2007, ces vagues n'étaient pas disponibles au moment où j'ai effectué le gros du travail de construction de ce dictionnaire. Le volume de travail que cela aurait comporté m'a portée à choisir, lorsque je terminais ma deuxième année de thèse, de ne pas traiter la nouvelle information disponible en raison du temps disponible. Cependant, toutes les procédures SAS décrites au cours de ce chapitre sont extensibles à d'autres périodes (Plus d'informations sur la page personnelle de [Cernicchiaro](#))

7. Plus d'informations sur ce Dictionnaire et sa Construction sur la page personnelle de [Cernicchiaro](#)

8. Cela signifie que lorsque deux variables collectent la même information pour deux membres distincts d'un ménage, l'information ici est comptée double.

- Variable incluse à un moment  $A$  compris entre 2000 et 2004 ( $A- >$ )
- Variable incluse après 2004<sup>9</sup> ( $| - >$ )
- Variable incluse de façon **Irregulière**

Un important tronc commun d’informations (i.e. 38,8% des 423 variables observées au moins une fois dans les tables annuelles) est remarquable. Le type d’information collecté, comme précédemment dit dans cette présentation, concerne aussi bien les caractéristiques des ménages (démographiques, socio-économiques), de leur environnement (location géographique) et de leur “comportement automobile” (utilisation et d’équipement).

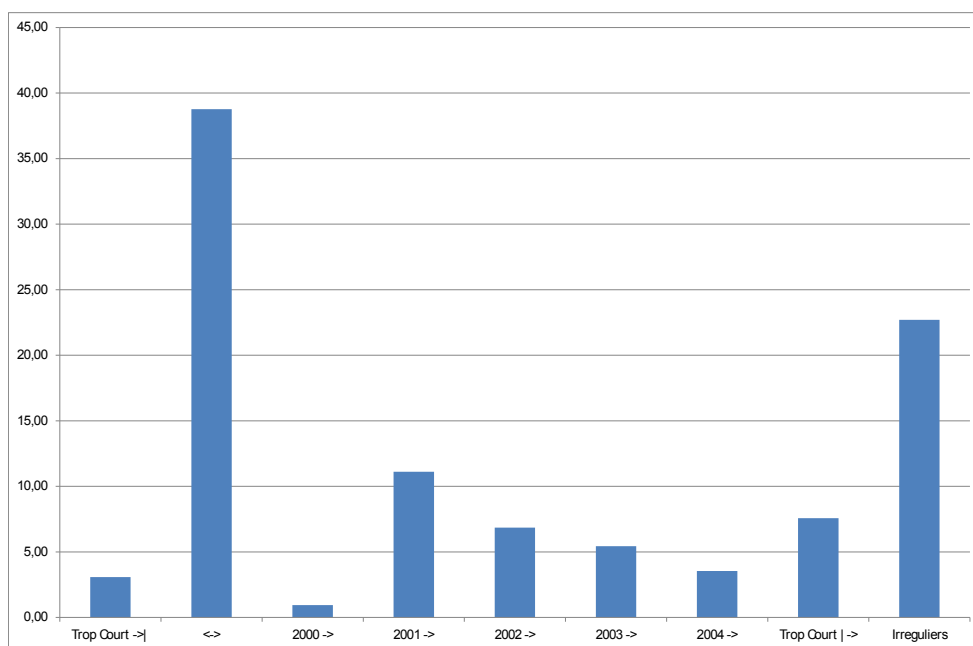


FIGURE 3.1 – Évolution de l’Information Collectée – Parc Auto

En 2001 un nouveau volet, concernant les comportements d’utilisation des transports en commun a été ajouté à l’enquête. Il s’agit de l’innovation la plus marquante dans la structure de l’enquête, suivie par l’introduction d’un complément à l’information concernant les membres adultes et l’utilisa-

---

9. notamment en 2007



tion automobile en 2007 (11,1% et 7,6% de l'information disponible totale respectivement).

Une grande partie de l'information, et notamment celle concernant les opinions, les préférences et les habitudes des ménages, n'est pas collectée de façon régulière. 22,7% des variables sont disponibles pour moins de 5 vagues. Cette information n'est donc pas exploitable en vue de l'objectif fixé dans ce travail.

### 3.1.2 DIREM

Il s'agit d'une base de données des prix pétroliers, gérée par la Direction Générale de l'Energie et du Climat (DGEC du ministère de l'Écologie du Développement Durable et de l'Énergie), qui couvre la France métropolitaine (hors Corse).

Chaque lundi, la DGEC reçoit les prix moyens hebdomadaires nationaux au détail à la pompe de chaque distributeur (compagnie pétrolière, centrale d'achat des grandes et moyennes surfaces, indépendants) qui commercialise auprès des consommateurs en métropole des carburants, du fioul domestique et du fioul lourd, pondérés par les volumes sur l'ensemble de son réseau de distribution. Il s'agit des prix réellement acquittés par le consommateur final. Les prix sont exprimés en euros courants par hectolitre ou en euros courants par tonne pour les fiouls lourds.

Chaque opérateur déclare un prix moyen national au détail pour chaque produit qu'il distribue. Pour chaque produit, la DGEC calcule un prix moyen national pondéré (par les parts de marché de chaque opérateur déclarant) qui est publié chaque lundi. Les prix mensuels, semestriels et annuels sont ensuite calculés par moyenne arithmétique.

La DGEC transmet ensuite ces prix aux services de la Commission européenne, dans le cadre de la mise en oeuvre des dispositions du Conseil européen 1999/280/CE du 22 avril 1999 et de la Commission européenne 1999/566/CE du 26 juillet 1999 concernant une procédure communautaire d'information et de consultation sur les coûts d'approvisionnement en pétrole brut et les prix à la consommation des produits pétroliers. Les Etats Membres doivent communiquer, au plus tard le mardi à midi, aux services de la Commission européenne les prix moyens hebdomadaires nationaux au détail des produits pétroliers.

TABLE 3.2 – Prix DIREM TTC

Date	SuperCarb.	ARS	Gazole	Super SP95	Super SP98	GPLc
1993	82,92	—	55,97	77,87	79,72	—
1994	85,54	—	58,81	80,30	81,90	—
1995	89,11	—	58,69	85,70	86,31	—
1996	94,69	—	65,33	90,91	91,71	—
1997	98,01	—	67,64	94,11	95,03	43,53
1998	95,80	—	64,24	91,86	92,78	42,01
1999	100,25	—	69,00	95,50	96,36	43,24
2000	—	116,89	84,68	109,17	110,93	52,38
2001	—	111,86	79,60	103,29	105,70	50,70
2002	—	109,52	77,24	101,46	103,65	49,71
2003	—	109,60	79,35	101,63	103,66	53,76
2004	—	114,10	88,47	106,03	108,27	56,61
2005	—	127,46	102,68	116,59	120,53	64,33
2006	—	135,40	107,75	123,68	127,43	71,05
2007	—	—	109,49	127,65	130,86	71,15
2008	—	—	126,71	135,38	139,29	76,27

Les prix reportés dans la base concernent plusieurs produits. Pour ce qui concerne les carburants, les produits suivants sont déclarés par les opérateurs :

- Supercarburant : L'essence plombée a été interdite à la vente le 1<sup>er</sup> janvier 2000 par la directive relative à la qualité des carburants, adoptée le 29 juin 1999 par le Conseil et le Parlement Européens. Pour pallier à la disparition du marché de ce type de carburant, encore utilisé par certains véhicules (les motos notamment), le carburant Anti Récession de Soupapes (ARS) a été introduit à sa place et est resté sur le marché jusqu'en 2006.
- Supercarburant sans plomb 95 et 98 : une directive européenne de mars 1985 a imposé l'introduction du supercarburant sans plomb en 1990. En accord avec cette directive, les véhicules produits à partir de cette date ont été conçus pour fonctionner avec de l'essence sans plomb. La

mise en circulation de véhicules neufs à essence ne pouvant fonctionner qu'avec un carburant plombé a été ensuite interdite à compter du 1er octobre 1990 par une seconde directive du 3 décembre 1987.

- Gazole
- Gaz de Pétrole Liquéfié (GPL) : utilisé depuis longtemps comme combustible, ce carburant a fait son entrée sur le marché automobile le 3 janvier 1997.

Les moyennes annuelles des prix du carburant, de 1993 à 2008, sont reportées par type en tab. 3.2.

### 3.1.3 Base des Données Automobile

Cette dernière table fournit une description détaillée de 7200 types de véhicules automobiles. Elle a été construite par Philippe Boursin, et gracieusement mise à disposition (Données et informations sur la page personnelle de P. [Boursin](#)). La liste complète des informations décrivant chaque véhicule est reportée en table 3.3.

De ces variables sont retenues celles relatives aux dimensions, au poids, à la description des performances du moteur et sa consommation. Elles seront utiles pour l'application présentée au cours du chapitre 6.

Les données sont fournies en format Excell et les récurrences (i.e. les types de véhicules) décrites sont (univoquement) identifiables par combinaison de leur marque, modèle et type. Comme on le verra par la suite (au cours de la section 4.2.2), ces informations doivent être appariées avec la population des véhicules observés au cours de l'enquête Parc Auto. Or, pour ces véhicules seulement la marque et le modèle sont collectés.

Le fichier Excell est alors à mettre en forme avant exploitation. Cette opération est réalisée par application successive de 12 macros, écrites en langage Visual Basic, que je vais présenter ici de façon synthétique<sup>10</sup>.

---

10. Pour plus d'informations sur le fichier Excell prenant en charge les macros construit à cette fin, le lecteur peut visiter la page personnelle de [Cernicchiaro](#)

TABLE 3.3 – Caractéristiques Vehicules

<b>Moteur</b>	Type Distribution Cylindrée (cm <sup>3</sup> ) Nb. Cylindres Alésage (mm) Course (mm) Rapport Alésage/Course Régime (tr/mn) Puissance (ch, kW, tr/mn) Psp (ch/l) P utile (kW) Pression ME (bar) Rapport Volumétrique Rempl Couple (mkg, Nm, tr/mn) Elast	<b>Transmission</b>	Type Boite de Vitesses étagement
		<b>Châssis</b>	Type
		<b>Aérodynamique</b>	Coeff. de Traînée ( $C_x$ ) Surface de Traînée ( $SC_x$ )
		<b>Suspension</b>	Avant Arrière
		<b>Direction</b>	Type Assistance Trs Volant d braquage (m)
		<b>Freins</b>	Avant Arrière Assistances
		<b>Pneumatiques</b>	
<b>Dimensions</b>	Alimentation Empattement (mm) Voie Avant (mm) Voie Arrière (mm) Longueur (mm) Largeur (mm) Hauteur (mm)	<b>Consommations</b>	l/100km ACEA Rdt
		Urbaine	mesurée écart Wh/km
		Ville	ACEA Rdt
<b>Réservoir</b>	Carburant (dm <sup>3</sup> ) Coffre (dm <sup>3</sup> )	Extra-Urb.	mesurée écart Wh/km
<b>Poids</b>	Total (kg) Pds/Pu (kg/ch) Poids av (kg) CU (kg) Remorq (kg)	Route	ACEA Rdt
<b>Vitesse Maxi</b>		Autoroute	mesurée écart Wh/km
<b>Résistances</b>	Roulmt (ch) Air (ch) Rdt	Mixte (km/h) g CO <sub>2</sub> Moyenne	ACEA Rdt /100 km mesurée écart Wh/km
<b>Accel lat.</b>	r. 61 m (m/s <sup>2</sup> )	SAE J1364	/100 km
<b>Départ Arrête</b>	100 km/h ( s ) 200 km/h ( s ) 400 m ( s ) 1000 m ( s )	CEE	l/100 km
		<b>Autonomie</b>	km
		<b>Type de Carrosserie</b>	

### Traitement Préliminaire

Plusieurs opérations sont à effectuer si l'on veut, comme c'est le cas dans ce travail, utiliser l'information apportée par cette base pour compléter celle fournie par le panel des véhicules observés dans Parc Auto.

L'objectif de cette phase préliminaire de traitement est alors de codifier les récurrences de la table Excell selon la définition utilisée dans l'enquête (variable `marmod`). Un fichier Excell détaillant cette codification est mis à disposition chaque année par L. Hivert.

Une analyse minutieuse des fichiers Excell, par observation aussi bien directe qu'automatisée, a permis de rendre comparables les écritures des marques et modèles en clair. Une série de corrections systématiques est alors appliquée aux étiquettes (i.e. information en clair) utilisées dans les deux, dans le but de les homogénéiser. L'application aux 2 fichiers d'un certain nombre de procédures de comparaison, qui ne seront pas détaillées ici, permet alors la construction d'une table de descriptions techniques pour les "types de véhicules `marmod`".

Puisque comme dit plus haut, l'information est détaillée par type dans la base de données automobile mais pas dans la codification Parc Auto, plusieurs descriptions peuvent correspondre à un même code `marmod`. De plus, certains de ces codes sont définis de façon large (ex. ClioI/II/III). Pour conclure le traitement préliminaire de cette table, un indicateur de fiabilité de la codification est construit. Comme on le verra par la suite, un tel indicateur sert aussi de support à la sélection des informations les plus cohérentes pour les véhicules (chapitre 4).

## 3.2 Analyse des Prix du Carburant

Comme on l'a vu, la base Parc Auto fournit une description représentative des ménages français depuis 1996<sup>11</sup>. On va donc analyser les évolutions des prix du carburant sur cette période, afin de sélectionner le sous-échantillon le plus approprié.

Les véhicules utilisant un carburant autre que Essence ou Diesel sont très peu représentés dans la base utilisée (le lecteur est renvoyé à la section 5.3.2

---

11. Entre 1993 et 1996 le prix d'achat des véhicules n'étant pas renseigné, la description fournie pour ces vagues n'est pas exploitable pour cette application.

pour plus d'informations.). Ce sont donc seulement ces deux types de carburants qui seront considérés dans ce travail.

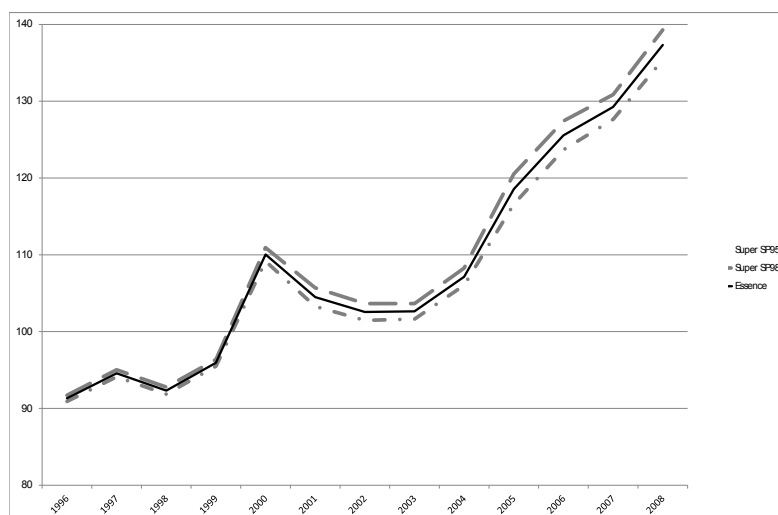


FIGURE 3.2 – évolutions : Essence

Comme expliqué en introduction, il existe deux types de supercarburant sans plomb : Essence SP95 et SP98.

La fig. 3.2 montre que l'évolution des prix pour ces deux types de carburant est très proche. Le prix de l'essence est donc défini ici comme la moyenne annuelle de ces deux prix.

Les évolutions des deux types de carburant de 1996 à 2008 sont reportées en fig. 3.3 et tab. 3.4. Il ressort que les deux prix ont connu une évolution similaire au cours de cette période, bien que légèrement plus rapide pour le Diesel.

Cela peut être expliqué par une revue synthétique des politiques de taxation du carburant prises en France par les gouvernements successifs depuis 1990.

En effet, un taux de taxe réduit d'environ 30% sur le prix du gasole, par rapport à celui du super sans plomb 95, a toujours été adopté dans le but de maintenir le prix du premier plus bas. Ce taux a peu varié au fil du temps, en passant de 25% à 35%. Les prix à la production (Hors Taxe) de ces deux carburants étaient sensiblement les mêmes. Cependant, la demande mondiale pour le super sans plomb 95 devenant de plus en plus forte, le ga-

TABLE 3.4 – Prix du Carburant

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Essence	0,95	0,92	0,96	1,10	1,04	1,03	1,03	1,07	1,19	1,26	1,29	1,37
$\Delta$	0,03	-0,02	0,04	0,14	-0,06	-0,02	0,00	0,05	0,11	0,07	0,04	0,08
Diesel	0,68	0,64	0,69	0,85	0,80	0,77	0,79	0,88	1,03	1,08	1,09	1,27
$\Delta$	0,02	-0,03	0,05	0,16	-0,05	-0,02	0,02	0,09	0,14	0,05	0,02	0,17

zole a récemment connu une progressive augmentation de prix (le premier étant par conséquent devenu moins cher à la production que ce dernier) en rapprochant progressivement les prix réels de marché (Taxes Comprises).

Le gasole connaît donc des taux d'augmentation annuelle qui sont, la plupart du temps, plus importants que ceux observés pour l'essence. Cependant le prix annuel du Diesel n'a jamais dépassé celui de l'essence au cours de la période considérée.

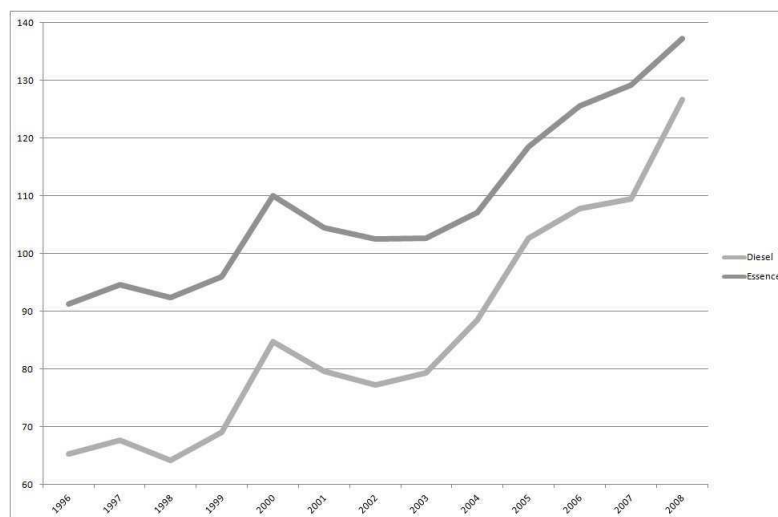


FIGURE 3.3 – évolutions des Carburants

Le Diesel n'a rattrapé l'Essence qu'au cours de l'année 2008. Au premier semestre 2008, le sans plomb 95 n'est que 7% plus cher que le gazole, alors qu'en 1990, il était encore 47% plus cher. Ce rapprochement pourrait alors

expliquer la baisse qui semble toucher dernièrement l'attractivité des motorisations diesel. Malgré cette baisse, cette demande reste majoritaire par rapport aux autres types de moteurs, son rendement restant toujours meilleur que celui des moteurs à essence (Comme discuté aux chapitres 7 et 8).

### 3.3 Choix de la Période.

La période qui semble être la plus adaptée, en accord avec l'objectif annoncé, est celle qui va de 2002 à 2008. Plusieurs raisons motivent ce choix, comme on a pu le voir au cours de ce chapitre.

Avant tout, les évolutions de ces prix ont connu une augmentation continue qui permet de supposer une dynamique stationnaire pour ces prix. De plus, on a observé avant 2002 plusieurs changements de tendance qui empêchent de faire une hypothèse simple pour les dynamiques que ces prix suivent.

Une autre raison, discutée en 3.1.1, découle de l'information disponible. Si l'on souhaite utiliser une période suffisamment longue et couvrant un champ d'information satisfaisant il semble optimal de faire débuter l'observation après 2000. Cela permet en effet d'inclure dans l'information exploitable les variables d'utilisation des transports en commun<sup>12</sup>. L'importance de ces variables dans l'étude de la mobilité automobile, comme on le verra au cours de la troisième partie, m'a motivée à inclure leur préparation dans la procédure construite (présentée en introduction du chapitre 5) en vue d'évolutions futures du travail effectué au cours de ma thèse.

Le choix d'arrêter le traitement de la base en 2008 est pris par simplicité. Les changements de tendances connues en France depuis cette date, aussi bien par les évolutions des prix que par celles des comportements de mobilité de la population, rendent peu attractive l'utilisation de cette dernière période pour l'application. En effet, cela demanderait de pouvoir estimer des transitions pour des tendances dont le traitement est délicat. Le temps à ma disposition n'étant pas suffisant pour une telle opération, j'ai préféré ne pas me consacrer au traitement de ces vagues.

---

12. Cette information ne sera finalement pas utilisée dans l'application en raison du temps et des ressources qui auraient du être employées pour en estimer les transitions (ex. données territoriales concernant l'offre en transport en commun).



Pour conclure, deux raisons supplémentaires découlent du traitement des données.

En premier lieu, la procédure d'appariement présentée au cours du chapitre qui suit (4) permet d'étudier les "cycles de vie"<sup>13</sup> des véhicules déclarés par les vagues successives de l'enquête Parc Auto. Bien que cette procédure soit construite de façon à réduire la perte d'information qui découle du souhait d'observer des parcs complets (i.e. des parcs tels que les véhicules qui les composent sont identifiés tout au long de l'observation et leur description est cohérente), cette perte n'est toutefois pas inexistante. La taille de l'échantillon exploitable, après traitement des informations collectées, est satisfaisante en utilisant la période 2002 – 2008.

Si les ménages sont identifiables de façon univoque, cela n'est pas de même pour chaque membre. En effet, ces récurrences sont identifiées par un rang déclaré (comme c'est le cas pour les véhicules, 4.2). Les changements dans la composition sont analysés à l'aide d'une procédure SAS, qui sera présentée par la suite (5.1).

En appliquant la procédure deux fois, une à partir de 1999 et une de 2002<sup>14</sup>, il apparaît que le choix d'utiliser une période plus longue augmente la probabilité de rencontrer des ménages connaissant un "changement majeur de structure" (changement du chef, sortie d'enfants, naissances, entrées/sorties d'adultes, etc.). Il est vraisemblable qu'un événement tel que une mise en couple, une séparation ou un décès puisse impliquer une réorganisation de la vie du ménage et donc par exemple un changement dans le regard qu'il porte sur son parc automobile. Il est alors préférable de choisir une période qui permet de réduire la part de ménages concernés par ces changements. De plus, une période plus longue augmente la probabilité de se confronter au cas de ménages ne déclarant pas le revenu annuel au moins une fois.

Pour toutes ces raisons, la période ici utilisée est celle couvrant les vagues observées entre 2002 et 2008. Comme dit plus haut, les observations résumées ici pour ce qui concerne la taille de l'échantillon retenu seront analysées plus dans le détail au cours des prochains chapitres de cette partie.

---

13. La part de ce travail explique plus dans le détail ce que je définis comme le cycle de vie d'un véhicule : l'utilisation qu'en ménage en fait pendant toute la période de possession.

14. Seulement cette deuxième application sera commentée au chapitre 5. Pour plus d'informations le lecteur peut visiter la page personnelle de [Cernicchiaro](#).

# Chapitre 4

## La Procédure d'Appariement des Véhicules

Au cours du chapitre précédent j'ai présenté les raisons qui me poussent à créer une "procédure d'appariement", visant à permettre le traitement du panel au niveau véhicule.

Au cours de ce chapitre je vais présenter cette procédure. Un volume important de travail préliminaire a été réalisé afin de la définir<sup>1</sup>.

J'ai en premier lieu procédé à une analyse directe des tables, en me servant des techniques mixtes automatiques/manuelles<sup>2</sup> utilisées par exemple pour l'identification des erreurs de déclaration et saisie dans le cadre de l'Enquête Nationale Déplacement Transport (ENTD IFSTTAR). En raison du volume d'informations à traiter, j'ai restreint l'analyse aux ménages "mono-équipés" au cours de toute l'observation. Ce cas est le plus simple à traiter. Le ménage et le véhicule forment un individu unique et d'une année sur l'autre le véhicule peut être, soit le même que l'année précédente, soit un nouvel achat, soit être introduit par un nouveau membre (ou hérité, reçu en cadeau, mis à disposition par l'employeur, ...).

L'information concernant le type d'entrée est fournie par une variable collectant l'année d'achat, une deuxième pour le type d'entrée<sup>3</sup> et une indi-

---

1. Ce travail préliminaire est ici décrit de façon succincte. Seulement la procédure finale sera détaillée dans ce chapitre. Plus d'informations sur la page personnelle de [Cernicchiaro](#).

2. Observation directe des récurrences composant les tables, par sélection automatisée.

3. Cette variable n'est pas utilisée dans la procédure d'appariement qui sera présentée en 4.2. Bien que son utilisation pourrait apporter plus de précision, la marginalité du cas et le fait que cela ne donne pas d'indications sur le moment de l'entrée, me fait penser qu'il ne s'agit pas d'une variable principale pour l'objectif ici visé.

catrice qui signale si le véhicule a été introduit par un nouveau membre du foyer. Le fait que le véhicule puisse rentrer sans passer par un achat rend ce cas aussi délicat à traiter. Cependant, le cas d'un véhicule qui rentre d'une façon autre que l'achat est marginal (les véhicules déclarés qui n'avaient pas été achetés au moment de l'entrée représentent entre 3–6%<sup>4</sup> de 1996 à 2008, et moins de 0,5%<sup>5</sup> des déclarations annuelles concernent un véhicule introduit par un nouveau membre). L'analyse permet de remarquer que l'absence d'un indicateur d'achat est un problème bien plus important à traiter. En effet, cela permettrait de confirmer l'information collectée par la date d'achat qui est souvent touchée par les erreurs de mémoire.

L'observation directe au niveau véhicule des vagues Parc Auto permet de reconnaître les variables qui, à moins d'erreurs plus au moins systématiques, donnent une description succincte mais suffisante pour reconnaître le véhicule d'une année sur l'autre. Il s'agit des 5 variables suivantes :

- **“Code national d'identification du type”** (Champ D.2.1 de la Carte Grise, CNIT<sup>6</sup>).
- **Année d'achat** et de **première mise en circulation** du véhicule.
- **Marque** et **Modèle** du véhicule.

L'analyse préliminaire met en lumière plusieurs incohérences, vraisemblablement dues au type de questionnaire utilisé (i.e. Enquête papier auto-gérée, envoyée par voie postale).

On devrait en effet, pour un véhicule fixé, observer la même valeur de ces variables tout au long de l'observation mais ce n'est pas toujours le cas. Plusieurs types d'erreurs sont détectables, comme les erreurs de mémoire entachant les dates déclarées ou les erreurs de saisie qui touchent la variable Marque-Modèle (**marmod**) codée *a posteriori*<sup>7</sup> et le Code National du type, enregistrés par lecture optique.

Sur ces bases une première procédure applicable aux ménages mono-équipes est réalisée. Les données issues de cette procédure sont utilisées pour l'application présentée au chapitre 7.

J'ai alors entamé une analyses sur la base complète, sans restrictions sur

---

4. 4 – 8% année d'achat déclarée égale à l'année observée.

5. Non réponse : 3 – 9%.

6. Dans l'enquête il est appelé Type Mines, comme avant 1995.

7. Les ménages sont invités à indiquer la marque et le modèle en clair. Au moment de la saisie un code résumant les deux informations est alors assigné à chaque véhicule au moment de la saisie.

la taille du parc.

À cette étape j'ai pu m'apercevoir du fait que, contrairement à ce qui a été observé au chapitre précédent (section 3.3) au niveau ménage, l'utilisation d'une période longue semble être plus efficace lorsque l'objectif est d'identifier et de décrire les véhicules<sup>8</sup>.

La période à utiliser pour l'application de la procédure est alors choisie en analysant l'information disponible, comme fait au chapitre précédent pour les ménages<sup>9</sup>.

Le tronc commun est plus important que celui observé au chapitre précédent (46% des variables incluses au moins une fois au niveau véhicule, contre 38,8% au niveau ménage). Cependant plusieurs informations ne sont pas exploitables sur toute la période disponible.

Bien que la base Parc Auto conservé à l'IFSTTAR fournisse un panel représentatif de la France entière depuis 1993<sup>10</sup>, c'est seulement à partir de 1996 que le prix d'achat des véhicules possédés par le ménage<sup>11</sup> est collecté. D'autres informations ne sont pas toujours disponibles, comme certaines variables concernant les décisions d'assurance qui ne sont plus observées après 1999 ou celles d'équipement (gps, autoradio, etc.) observées à partir de 2003. Cependant, ces variables n'ont pas des propriétés telles de permettre l'identification du véhicule et leur absence n'a pas d'effet sur le fonctionnement de la procédure.

Une information importante est fournie par un groupe de variables qui ont pour but la description du véhicule que celui déclaré substitue dans le parc (si c'est le cas). Cependant cette information n'est pas collectée après 2002.

La période choisie est alors la plus longue possible, c'est-à-dire celle couvrant les années comprises entre 1996 et 2008.

Les traitements préliminaires nécessaires pour l'exploitation des données, sur les variables et les récurrences, sont présentés au cours de la section 4.1. La procédure, présentée ensuite (section 4.2) est composée de deux étapes.

---

8. Les données utilisées pour l'application présentée au chapitre 8 de la prochaine section sont issues d'une procédure construite à cette étape et appliquée seulement aux données observées entre 2002 et 2008.

9. Construction d'un dictionnaire unique

10. Les informations collectées par les vagues précédentes ne sont pas disponibles sur le serveur auquel j'ai eu accès pour ce travail.

11. Une variable clé pour l'application présentée au cours de la prochaine partie

Avant tout les véhicules sont identifiés d’une année sur l’autre (4.2.1). Dans un deuxième temps (4.2.2) les informations sensées être propres au véhicules et indépendantes du temps (comme par exemple son kilométrage à l’achat) sont reconstituées par observation complète de la période de déclaration du véhicule. L’information concernant l’utilisation du véhicule (i.e. kilométrage déclaré et au compteur) est aussi traitée à cette étape.

Le panel de véhicule que cette procédure permet de construire est brièvement analysé en conclusion de ce chapitre (section 4.3).

## 4.1 Traitement Préliminaire

En appariant les vagues annuelles on observe sur les 29128 ménages ayant répondu au moins une fois à l’enquête entre 1996 et 2008, que 3206 se sont “absentés” du panel au moins un an. Je définis alors une *période d’observation* une ou plusieurs années consécutives pendant lesquelles le ménage répond à chaque vague de l’enquête.

La table 4.1 reporte la répartition des ménages enquêtés selon le nombre de périodes pendant lesquels ils sont continuellement observés.

TABLE 4.1 – Distribution des Ménages par Nombre de Périodes d’Observation

Nb Per. Obs. <sup>a</sup>	Fréquence	Cumulée
1	25922	25922
2	2669	28591
3	459	29050
4	69	29119
5	9	29128

<sup>a</sup>. Nombre de fois que le ménage entre dans le Panel (soit pour la première fois, soit après une ou plusieurs années non observées).

Puisque, pour ces ménages, on ne dispose d’aucune information entre une période d’observation et la suivante, je traiterai les 32958 périodes d’observation séparément. La clé d’identification (i.e. l’identifiant du ménage) est alors complétée avec la “position” de l’observation (i.e. 1 si le ménage est observé pour la première fois, 2 si le ménage a déjà été observé au cours d’une

première période et n'a pas répondu au moins un an à l'enquête, 3 si c'est la deuxième fois que le ménage recommence à répondre après s'être absenté de l'enquête, etc.).

La répartition des 32958 périodes continues d'observation par le nombre d'années qui les composent est reportée en table 4.1.

TABLE 4.2 – Nombre de Périodes Observées par durée

1	16072		
2	6410	8	545
3	3192	9	385
4	2054	10	305
5	1504	11	253
6	1024	12	166
7	709	13	339

On observe que seulement 51,2% compte au moins deux années observées de suite.

Si l'on souhaite couvrir une période suffisamment longue, la taille de l'échantillon utilisable est réduite. Par exemple 9664 ménages sont observables au moins 2 années consécutives entre 2002 et 2008. Si en plus on veut un sous panel cylindré cette taille est moins importante. Si l'on veut observer les ménages de 2002 à 2008 l'échantillon comptera alors 1627 ménages.

Comme dit en introduction de cette partie, si les ménages sont identifiables par une clé unique, les véhicules composant la flotte se voient attribuer un rang. Puisque les ménages sont libres de définir ces rangs, on ne peut pas être sûr qu'ils identifient de façon unique ces véhicules. La création d'un identifiant plus approprié est alors nécessaire.

La procédure qui suit est construite dans le but d'obtenir un tel identifiant de façon à perdre le moins d'information possible.

## 4.2 La Procédure

Pour pouvoir étudier de façon longitudinale les véhicules, il faut pouvoir les appairer d'une année sur l'autre. Je ne considérerai donc pas les 23064

véhicules possédés par les ménages au cours des 16072 périodes<sup>12</sup> pour lesquelles on dispose d'une seule année d'observation.

Des 16886 périodes d'une durée supérieure ou égale à 2 ans, 15177 font apparaître au moins une année où le ménage possède au moins un véhicule. Ce sont ces périodes qui sont utilisées, afin de construire une base décrivant de manière cohérente les véhicules possédés par le ménage pendant leur durée d'observation, des dates d'achat aux dates d'abandon.

On dispose de 86590 observations annuelles de véhicules, pour 61736 observations annuelles de ménages sur les 15177 périodes<sup>13</sup>. Puisque la première année de chaque période d'observation est utilisée comme référence (i.e. Un identifiant, défini égal au rang déclaré, est assigné aux 21999 véhicules possédés au cours de la première année observée pour chaque période), 64591 véhicules sont à identifier sur 46559 années.

L'identification se fait de la façon suivante.

### 4.2.1 Identification des Véhicules

Pour chaque année observée, les véhicules composant le parc automobile décrit par un ménage sont confrontés, un par un, à ceux que celui-ci déclare l'année suivante. Pour chaque couple de véhicules est alors défini un degré de ressemblance, sur la base des 5 variables identifiées plus haut.

La décision de définir un degré de ressemblance est nécessaire dans le but de minimiser la perte d'observation. Comme dit en introduction, la taille de l'échantillon disponible est réduite de façon non négligeable lorsque l'on souhaite considérer des ménages observés continuellement au cours d'une période suffisamment longue. Si l'on ne se limitait qu'aux ménages pour lesquels aucun doute possible entache l'appariement des véhicules (i.e. chaque véhicule déclaré correspond exactement<sup>14</sup> à un seul véhicule précédemment décrit ou ne correspond à aucun autre véhicule mais les déclarations confirment sans doute son entrée dans le parc au cours de l'année) l'échantillon exploitable serait alors sensiblement petit.

---

12. Voir table 4.3, à la section suivante

13. 33220 véhicules pendant toute la période observée, 14459 au moins un véhicule mais une seule année observée

14. Les valeurs déclarées pour les 5 variables sont identiques d'une année sur l'autre

Les 5 variables sur lesquelles se base la construction de l'identifiant présentent différents types d'erreurs. Je vais donc définir dans un premier temps un degré de ressemblance pour chacune de ces variables.

Le **Code Nationale d'Identification du Type**, qui a progressivement remplacé depuis 1995 le type Mines selon les nouvelles réglementations européennes, permet de définir les caractéristiques essentielles succinctes d'un véhicule.

Plusieurs problèmes sont rencontrés en traitant une telle variable. En premier lieu on remarque que les ménages ont tendance à omettre l'information, notamment lorsqu'ils ont déjà déclaré le véhicule auparavant. En outre, comme dit précédemment, cette information est renseignée à la main par les enquêtés et saisie ensuite par lecture optique. L'analyse manuelle des données permet de remarquer que cette procédure de saisie cause souvent la modification (ex. I → 1) ou la perte de certains caractères dans l'information enregistrée. Pour conclure, on trouve certains CNIT qui ne sont pas complets et comptent un nombre de caractères inférieur aux 12 définissant le code<sup>15</sup> (ou sont complétés par une queue de 0, dans ce cas la procédure efface cette queue avant traitement.). L'observation directe des récurrences permet de s'apercevoir que souvent ces CNIT sont une troncature d'une déclaration précédente.

Je définis alors 3 degrés de ressemblance, lorsque les deux CNIT ne sont pas exactement égaux, par comparaison caractère par caractère :

1. Très ressemblants :
  - Tous les caractères sont au moins ressemblants
  - Tous les caractères du CNIT le plus court sont identiques à celles d'une sous-string de l'autre
  - Le CNIT plus court est égal à l'autre moins un caractère
2. Ressemblants :
  - Un seul caractère est différent
  - Tous les caractères du CNIT le plus court sont au moins ressemblants à ceux d'une sous-string de l'autre
  - Le CNIT plus court est ressemblant à l'autre moins un caractère
3. Peu ressemblants :
  - Le CNIT plus court diffère d'une sous-string de l'autre pour un caractère

---

15. Le CNIT a connu en 1998 un changement de définition, afin d'obtenir un code plus stable, qui réduit le nombre de caractères. On rencontre donc aussi, dans la base, des CNIT plus anciens comptant un nombre de caractères supérieur à 12.



- Le CNIT plus court diffère de "l'autre moins un caractère" pour un caractère

La ressemblance des autres variables est définie comme suit.

- **Année d'achat**
  - Très ressemblants : les années d'achat se suivent
  - Ressemblants : les années d'achat ont une différence  $> 1$  mais  $\leq 5$
  - Peu ressemblants : les années d'achat ont une différence = 10
- **Année de première mise en circulation**
  - Très ressemblants : les années de construction se suivent
  - Ressemblants : les années de construction ont une différence  $> 1$  mais  $\leq 5$
  - Peu ressemblants : les années de construction ont une différence = 10
- **Marque**
  - Très ressemblants : les codes de marque se suivent dans la liste mar-mod
  - Ressemblants : les codes de marque ont une différence = 10
  - Peu ressemblants : l'un des codes de marque est l'unité de l'autre
- **Modèle**
  - Très ressemblants : les codes de modèle se suivent dans la liste mar-mod
  - Ressemblants : les codes de modèle ont une différence = 2
  - Peu ressemblants : les codes de modèle ont une différence = 10

Ces indicateurs de ressemblance sont construits de façon à indiquer aussi les cas où l'information est manquante ou incomplète. Dans ces cas les variables assument des valeurs négatives (selon le type de problème rencontré).

Pour chaque couple de véhicules est alors défini un "indicateur de fiabilité d'appariement", basé sur l'évaluation simultanée des niveaux de ressemblance des 5 variables.

Un indicateur de nouvelle entrée (i.e. véhicule entrant dans le parc du ménage, que ce soit par l'achat ou par introduction d'un nouveau membre) est aussi construit pour chaque véhicule.

- -1 = l'année d'achat n'est pas renseignée
- 0 = l'année d'achat est différente de l'année observée
- 1 = l'année d'achat est l'année observée
- 2 = l'année d'achat est l'année précédente celle observée
- 3 = l'année d'achat précède de 10 ans celle observée

À cette étape on observe un nouveau problème. Effectivement, le choix d'utiliser un indicateur de vraisemblance implique que la possibilité d'un appariement sera indiquée même lorsque deux véhicules ont quelque caractéristique commune (par exemple, achetés la même année). La procédure traite alors de façon croisée toutes les ressemblances possibles pour les véhicules appartenant à un même parc à l'aide de l'indicateur de fiabilité et celui d'achat, de façon à déterminer le plus large nombre d'appariements fiables <sup>16</sup>.

Pour conclure la phase d'identification, une attention particulière est portée à la suite d'années d'achats déclarés.

En effet la procédure jusqu'ici décrite ne permet pas toujours d'identifier un achat lorsque celui-ci concerne un véhicule qui est très ressemblant à un autre abandonné à la fin de la période précédente et/ou l'information concernant l'année d'achat n'est pas précise au moment où il a lieu. Pour cette raison la procédure réalisée identifie les cas où ce type de "appariement forcé" a lieu, c'est-à-dire lorsqu'à partir d'une certaine date l'année d'achat déclarée change (et continue à être déclarée de cette nouvelle façon par la suite). Dans ces cas l'achat est reconnu et la suite d'observations est décomposée, décrivant ainsi deux véhicules différents.

L'identifiant des véhicules a ainsi été créé. Pour ces véhicules l'information est sélectionnée comme suit.

### 4.2.2 Sélection des Informations

Une part des questions portant sur la description de la flotte, concernent des informations qui ne varient pas avec le temps. Ces caractéristiques sont alors ajoutées à la base.

On devrait pour ces véhicules observer la même valeur de certaines variables au cours de toute leur durée d'observation. Ce n'est pas toujours le cas. Plusieurs types d'erreurs sont détectables en observant les données. Comme dit plus haut, les erreurs de mémoire entachent les dates déclarées

---

16. Plus d'informations sur la page personnelle de [Cernicchiaro](#)

et les erreurs de saisie touchent des variables codés *a posteriori* (comme par exemple le code de Marque-Modèle, `marmod`). De plus les déclarations concernant les quantités, comme les prix ou les kilométrages, sont souvent arrondies.

L'objectif est alors de corriger ces erreurs de manière à disposer de données plus fiables pour chaque véhicule. À ce titre sont utilisées les informations suivantes, présentes dans la base :

### Conditions à l'Achat

- Date d'Achat du Véhicule : Année et Mois.
- Date de Première Mise en Circulation : Année et Mois.
- Indicateur du Type d'Achat : Neuf ou d'Occasion.
- Kilométrage du Véhicule à l'Achat : en kilomètres (km).
- Prix du Véhicule à l'Achat : en euro (€) <sup>17</sup>.

### Description du Véhicule

- Marque et Modèle du Véhicule : codés selon la classification Parc Auto.
- Type de Carburant : Essence, Diesel ou Autre.
- Type de Carrosserie :
  1. *Berline* : une automobile fermée par un toit rigide fixe, avec un pare-brise fixe, quatre portes latérales, et quatre fenêtres latérales.
  2. *Break* : un type de carrosserie automobile similaire à une berline, mais avec un toit qui se prolonge jusqu'à l'aplomb de l'arrière du véhicule.
  3. *Coupé* : une voiture fermée à deux ou trois portes et possédant généralement deux ou quatre places;  
*Cabriolet* : une automobile ouverte et découvrable grâce à une capote ou un toit escamotable, avec un pare-brise fixe, et des portes sans encadrement supérieur.
  4. *Monospace* : un véhicule automobile ne comportant généralement qu'un seul volume, destiné avant tout à un usage familial et bénéficiant d'une carrosserie plus volumineuse que celle d'un break ordinaire. Il comporte au minimum 5 sièges, souvent distincts, et peut aller jusqu'à 8 ou 9 sièges;

---

17. Avant 2002 les données sont collectées en Francs. Une conversion est effectuée afin de homogénéiser les valeurs.

*Van* : un véhicule utilitaire léger dont l'espace intérieur a été aménagé pour le transport de personnes.

5. *Camping-car* : un véhicule habitable ;  
*Mobil-home* : une habitation construite sur un châssis remorquable.
  6. *Utilitaire léger (Fourgonnette)* : un véhicule de transport automobile, circulant sur les routes et principalement destiné aux marchandises ;
  7.  $4 \times 4$  : un véhicule tout-terrain.
- Consommation moyenne des Véhicules : en litres pour 100 kilomètres (l/100km).
  - Puissance Fiscale : en chevaux fiscaux.

En observant simultanément toutes les variables d'un même groupe (conditions à l'achat ou description du véhicule), on peut évaluer la cohérence entre les déclarations annuelles.

Pour ce qui concerne les informations décrivant l'achat, les conditions de cohérence sont dictées par les liens logiques existant entre ces variables (Ex. Un véhicule acheté neuf ne peut pas avoir été immatriculé avant l'achat).

La cohérence entre les caractéristiques descriptives du véhicule est moins évidente à évaluer. Comme expliqué en introduction, est ici disponible une base décrivant de façon détaillée un grand nombre de différents types de véhicules. Celle-ci peut alors être utilisée pour la création d'un "indicateur de cohérence" de la combinaison de Marque, Modèle, Carburant et Carrosserie (selon que celle-ci existe ou non dans la base externe)<sup>18</sup>.

En joignant à l'observation de la fiabilité totale des déclarations annuelles le nombre de fois qu'une même valeur est déclarée pour la variable au cours de la période d'observation, on peut mieux identifier les déclarations incohérentes et sélectionner donc la valeur la plus fiable.

Lorsque le type de véhicule retenu existe dans la base, les variables suivantes sont ajoutées à la base des véhicules observés dans Parc Auto, afin d'élargir le nombre d'informations disponibles pour la description des véhicules.

- Consommation moyenne des Véhicules<sup>19</sup> : en litres pour 100 kilomètres (l/100km).

---

18. Plus d'informations sur la page personnelle de [Cernicchiaro](#)

19. Utilisée pour compléter les valeurs manquantes dans la base obtenue de Parc Auto.

- Puissance Nominale : en milliers de kW.
- Poids du Véhicule : en milliers de kg.
- Dimensions du Véhicule (Hauteur, Largeur et Longueur) : en mètres.
- Consommation sur Routes urbaines.
- Consommation sur Routes extra-urbaines.
- Autonomie du Véhicule : en kilomètres (km).
- Vitesse Maximale du Véhicule : en kilomètres par heure (km/h).

Puisque seuls la marque et le modèle du véhicule, et non le type, sont collectés pour les véhicules observés au cours de l'enquête "Parc Auto" et que plusieurs types de véhicules peuvent correspondre à un même couple marque/modèle, les valeurs des variables descriptives ici considérées sont obtenues comme moyenne des valeurs observées par marque, modèle, type de carburant (en distinguant entre essence, diesel et autre) et carrosserie, dans la base des Données Automobiles.

Pour conclure, l'information concernant l'utilisation du véhicule (i.e. kilométrage déclaré et au compteur) est analysée et assignée/corrigée, lorsque besoin il y a et que cela est possible (i.e. seulement une des 3 informations "Kilométrage en Début de période", "Kilométrage parcouru" et "Kilométrage au compteur" n'est pas disponible et les autres sont cohérentes entre elles)<sup>20</sup>.

### 4.3 La Taille de l'Échantillon des Véhicules Identifiés

On peut maintenant évaluer la taille de l'échantillon effectivement disponible, après identification des véhicules. Je n'analyse ici que les effets de la première étape de la procédure, c'est-à-dire les exclusions de ménages dues à la difficulté rencontrée en essayant de déterminer les appariements de façon automatique.<sup>21</sup>

En table 4.3 sont reportés les mouvements annuels des vagues répondant à l'enquête. Je ne parle pas ici des vagues annuelles (comme ça a été le cas

---

20. La table descriptive des véhicules ainsi obtenue sera utilisée notamment pour l'application présentée au cours du chapitre 7 de la prochaine partie, pour l'étude du prix des véhicules.

21. D'autres conditions, résultant des observations formulées à l'aide de la deuxième étape de cette procédure, seront commentées au chapitre suivant (5).

### 4.3. LA TAILLE DE L'ÉCHANTILLON DES VÉHICULES IDENTIFIÉS<sup>63</sup>

jusqu'ici) mais des “vagues entrantes” dans la base. Une vague comprend alors tous les ménages observés pour la première fois (entre 1996 et 2008<sup>22</sup> la même année. La répartition de ces vagues est reportée en ligne par année de dernière observation. En colonne sont donc reportés les volumes des ménages sortant de l'enquête<sup>23</sup>.

Puisque ce tableau n'est pas forcément de simple lecture, je vais présenter de façon plus détaillée les informations fournies (tab. 4.3).

Pour chaque “vague entrante” trois volumes sont reportés. La ligne centrale (en gras) indique le nombre de ménages motorisés, sortant l'année correspondant à la colonne, pour lesquels tous les véhicules déclarés ont été identifiés par la procédure. Les nombres des ménages pour lesquels au moins un véhicule ne l'a pas été sont reportés en troisième ligne. La première indique le nombre de ménages sortant qui n'ont jamais déclaré de véhicules (parce que non-motorisés tout au long de l'observation). Les totaux pour ces 3 informations sont reportés en avant dernière colonne. La dernière colonne reporte enfin le nombre total des ménages motorisés (au moins une fois) composant la vague correspondante à la ligne, ainsi que le pourcentage sur ce total des ménages pour lesquels la procédure a pu identifier tous les véhicules.

La dernière ligne indique les nombres totaux, toute “vague entrante” confondue, des “ménages multi-observés” (i.e. les ménages répondant à l'enquête au moins deux années de suite) sortant une même année (correspondante à la colonne)<sup>24</sup>.

Pour conclure, les informations concernant les “ménages mono-observés” sont reportées en bas de la page. Pour ces ménages l'année d'entrée correspond à celle de sortie. Les totaux annuels sont alors reportés par vague (en colonne) en distinguant entre ceux qui possèdent au moins un véhicule cette année et les non-motorisés (en ligne).

La dernière ligne reporte le nombre total de ménages sortant de l'enquête chaque année.

846 périodes de l'échantillon présentent des incohérences entre les parcs déclarés deux années de suite au moins une fois au cours de l'observation. Ne disposant pas d'un identifiant pour toutes les véhicules déclarés, les ménages

---

22. Les ménages ayant répondu pour la première fois avant 1996 sont considérés comme entrant en début de la période ainsi définie (1996 – 2008).

23. Ici aussi, puisque la période est tronquée, les ménages ayant continué à répondre après 2008 sont considérés sortir à la fin de l'observation.

24. Les derniers 2 totaux de cette ligne ont déjà été commentés (en introduction de 4.2).

concernés par ce problème ne peuvent pas être utilisés. La base obtenue contient alors 29341 véhicules, observés sur un total de 14331 ménages (répondant au moins deux années de suite à l'enquête).

La perte d'information semble être suffisamment contenue par la procédure. En effet moins de 10% des "ménages multi-observés" motorisés sortants chaque année est touché par des problèmes d'identification de ses véhicules. Le pourcentage des ménages complètement identifiés varie entre 91,6% et 96,9% au cours de la période de 1996 à 2008.

La période à utiliser pour l'application a été définie au cours du chapitre précédent (3.3) comme 2002 – 2008. Comme expliqué à cette occasion, puisque l'objectif du travail est d'étudier les dynamiques du comportement automobile des ménages, il faut une période suffisamment longue qui permet de faire des hypothèses simples sur leur évolution au fil du temps. La taille de l'échantillon des "ménages multi-observés" motorisés exploitables est alors obtenue par somme des valeurs en gras sur les colonnes de 2003 à 2008, et compte 8017 ménages. La période cylindrée compte 1266 ménages. La procédure semble alors produire les résultats souhaités et la période d'observation, choisie au cours de la section précédente, peut être utilisée pour l'application avec la base ainsi obtenue.

TABLE 4.3 – Répartition par Année de Dernière Observation des “Vagues Entrantes” entre 1996 et 2008

		1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	Tot MI–Obs <sup>d</sup>	Tot MI–Obs M <sup>b</sup>
1996	JM <sup>c</sup>	160	87	68	72	29	13	25	14	14	11	11	34	538	<b>3953</b> (95,6) <sup>f</sup>
	TI <sup>d</sup>	<b>1062</b>	<b>700</b>	<b>522</b>	<b>539</b>	<b>233</b>	<b>129</b>	<b>111</b>	<b>74</b>	<b>78</b>	<b>38</b>	<b>26</b>	<b>266</b>	<b>3778</b>	
	NI <sup>e</sup>	21	17	11	19	13	12	13	7	6	15	2	39	175	
1997	JM	30	16	13	8	7	7	6	3	6	3	6	0	19	115
	TI	<b>388</b>	<b>169</b>	<b>150</b>	<b>73</b>	<b>68</b>	<b>36</b>	<b>29</b>	<b>30</b>	<b>18</b>	<b>4</b>	<b>84</b>	<b>1049</b>	<b>1106</b>	
	NI	6	5	5	2	1	8	3	1	0	2	24	57	94,8	
1998	JM	30	19	9	2	4	6	2	3	2	16	3	16	93	<b>1113</b> (93,6)
	TI	<b>353</b>	<b>218</b>	<b>100</b>	<b>80</b>	<b>49</b>	<b>30</b>	<b>35</b>	<b>22</b>	<b>9</b>	<b>146</b>	<b>1042</b>	<b>1042</b>		
	NI	12	7	2	5	5	7	5	3	4	21	71	71		
1999	JM	28	21	8	8	5	1	3	6	19	99	19	99	1136	<b>1136</b> (93,5)
	TI	<b>381</b>	<b>197</b>	<b>124</b>	<b>81</b>	<b>63</b>	<b>47</b>	<b>20</b>	<b>17</b>	<b>132</b>	<b>1062</b>	<b>1062</b>	<b>1062</b>		
	NI	10	6	13	5	8	7	4	4	17	74	74	74		
2000	JM	18	23	7	6	11	5	4	26	100	100	100	100	886	<b>934</b> (94,9)
	TI	<b>323</b>	<b>155</b>	<b>98</b>	<b>54</b>	<b>55</b>	<b>21</b>	<b>18</b>	<b>162</b>	<b>886</b>	<b>886</b>	<b>886</b>	<b>886</b>		
	NI	7	8	1	8	3	3	5	13	48	48	48	48		
2001	JM	52	32	19	18	8	11	40	180	180	180	180	180	1084	<b>1084</b> (94,1)
	TI	<b>350</b>	<b>186</b>	<b>132</b>	<b>90</b>	<b>41</b>	<b>16</b>	<b>205</b>	<b>1020</b>	<b>1020</b>	<b>1020</b>	<b>1020</b>	<b>1020</b>		
	NI	13	7	8	10	6	3	17	64	64	64	64	64		
2002	JM	39	17	17	11	7	39	130	130	130	130	130	130	1152	<b>1152</b> (93,0)
	TI	<b>362</b>	<b>171</b>	<b>139</b>	<b>76</b>	<b>52</b>	<b>271</b>	<b>1071</b>	<b>1071</b>	<b>1071</b>	<b>1071</b>	<b>1071</b>	<b>1071</b>		
	NI	6	9	15	8	6	37	81	81	81	81	81	81		
2003	JM	43	16	13	7	37	116	116	116	116	116	116	116	927	<b>927</b> (91,6)
	TI	<b>286</b>	<b>183</b>	<b>72</b>	<b>33</b>	<b>275</b>	<b>849</b>	<b>849</b>	<b>849</b>	<b>849</b>	<b>849</b>	<b>849</b>	<b>849</b>		
	NI	17	14	5	4	38	78	78	78	78	78	78	78		
2004	JM	25	9	4	30	68	68	68	68	68	68	68	68	774	<b>774</b> (92,5)
	TI	<b>332</b>	<b>100</b>	<b>55</b>	<b>229</b>	<b>716</b>	<b>716</b>	<b>716</b>	<b>716</b>	<b>716</b>	<b>716</b>	<b>716</b>	<b>716</b>		
	NI	17	7	8	26	58	58	58	58	58	58	58	58		
2005	JM	26	9	47	82	82	82	82	82	82	82	82	82	692	<b>692</b> (92,5)
	TI	<b>185</b>	<b>100</b>	<b>355</b>	<b>640</b>	<b>640</b>	<b>640</b>	<b>640</b>	<b>640</b>	<b>640</b>	<b>640</b>	<b>640</b>	<b>640</b>		
	NI	11	7	34	52	52	52	52	52	52	52	52	52		
2006	JM	20	66	86	86	86	86	86	86	86	86	86	86	897	<b>897</b> (95,1)
	TI	<b>277</b>	<b>576</b>	<b>853</b>	<b>897</b>	<b>897</b>	<b>897</b>	<b>897</b>	<b>897</b>	<b>897</b>	<b>897</b>	<b>897</b>	<b>897</b>		
	NI	9	35	44	44	44	44	44	44	44	44	44	44		
2007	JM	102	102	102	102	102	102	102	102	102	102	102	102	1409	<b>1409</b> (96,9)
	TI	<b>1365</b>	<b>1365</b>	<b>1365</b>	<b>1365</b>	<b>1365</b>	<b>1365</b>	<b>1365</b>	<b>1365</b>	<b>1365</b>	<b>1365</b>	<b>1365</b>	<b>1365</b>		
	NI	44	44	44	44	44	44	44	44	44	44	44	44		
<b>Tot. Sort. MI–Obs<sup>g</sup></b>		<b>1243</b>	<b>1228</b>	<b>1186</b>	<b>1461</b>	<b>1041</b>	<b>1063</b>	<b>1090</b>	<b>1022</b>	<b>1174</b>	<b>750</b>	<b>742</b>	<b>4886</b>	<b>16886</b>	<b>15177</b>
														Tot Mn–Obs <sup>h</sup>	
Tot Mn–Obs M <sup>a</sup>	2314	958	821	1083	742	972	801	858	899	1113	651	1105	2142	14459	16072
Tot Mn–Obs nMI <sup>j</sup>	288	92	86	107	96	133	91	111	86	86	82	113	242	1613	
<b>TOT</b>	<b>2602</b>	<b>2293</b>	<b>2135</b>	<b>2376</b>	<b>2299</b>	<b>2146</b>	<b>1955</b>	<b>2059</b>	<b>2007</b>	<b>2373</b>	<b>1483</b>	<b>1960</b>	<b>7270</b>	<b>TOTAL :</b>	<b>32958</b>

a. Multi–Observation

b. Multi–Observation Motorisés

c. Jamais Motorisé

d. Tous les véhicules Identifiés

e. Au moins un véhicule non identifié

f. Pourcentage Périodes MI–Obs M. Complètement Identifiés

g. Total Ménages Multi–Observation En Sortie

h. Mono–Observation

i. Mono–Observation Motorisés

j. Mono–Observation Non–Motorisés





# Chapitre 5

## L'échantillon Retenu

Au cours de cette part les bases de données et leurs préparations ont été présentées. Ce chapitre conclut la part, par une présentation de l'échantillon obtenu.

Dans un premier temps (section 5.1) les dernières manipulations, à effectuer au niveau ménage sont présentées. L'échantillon retenu est alors construit et analysé, dans le but de vérifier les conséquences que les traitements peuvent avoir eu au cours de la section 5.2. Pour conclure une analyse de la problématique est menée, sur l'échantillon retenu, en section 5.3.

Quelque opération préliminaire est à mener pour passer du niveau véhicule au niveau ménage.

Avant tout, les informations concernant le parc total (ex. le nombre d'achats/abandons, le kilométrage total, l'âge moyen, etc.) sont obtenus par agrégation de l'information reconstruite au niveau véhicule, par le biais de la procédure présentée au chapitre précédent (4.2).

De la même façon sont obtenues les informations agrégées d'utilisation des transports en commun du ménage. Ces informations sont en effet collectées par individu dans l'enquête. Cette information est collectée par deux variables différentes avant et après 2004 (la première définie en Francs et la deuxième en €). La définition est donc homogénéisée au préalable.

## 5.1 L'appariement au Niveau Ménage

Comme on l'a vu pour les véhicules, les membres du ménage se voient aussi attribuer un rang.

À la différence de ceux déclarés pour les véhicules, ces rangs doivent être attribués par les ménages selon une règle précise, spécifiée dans le questionnaire. Les membres sont codés dans la base par ordre décroissant d'âge.

La difficulté à appairer ces informations est alors moindre par rapport à celle expliquée au chapitre précédent. Cependant, cette opération est très utile pour observer les mouvements internes au ménage.

Dans ce travail le ménage est supposé être "prévoyant" lorsqu'il effectue ses choix de mobilité. Il est vraisemblable qu'un changement "majeur", comme une mise en couple, une séparation ou un décès puisse impliquer une réorganisation de la vie du ménage et donc par exemple un changement dans le regard qu'il porte sur son parc automobile. Pour cette raison, il me semble important de pouvoir connaître la part de ménages connaissant une telle situation.

Les informations ici considérées concernent les membres des 9664 ménages observés au moins 2 fois de suite entre 2002 et 2008<sup>1</sup>. Pour chaque membre sont déclarés le sexe, l'âge et la catégorie socio-professionnelle (5 modalités : Employé ou Ouvrier, Profession Intermédiaire, Cadre, Inactif, Retraité).

Comme dit en introduction, je vais définir ici une variable indiquant le type de chef du ménage.

Un ménage peut avoir un seul chef, homme ou femme, ou un couple de chefs des deux sexes. On peut dans un premier temps appairer les informations les concernant.

En codant les 3 configurations possibles, il est possible de définir, par comparaison avec la configuration de l'année précédente, quel rôle (Chef Homme, Chef Femme, Chef avec Maitresse de maison<sup>2</sup> ou Maitresse) avait l'individu considéré. Cette information est alors utilisée pour l'appariement et sert à tester la cohérence des âges déclarés pour l'individu au cours de l'observation.

Les informations concernant les autres membres du ménage sont alors incluses. Plusieurs variables décrivant la structure du ménage sont alors

---

1. Comme dit en introduction, se limiter à une période plus courte réduit la part de ménages observant un changement.

2. En reprenant la terminologie Parc Auto, lorsque le ménage a un couple de chefs on appellera ici "Maitresse" la femme.

construites<sup>3</sup> par agrégation des informations collectées au niveau individu (ex. Nombre de Membres par catégorie socio-professionnelle). Des variables collectant l'information retardée sont aussi construites à cette étape.

Une procédure de traitement transversale des informations collectées est alors construite dans le but d'identifier les changements majeurs définis plus haut.

On peut observer que 2,8% des ménages enquêtés connaissent un changement de chef et dans 1% des cas des enfants (mineurs) sortent du ménage au cours de l'enquête<sup>4</sup>.

Une autre remarque qui peut être faite est que la plupart du temps le ménage arrête de répondre à l'enquête après des tels changements, comme le montre table 5.1.

TABLE 5.1 – Nombre d'Années Observées Après Changement

Type	Tot	1	2	3	4	5	6
Chef	278	41,73	17,27	14,03	9,71	11,87	5,39
Enfants	207	42,03	24,15	14,49	9,18	6,76	3,38

Une telle observation pourrait confirmer l'hypothèse qu'un tel changement implique des variations de comportement automobile<sup>5</sup>.

Une condition excluant ces ménages est alors considérée dans la sélection de l'échantillon retenu, présentée au cours de la section qui suit.

## 5.2 L'Analyse de l'Échantillon Retenu

17357 ménages ont répondu à l'enquête au cours de 19107 périodes entre 2002 et 2008. Par simplicité je parlerai dorénavant de ménage pour indiquer une période continue d'observation.

Deux procédures, présentées respectivement en 4.2 et 5.1, permettent la

---

3. seulement pour 3 périodes est observée l'absence d'information pour au moins un membre

4. D'autres changements, comme par exemple les naissances, ont été analysés mais ne sont pas reportés ici.

5. i.e. Le ménage ne se sentant plus "cohérent" arrête de répondre

construction d'un échantillon répondant aux besoins du travail.

Le sujet ici étudié est le comportement d'équipement automobile des ménages. Les informations concernant le parc automobile du ménage et l'utilisation qu'il en fait doivent alors être traitées avec attention lorsque l'on construit l'échantillon pour l'application.

À cet effet, on peut définir les 5 conditions suivantes pour que le ménage soit retenu :

1. Le ménage ne possède jamais plus de 3 véhicules entre 2002 et 2008.
2. Tous les véhicules ont été identifiés par la procédure (i.e. Les variables d'identification 4 sont suffisantes pour reconnaître les véhicules d'une année sur l'autre tout au long de l'observation.).
3. Tous les appariements sont fiables (i.e. Une seule date d'achat cohérente pour tous ses véhicules).
4. La suite des kilométrages annuels est définie de façon cohérente pour chaque véhicule observé au sein du ménage pendant l'enquête.
5. Le type de carburant est identifié pour tout véhicule du ménage.

Deux conditions supplémentaires sont à imposer, à mon sens, dans la construction de l'échantillon.

En premier lieu, comme dit en 5.1, le ménage ne doit pas connaître de "changements démographiques majeurs" au cours de l'observation. Pour conclure, le revenu doit avoir été déclaré chaque année. 9,66% des ménages observées sont dans ce cas (i.e. ne déclarent pas leur revenu au moins une fois).

Le choix de ne pas utiliser ces observations plutôt qu'en imputer les informations est critiquable. Cependant on peut remarquer que dans environ la moitié de ces cas (revenu manquant au moins une fois), au moins une des autres conditions est violée. Il se pourrait alors que l'absence du revenu déclaré reflète une réponse annuelle peu fiable, je préfère donc ne pas

considérer ces ménages.

Sous ces conditions 13132 ménages sont retenus dans l'échantillon. De ces ménages, 6133 ont été observés au moins deux années de suite (Voir tab. 5.2).

TABLE 5.2 – Conditions de Construction

Répartition par "Respect les conditions" et type d'observation						
	Parc <sup>a</sup>		Démographie		Revenu	
	Non	Oui	Non	Oui	Non	Oui
Total	4705	14402	371	18736	1845	17262
PANEL <sup>b</sup>	2603	7061	371	9293	1264	8400
CROSS <sup>c</sup>	2102	7341		9443	581	8862

Conditions Parc et Demo <sup>d</sup>

Demo	Parc		
	Non	Oui	Total
Non	32	25	57
Oui	747	1041	1788
Total	779	1066	1845

Tailles échantillons

	NON	OUI	Total
PANEL	3531	6133	9664
CROSS	2444	6999	9443
Total	5975	13132	19107

Nb. Réponses Annuelles

	2002	2003	2004	2005	2006	2007	2008	Total	
Parc Auto		6466	6523	6291	6257	5600	6846	7270	45253
Retenus		4294	3939	3763	3788	3385	4416	4900	28485
		66,4%	60,4%	59,8%	60,5%	60,4%	64,5%	67,4%	62,9%
PANEL		2714	3307	3089	2938	2859	3512	3067	21486
CROSS		1580	632	674	850	526	904	1833	6999

a. Défini par les conditions 5.2

b. Ménages observés plusieurs années de suite.

c. Ménages observés une seule fois

d. lorsque Revenu non respecté

### 5.2.1 Analyse Socio-Démographique

Je vais maintenant présenter une brève analyse statistique des caractéristiques socio-démographiques du ménage, dans le but d'observer les conséquences de l'utilisation des procédures décrites plus haut. De plus, souligner les relations reliant les caractéristiques du ménage à celles du parc automobile qu'il possède, pourrait permettre de mieux interpréter les résultats par la suite.

Puisque ne sont retenus ici que les ménages n'ayant pas connu de changements démographiques majeurs au cours de la période d'observation, l'analyse reportée en 5.3 concerne les conditions initiales (i.e. structure socio-démographique du ménage la première année observée) de tous les ménages retenus, qu'ils soient observés sur une ou plusieurs périodes. Je choisis de considérer la totalité de l'échantillon retenu puisque par la suite les estimations initiales (i.e. estimations statiques des modèles étudiés) seront calculées pour cet échantillon. De plus, les observations que l'on peut faire sur l'échantillon complet ne diffèrent pas de façon significative de celles que l'on peut faire en étudiant le sous-échantillon panel<sup>6</sup>.

Les variables analysées sont celles présentées en introduction de cette section :

- *Type de Chef(s)*;
- *Âge du Chef Principal*;
- *Nombre de Membres*;
- *Classe Socio-Professionnelle du Chef Principal*;

**EO** : Employé / Ouvrier,

**PI** : Professions Intermédiaires,

**CA** : Cadre,

**IA** : Inactif,

**RT** : Retraité.

Les relations reliant ces caractéristiques définissent le cadre du ménage, i.e. les conditions dans lesquelles il se trouve pour la période, ses besoins et ses ressources.

Les données reflètent certaines tendances démographiques bien connues (le lecteur peut se référer à [Pla and Beaumel \(2011\)](#) et [Mazuy and Barbieri](#)

---

6. Plus d'informations sur la page personnelle de [Cernicchiaro](#)

(2011) pour plus d'informations) et rien de surprenant est à remarquer,

La probabilité d'appartenir à un couple de chef (tab. 5.3) est la plus élevée pour les adultes d'une âge comprise entre 40 et 60 ans. L'âge moyen au moment de la première "mise en couple", que ce soit par le mariage ou par le pacs, bien-qu'en hausse continue, se place en effet toujours aux alentours de 30 ans.

L'importance de la probabilité que ces unions soient dissoutes peut expliquer en partie le retour au célibat observé pour les classes plus élevées. La mortalité en est une autre cause principale.

L'âge au décès, aussi en hausse pendant la période observée, se place autour de 80 ans. L'espérance de vie étant plus élevée pour les femmes que pour les hommes, ceci n'est pas étonnant.

En analysant le croisement de ces deux variables avec le nombre de membres, d'une part, et la classe socio-professionnelle du chef principal de l'autre, les conjectures formulées plus haut sur leurs liens sont confirmées. Les analyses croisées sont reportées en tab. 5.4.

Comme on pouvait s'y attendre, les ménages ayant un couple de chefs comptent plus souvent un ou plusieurs autres membres. Bien que la proportion d'individus vivant seuls (i.e. part des "aucun autre membre" pour les chefs seuls) est toujours plus importante que celle des "couples seuls" (i.e. part de "aucun autre membre" pour les couples de chefs), la première baisse lorsque l'âge augmente, contrairement à la deuxième qui est amplifiée.

La relation entre la retraite et l'âge est logique et dictée par la législation française. Pour ce qui concerne les non-retraités, plusieurs tendances peuvent être mises en lumière. Avant tout les femmes seules connaissent la plus forte probabilité, indépendamment de leur âge, à être inactives.

On observe principalement des niveaux socio-professionnels plus bas, tous âges et types de chef confondus. Pour les hommes en couples comme pour les femmes seules, la probabilité d'être "de classe basse" est moins importante lorsque l'âge du chef est supérieur à 40 ans, alors que pour les hommes seuls on observe l'effet inverse.



TABLE 5.3 – Distribution Type Chef par Age du Chef Principal

	[18; 40]	]40; 60]	]60; 70]	]70; 99]	Total
Couple de Chefs <sup>a</sup>	67,01	69,20	66,19	55,32	8699
Cehf Femme	20,50	19,05	22,49	31,10	2833
Chef Homme	12,50	11,75	11,32	13,57	1600
Total	5098	4400	2094	1540	13132

a. En accord avec la nomenclature de l'enquête, lorsqu'il s'agit d'un couple de chefs, l'homme est ici considéré comme le chef principal.

TABLE 5.4 – Analyse par Type de Chef et Age du Chef Principal

Distribution par Nombre de Membres autres que Chef(s)

	[18; 40]			]40; 60]			]60; 70]			]70; 99]			Total
	Coup.	F.	H.	Coup.	F.	H.	Coup.	F.	H.	Coup.	F.	H.	
0	33,75	69,38	91,05	36,75	56,92	76,98	88,24	91,51	90,72	94,13	93,11	89,00	7755
1	27,11	17,22	7,38	23,94	26,49	16,05	9,45	7,01	8,02	5,05	5,85	11,00	2464
2	28,69	9,95	1,41	26,24	11,93	6,00	1,73	1,06	0,84	0,59	0,63	0,00	2062
>= 3	10,45	3,44	0,16	13,07	4,65	0,97	0,58	0,42	0,42	0,23	0,42	0,00	851

Distribution par Classe Socio-Professionnelle du Chef Principal

	[18; 40]			]40; 60]			]60; 70]			]70; 99]			Total
	Coup.	F.	H.	Coup.	F.	H.	Coup.	F.	H.	Coup.	F.	H.	
EO	49,39	44,78	34,54	38,36	40,93	37,33	0,43	3,61	2,11	0	0	0	4107
PI	28,54	25,84	29,67	27,59	27,33	26,31	1,23	1,27	2,95	0	0	0	2669
CA	20,29	13,78	26,69	19,90	13,96	19,73	2,16	0,85	1,27	0	0	0	1869
IA	1,73	15,41	8,95	2,07	9,31	5,03	0,36	9,98	1,27	0,23	13,15	0	564
RT	0,06	0,19	0,16	12,09	8,47	11,61	95,82	84,29	92,41	99,77	86,85	100,00	3923

### 5.2.2 Analyse par Comparaison avec les Données Brutes

Le processus de construction de l'échantillon n'ayant pas pour objectif d'en conserver la représentativité, il me semble important étudier sa structure, avant d'analyser son contenu.

Dans ce but, les distributions annuelles de plusieurs variables caractérisant le ménage (5 variables) et sa mobilité automobile (2 variables) sont comparées à celles observées pour l'échantillon complet (représentatif).

Afin de mesurer les parts de perte de représentativité, lorsque celle-ci est observée, dues respectivement à la procédure de sélection et au caractère cross-section de l'observation du ménage (i.e. ménage répondant à l'enquête une seule année), sont aussi reportées les distributions obtenues dans l'échantillon panel complet (i.e. tout ménage ayant été observé au moins 2 années de suite) et celui retenu complet (i.e. tout ménage répondant aux conditions de sélection).

Les variables considérées sont le revenu (tab. 5.5) et la taille (tab. 5.6) du ménage, le type de chef (tab. 5.7), la classe d'âge du chef principal (tab. 5.8) et sa classe socio-professionnelle (tab. 5.9) pour la description démographique, et le nombre de véhicules composant le parc automobile du ménage (tab. 5.10) et sa classe de demande de kilométrage (tab. 5.11) pour la mobilité automobile.

On remarque que les ménages appartenant aux classes "basses" de revenu ( $\leq 26700\text{€}$ ) sont légèrement sur-représentés dans les échantillons retenus (tab. 5.5). Cependant les répartitions annuelles restent assez proches de celles observées en France, et la perte de représentativité ne semble pas significative.

Les conditions imposées au cours de la construction de l'échantillon impactent par contre la répartition des ménages selon leur structure et leur parc automobile, comme on pouvait s'y attendre,

Le "petit ménage" (i.e. moins de 3 membres) est sur-représenté dans l'échantillon retenu. Cela pourrait signifier que lorsque le nombre de membres est plus élevé le ménage a tendance à moins bien répondre. En regardant plus dans le détail on s'aperçoit que ce déséquilibre est plus important lorsque l'on considère toutes les observations panélistées. Ce type de ménage pourrait donc préférer ne pas répondre à l'enquête, après avoir constaté une première fois le nombre d'informations à fournir. Les deux effets allant dans le même sens (i.e. petit ménage sur-représenté dans les deux cas) la perte est amplifiée dans l'échantillon retenu.

Des considérations similaires peuvent être faites en comparant les distribu-

tion par type de chef. Cela n'est pas surprenant si, comme on pourrait si attendre, les ménages ayant un couple de chefs comptent plus souvent un nombre important de membres.

En observant l'échantillon selon l'âge du chef principal on remarque une sur-représentation de "vieux chef" ( $> 70$  ans). Les distributions partielles montrent que la cause principale de cette perte de représentativité est la sortie plus importante des "jeunes chefs" après une année de participation à l'enquête. Ces derniers répondent par contre légèrement mieux (i.e. sur-représentés dans l'échantillon retenu complet). La perte de représentativité est alors atténuée dans l'échantillon retenu par l'interaction de ces deux effets.

Les mêmes effets se reflètent, logiquement, sur la distribution par catégorie socio-professionnelle qui voit une sur-représentation des chefs retraités.

La catégorie socio-professionnelle est définie comme défini plus haut (**EO** : Employé / Ouvrier, **PI** : Professions Intermédiaires, **CA** : Cadre, **IA** : Inactif, **RT** : Retraité).

Aucune variation remarquable ne semble affecter par contre la distribution par classe socio-professionnelle pour les chefs non retraités.

On peut maintenant s'intéresser aux variables portant sur l'utilisation automobile.

Avant d'analyser les distributions par nombre de véhicules composants le parc automobile annuel, il faut rappeler que les ménages "super-équipés" (i.e. plus de 3 véhicules) ne sont pas retenus dans l'échantillon. Cette décision découle du fait qu'il est demandé aux enquêtés de ne décrire que 3 véhicules maximum. Puisque le kilométrage total, qui comme on verra au cours des chapitres suivants joue un rôle clé dans les processus étudiés, n'est pas collecté directement mais doit être calculé comme somme des kilométrages parcourus avec chaque véhicule, les ménages "super-équipés" ne sont pas exploitables pour les besoins du travail. En observant les distributions 5.10 on voit que de toute façon ce type de parc est très peu présent dans l'échantillon représentatif ( $< 1\%$ ) et son exclusion n'a pas de répercussions remarquables.

TABLE 5.5 – Comparaison Répartition Revenu

Classe en k€	ECHANTILLON 2002 – 2008								ECHANTILLON 2002 – 2008 retenu <sup>a</sup>							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
< 15, 2	25,56	25,42	26,70	23,42	24,85	20,57	19,07	25,31	24,30	25,30	22,60	24,05	19,50	18,57		
15, 2 – 19, 1	13,28	12,89	12,91	12,67	12,76	11,24	10,85	13,51	13,40	12,81	12,70	12,82	11,57	10,73		
19, 1 – 22, 9	11,94	11,24	11,29	11,96	11,41	11,85	12,31	12,13	11,63	11,56	12,80	11,91	11,71	12,49		
22, 9 – 26, 7	10,39	10,15	11,17	11,02	10,74	11,88	10,77	10,88	10,26	11,48	11,22	11,26	12,09	10,90		
26, 7 – 30, 5	9,56	9,87	7,56	8,49	8,42	8,28	8,68	9,34	9,93	7,63	8,16	8,77	8,20	8,71		
30, 5 – 38, 1	12,90	13,45	12,74	13,47	13,61	14,97	15,27	13,20	13,63	13,45	13,31	13,00	15,15	15,14		
38, 1 – 45, 7	6,76	7,62	7,59	8,24	7,59	8,50	9,57	6,17	7,01	7,52	8,24	7,71	8,72	9,49		
45, 7 – 61	6,39	5,93	6,38	6,40	6,79	7,88	8,12	6,15	6,37	6,59	6,36	6,74	7,88	8,47		
> 61	3,23	3,42	3,65	4,33	3,84	4,84	5,36	3,31	3,48	3,67	4,62	3,75	5,19	5,49		
Missing	327	473	429	369	340	272	333									

Classe en k€	ECHANTILLON PANEL <sup>b</sup> 2002 – 2008								ECHANTILLON PANEL 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
< 15, 2	26,54	25,86	27,09	24,27	25,50	20,52	19,24	26,23	24,83	25,41	23,25	24,52	19,19	18,13		
15, 2 – 19, 1	13,94	13,38	13,23	13,20	13,16	11,99	11,52	14,70	14,03	13,27	13,27	13,36	12,44	11,57		
19, 1 – 22, 9	12,07	11,45	11,47	12,03	11,62	12,39	13,04	12,60	11,82	11,85	13,04	12,10	12,19	13,53		
22, 9 – 26, 7	10,21	10,27	11,29	10,85	10,88	11,82	10,81	10,69	10,43	11,75	11,27	11,44	12,24	11,18		
26, 7 – 30, 5	9,63	9,80	7,55	8,59	8,41	8,18	8,77	9,17	9,89	7,77	8,51	8,85	8,17	8,90		
30, 5 – 38, 1	12,27	12,98	12,42	13,37	13,29	14,64	15,08	12,53	13,15	12,95	13,17	12,56	14,81	15,06		
38, 1 – 45, 7	6,44	7,50	7,63	8,03	7,29	8,13	8,67	5,53	6,77	7,54	7,86	7,45	8,37	8,38		
45, 7 – 61	6,28	5,67	5,95	6,04	6,46	7,74	7,81	5,93	5,96	6,02	5,79	6,37	7,55	8,02		
> 61	2,62	3,09	3,36	3,63	3,39	4,60	5,06	2,62	3,11	3,43	3,85	3,36	5,04	5,22		
Missing	193	383	364	286	299	213	224									

a. Ménages ayant déclaré leur revenu au cours de toute l'observation, possédant < 4 véhicules identifiables et qui ne connaît pas de changements démographiques majeurs

b. Ménages ayant répondu au moins 2 années de suite

TABLE 5.6 – Comparaison Distribution Taille Ménage

	ECHANTILLON 2002 – 2008								ECHANTILLON 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
1	26,43	26,90	28,44	27,98	30,14	28,29	28,01	28,57	29,96	31,33	30,39	32,91	30,41	30,57		
2	40,98	41,59	39,48	40,42	41,91	39,35	39,60	40,68	41,51	39,38	40,23	42,39	39,54	39,22		
3	14,07	13,17	13,42	13,28	12,52	13,83	13,73	13,41	11,96	12,25	12,54	11,11	13,09	13,24		
4	12,30	12,52	13,19	13,39	10,59	13,16	13,54	11,48	11,40	12,25	12,35	9,31	12,39	12,57		
5	5,12	4,72	4,61	4,01	4,05	4,34	4,25	4,98	4,32	4,04	3,91	3,72	3,83	3,73		
6	0,90	0,94	0,67	0,77	0,68	0,88	0,66	0,72	0,79	0,61	0,50	0,50	0,63	0,49		
7	0,17	0,11	0,16	0,10	0,07	0,10	0,19	0,14	0,08	0,13	0,03	0,03	0,07	0,12		
8	0,02	0,03	0,02	0,03	0,02	0,04	0,03	0,00	0,00	0,00	0,05	0,03	0,05	0,04		
9	0,02	0,02	0,02	0,02	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		

	ECHANTILLON PANEL 2002 – 2008								ECHANTILLON PANEL 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
1	28,51	28,39	29,93	30,60	31,54	30,26	30,64	31,95	31,57	33,02	33,42	34,31	32,97	33,39		
2	43,18	42,53	41,01	42,07	42,72	40,72	41,44	42,93	42,27	40,89	42,44	43,48	41,06	41,41		
3	12,50	12,15	12,23	11,49	11,55	12,63	12,20	11,27	10,92	10,84	10,38	9,86	11,56	10,92		
4	10,37	11,47	11,99	11,43	9,92	11,73	11,15	9,17	10,43	11,10	10,04	8,81	10,68	10,34		
5	4,46	4,47	4,03	3,66	3,55	3,80	3,79	4,09	4,05	3,46	3,27	3,11	3,22	3,52		
6	0,84	0,86	0,68	0,63	0,62	0,71	0,63	0,55	0,70	0,62	0,37	0,38	0,43	0,33		
7	0,09	0,09	0,09	0,08	0,06	0,09	0,12	0,04	0,06	0,06	0,03	0,00	0,03	0,07		
8	0,02	0,02	0,02	0,02	0,02	0,05	0,02	0,00	0,00	0,00	0,03	0,03	0,06	0,03		
9	0,02	0,02	0,02	0,02	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		

TABLE 5.7 – Comparaison Distribution Type de Chef

	ECHANTILLON 2002 – 2008								ECHANTILLON 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
Couple	66,18	66,00	65,20	65,35	63,11	64,75	64,72	64,32	63,32	62,77	63,33	60,65	62,75	62,12		
Chef F,	22,18	22,17	21,92	22,04	23,50	22,80	23,34	23,10	23,25	22,56	22,41	24,14	23,32	24,71		
Chef H,	11,65	11,84	12,88	12,61	13,39	12,45	11,94	12,58	13,43	14,67	14,26	15,21	13,93	13,16		

	ECHANTILLON PANEL 2002 – 2008								ECHANTILLON PANEL 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
Couple	63,95	64,37	63,64	62,81	61,70	62,95	62,73	60,94	61,60	61,15	60,69	59,15	60,36	60,32		
Chef F,	23,48	23,24	23,05	23,67	24,64	23,97	23,99	25,02	24,40	23,66	24,03	25,32	24,80	24,75		
Chef H,	12,57	12,39	13,31	13,52	13,66	13,08	13,28	14,04	14,00	15,18	15,28	15,53	14,83	14,93		

TABLE 5.8 – Comparaison Distribution Age du Chef

	ECHANTILLON 2002 – 2008								ECHANTILLON 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
[18; 40]	26,12	25,42	27,12	26,27	23,96	27,87	27,51	28,55	27,29	30,14	28,83	26,74	31,16	30,27		
]40; 60]	34,50	33,54	32,92	32,70	32,77	33,27	34,32	33,65	32,62	31,49	31,57	31,34	31,93	33,61		
]60; 70]	21,45	21,11	19,92	19,98	20,16	18,67	18,46	21,63	21,58	20,30	20,22	20,44	18,59	18,31		
]70; 99]	17,92	19,93	20,04	21,05	23,11	20,19	19,71	16,16	18,51	18,07	19,38	21,48	18,32	17,82		

	ECHANTILLON PANEL 2002 – 2008								ECHANTILLON PANEL 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
[18; 40]	21,75	22,51	22,13	21,15	20,79	23,17	20,73	23,03	23,95	23,76	22,57	22,42	25,60	22,89		
]40; 60]	33,36	33,35	33,19	32,25	32,36	33,65	32,69	32,68	32,57	32,37	31,14	31,51	32,66	31,86		
]60; 70]	24,70	22,49	22,09	22,26	21,70	20,43	21,69	25,87	23,22	23,11	23,21	22,39	20,73	22,24		
]70; 99]	20,20	21,66	22,60	24,34	25,15	22,74	24,89	18,42	20,26	20,75	23,08	23,68	21,01	23,02		

TABLE 5.9 – Comparaison Distribution Classe Socio-Professionnelle du Chef

	ECHANTILLON 2002 – 2008								ECHANTILLON 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
EO	28,75	27,52	27,32	27,55	26,18	28,24	29,37	28,55	26,53	26,52	27,03	25,26	27,65	28,94		
PI	16,78	16,97	16,77	15,97	14,66	16,33	16,44	17,37	17,36	17,22	16,18	15,01	16,73	17,16		
CA	9,70	9,37	9,54	9,92	9,89	11,28	11,60	10,53	10,48	10,71	11,46	11,05	12,73	12,96		
IA	4,59	5,03	3,07	3,12	3,27	3,52	3,09	4,31	4,72	3,14	3,01	3,43	3,69	3,31		
RT	40,18	41,12	43,30	43,44	46,00	40,64	39,50	39,24	40,90	42,41	42,32	45,26	39,20	37,63		

	ECHANTILLON PANEL 2002 – 2008								ECHANTILLON PANEL 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
EO	26,18	26,90	25,69	25,44	25,25	26,71	25,21	25,39	25,82	24,86	24,54	24,31	25,68	24,45		
PI	15,32	15,68	15,04	14,25	13,52	14,98	14,24	15,81	16,06	14,89	14,19	13,29	15,21	14,41		
CA	8,31	8,66	8,14	8,09	8,47	10,16	9,82	8,84	9,65	9,36	9,02	9,44	11,59	11,22		
IA	4,88	4,90	2,96	3,02	3,14	3,32	2,99	4,50	4,57	2,75	2,89	3,32	3,42	3,16		
RT	45,31	43,86	48,17	49,19	49,62	44,83	47,73	45,47	43,91	48,14	49,35	49,63	44,11	46,76		

TABLE 5.10 – Comparaison Distribution Taille Parc

	ECHANTILLON 2002 – 2008								ECHANTILLON 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
0	13,39	14,12	13,42	13,01	13,96	11,74	11,80	15,00	16,12	15,68	15,15	16,51	13,45	13,37		
1	51,25	50,41	49,02	49,02	49,75	47,68	47,39	53,70	54,02	53,39	52,61	53,44	51,15	50,96		
2	29,74	30,26	31,84	32,11	30,71	34,28	34,65	28,20	27,49	28,20	29,14	27,09	32,27	32,63		
3	4,96	4,77	5,18	5,26	5,07	5,39	5,21	3,10	2,36	2,74	3,09	2,95	3,13	3,04		
> 3	0,65	0,44	0,54	0,61	0,50	0,91	0,95									

	ECHANTILLON PANEL 2002 – 2008								ECHANTILLON PANEL 2002 – 2008 retenu							
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008		
0	14,48	14,58	14,29	14,39	14,38	12,28	12,61	17,06	16,90	16,74	16,95	16,93	14,04	14,22		
1	54,11	52,14	50,32	50,87	50,65	49,56	49,88	56,78	55,43	54,35	53,95	54,21	53,25	53,51		
2	26,85	28,63	30,10	29,62	29,85	32,34	32,05	24,06	25,55	26,32	26,55	26,30	30,04	29,80		
3	4,21	4,39	4,94	4,88	4,73	5,03	4,77	2,10	2,12	2,59	2,55	2,55	2,68	2,48		
> 3	0,35	0,25	0,36	0,24	0,39	0,80	0,70									

On observe tout de même un déséquilibre dans la distribution retenue par rapport à la référence. En effet les “non-équipés” et “mono-équipés” sont sur-représentés que l’on observe que les “répondants multiples”, que les “bons répondants” ou le croisement des deux.

TABLE 5.11 – Comparaison Distribution Kilométrage

	ECHANTILLON 2002 – 2008							ECHANTILLON 2002 – 2008 retenu						
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008
0	15,74	17,22	16,71	16,14	18,03	15,30	15,42	15,49	16,68	17,01	16,37	18,61	15,76	15,47
< 2500	4,20	4,73	5,06	4,77	4,95	4,42	5,21	3,89	4,39	4,78	4,59	4,87	4,26	5,20
[2500, 5000[	5,39	5,05	5,10	5,87	5,89	6,03	5,99	5,45	5,23	5,31	5,68	6,00	5,80	5,88
[5000, 10000[	14,13	14,71	14,38	14,74	15,11	14,45	14,68	14,39	15,41	14,72	15,55	15,24	15,24	14,69
[10000, 15000[	16,99	15,68	15,72	16,19	15,64	15,34	15,97	17,61	16,50	16,02	16,82	16,40	15,49	16,63
[15000, 20000[	13,13	12,51	12,40	12,89	12,92	12,40	12,10	13,69	12,92	13,45	12,67	12,41	12,77	12,29
[20000, 35000[	20,72	20,57	20,30	19,40	18,63	20,45	20,21	20,38	20,36	19,74	19,35	18,55	20,52	20,43
>= 35000	9,68	9,53	10,34	9,99	8,82	11,61	10,43	9,11	8,50	8,96	8,98	7,92	10,17	9,41
Missing	756	982	838	804	748	763	797							

	ECHANTILLON PANEL 2002 – 2008							ECHANTILLON PANEL 2002 – 2008 retenu						
	2002	2003	2004	2005	2006	2007	2008	2002	2003	2004	2005	2006	2007	2008
0	17,02	17,62	17,35	17,43	18,61	15,81	16,37	17,43	17,36	17,51	17,70	19,13	16,23	16,07
< 2500	4,72	5,04	5,35	5,10	5,28	4,72	5,35	4,31	4,72	5,11	4,97	5,25	4,56	5,28
[2500, 5000[	5,81	5,36	5,70	6,53	6,42	6,50	6,95	6,04	5,59	6,05	6,47	6,65	6,26	6,81
[5000, 10000[	15,48	15,49	15,43	15,70	15,97	15,29	16,55	16,21	16,27	15,83	16,81	16,37	16,29	16,86
[10000, 15000[	17,60	15,91	15,96	16,25	15,97	15,93	16,71	18,39	16,57	16,12	16,75	16,75	16,34	17,70
[15000, 20000[	12,95	12,64	12,17	12,95	12,57	12,20	11,27	13,41	13,06	13,31	12,73	11,75	12,44	11,54
[20000, 35000[	18,84	19,22	19,07	17,79	17,52	19,27	17,59	17,69	18,81	18,42	17,36	17,35	19,16	17,67
>= 35000	7,58	8,73	8,98	8,26	7,65	10,27	9,20	6,52	7,62	7,64	7,22	6,75	8,71	8,05
Missing	551	814	706	663	660	645	572							

Pour conclure cette analyse préliminaire on va s'intéresser à la classe de kilométrage annuel à parcourir en voiture. Mise à part une légère sous-représentation des deux classes supérieures, l'échantillon retenu semble rester représentatif pour ce qui concerne la demande de kilométrage.

### 5.3 L'Analyse Préliminaire de la Problématique

Je vais présenter ici l'analyse descriptive de l'utilisation automobile des ménages, selon leurs caractéristiques. Les observations ici utilisées sont toutes les 21486 disponibles pour l'échantillon panel retenu, analysées de façon intertemporelle.

Les distributions par nombre de véhicules utilisées pendant une année et classe de demande de kilométrage selon les caractéristiques socio-démographiques précédemment étudiées, sont reportées respectivement en tab. 5.13 et 5.12. Les observations possibles sont plutôt intuitives et en ligne avec ce que l'on pourrait attendre.

L'équipement en automobile augmente lorsque le budget est plus important, aussi bien qu'en fonction de la taille du ménage. Une observation intéressante est que la portion de ménages non-motorisés diminue en considérant des classes de revenu de plus en plus élevées, mais recommence à s'accroître au-delà d'un revenu annuel de 45700 €, sans pour autant dépasser celle des multi-équipés.

La probabilité de non-motorisation la plus importante est observée pour les ménages ayant un chef femme. En général, les mono-chefs semblent être plus souvent non-équipés ou mono-équipés que les couples de chefs.

L'âge du chef principal cause aussi des disparités dans les distributions. Les chefs les plus âgés ( $> 70$  ans) connaissent la non-motorisation plus souvent que les autres, alors que la classe "adulte" ( $[40; 60]$ ) est plus susceptible d'être "multi-motorisée" (i.e. plus d'un véhicule utilisé pendant l'année).

Lorsque le chef principal n'est pas actif (i.e. inactif ou retraité) son parc a tendance à être petit ou inexistant. Ce deuxième cas est plus souvent vérifié pour les chefs inactifs plutôt que pour les retraités. En se rappelant qu'il est plus probable pour un chef femme d'être inactif on comprend mieux la distribution observée pour les ménages ayant ce type de chef.

Une autre observation est que les chefs cadres sont caractérisés par la plus élevée portion de non-motorisés parmi les ménages à chef principal actif. Cela apporte plus de précision aux observations faites sur le revenu. Un cadre est en effet plus susceptible de se voir mis à disposition d'autres moyens de transport, lui permettant de ne pas se motoriser.

Des tendances tout-à-fait similaires sont mises en lumière par l'analyse du volume de kilométrage (voir tab. 5.12).

TABLE 5.12 – Distribution par Volume de Kilométrage

		Revenu								
	< 15, 2	15, 2 – 19, 1	19, 1 – 22, 9	22, 9 – 26, 7	26, 7 – 30, 5	30, 5 – 38, 1	38, 1 – 45, 7	45, 7 – 61	> 61	
Total	4932	2837	2671	2429	1879	2903	1602	1407	826	
0	37,75	19,60	13,70	11,20	8,52	7,44	5,87	7,46	10,53	
< 2500	7,46	6,91	5,99	4,49	3,35	2,34	2,56	2,13	1,69	
[2500, 5000[	7,64	9,41	7,49	5,80	5,22	4,65	3,12	2,77	4,60	
[5000, 10000[	16,52	19,60	19,21	18,24	16,34	14,67	13,61	11,16	10,05	
[10000, 15000[	12,41	17,87	18,16	20,75	19,43	18,26	17,54	16,42	14,29	
[15000, 20000[	7,16	11,28	14,00	13,01	14,90	16,29	16,42	16,06	12,47	
[20000, 35000[	7,58	12,13	17,07	19,88	23,42	24,91	26,84	27,65	30,63	
$\geq 35000$	3,47	3,21	4,38	6,63	8,83	11,44	14,04	16,35	15,74	

TABLE 5.13 – Distribution par Taille du Parc

		Revenu									
		< 15, 2	15, 2 – 19, 1	19, 1 – 22, 9	22, 9 – 26, 7	26, 7 – 30, 5	30, 5 – 38, 1	38, 1 – 45, 7	45, 7 – 61	> 61	
Total		4932	2837	2671	2429	1879	2903	1602	1407	826	
0		36,96	18,68	12,02	9,72	7,34	6,20	4,18	6,18	8,23	
1		54,08	65,91	66,15	60,11	52,58	50,50	42,88	35,61	35,47	
2		8,54	14,28	20,59	28,41	37,20	39,48	47,88	51,95	48,79	
3		0,43	1,13	1,24	1,77	2,87	3,82	5,06	6,25	7,51	

		Taille du Ménage							
		1	2	3	4	5	6	7	8
Total		7076	9028	2330	2176	758	104	9	5
0		35,73	7,24	7,21	3,31	2,77	5,77	11,11	0
1		62,18	60,41	37,34	32,44	29,68	36,54	44,44	40,00
2		1,98	30,67	47,94	58,09	60,42	54,81	44,44	60,00
3		0,11	1,67	7,51	6,16	7,12	2,88	0	0

		Âge du Chef			
		[18; 40]	]40; 60]	]60; 70]	]70; 99]
Total		5056	6904	4916	4610
0		14,10	13,53	11,41	26,94
1		45,51	47,86	64,85	63,04
2		39,04	33,20	22,31	9,74
3		1,34	5,42	1,42	0,28

		Type de Chef		
		Couple de Chefs	Chef Femme	Chef Homme
Total		13024	5277	3185
0		4,43	40,80	22,61
1		49,58	56,60	70,80
2		42,13	2,33	6,34
3		3,86	0,27	0,25

		CSP du Chef				
		EO	PI	CA	RT	IA
Total		5379	3193	2134	10025	755
0		13,46	9,93	13,96	17,08	52,85
1		47,07	46,88	44,05	64,12	40,00
2		36,01	38,71	38,10	17,70	6,89
3		3,46	4,48	3,89	1,11	0,26



Taille du Ménage

	1	2	3	4	5	6	7	8
Total	7076	9028	2330	2176	758	104	9	5
0	36,98	8,60	8,41	4,23	3,96	5,77	11,11	0,00
< 2500	7,53	4,65	2,06	1,56	1,58	0,96	11,11	0,00
[2500, 5000[	8,66	6,69	2,75	2,30	1,58	1,92	0,00	0,00
[5000, 10000[	18,22	19,32	9,31	8,92	8,58	6,73	0,00	40,00
[10000, 15000[	12,22	21,98	14,81	13,97	16,09	11,54	0,00	20,00
[15000, 20000[	7,19	15,08	16,01	15,81	13,72	15,38	0,00	20,00
[20000, 35000[	7,17	17,84	30,21	35,39	34,96	29,81	33,33	20,00
>= 35000	2,02	5,85	16,44	17,83	19,53	27,88	44,44	0,00

Âge du Chef

	[18; 40]	]40; 60]	]60; 70]	]70; 99]
Total	5056	6904	4916	4610
0	15,45	14,75	12,84	27,94
< 2500	2,25	3,75	5,17	9,15
[2500, 5000[	3,13	4,42	6,71	11,97
[5000, 10000[	10,64	13,57	19,87	23,12
[10000, 15000[	14,95	16,57	21,24	14,95
[15000, 20000[	15,15	13,06	14,81	6,77
[20000, 35000[	25,20	23,13	16,11	4,97
>= 35000	13,23	10,75	3,25	1,13

Type de Chef

	Couple de Chefs	Chef Femme	Chef Homme
Total	13024	5277	3185
0	5,61	41,99	24,21
< 2500	3,41	7,90	5,90
[2500, 5000[	4,98	8,19	8,32
[5000, 10000[	15,39	16,92	19,47
[10000, 15000[	19,33	11,7 9	15,48
[15000, 20000[	15,68	6,27	10,52
[20000, 35000[	24,74	4,96	12,81
>= 35000	10,86	1,97	3,30

CSP du Chef

	EO	PI	CA	RT	IA
Total	5379	3193	2134	10025	755
0	14,45	11,40	15,75	18,31	53,64
< 2500	3,25	2,63	2,53	6,90	5,83
[2500, 5000[	3,89	3,38	3,94	8,99	5,70
[5000, 10000[	13,11	12,15	9,93	21,21	11,52
[10000, 15000[	16,36	14,91	15,18	18,83	8,61
[15000, 20000[	13,87	15,50	12,42	11,52	6,23
[20000, 35000[	24,43	25,62	25,26	11,75	5,70
>= 35000	10,65	14,41	15,00	2,48	2,78

### 5.3.1 Les Ménages Mono-équipés

Sur les 19107 ménages observés, on compte 40,41% de ménages *Mono-équipés*, c'est-à-dire possédant un seul véhicule par an au cours de la période. L'importance de ce pourcentage suggère l'intérêt que l'analyse d'une telle population pourrait avoir. Son comportement sera étudié au chapitre suivant.

	0 ou Multi	Mono
1Obs	57,88	42,12
+Obs	59,29	40,71
Total	11196	7911

Afin de pouvoir effectuer une analyse transversale, ne sera considéré ici que le sous-panel retenu cylindré, des ménages observés continuellement de 2002 à 2008, comptant 773 ménages (sur les 1627 de la base complète).

Les ménages retenus sont des ménages qui ne connaissent pas de changement de chef ni de sorties d'enfants. Le nombre de membres aussi bien que la classe socio-professionnelle du chef peuvent par contre varier. Les graphiques 5.1, pour le sous-panel cylindré complet, montrent la présence de passages à la retraite des chefs et de sorties de membres adultes.

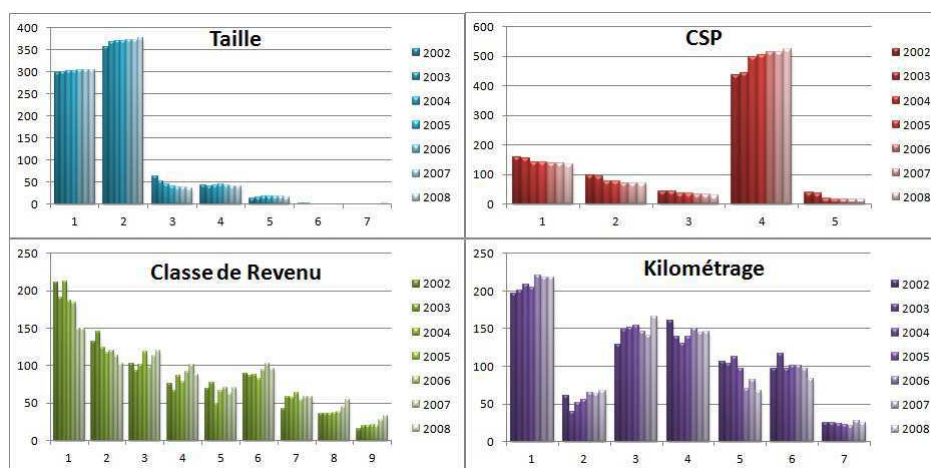


FIGURE 5.1 – Distribution échantillon cylindré

De ces 773 ménages, 423 sont mono-équipés au cours de la période. Pour cet échantillon le véhicule et le ménage sont alors un seul individu observé. La

table 5.16 reporte les mouvements des 691 véhicules possédés par ces ménages entre 2002 et 2008. Sur la diagonale on observe le nombre d'achat par année. La somme des différences par colonne est le nombre de véhicules abandonnés.

TABLE 5.14 – Mouvement des Véhicules de l'échantillon Mono-équipés

	2002	2003	2004	2005	2006	2007	2008	Total
2002	423	377	335	291	257	227	204	2114
2003	0	46	42	36	35	32	28	219
2004	0	0	46	43	40	38	37	204
2005	0	0	0	53	49	47	41	190
2006	0	0	0	0	42	38	36	116
2007	0	0	0	0	0	41	37	78
2008	0	0	0	0	0	0	40	40

Les distributions observées pour le sous-échantillon des mono-équipés (fig. 5.2) ne diffèrent pas significativement, mise à part qu'il inclut des ménages de taille tendanciuellement plus réduite.

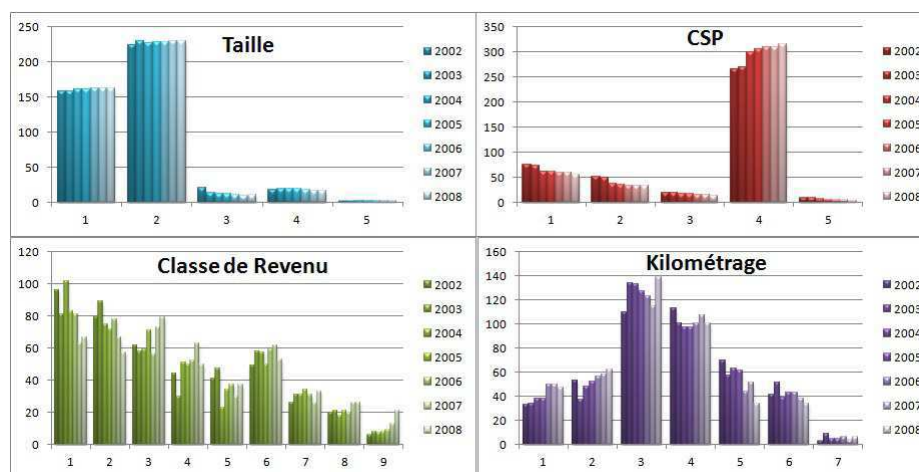


FIGURE 5.2 – Distribution échantillon cylindré Mono-équipés

De ces ménages, 219 abandonnent 268 véhicules au cours de l'observation. Leurs mouvements sont reportés en table 5.16.

TABLE 5.15 – Mouvement des Véhicules Abandonnés (Mono-équipés)

	2002	2003	2004	2005	2006	2007	2008	Total
2002	219	173	131	87	53	23	686	
2003	0	18	14	8	7	4	51	
2004	0	0	9	6	3	1	19	
2005	0	0	0	12	8	6	26	
2006	0	0	0	0	6	2	8	
2007	0	0	0	0	0	4	4	
Total	219	191	154	113	77	40	794	

On observe 160 abandons d'essence et 108 de diesel.

Confrontant les caractéristiques observées pour ces véhicules au moment de leur abandon (tab. 5.16) avec celles observées dans l'échantillon retenu (voir section 5.3.2), on remarque que ces ménages ont tendance à acheter des véhicules plus jeunes. Par conséquent le prix d'achat est légèrement plus élevé. L'âge à l'abandon est aussi plus élevée, mais la durée de possession ne varie pas significativement comme son utilisation (kilométrage au compteur).

TABLE 5.16 – Caractéristiques des Véhicules Abandonnés (Mono,cylindré)

	Carb.	Moyenne	Ecart-type	Minimum	Maximum
Diesel	Âge Abd.	7,07	4,80	1	20
	Compteur	115663,60	84655,62	950	396000
	Consommation	6,33	0,98	4	9
	Âge Ach.	1,31	2,59	0	12
	Prix Ach.	15285,84	6768,56	1500	34000
Essence	Âge Abd.	8,79	4,47	1	21
	Compteur	100993,52	62549,08	0	427274
	Consommation	7,68	1,42	4,5	15
	Âge Ach.	1,47	2,89	0	14
	Prix Ach.	11726,81	5884,21	630	36900



En table 5.17 sont reportés les mouvements annuels pour chaque cohorte de véhicules, une cohorte étant formée par les véhicules observés pour la première fois la même année.

Sur la diagonale du tableau des Observations on observe alors la totalité des entrées observées chaque année. Les sorties sont obtenue par différence de deux colonnes adjacents.

La part des entrés due à l'achat est reportée au tableau suivant (Achats). Le tableau final (Abandons) présente la part des sorties due à la décision du ménage de se séparer du véhicule.

Une première observation est que les véhicules utilisant un type de carburant autre que diesel ou essence ne sont presque pas représentés ici (0,9% de l'échantillon retenu). Seulement ces deux types de carburant seront alors pris en considération dans ce travail.

TABLE 5.18 – Conditions à l'Acaht

	Diesel				
	Nb.	Moy.	Ec.-type	Min.	Max.
Âge à l'achat	4709	2,42	3,75	0	25
Prix à l'achat	4272	14125,77	7676,47	600	72000
Consommation	4683	6,53	1,26	2,00	13,90

	Essence				
	Nb.	Moy.	Ec.-type	Min.	Max.
Âge à l'achat	5210	3,34	4,86	0	43
Prix à l'achat	4756	9477,00	5979,09	525	70000
Consommation	5181	7,42	1,45	3	17

	Autre				
	Nb.	Moy.	Ec.-type	Min.	Max.
Âge à l'achat	87	3,08	4,67	0	19
Prix à l'achat	73	13666,32	8707,86	760	31198
Consommation	83	8,22	2,37	4	13

Plusieurs différences peuvent être remarquées entre les véhicules diesel et à essence.

En observant les caractéristiques à l'achat moyennes par type de carburant, reportées en table 5.18, on voit déjà que les essences sont achetés plus vieilles

mais moins chères. Celles-ci sont aussi caractérisées par une consommation plus élevée, environ 1 litre en plus que les diesel tout les 100 kilomètres.

On voit aussi, en table 5.19, que si les voitures essence sont conservées plus longtemps elles sont en même temps moins utilisées au cours de la possession.

TABLE 5.19 – Conditions à l'Abandon

	Diesel				
	Nb.	Moy.	Ec.-type	Min.	Max.
Durée de Possession	1172	4,79	3,63	1	19
Âge à l'abandon	1172	7,35	5,11	1	26
Kilométrage Cumulé	1168	119607,83	87526,54	0	458838
	Essence				
	Nb.	Moy.	Ec.-type	Min.	Max.
Durée de Possession	1343	6,50	4,48	1	23
Âge à l'abandon	1343	10,56	5,94	1	47
Kilométrage Cumulé	1338	105128,30	63580,37	0	427274
	Autre				
	Nb.	Moy.	Ec.-type	Min.	Max.
Durée de Possession	31	5,06	4,31	1	19
Âge à l'abandon	31	8,97	6,19	1	22
Kilométrage Cumulé	31	102051,74	70378,60	0	276106

Une différence moyenne d'environ 14500 km est observable au compteur entre les véhicules abandonnés diesel et essence. Il est alors intéressant d'observer si l'utilisation que l'agent fait au fil du temps d'un véhicule suit ou pas la même tendance pour les diesel et les essences.

En observant table 5.20, on voit dans un premier temps que l'utilisation annuelle est logiquement aussi plus importante pour les diesel que pour les essences. En deuxième lieu, l'utilisation n'est pas la même selon la "phase de vie" du véhicule. Les véhicules semblent être moins utilisés en début de possession (première année). L'utilisation semble s'intensifier par contre en fin de possession (dernière année).

On peut remarquer aussi que ces différences d'utilisation ne semblent pas

TABLE 5.20 – Utilisation

	Diesel				
	Nb.	Moy.	Ec.-type	Min.	Max.
Moyenne Annuelle	11078	15792,37	14898,85	0	326000
Première Année (Achat)	1644	11487,49	17056,71	0	326000
Dernière Année (Abandon)	1172	17982,64	16940,48	0	325000

	Essence				
	Nb.	Moy.	Ec.-type	Min.	Max.
Moyenne Annuelle	13642	9773,74	11193,85	0	298512,00
Première Année (Achat)	1244	7673,42	14283,99	0	204670,00
Dernière Année (Abandon)	1343	10945,41	9983,44	0	141487

	Autre				
	Nb.	Moy.	Ec.-type	Min.	Max.
Moyenne Annuelle	178	10350,64	9640,89	0	80000,00
Première Année (Achat)	22	4179,05	5409,79	0	20000,00
Dernière Année (Abandon)	31	11142,87	11517,33	0	41074

être les mêmes pour les diesel et les essences. Les deux séries de kilométrages annuels, par durée de possession, ainsi que celle des moyennes tout type de carburant confondu, sont représentés en fig. 5.3. Ce que les kilométrages moyens suggéraient y est confirmé.

TABLE 5.21 – Distribution par Durée de Possession, selon le Carburant

	1	2	3	4	5	6	7	8	9	10	> 10	Total
Diesel	16,38	16,98	14,93	10,24	7,94	8,45	4,52	4,52	3,33	3,50	9,22	1172
Essence	10,35	9,61	10,65	10,65	8,19	8,41	7,00	8,12	4,10	4,39	18,54	1343
Autre	25,81	9,68	9,68	9,68	6,45	9,68	3,23	9,68	3,23	3,23	9,68	31
Total	339	331	321	266	205	215	148	165	95	101	360	2546

Les variations observées en fin de série pour les diesel peuvent venir de la faible présence de véhicules de ce type conservés à très longue durée (distributions reportées en tab. 5.21).



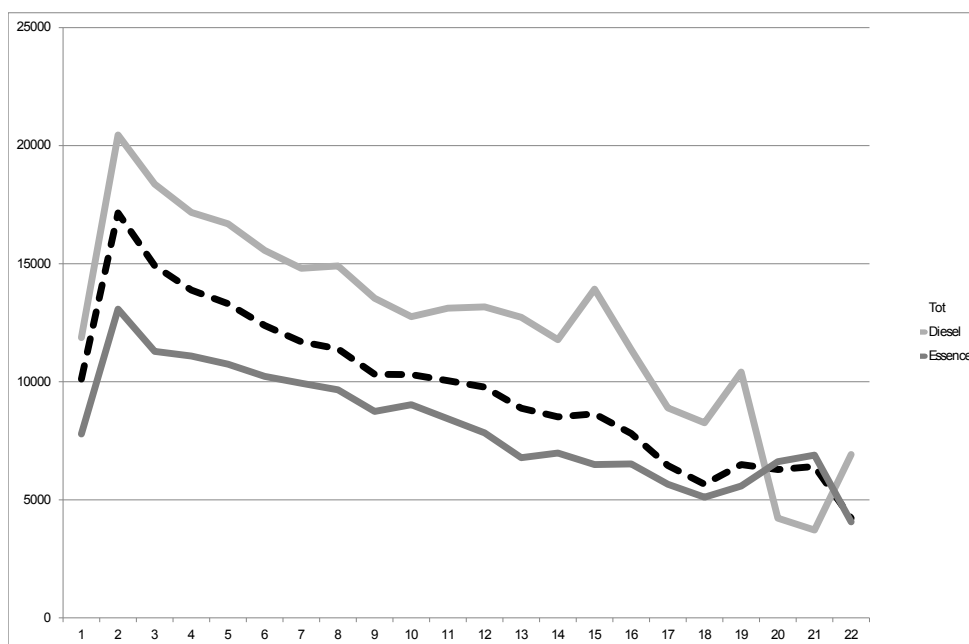


FIGURE 5.3 – Évolutions de l'Utilisation d'un Véhicule au cours de la Possession par Type de Carburant

TABLE 5.22 – Évolution de l'Utilisation Automobile

	2002	2003	2004	2005	2006	2007	2008
Parc	3018	3733	3545	3370	3273	4262	3697
Taille Moy. <sup>a</sup>							
Km. Moy. <sup>b</sup>	12545,65	12591,21	12627,23	12240,22	12132,45	12450,01	12570,56
écart type	12393,70	10786,65	14049,98	11482,85	15036,37	13009,96	15635,70

<sup>a</sup>. Par ménage

<sup>b</sup>. Par véhicule

TABLE 5.23 – évolution des Conditions à l'Achat

	2002	2003	2004	2005	2006	2007	2008
Âge	2,93	3,36	3,01	3,43	3,66	3,48	3,20
	(338)	(480)	(437)	(400)	(330)	(378)	(547)
Prix	12016,93	12359,31	13281,99	12749,73	13851,48	13706,24	12726,90
	(312)	(430)	(401)	(357)	(301)	(341)	(463)

TABLE 5.24 – évolution des Conditions à l'Abandon

	2002	2003	2004	2005	2006	2007
Durée de Poss.	5,46	5,70	5,71	5,83	5,78	5,74
Age	8,86	8,73	9,01	8,91	9,56	9,27
	(468)	(416)	(393)	(318)	(381)	(570)
Kilométrage Cum.	109014,58	105478,40	113748,60	111141,02	116426,16	114442,42
	(467)	(414)	(391)	(318)	(380)	(567)

Les caractéristiques des véhicules, aussi bien à l'achat qu'à l'abandon ainsi que l'utilisation, restent stables au cours de la période.

Comme on l'a vu plus haut, on observe un âge moyen à l'achat d'environ 3 ans (répartition reportée en tab. 5.25). Or, si l'on analyse cet âge par le biais des caractéristiques socio-démographiques du ménage on s'aperçoit de l'existence de tendances plutôt différentes.

TABLE 5.25 – Répartition par Âge à l'Achat

	Freq	Prc. %	Prc. % Cum		Freq	Prc. %	Prc. % Cum
0	12654	50,82	50,82	15	147	0,59	98,20
1	2471	9,92	60,75	16	112	0,45	98,65
2	1770	7,11	67,86	17	117	0,47	99,12
3	1549	6,22	74,08	18	70	0,28	99,40
4	1203	4,83	78,91	19	43	0,17	99,57
5	879	3,53	82,44	20	28	0,11	99,68
6	748	3,00	85,44	21	20	0,08	99,76
7	580	2,33	87,77	22	20	0,08	99,84
8	541	2,17	89,95	23	8	0,03	99,88
9	443	1,78	91,73	24	13	0,05	99,93
10	387	1,55	93,28	25	5	0,02	99,95
11	352	1,41	94,69	27	6	0,02	99,97
12	296	1,19	95,88	28	3	0,01	99,98
13	241	0,97	96,85	31	2	0,01	99,99
14	188	0,76	97,61	43	2	0,01	100,00

TABLE 5.26 – Analyse de l'Âge à l'Achat, Selon les Caractéristiques du Ménage

	Âge du Chef Principal				
	Nb.	Moy.	Ec.-type	Min.	Max.
[18; 40]	3210	3,75	4,64	0	31
]40; 60]	3799	3,10	4,60	0	43
]60; 70]	1951	1,97	3,78	0	28
]70; 99]	1046	1,34	2,94	0	18

	Csp du Chef Principal				
	Nb.	Moy.	Ec.-type	Min.	Max.
EO	3050	4,42	5,08	0	43
PI	2073	2,96	4,27	0	31
CA	1334	1,95	3,44	0	20
IA	196	3,43	4,75	0	20
RT	3353	1,84	3,63	0	28

	Nombre de Membres				
	Nb.	Moy.	Ec.-type	Min.	Max.
1	1699	2,55	4,05	0	23
2	4289	2,35	4,04	0	31
3	1699	3,45	4,55	0	25
4	1656	3,74	4,92	0	43
5	573	3,83	4,83	0	22
6	81	3,57	4,85	0	24
7	5	11,80	10,11	2	27
8	4	7,00	9,90	0	21

	Type de Chef				
	Nb.	Moy.	Ec.-type	Min.	Max.
Couple	7740	2,90	4,39	0	43
Femme	1284	2,99	4,47	0	22
Homme	982	2,80	4,26	0	23

Les âges moyens selon l'âge du chef principal, sa classe socio-professionnelle, le nombre de membres et le type de chef sont reportés en tab, [5.26](#).

Les ménages ayant un chef jeune, comptant beaucoup de membres, comme ceux dont le chef est inactif ou de classe socio-professionnelle basse semblent acheter plus vieux que les autres.

Aucune différence significative est remarquable selon le type de chef.



## Troisième partie

### Les Applications.

Les choix d'équipement automobile  
des ménages français



# Chapitre 6

## Les Prix Hédonistes des Véhicules

Au cours des dernières décades le pourcentage de ménages français choisissant la motorisation n'a pas cessé d'augmenter, faisant de la demande automobile un des sujets porteurs des recherches actuelles.

Le véhicule est un bien complexe dont le prix reflète la valeur de ces caractéristiques, comme la puissance du moteur, la consommation de carburant, les dimensions de l'habitacle, le poids du véhicule, etc. Lorsque l'on veut étudier la demande d'un tel bien il faut alors pouvoir distinguer entre ces différents effets. Or, seul le prix total est observable et les valeurs spécifiques de chaque caractéristique restent implicites. Mais puisque ces caractéristiques changent d'un véhicule à l'autre, il est nécessaire d'en connaître la valeur propre si l'on veut être capables de comparer entre eux des véhicules différents.

Des méthodes spécifiques sont donc à utiliser lorsque l'on s'attaque à une telle problématique.

Les travaux portant sur l'étude de tels prix ont été commencés par [Court \(1941\)](#), [Lancaster \(1966\)](#) et [Rosen \(1974\)](#). Ce dernier proposa de les considérer comme composition des valeurs spécifiques aux caractéristiques du bien.

Ce chapitre présente une application de la méthode des prix hédonistes aux prix d'achat neuf des véhicules. Je propose aussi l'estimation des prix d'occasion des véhicules comme fonction géométrique décroissante de l'usure du bien.

Le modèle des prix hédonistes se compose de trois étapes. Une trans-



formation Box-Cox de la consommation moyenne de carburant et du prix d'achat est estimée en premier lieu. Dans un deuxième temps la consommation transformée est estimée comme fonction de certaines caractéristiques physiques du véhicule. La fonction de prix est donc estimée pour les véhicules neufs comme fonction de la consommation et des caractéristiques du véhicule. Le prix de vente d'un véhicule usagé est alors estimé comme fonction de son prix neuf, obtenu par le biais du modèle construit à l'étape précédente.

Ce chapitre est structuré comme suit. Une description des données utilisées pour l'étude est avant tout fournie (Sec. 6.1). Dans un deuxième temps la spécification du modèle est présentée (Sec. 6.2). Les résultats obtenus sont donc commentés (Sec. 6.3) et les conclusions tirées (Sec. 6.4).

## 6.1 Les Données

L'échantillon utilisé est tiré du panel "Parc Auto", une base de données propriété de TNS-Sofres et conservée sur un système Unix à l'IFSTTAR, qui a assuré le traitement des fichiers.

L'objectif d'un tel dispositif est de fournir une source importante d'informations sur le volume et la structure du marché (neuf et occasion) et du parc automobile français. Les comportements d'usage (kilométrages et consommations), les attitudes et les opinions des ménages sont aussi capturés par l'enquête.

Les vagues annuelles du panel sont décrites à l'aide d'une enquête auto-administrée conçue par TNS-Sofres, avec le conseil scientifique de l'IFSTTAR. L'échantillon panélisé de ménages, suivi depuis 1983, est constitué de 10000 ménages représentatifs de la France entière. Il est renouvelé au taux d'environ 1/3 par an. L'enquête est menée chaque début d'année. Elle porte sur les 12 mois écoulés au cours de l'année civile précédente. Le taux de réponse annuel se situe entre 65% et 70%<sup>1</sup>.

Les ménages enquêtés sont invités chaque année à fournir, en s'aidant de

---

1. Chaque vague est pondérée après réception des questionnaires afin de conserver la représentativité de l'échantillon. Cependant l'échantillon utilisé dans ce travail est construit en utilisant une procédure d'appariement des véhicules (4.2) qui a pour objectif la minimisation de la perte d'information, sans tenir compte de la représentativité. Ne disposant pas des informations nécessaires à la construction d'un nouveau système de poids, aucune pondération est utilisée dans ce travail.

la carte grise, une brève description des véhicules composant sa flotte automobile<sup>2</sup> (marque, modèle, carburant, carrosserie, puissance fiscale, consommation moyenne), ainsi que de l'historique d'achat du véhicule (prix d'achat, année d'achat, année de première mise en circulation, achat neuf ou d'occasion, kilométrage à l'achat).

Les informations disponibles étant insuffisantes pour notre travail, la description des véhicules est à compléter par d'autres variables. Les caractéristiques physiques des véhicules, comme par exemple le poids ou la puissance maximale du moteur, sont alors obtenues par appariement du panel avec une base reportant les caractéristiques d'environ 7200 types de véhicules, construite par P. Boursin.

Puisque seule la marque et le modèle du véhicule, et pas le type, sont relevés pour les véhicules observés au cours de l'enquête TNS-SOFRES et que plusieurs types de véhicules peuvent correspondre à un même couple marque/modèle, les valeurs des variables descriptives ici considérées sont obtenues comme moyennes des valeurs observées par marque, modèle, type de carburant (en distinguant entre essence, diesel et autre) et carrosserie, dans la base descriptive.

Les types de carrosserie utilisés pour cette opération sont les suivantes :

1. *Berline* : une automobile fermée par un toit rigide fixe, avec un pare-brise fixe, quatre portes latérales, et quatre fenêtres latérales.
2. *Break* : un type de carrosserie automobile similaire à une berline, mais avec un toit qui se prolonge jusqu'à l'aplomb de l'arrière du véhicule.
3. *Coupé* : une voiture fermée à deux ou trois portes et possédant généralement deux ou quatre places ;  
*Cabriolet* : une automobile ouverte et découvrable grâce à une capote ou un toit escamotable, avec un pare-brise fixe, et des portes sans encadrement supérieur.
4. *Monospace* : un véhicule automobile ne comportant généralement qu'un seul volume, destiné avant tout à un usage familial et bénéficiant d'une carrosserie plus volumineuse que celle d'un break ordinaire. Il comporte au minimum 5 sièges, souvent distincts, et peut aller jusqu'à 8 ou 9 sièges ;

---

2. Les ménages possédant plus de 3 véhicules ne décrivent que les 3 les plus utilisés.

*Van* : un véhicule utilitaire léger dont l'espace intérieur a été aménagé pour le transport de personnes.

5. *Camping-car* : un véhicule habitable ;  
*Mobil-home* : une habitation construite sur un châssis remorquable.
6. *Utilitaire léger (Fourgonnette)* : un véhicule de transport automobile, circulant sur les routes et principalement destiné aux marchandises ;
7.  $4 \times 4$  : un véhicule tout-terrain.

Si l'observation du panel débute en 1983, c'est seulement à partir de 1996 qu'une question portant sur le prix d'achat des véhicules a été insérée dans l'enquête. Le sous-échantillon utilisé n'inclut donc que les 3345 ménages enquêtés au moins deux ans de suite entre 1996 et 2008<sup>3</sup>.

Il est composé des 4439 véhicules achetés neufs et 5472 véhicules achetés d'occasion, par les ménages retenus et dont l'achat est observé (i.e. le ménage a répondu à l'enquête l'année d'achat du véhicule).

Avant de continuer, une remarque importante est à faire sur les conséquences du type de données composant notre échantillon. En observant que des véhicules ayant été effectivement achetés l'on ne peut pas étudier l'offre et la demande séparément, mais seulement l'équilibre entre les deux. On va donc inclure, comme expliqué dans la section 6.2 des variables exogènes représentant la demande.

Après plusieurs test sont retenus pour la spécification du modèle de 9 variables continues et 4 dummies pour la description des véhicules, ainsi que 3 indicateurs concernant le ménage et son environnement<sup>4</sup>. 2 variables supplémentaires sont retenues pour l'estimation des prix d'occasion<sup>5</sup>.

#### *DESCRIPTION VEHICULE : Variables Continues*

- Consommation moyenne, en litres pour 100 kilomètres :

La consommation moyenne du véhicule est un facteur à traiter avec une attention particulière. En effet cette caractéristique, dont le prix propre participe de façon importante à la détermination du prix hédoniste du véhicule, dépend à son tour des caractéristiques du véhicule et de son moteur. La consommation observée est fonction de la consommation

---

3. Je considère ici que le risque de réponses peu fiables est élevé pour les ménages ne participant à l'enquête qu'une seule année, ils ne sont donc pas utilisés pour l'estimation

4. Statistiques descriptives : table 6.1

5. Statistiques descriptives : table ??

spécifique du moteur<sup>6</sup>, et donc de sa puissance nominale max, du poids total du véhicule, de son aérodynamisme, ainsi que de la puissance utilisée, toutes ces caractéristiques ayant une part dans la composition du prix hédoniste.

- Puissance nominale, en milliers de kW<sup>7</sup>.
- Poids du véhicule, en milliers de kg :  
La puissance nominale et le poids du véhicule sont deux facteurs cruciaux de la consommation, et ils affectent en même temps le prix d'équilibre. Il semble donc important de les inclure dans la spécification du modèle. En raison de leur corrélation importante<sup>8</sup> il faut quand même les traiter avec attention. On va donc utiliser en plus le rapport du poids sur la puissance.
- Dimensions du véhicule, en mètres :  
Selon les lois de la physique, les dimensions du véhicule (largeur, longueur, hauteur, volume et surface d'impact frontale<sup>9</sup>) ont un impact sur la consommation du véhicule. En même temps, la valeur attachée à ces caractéristiques joue aussi un rôle sur le prix d'équilibre et varie vraisemblablement avec le type de demande<sup>10</sup>.

#### *DESCRIPTION VEHICULE : Variables Dichotomiques*

- Nationalité du constructeur, en 2 classes :  
Par contrainte de confidentialité<sup>11</sup>, la marque et le modèle du véhicule, ne sont pas considérés dans ce travail. En admettant qu'il existe a minima des disparités internationales entre les producteurs (ne serait-ce que parce qu'ils cherchent à servir dans un premier temps leur marché local) et entre les consommateurs, la nationalité du constructeur est considérée dans la modélisation en distinguant entre les véhicules de production française et les autres.

---

6. Rapport entre le volume de carburant consommé et d'énergie produite.

7. On ne prends pas en compte les émissions de  $CO_2$ , le bonus-malus n'étant pas en vigueur pendant la période considérée. [Delucchi and McCubbinc \(2002\)](#) montrent de plus que le prix marginal de la qualité de l'aire est un facteur très délicat et difficile à traiter.

8. Coef. Pearson : 0,57

9. approximé avec largeur×hauteur

10. Voir variable Carrosserie

11. Vis-à-vis des constructeurs, qui auraient du être impliqués dans ce travail pour pouvoir mener une étude détaillée par firme.

- Type de carburant, en 2 classes :  
Le type de carburant est vraisemblablement un facteur non négligeable pour l'étude des prix des véhicules. En effet deux différences sont remarquables entre les véhicules diesel et à essence. D'une part les premières consomment en moyenne moins que les deuxièmes, de l'autre le prix du diesel est resté inférieur à celui de l'essence au cours de toute la période observée (et il l'est encore de nos jours). Ce pendant les prix des deux types de carburant, qui ont connu des évolutions au cours du temps caractérisées par une allure similaire (i.e. les périodes de hausse et de baisse correspondent pour les deux carburants), n'ont pas évolué à la même vitesse et le prix du diesel a "rattrapé" celui de l'essence au fur et à mesure.  
Compte tenu du faible nombre d'observations concernant les véhicules GPL, électriques ou hybrides, celles-ci sont exclues de l'analyse et je ne tiens compte que des véhicules à essence ou diesel.
- Type de carrosserie, en 3 classes :  
La carrosserie pourrait avoir un effet sur l'utilisation que l'on fait du véhicule. Il est alors raisonnable de penser que les prix propres à ses caractéristiques varient en fonction des services que le type de véhicule peut fournir (par exemple, la valeur attachée au volume du véhicule pourrait ne pas être la même s'il s'agit d'une berline ou d'une monospace). On va ici distinguer entre les berlines, les carrosseries proches des berlines (Break, Coupé et Cabriolet) et les véhicules volumineux (Monospace–Van, Fourgonnette et 4 × 4).  
Les Camping-Car et les Mobil-Home ne sont pas inclus dans l'analyse.
- Année d'Achat, en 2 classes :  
La période de construction apparaît aussi comme importante compte tenu des évolutions technologiques et des changements de la structure des préférences des consommateurs.  
Puisque l'on ne dispose pas ici de l'année de début de production pour les véhicules observés, j'utiliserai l'année d'achat, en distinguant entre deux périodes selon que le véhicule a été acheté avant ou à partir de 1999, en considérant 1999 comme l'année de la prise de conscience par rapport la raréfaction des ressources pétrolières et le prélude au phénomène de "dieselisation" du parc français.
- Climatisation, en 2 classes :  
Comme il a été montré par l'ADEME ([Barbusse and Gagnepain \(2003\)](#)), la climatisation augmente la consommation de carburant. On peut donc

s'attendre à observer un effet sur la consommation lorsque le véhicule est doté d'un système de climatisation.

#### *INDICATEURS DEMANDE*

- Taille du Ménage, discrète :  
Le nombre de personne composant le foyer a souvent un impact sur le type de service cherché et donc sur la valeur propre aux caractéristiques du véhicule que l'acheteur sera disposé à accepter. Il me semble donc important de l'inclure dans la spécification.
- Revenu annuel du ménage, en 3 classes :  
Le revenu d'un ménage détermine en partie sa contrainte budgétaire, par conséquent il peut affecter la valeur propre aux caractéristiques. Son effet ne peut donc pas être négligé lorsque l'on se trouve en situation d'équilibre.
- Localisation du ménage, en 4 classes :  
On peut supposer que l'utilisation des véhicules change lorsque le ménage est localisé en ville, ou en banlieue, plutôt que dans une zone rurale. Si la demande change, l'offre est aussi susceptible de changer. L'effet de cet indicateur est donc retenu.

#### *VENTES D'OCCASION*

Le véhicule étant un bien durable, ses performances déclinent avec l'usure. On peut alors considérer l'usure comme fonction des deux effets suivants : l'âge et le kilométrage cumulé du véhicule.

En effet, le temps qui passe, les saisons qui se succèdent, ont très vraisemblablement un rôle sur le processus de déclin du véhicule, et donc de sa valeur. L'utilisation du véhicule est aussi un facteur vraisemblablement important dans ce processus et les deux effets sont étroitement liés.

## **6.2 Le Modèle**

Comme annoncé au cours de la section précédente, on se trouve ici en Situation d'Équilibre.

Une telle situation a fait l'objet de nombreuses études dans la littérature

TABLE 6.1 – Statistique Descriptives

Variable	Freq.	Variable	Freq.
Type de Carburant		Nationalité	
Essence	2239	Française	2983
Diesel	2173	Autre Européenne	862
Autre	27	Hors Europe	594
Carrosserie		Période de Construction <sup>a</sup>	
NN	1	Avant 1999	3234
Berline	3362	À partir de 1999	1205
Break	274		
Coupé	21		
Cabriolet	18		
Monospace-Van	682		
CanpingCar-MobilHome	4		
Utilitaire leg.	67		
4 × 4	10		

Variable	N	Min	Max	Avg.
Consommation Moyenne <sup>b</sup>	4424	3,000	14,000	6,912
Puissance Nominale <sup>c</sup>	4436	0,013	0,191	0,076
Poids Total <sup>d</sup>	4384	0,600	2,842	1,231
Poids / Puissance	4384	7,190	81,538	16,641
Largeur <sup>e</sup>	4410	1,465	10,010	1,715
Hauteur	4364	1,235	2,680	1,478
Longueur	4410	3,246	5,070	4,129
Volume	4364	6,925	73,489	10,527
Surface d'impact	4364	2,013	16,817	2,539

Variable	Freq.
Taille du Ménage	
1	649
2	2064
> 3	1726
Revenu annuel du ménage	
NN	213
< 23000 <sup>f</sup>	1384
[23000, 38000]	1659
> 38000	1183
Localisation du ménage	
Ville	1209
Banlieue Intérieure	765
Banlieue Extérieure	658
Autre <sup>g</sup>	1807

*a.* Basée sur l'année de première mise en circulation

*b.* en l/100km

*c.* en milliers de kw

*d.* en milliers de kg

*e.* en mètres

*f.* en €

*g.* Couronne (Urbaine ou Rurale) et Frange Urbaine

(voir par exemple Rosen (1974), Epple (1987), Brown and Rosen (1982)). Regardons de plus près ce que cela signifie.

Lorsque l'on s'intéresse aux prix effectif de vente d'un bien, il faut prendre en compte les deux composantes du marché : l'offre et la demande.

Tout achat peut être considéré comme le résultat d'une négociation entre le vendeur et l'acheteur. Chaque partie a un prix de réserve qui correspond, pour l'acheteur au prix maximal  $P_{it}^{*d}$  qu'il est disposé à payer, et pour le vendeur au prix minimal  $P_{it}^{*o}$  auquel il est disposé à vendre.

On a donc le système d'équations suivant :

$$\begin{cases} P_{it}^{*o} = x_{it}\beta^o + \epsilon_{it}^o \\ P_{it}^{*d} = x_{it}\beta^d + \epsilon_{it}^d \end{cases}, \quad \begin{bmatrix} \epsilon_{it}^o \\ \epsilon_{it}^d \end{bmatrix} \sim N(0, \Sigma) \quad (6.1)$$

Une telle négociation, dépendante du rapport de force existant entre les deux parties, aboutit à un prix de transaction  $P_{it}$  qui est satisfaisant à la fois pour le vendeur (i.e.  $P_{it} \geq P_{it}^{*o}$ ) et pour l'acheteur (i.e.  $P_{it} \leq P_{it}^{*d}$ ). L'acheteur refusera toute offre supérieure à son prix de réserve, pour que la négociation ait lieu il faut donc que la relation  $P_{it}^{*o} \leq P_{it}^{*d}$  soit respectée<sup>12</sup>.

Comme on l'a remarqué au cours de la section précédente (Sec. 6.1), on n'observe pas tous les véhicules disponibles sur le marché mais seulement ceux qui ont été achetés. On est alors en situation d'équilibre entre offre et demande et l'on ne dispose que des prix de transactions  $P_{it}$ . De plus, aucune information ni sur les prix de réserve ni sur le type de rapport de force entre les deux négociateurs n'est disponible.

Or, lorsque l'on s'intéresse à un bien complexe (i.e. un bien composé de plusieurs caractéristiques susceptibles de se voir assigner une valeur spécifique aussi bien dans la définition de l'offre que de la demande), tel qu'un véhicule, cette négociation ne se fait pas sur la quantité de bien à acheter mais sur la quantité de chacune de ses caractéristiques. Dans notre cas on peut imaginer qu'au cours de la négociation le prix d'une caractéristique donnée variera, pour les deux acteurs, en fonction des autres éléments composant le bien. Considérons par exemple la consommation moyenne, on peut supposer que le même prix ne sera pas accepté, à parité de consommation, pour une berline ou pour une monospace.

---

12. Dans le cas contraire, l'achat est impossible



Cette situation présente donc une difficulté importante dans l'estimation qui mérite d'être traitée avec attention. En effet la simultanéité entre l'offre et la demande a dans ce cas une forme particulière qui ne peut pas être résolue en utilisant un système d'équation. Comme le dit [Rosen \(1974\)](#) “ *les consommateurs sont en situation concurrentielle bien que le prix marginal ne soit pas constant.*”

Il reprend la théorie de consommation de [Lancaster \(1966\)](#) et, en s'inspirant des travaux sur les biens complexes de [Court \(1941\)](#), propose une méthode adaptée à une telle situation, qu'il appelle la méthode des Prix Hédoniste. Dans ses travaux il propose de décomposer le prix d'un bien complexe pour étudier séparément l'effet de chacune de ses caractéristiques déterminant les prix hédonistes du bien.

[Cropper and McConnell \(1988\)](#) prouvent que le choix de formes fonctionnelles plus adaptées réduit l'erreur moyenne de l'estimation. Notamment, ils montrent que lorsque l'on est en présence de facteurs non observables ou représentés par des variables proxy les fonctions Box-Cox linéaires (voir [Box and Cox \(1964\)](#)) sont une solution satisfaisante.

La méthode des prix hédonistes à été appliquée dans plusieurs domaines, comme l'étude du prix du logement ([Harrison and Rubinfeld \(1978\)](#), [Kim and Anselinc \(2003\)](#), [Sheppard \(1999\)](#) et plus récemment [Cavailhès \(2005\)](#)) du marché du coton ([Bowman and Ethridge \(1992\)](#)) ou du marché automobile ([Cowling and Cubbin \(1972\)](#), [Goodman \(1983\)](#) and [Delucchi and McCubbinc \(2002\)](#)). [Goodman \(1983\)](#) s'intéresse au prix hédoniste du rendement du véhicule. [Delucchi and McCubbinc \(2002\)](#) essaient, à l'aide de la méthode des prix hédonistes, d'étudier le prix spécifique de la pollution dans la détermination du prix d'un véhicule.

[Cavailhès \(2005\)](#) suggère que l'utilisation de certaines variables exogènes au prix du bien, mais ayant un effet sur la forme de la fonction de prix hédonistes, améliore l'estimation. Comme précédemment annoncé, on utilise alors trois variables concernant le ménage et son environnement comme instruments dans l'estimation.

Avant de présenter la méthodologie adoptée dans ce travail, il faut prendre en compte encore une dernière difficulté.

On a vu au cours de la section précédente qu'en utilisant la consommation comme variable explicative pour le prix, on est confronté à un problème d'endogénéité. En effet, la consommation dépend d'autres caractéristiques affectant aussi le prix, comme le type de moteur, sa puissance maximale ou le poids du véhicule.

En raison de l'importance de l'effet de la consommation sur le prix, on ne peut pas éviter de prendre en compte cette caractéristique dans l'estimation du prix. On est donc confronté à un biais de simultanéité qui peut être corrigé en estimant le système d'équation suivant :

$$\begin{aligned} p &= \beta_c c + \beta_1 x_1 \\ c &= \beta_2 x_2 \end{aligned} \quad (6.2)$$

où  $p$  est le prix d'achat neuf du véhicule,  $c$  sa consommation moyenne,  $x_1$  le vecteur des variables explicatives et  $x_2$  les instruments.

La méthodologie se compose alors en trois phases.

Avant tout une transformation Box-Cox linéaire est effectuée sur le prix et la consommation. Cela consiste à estimer par la méthode des Moindres Carrés Ordinaires (MCO) le  $\lambda$  qui optimise la fonction de transformation suivante<sup>13</sup> :

$$\tilde{y} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(y) & \text{if } \lambda = 0 \end{cases} \quad (6.3)$$

où  $\tilde{y}$  est la valeur transformée de la variable  $y$ . En utilisant les variables transformées les formes fonctionnelles qui spécifient le prix et la consommation sont plus adaptées.

Une fois les deux variables transformées, on les insère dans le système d'équations présenté en 6.5 et le système est résolu en utilisant l'estimateur MCO à 2 étapes (2SLS)<sup>14</sup>.

Dans un premier temps la consommation est estimée par les MCO comme :

$$\tilde{c} = \beta_0 + \beta_2 x_2 \quad (6.4)$$

---

13. Le lecteur peut se référer aux travaux de [Box and Cox \(1964\)](#) et [Cropper and McConnell \(1988\)](#) pour une présentation plus approfondie.

14. Le modèle ne manifestant pas de problèmes d'hétéroskedasticité, les estimateurs 3SLS et 2SLS produisent les mêmes résultats.

où  $\tilde{c}$  est une transformation Box-Cox de la consommation,  $\beta_0$  est un effet constant,  $\beta_2$  un vecteur de paramètres pour les instruments  $x_2$ .

La consommation  $\hat{c}$  ainsi estimée est alors utilisée dans l'estimation de la fonction de prix :

$$\tilde{p} = \beta_c \hat{c} + \beta_1 x_1 \quad (6.5)$$

où  $\tilde{p}$  est le prix transformé et  $\beta_1$  le vecteur des effets des caractéristiques  $x_1$ <sup>15</sup>.

Comme expliqué dans la section précédente (Sec. 6.1), les vecteurs  $x_1$  des variables explicatives et  $x_2$  des instruments peuvent être décomposés en deux sous-vecteurs :

$$x_i = (x_i^v, x_i^m), \quad i = 1, 2 \quad (6.6)$$

où  $x_i^v$  est un vecteur de caractéristiques du véhicule et les  $x_i^m$  décrivent le ménage et son environnement. Puisque plusieurs facteurs n'ont vraisemblablement pas d'effet linéaire aussi bien sur la consommation que sur le prix, plusieurs variables croisées sont utilisées dans le modèle. Les effets estimés ne sont donc pas constants mais dépendent de la variation des autres caractéristiques du véhicule.

Les paramètres ainsi obtenus sont alors réestimés en utilisant la méthode des Moindres Carrés Généralisés, une méthode plus robuste que les MCO qui corrige l'hétéroschedasticité et l'autocorrélation éventuelles.

Le prix d'achat d'un véhicule d'occasion peut être estimé comme un pourcentage décroissant du prix d'achat neuf. Il décroît d'une part à cause de l'augmentation du niveau technologique du marché (ici considéré, en formulant une hypothèse très réductrice, comme dépendant du seul facteur temps) et de l'autre à cause de la dépréciation due à son utilisation.

On peut supposer alors ce prix comme étant une fonction décroissante géométrique de l'âge et du kilométrage cumulé

---

15. Le vecteur  $x_1$  inclut une constante.

$$p_t = \frac{p^{\alpha_p}}{(1 + a_t)^{\alpha_a} f(km_t, \alpha_{km})} \quad (6.7)$$

où  $p$  est le prix neuf estimé.

En considérant le marché d'occasion comme un marché parfait, le prix d'achat d'un véhicule d'occasion est égal au prix de revente d'un véhicule possédé par le ménage. On pourrait donc utiliser la relation estimée pour définir le "scrap-value" des véhicules composant la flotte, à chaque période de possession  $t$  observée.

### 6.3 Les Résultats

Comme on l'a vu au cours de la section précédente, la première étape consiste à estimer le paramètre  $\lambda$  permettant une transformation Box-Cox adaptée des variables endogènes, prix et consommation, du modèle.

En utilisant le test de Ramesy, on voit que l'hypothèse de régression linéaire simple est rejetée. Il faut donc trouver une meilleure forme fonctionnelle pour la spécification du modèle. La transformation au préalable des variables expliquées par la méthode Box-Cox permet de résoudre le problème du choix de cette forme fonctionnelle.

Les paramètres de transformation estimés sont respectivement  $\lambda_c = -0,25$  et  $\lambda_p = 0,5$  pour la consommation et le prix. Les variables transformées, utilisées par la suite, sont donc :

$$\tilde{c} = \frac{c^{-0,25} - 1}{-0,25} \quad \text{et} \quad \tilde{p} = \frac{p^{0,5} - 1}{0,5} \quad (6.8)$$

En effectuant à nouveau le test de Ramesy avec les variables ainsi transformées, l'hypothèse n'est plus rejetée et la spécification peut être retenue.

La deuxième étape consiste à estimer le système d'équation 6.5 en utilisant les variables transformées 6.8. Deux tests sont effectués pour confirmer la nécessité d'estimer simultanément ces deux variables.

Dans un premier temps le prix est régressé sur la consommation. L'effet de la deuxième sur le premier, comme supposé, est significativement différent

de 0<sup>16</sup> et explique à lui seul 6,4% du prix. La consommation ne peut donc pas être exclue de l'estimation. L'hypothèse d'endogénéité de la consommation du véhicule est alors vérifiée par le biais d'un test de Hausman. L'hypothèse n'étant pas rejetée au seuil de 1%<sup>17</sup>, les paramètres  $\beta$  du système sont obtenus par l'estimateur 2SLS.

Plusieurs formes ont été testées pour la spécification observable de la consommation et du prix. Les variables ont été retenues au seuil de 10%.

La consommation du véhicule dépend<sup>18</sup> du type de carburant, de la période de construction<sup>19</sup>, du poids, de la puissance maximale du moteur et des dimensions du véhicule. Il apparaît que ces caractéristiques n'ont pas que des effets directs mais qu'elles interagissent entre elles.

Le poids et la puissance, aussi bien que la hauteur et la largeur du véhicule, s'avèrent avoir des effets croisés. L'introduction de variables croisées dans le modèle implique que les effets estimés ne sont pas linéaires mais fonction d'autres caractéristiques. L'effet total de ces caractéristiques s'écrit donc comme<sup>20</sup> :

$$\beta_i^{Tot} = \beta_i^{const.} + \beta_{i,j}^{interact.} * f_i(j) \quad (6.9)$$

où  $i$  est l'indicateur de la caractéristique,  $\beta_i^{const.}$  un effet constant et  $\beta_{i,j}^{interact.}$  l'effet joint des caractéristiques  $i$  et  $j$ . L'effet de la variable jointe varie selon les caractéristiques.

On observe un effet positif direct de la hauteur sur le coefficient de la largeur ( $f_{Largeur}(j) = Hauteur$ ) et vice versa ( $f_{Hauteur}(j) = Largeur$ ). Les effets constants des deux caractéristiques étant négatif, les effets totaux de la largeur et de la hauteur changent de signe en fonction de la variation de l'autre caractéristique. Il existent donc une hauteur et une largeur critiques<sup>21</sup> tel que

16. Au seuil de 1%

17. Valeur du test t du paramètre estimé pour l'effet du résidu de la consommation sur le prix significatif : 3,44)

18. Les résultats de l'estimation pour la consommation et le prix sont reportés en tab. 6.2 et 6.3 respectivement.

19. Représenté ici par l'année de première mise en circulation

20. L'effet total d'une caractéristique  $x_i$  sur la variable estimée  $y$  est obtenu en dérivant  $y$  respect  $x_i$  ( $= \frac{\partial y}{\partial x_i}$ ).

21. Respectivement 1,66m et 1,76m

si la quantité d'une de ces deux caractéristiques est inférieure à sa quantité critique l'effet de l'autre sur la consommation est négatif. Si l'on considère les deux en même temps, on voit alors que si la quantité des deux caractéristiques dépasse le seuil, la somme des effets totaux est positive, et inversement la consommation diminue lorsque les deux quantités sont inférieures aux seuils. Or si les deux quantités s'éloignent en direction opposée par rapport aux points critiques respectifs, si cet éloignement suit le même pas des deux côtés ou la différence entre la quantité inférieure au seuil et ce même seuil est supérieur à cette différence pour l'autre caractéristique, la somme des deux effets totaux est négative. Si, en revanche, une des deux quantités est "plus supérieure" à son seuil que l'autre est inférieure, la consommation augmente.

Pour interpréter ces effets il faut prendre en compte le fait que la variation de ces deux caractéristiques joue sur l'aérodynamisme du véhicule.<sup>22</sup> On peut alors supposer que si le véhicule est "suffisamment bas" et large son aérodynamisme est supérieur<sup>23</sup> et par conséquent consomme moins.

Puisque l'on observe que des véhicules qui ont été achetés, cela pourrait aussi signifier que lorsque l'agent veut acheter un petit véhicule il accepte des dimensions plus importantes, tant qu'elles restent en dessous du seuil critique, à condition que la consommation diminue, alors que quand le véhicule cherché est un grand véhicule, l'agent accepte plus facilement une consommation plus élevée.

Les effets du poids et de la puissance ont été estimés différemment en fonction du type de carburant et de la période de construction. Avant tout on observe que l'effet de la période de construction n'est significatif que pour les véhicules diesel. Le choix fait en construisant les périodes de construction est celui de faire la différence entre les véhicules mis en circulation avant le début du phénomène de diésélisation et les autres. On observe alors que les diesel consomment moins que les véhicules à essence, et s'ils ont été achetés après 1999 consomment moins qu'avant.

Pour les véhicules à essence seulement le poids, et pas la puissance du moteur, semble avoir un effet significatif. Une augmentation du poids pour un véhicule à essence cause une hausse de la consommation. Pour les diesel il existe aussi bien des liens plus complexes entre ces deux effets qu'une

---

22. On utilise ici la surface frontale, approximée par le produit de la largeur fois la hauteur, comme proxy de l'aérodynamisme du véhicule.

23. En effet si la surface d'impact frontale augmente on peut supposer que le facteur de perforation de l'air aussi. Le véhicule est alors probablement plus aérodynamique.

différence entre les effets selon la période de construction. Puisque l'effet du rapport  $\frac{Poids}{Puissance}$  est inclus dans l'estimation, la puissance a un effet inverse sur le poids ( $f_{Poids}(j) = \frac{1}{Puissance}$ ) et l'effet total de la puissance dépend à la fois du poids et de la puissance elle-même, selon la relation  $f_{Puissance}(i, j) = -\frac{Poids}{Puissance^2}$ . Dans ce cas aussi, comme on l'a vu pour la hauteur et la largeur, les effets changent de signe en fonction des variations des deux variables. En analysant ces variations on pourrait y voir l'effet de la manière dont la consommation spécifique du moteur évolue en fonction de sa puissance maximale.

En effet celle-ci diminue en fonction de la puissance nécessaire pour rouler<sup>24</sup> tant que cette deuxième reste suffisamment inférieure à la puissance maximale du moteur<sup>25</sup>. À partir de ce point, plus la puissance nécessaire approche la puissance maximale, plus la consommation spécifique du moteur est importante. Puisque la puissance nécessaire augmente en fonction du poids du véhicule, à puissance (maximale) fixée, le "poids critique" que les résultats semblent indiquer pourrait correspondre au point de minimum de la courbe d'évolution de la consommation spécifique.

Rappelons que les données utilisées pour l'estimation concernent des véhicules pour lesquels la transaction a bien eu lieu. Comme expliqué dans la section précédente (Sec.6.2), pour pallier à ce problème j'utilise certaines variables représentant le ménage et son environnement comme instruments dans l'estimation. À cet effet, le revenu annuel du ménage<sup>26</sup>, sa taille et sa zone de résidence ont un effet significatif et sont retenus dans l'estimation.

Le prix de transaction du véhicule semble dépendre de la consommation, du volume, du poids, de la puissance et de la hauteur, ainsi que du type de carburant et de carrosserie du véhicule et du type d'acheteur, caractérisé par la taille du ménage, la classe de revenu annuel et la zone de résidence<sup>27</sup>.

Les véhicules diesel sont vendus à un prix supérieur aux véhicules essence, indépendamment du type de carrosserie. Cependant les véhicules ayant une "grande carrosserie", comme les monospaces, se vendent moins chers que les

---

24. qui dépend du poids et des dimensions du véhicule, pour une vitesse de marche fixée

25. Un graphique montrant ce rapport dans le cas d'un moteur à essence est présenté à l'adresse suivante : [http://www.edurete.org/pd/sele\\_art.asp?ida=2931](http://www.edurete.org/pd/sele_art.asp?ida=2931)

26. L'information ici utilisée est celle concernant le revenu net, déclaré en tranches par le ménage annuellement au cours de l'enquête.

27. L'effet de la nationalité du constructeur a été aussi testé, comme annoncé, mais lorsque les caractéristiques du véhicule et de son moteur sont utilisées dans l'estimation ceci n'est plus significatif.

véhicules du type break, coupé ou cabriolet et plus chers que les berlines. La taille du ménage a un effet négatif sur le prix d'achat du véhicule. Les ménages nombreux semblent acheter des véhicules moins onéreux. Le revenu annuel ne semble pas avoir d'effet direct sur le prix d'achat du véhicule mais un impact sur les effets de la taille du ménage et de la consommation du véhicule apparaît et ces variables sont donc incluses dans l'estimation.

La consommation a un effet négatif sur le prix, comme on pouvait si attendre, sauf pour les véhicules du type monospace, 4×4 ou fourgonnette. Pour ces types de véhicules, contrairement aux autres, le prix augmente lorsque la consommation augmente. Ceci pourrait signifier que lorsque l'on achète un gros véhicule d'autres caractéristiques, comme ses dimensions par exemple, ont un poids plus important dans le processus de négociation.

Si l'on observe le volume<sup>28</sup> on voit que son effet sur les prix de transaction est positif mais moins important pour les véhicules diesel et encore moins pour les monospaces. Pour conclure, l'effet positif du volume semble être de plus en plus grand au fur et à mesure que la taille du ménage augmente.

Le poids, comme le volume, a un impact positif sur le prix, plus important pour les essences que pour les diesel. La puissance ne semble pas avoir d'effet sur les véhicules à essence mais a un impact positif sur les prix de transaction des diesel.

Il a été dit au cours de la section précédente, que le prix d'achat sur le marché d'occasion est ici supposé être fonction décroissante géométrique de l'âge et du kilométrage cumulé. Pourtant aucune forme fonctionnelle a été proposée pour l'effet de ce deuxième.

Une première hypothèse peut être que le kilométrage entre linéairement dans la fonction géométrique (table 6.4, (1)), c'est-à-dire que la baisse de prix causé par l'augmentation du kilométrage ne dépend pas du niveau de

---

28. Approximé par le produit des 3 dimensions du véhicule.



TABLE 6.2 – Estimation : Consommation Véhicules Neufs

	Label	Estimate	Std. Dev.	T-stat.
Intercept		3,073694	0,481864	6,38
Diesel		-0,59035	0,098244	-6,01
Avant 1999 <sup>a</sup>		-0,14130	0,088440	-1,60
Diesel	Avant 1999	1,030374	0,166054	6,21
Poids		0,207812	0,046573	4,46
	Avant 1999	0,031320	0,063819	0,49
	Diesel	-0,26508	0,069281	-3,83
	Avant 1999 Diesel	0,450716	0,134164	3,36
Puissance <sup>b</sup>		-0,27785	0,519485	-0,53
	Avant 1999	0,103418	0,881957	0,12
	Diesel	4,288404	1,138956	3,77
	Avant 1999 Diesel	-10,0192	2,254842	-4,44
$\frac{Poids}{Puissance}$		-0,00114	0,003146	-0,36
	Avant 1999	0,007228	0,005168	1,40
	Diesel	0,026057	0,005362	4,86
	Avant 1999 Diesel	-0,04871	0,008979	-5,43
Largeur		-1,22922	0,295025	-4,17
Hauteur		-1,30261	0,300666	-4,33
Longueur		0,106692	0,012276	8,69
Impact Frontale <sup>c</sup>		0,740206	0,175778	4,21
Taille du ménage		0,003880	0,001396	2,78
Revenu Annuel				
	< 23000 <sup>e</sup>	-0,01713	0,003929	-4,36
	[23000; 38000] <sup>e</sup>	-0,01257	0,003643	-3,45
Zone de résidence				
	Centre Ville	0,018880	0,003569	5,29
	Banlieue Interieure	0,023229	0,004148	5,60
	Banlieue Exterieure	0,017556	0,004426	3,97
Number of Observations	4116	DF <sup>d</sup>	25	
SSR <sup>e</sup>	20,50488			
F	108,88	Pr > F	< .0001	
R-Square	0,42610	Adj R-Sq	0,42219	

DATA SOURCE : PARC AUTO

SOFTWARE : SAS

Effets Totaux			
1ère Mise En Circ.	Carb.	Const.	Interact.
<i>PoidsTot</i>			
Avant 1999	diesel	0,424768	-0,016565
Avant 1999	essence	0,239132	0,006088
[1999; 2008]	diesel	-0,057268	0,024917
[1999; 2008]	essence	0,207812	-0,00114
<i>PuisKw</i>			
Avant 1999	diesel	-5,905228	-0,016565
Avant 1999	essence	-0,24653	0,006088
[1999; 2008]	diesel	4,010554	0,024917
[1999; 2008]	essence	-0,27785	-0,00114

a. 1ère mise en circulation

b. En kw

c. Largeur \* Hauteur

d. Degrees of Freedom

e. Sum of Squared Residuals

TABLE 6.3 – Estimation 2SLS : Prix Véhicules Neufs

	Label	Estimate	Std. Dev.	T-stat.	
Intercept		7,192093	1,206353	5,96	
Berline		-4,06472	0,626348	-6,49	
Monospace <sup>a</sup>		-2,69312	0,814585	-3,31	
< 23000e <sup>b</sup>		0,440625	0,480496	0,92	
[23000; 38000]e		0,422440	0,464841	0,91	
Diesel		0,782833	0,312291	2,51	
	Berline	0,419506	0,130668	3,21	
	Monospace	1,025455	0,161260	6,36	
Taille du ménage		-0,44042	0,119435	-3,69	
	< 23000e	-0,04725	0,034321	-1,38	
	[23000; 38000]e	-0,06628	0,033839	-1,96	
	Berline	0,225951	0,053993	4,18	
	Monospace	0,267236	0,061789	4,32	
Consommation <sup>c</sup>		-1,70710	0,703678	-2,43	
	Berline	1,302642	0,267902	4,86	
	Monospace	3,355454	0,437138	7,68	
	< 23000e	-0,88409	0,323419	-2,73	
	[23000; 38000]e	-0,66703	0,305474	-2,18	
	Berline	< 23000e	0,431203	0,100716	4,28
	Berline	[23000; 38000]e	0,407478	0,092767	4,39
	Monospace	< 23000e	0,489636	0,113464	4,32
	Monospace	[23000; 38000]e	0,473548	0,103075	4,59
Volume		0,467632	0,069123	6,77	
	Diesel	-0,14135	0,055161	-2,56	
	Berline	0,066109	0,045745	1,45	
	Monospace	-0,35014	0,053165	-6,59	
	Taille du ménage	0,017742	0,009465	1,87	
Poids		2,575507	0,452542	5,69	
	Diesel	-1,19314	0,528929	-2,26	
Puissance Nominale <sup>d</sup>		-2,77990	1,785690	-1,56	
	Diesel	28,09685	3,404951	8,25	
Hauteur		-3,34293	0,340895	-9,81	
System Weighted MSE	0,8527				
Degrees of freedom	3660				
System Weighted R-Square	0,6658				

DATA SOURCE : PARC AUTO

SOFTWARE : SAS

a.  $-4 \times 4$  – *Furgonette*

b. Revenu Annuel

c. Transformation Box-Cox

d. En kw

TABLE 6.4 – Estimation : Prix d'Achat Occasion (log)

(1) $f(km_t, \alpha_{km}) = (1 + km_t)^{\alpha_{km}}$			
Label	Estimate	Std. Dev.	T-stat.
Intercept	0,38430	0,17041	2,26
Prix neuf	1,00653	0,01736	57,99
log(1+Age)	-0,63467	0,01188	-53,43
log(1+Kilométrage)	-0,02597	0,00698	-3,72
(2) $f(km_t, \alpha_{km}) = exp(\alpha_{km} km_t)$			
Label	Estimate	Std. Dev.	T-stat.
Intercept	-0,30458	0,16495	-1,85
Prix neuf	1,05091	0,01683	62,44
log(1+Age)	-0,51901	0,01082	-47,95
Kilométrage <sup>a</sup>	-0,00274	0,00014352	-19,07
Number of Observations <sup>b</sup>	5472		
DF	3		
	(1)	(2)	
SSR	2454,84978	2509,88499	
F	4879,27	5307,17	
Adj R-Sq	0,7279	0,7442	

DATA SOURCE : PARC AUTO

SOFTWARE : SAS

<sup>a</sup>. en milliers<sup>b</sup>. Used Data : véhicules achetés d'occasion

celui-ci,

Une deuxième hypothèse, qui semble à priori plus crédible, est que la forme fonctionnelle correcte soit l'exponentielle (table 6.4, (2)). Ceci revient à dire que plus le kilométrage est élevé, plus la baisse de son prix d'occasion est importante,

En observant les résultats de l'estimation on peut conclure que la forme à retenir est  $\exp(\beta km)$ . En effet le paramètre estimé est sensiblement plus significatif dans ce deuxième cas, et l'on observe aussi un  $R^2$  un peu plus important (74% de l'information est expliquée). De plus le paramètre lié au prix neuf est bien plus proche de 1 dans le premier modèle.

On peut tester cette impression ( $\alpha_p = 1$ ) pour les deux modèles. Dans le premier cas l'hypothèse n'est pas rejetée ( $F = 0.14$ ,  $Pr > F = 0,7068$ ), alors que dans le deuxième cas la probabilité est inférieure à 1% et l'égalité est rejetée ( $F = 9,15$ ,  $Pr > F = 0,0025$ ).

Considérer un paramètre unitaire pour  $p$  revient à dire que lorsque l'on achète un véhicule neuf et on le revend sans l'avoir utilisé on obtient le même prix payé pour l'achat<sup>29</sup>. Or, bien que cette hypothèse simplifie le problème, il est plus crédible de considérer que l'achat en lui même implique une dépréciation du véhicule.

Une fois la forme fonctionnelle définie, on peut s'interroger sur l'hypothèse d'indépendance de l'évolution de la valeur du véhicule de ses caractéristiques, formulée implicitement en estimant un seul paramètre pour l'âge ( $\alpha_a$ ) et le kilométrage ( $\alpha_{km}$ ).

Une première idée pourrait être que l'effet de l'âge sur la dévalorisation du véhicule n'est pas le même pour toutes les périodes considérés. Une telle hypothèse semble crédible si l'on considère que l'évolution technologique du marché automobile n'est pas constante au cours du temps.

On peut aussi vraisemblablement penser que la dépréciation d'un véhicule essence ne dépende pas du kilométrage cumulé de la même façon qu'un diesel. Dans la base "Parc Auto" complète, on observe en fait que ces derniers sont caractérisés par un kilométrage moyen à l'abandon de 115289 km et un niveau d'utilisation de 84739 km<sup>30</sup>, contre 84782 km et 68866 km respectivement pour les premiers.

On pourrait aussi penser que l'impact du prix neuf diffère s'il s'agit d'un véhicule de construction française ou pas.

---

29. Marché du neuf parfait

30. Voir table 8.2 et Rapport Intermédiaire 1 pour plus de précisions

On substituera alors dans l'équation 6.7

$$\begin{aligned}
 \alpha_p &= \alpha_{p_{fr}} * I(fr) + \alpha_{p_{et}} * I(et) \\
 \alpha_a &= \alpha_{a_0} * I(per0) + \alpha_{a_1} * I(per1) + \alpha_{a_2} * I(per2) + \alpha_{a_3} * I(per3) + \alpha_{a_4} * I(per4) \\
 \alpha_{km} &= \alpha_{km_d} * I(dies) + \alpha_{km_e} * I(ess)
 \end{aligned}
 \tag{6.10}$$

où  $I()$  est la fonction indicatrice, en obtenant les estimation reportées en table 6.5.

TABLE 6.5 – Estimation : Prix d'Achat Occasion (log), 1

(1) $f(km_t, \alpha_{km}) = (1 + km_t)^{\alpha_{km}}$			
Label	Estimate	Std. Dev.	T-stat.
Intercept	1,15350	0,17219	6,70
Prix neuf <sup>a</sup> _fr	0,88842	0,01785	49,76
Prix neuf _et	0,88995	0,01781	49,98
log(1+Age) _Construction avant 1995	-0,48051	0,01089	-44,12
log(1+Age) _Construction [1995; 1999[	-0,37307	0,01318	-28,31
log(1+Age) _Construction [1999; 2002[	-0,36861	0,01454	-25,34
log(1+Age) _Construction [2002; 2005[	-0,37480	0,01577	-23,76
log(1+Age) _Construction [2005; 2008]	-0,30919	0,02337	-13,23
Kilométrage <sup>b</sup> _dies	-0,00229	0,00013775	-16,65
Kilométrage _ess	-0,00486	0,00018988	-25,59
Number of Observations <sup>c</sup>	5472		
DF <sup>d</sup>	9		
SSR <sup>e</sup>	2605,70244		
F	2064,01	Pr > F	<, 0001
R-Square	0,7728	Adj R-Sq	0,7724

DATA SOURCE : PARC AUTO

SOFTWARE : SAS

- 
- a. estimé
  - b. en milliers
  - c. Used Data : véhicules achetés d'occasion
  - d. Degrees of Freedom
  - e. Sum of Squared Residuals

On peut tester les hypothèses formulées, pour en vérifier la pertinence. L'hypothèse d'égalité entre les effets du kilométrage pour les moteurs essence

et diesel est rejeté au seuil de 1% ( $F = 344,48, Pr > F <,0001$ ). Le pays constructeur ne semble pas avoir d'effet sur la dépréciation due au premier achat ( $F = 1,80, Pr > F = 0,1801$ ), comme les estimations le suggèrent. Pour ce qui concerne l'effet de la période de construction sur la part de dépréciation causée par l'âge du véhicule, l'hypothèse est rejetée lorsque l'on considère les 5 périodes, mais le test assume une probabilité de 0,9015 lorsque l'hypothèse nulle est l'égalité entre les effets de 1995 à 2005.

Les résultats du modèle, estimé en tenant compte de ces observations, sont reportés de suite (table 6.6).

TABLE 6.6 – Estimation : Prix d'Achat Occasion (log), 2

(1) $f(km_t, \alpha_{km}) = (1 + km_t)^{\alpha_{km}}$			
Label	Estimate	Std. Dev.	T-stat.
Intercept	1,14226	0,16712	6,83
Prix neuf <sup>a</sup>	0,88997	0,01729	51,48
log(1+Age)_Construction avant 1995	-0,48026	0,01084	-44,29
log(1+Age)_Construction [1995; 2005[	-0,37159	0,01206	-30,81
log(1+Age)_Construction [2005; 2008]	-0,30780	0,02332	-13,20
Kilométrage <sup>b</sup> _dies	-0,00229	0,00013712	-16,71
Kilométrage <sup>c</sup> _ess	-0,00485	0,00018840	-25,76
Number of Observations <sup>d</sup>	5472		
DF <sup>e</sup>	6		
SSR <sup>f</sup>	2605,42002		
F	3096,24	Pr > F	<,0001
R-Square	0,7727	Adj R-Sq	0,7724

DATA SOURCE : PARC AUTO

SOFTWARE : SAS

---

*a.* estimé

*b.* en milliers

*c.* en milliers

*d.* Used Data : véhicules achetés d'occasion

*e.* Degrees of Freedom

*f.* Sum of Squared Residuals

On a dit précédemment que l'on peut utiliser ces résultat pour estimer le scrap-value (valeur de revente) de chaque véhicule composant la flotte automobile d'un ménage au moment  $t$ .

Pour les véhicules achetés neufs on définit donc :

$$\begin{aligned}
sv_t &= \frac{p^{0,9}}{(1+a_t)^{(0,5*I(ac<1995)+0,4*I(1995\leq ac<2005)+0,3*I(ac\geq 2005))}} \\
&\quad * \frac{1}{\exp((0,002*I(dies)+0,005*I(ess))km_t)}
\end{aligned} \tag{6.11}$$

Lorsque l'on veut calculer cette valeur pour un véhicule acheté d'occasion, on ne dispose pas de l'information concernant le prix neuf du véhicule mais seulement le prix auquel le ménage l'a acheté. On a donc deux solutions, soit on estime le prix hédonique neuf à l'équilibre en utilisant le modèle 6.5 ici présenté, soit on récrit le scrap-value au moment  $t_1 = t + n$  (où  $t$  est l'année d'achat d'occasion) comme suit :

$$\begin{aligned}
p_{t1} &= \frac{p^{\alpha_p}}{(1+a_{t1})^{\alpha_a} \exp(\alpha_{km} * km_{t1})} \\
&= \frac{p_t * ((1+a_t)^{\alpha_a} \exp(\alpha_{km} * km_t))}{(1+a_t+n)^{\alpha_a} \exp(\alpha_{km} * (km_t + km_{t->n})} \\
&= \exp \left[ (\ln(p_t) + \alpha_a * \ln(1 + a_t) + \alpha_{km} * km_t) - \right. \\
&\quad \left. (\alpha_a * \ln(1 + a_t + n) + \alpha_{km} * (km_t + km_{t->n})) \right] \\
&= \exp \left[ (\ln(p_t) - \alpha_a * \ln(1 + \frac{n}{1+a_t}) - \alpha_{km} * km_{t->n}) \right] \\
&= \frac{p_t}{(1 + \frac{n}{1+a_t})^{\alpha_a} \exp(\alpha_{km} * km_{t->n})}
\end{aligned} \tag{6.12}$$

où

$$\begin{aligned}
\alpha_a &= 0,5 * I(ac < 1995) + 0,4 * I(1995 \leq ac < 2005) + 0,3 * I(ac \geq 2005) \\
\alpha_{km} &= 0,002 * I(dies) + 0,005 * I(ess)
\end{aligned} \tag{6.13}$$

et  $km_{t->n}$  est l'utilisation du véhicule faite par le ménage depuis l'achat.

## 6.4 Conclusion

Au cours des sections précédentes ont été étudiés les différents facteurs qui composent le prix de transaction d'un véhicule. Le prix de transaction est le fruit d'une négociation entre les deux acteurs, l'offre et la demande.

La méthode des prix hédoniste est ici choisie pour étudier un tel type de prix.

L'hypothèse d'erreurs i.i.d. normales a été relâchée en estimant des paramètres de transformation *Box-Cox* adaptés pour la consommation moyenne du véhicule et son prix de transaction. La consommation résultant être endogène au modèle de prix, un système d'équations simultanées est estimé. L'estimateur 2SLS est utilisé à cet effet.

La consommation du véhicule résulte dépendre du type de carburant, de la période de construction<sup>31</sup>, du poids, de la puissance maximale du moteur et des dimensions du véhicule. Le poids et la puissance, aussi bien que la hauteur et la largeur, ont des effets qui interagissent dans la détermination du prix. Certaines caractéristiques du ménage sont aussi utilisées comme instruments.

Le prix de transaction du véhicule semble dépendre de la consommation, du volume, du poids, de la puissance et de la hauteur, ainsi que du type de carburant et de carrosserie du véhicule et du type d'acheteur, caractérisé par la taille du ménage, la classe de revenu annuel et la zone de résidence.

La consommation a un effet négatif sur le prix, inversement à ses dimensions. Des différences sont remarquées selon le type de carburant et de carrosserie du véhicule. Le type d'acheteur, et notamment sa taille, a un effet négatif significatif sur le prix.

Une possible évolution de ce travail consiste à estimer les prix de catalogue des véhicules pour déterminer le prix de réserve de l'offre et les utiliser, joints au modèle ici présenté, pour simuler le processus de négociation et estimer ainsi le prix de réserve de l'acheteur.

Pour conclure, le prix d'occasion est fonction géométrique décroissante du prix neuf.

---

31. Représenté ici par l'année de première mise en circulation





# Chapitre 7

## Les Décisions de Possession et Utilisation des Véhicules : Ménages Mono-Equipés

Ce travail ayant fait l'objet d'une publication ([de Lapparent and Cernicchiaro \(2012\)](#)) je trouve opportun de présenter ici ces résultats, bien que quelques remarques soient à formuler sur l'échantillon retenu pour leur estimation.

### Une Brève Critique des Données

L'échantillon est issu d'une première version du processus d'appariement présenté plus haut (sec. 4.2) qui, par simplicité, ne considérait qu'un sous panel cylindré de Parc Auto. La base 2008 n'étant pas encore disponible à ce moment, l'échantillon cylindré va de 2000 à 2007.

Sur cette période, l'hypothèse faite sur les prix du carburant est moins crédible, cependant seulement la première année est touchée par une variation négative et cette variation est peu importante par rapport à celles connues par les deux carburants sur le reste de la période. Le choix de l'inclure quand même est dicté par le besoin d'utiliser un échantillon dont la taille soit significative.

En comparant les caractéristiques des véhicules possédés par ces ménages à celles observées pour les mono-équipés de la base (sec. 5.3) on observe un âge à l'achat légèrement moins élevé. Cela pourrait signifier que les véhicules, et de façon plus importante les diesel, sont achetés plus vieux depuis 2001. En même temps la procédure utilisée pour la construction de l'échantillon utilisé dans cette application n'incluait pas la sous procédure d'identification des appariements forcés (voir sec. 4.2) mais se limitait à sélectionner l'année d'achat cohérente le plus souvent déclarée au cours de la période appariée.

L'automobile, une des inventions technologiques majeures du *XIX<sup>ème</sup>*, s'est progressivement imposée au cours du siècle dernier, en devenant le principal mode de transport des pays développés dans les années '70. De nos jours ce bien peut être considéré comme l'un des éléments incontournables de notre société.

Les chiffres publiés chaque année par le Comité des Constructeurs Français d'Automobile ([CCFA \(2006\)](#)) montrent comment le pourcentage de ménages français choisissant la motorisation n'a cessé d'augmenter, en passant de

70,8% en 1980 à 83,2% en 2009. Si l'on s'intéresse en outre au choix de taille du parc automobile, on observe depuis dix ans une diminution de la part de ménages "monoéquipés", qui constitue encore aujourd'hui le pourcentage le plus important, en faveur principalement des ménages "biéquipés".

L'automobile étant un "bien durable", c'est à dire que son utilisation en cause la dépréciation, la décision de posséder un véhicule ne peut pas être dissociée du choix de l'usage que l'on souhaite en faire. L'achat d'un véhicule constitue pour les ménages un investissement qu'ils doivent rembourser pendant la période de possession, période qui, au cours de la dernière décade, devient de plus en plus longue, comme le montrent les chiffres publiés par le [CCFA \(2006\)](#).

L'amélioration des caractéristiques techniques des véhicules ainsi que l'augmentation des prix de vente peuvent expliquer ce phénomène ([Julliard \(2007\)](#)). D'un côté la durée de vie de ces biens augmente en permettant une utilisation plus longue, de l'autre, l'investissement se fait de plus en plus important. En même temps, la possession et l'utilisation d'un véhicule comportent un coût qui augmente au fur et à mesure que le bien se déprécie ([Baron \(2002\)](#)) et qui dépend aussi des évolutions du prix des carburants ([Julliard \(2007\)](#), [Graham and Glaister \(2002\)](#)). Un ménage qui choisit de se motoriser doit donc évaluer en même temps les dépenses qui en découlent ainsi que les bénéfices qu'il peut obtenir, pour décider quelles sont la durée de possession et l'utilisation optimales de son véhicule, ainsi que la taille de son parc.

Les caractéristiques socio-démographiques du ménage et sa localisation géographique, ainsi que sa contrainte budgétaire, jouent un rôle important dans la définition de ses nécessités en terme d'équipement et usage automobile. Il est donc de toute importance de les prendre en compte en même temps que les évolutions du marché automobile et du pétrole, lorsque l'on veut étudier les comportements de choix dans ce domaine.

La place importante qui est réservée à l'automobile dans notre société fait de la modélisation de la demande de possession et d'utilisation un élément nécessaire dans de nombreux domaines, comme par exemple l'organisation des transports en commun, la gestion de l'énergie, la définition des politiques de l'industrie automobile, le contrôle de la qualité de l'air.

En parcourant la littérature on retrouve plusieurs études à ce sujet qui peuvent être classées en fonction de la méthodologie et des données utilisées, selon le niveau de désagrégation, le type de choix étudié (réel ou hypothétique) et les dimensions de la demande considérées, la prise en compte du caractère compensatoire des facteurs influant sur le choix, l'équilibre entre

offre et demande et l'inclusion de la dynamique temporelle.

Dans mon étude je m'intéresse aux méthodes désagrégées compensatoires, basées sur des choix réels, approche utilisée dans les travaux de [Lave and Bradley \(1980\)](#), [Hoeherman et al. \(1983\)](#), [Mannering and Winston \(1985\)](#), [Berkovec and Rust \(1985\)](#), [Train \(1986\)](#), [Rust \(1987\)](#), [Hensher et al. \(1990\)](#), [Berkowitz et al. \(1990\)](#), [de Jong \(1990, 1996\)](#) et plus récemment dans les études menées par [Kim \(2002\)](#), [de Jong and Kitamura \(2009\)](#) et [Bento et al. \(2009\)](#).

L'idée commune à ces études est l'utilisation de la théorie des utilités. Le ménage est supposé avoir pour objectif de choisir, parmi les alternatives qui se présentent à lui, celle qui lui apporte la plus grande utilité, définie par agrégation des valeurs que le ménage associe à chaque composant du choix. La demande d'équipement est composée de plusieurs dimensions, comme le choix du nombre de véhicules à posséder, le type de ces véhicules et leur utilisation, qui sont à considérer simultanément. La difficulté que cela comporte porte les chercheurs à ne pas prendre en compte la totalité de ces dimensions, voire à en considérer une seule.

Lorsque plusieurs dimensions et leurs interactions sont considérées, la presque totalité des études utilise un modèle GEV (logit emboîté) pour l'estimation simultanée des composants discrets du choix, à l'exception de [Kim \(2002\)](#) qui estime la probabilité de choix des modalités disponibles à l'aide d'un probit polytomique.

Les choix du nombre et du type de véhicule étant à caractère discret, contrairement au choix d'utilisation, qui est de nature continue, c'est seulement dans quelques études qu'ils sont estimés simultanément. Les auteurs qui traitent le sujet (ex. [Mannering and Winston \(1985\)](#), [Train \(1986\)](#), [Hensher et al. \(1990\)](#)) se basent sur les travaux de [Dubin and McFadden \(1984\)](#) qui proposent l'application de l'identité de Roy à la fonction d'utilité indirecte conditionnelle au choix du véhicule, permettant de dériver la demande Marshallienne pour son usage. Ceci impose l'utilisation d'une fonction d'utilité non linéaire.

Le choix du type de véhicule pose aussi plusieurs problèmes et il n'est donc pas souvent inclus. Lorsqu'il est pris en compte (ex. [Lave and Bradley \(1980\)](#), [Mannering and Winston \(1985\)](#), [Berkovec and Rust \(1985\)](#), [Train \(1986\)](#), [Kim \(2002\)](#)), les comportements d'achat et le choix de conserver ou revendre un véhicule déjà possédé sont rarement considérés à la fois. Par exemple, Mannering et Winston ne s'occupent que des achats neufs, dans le but de définir l'impact que " l'affection " pour une marque automobile a sur les comportements d'achat. En outre, la difficulté à définir un ensemble de type de

véhicules qui soit à la fois exhaustif et d'une dimension qui la rende maniable, poussent les auteurs à ne considérer qu'un sous-ensemble des choix réels. Pour citer quelques exemples, Lave et Bradley distinguent les véhicules selon qu'ils soient de production locale ou étrangère, alors que Train définit l'ensemble de marques et modèles disponibles pour un ménage en incluant les types de véhicules qu'il a possédés dans le passé plus un petit nombre d'autres types tirés au hasard.

Une autre limite, commune à la quasi totalité des études citées, est l'absence de la dynamique temporelle dans l'estimation des choix. Lorsqu'elle est présente, comme par exemple dans les travaux de [de Jong \(1990, 1996\)](#), il s'agit généralement d'une dynamique tournée vers le passé.

Le caractère durable de l'automobile que l'on a mis en lumière précédemment, me porte à penser, contrairement aux études existantes, que l'agent base ses décisions sur l'anticipation des évolutions futures de son environnement (évolution des prix du carburant, changements dans le marché automobile, modifications de la structure du ménage et des caractéristiques découlant de sa localisation géographique), en accord avec la critique de [Lucas \(1976\)](#).

[Rust \(1987\)](#) propose un modèle dynamique tourné vers le futur dont je m'inspire dans mon étude. Cependant il ne considère, dans ses travaux, que le choix de conserver ou abandonner un véhicule et combien l'utiliser, sans prendre en compte les autres dimensions du choix d'équipement.

Dans ce chapitre je vais présenter un programme optimal stopping à horizon fini, en présence d'incertitude, visant à décrire les comportements d'utilisation automobile des ménages. Par comportements d'utilisation automobile j'entends les choix concernant la durée optimale de possession d'un véhicule et le kilométrage à parcourir. Je suppose que le ménage anticipe les évolutions futures de son état, ici représenté par son revenu, ainsi que de son environnement, ici les prix du carburant, lorsqu'il prend ses décisions. L'abandon d'un véhicule peut être alors représenté par un problème optimal stopping, dont la règle d'arrêt est la solution d'un problème de programmation dynamique stochastique structurale. L'idée sous-jacente à la solution théorique d'un tel problème est qu'il existe un "seuil de satisfaction". En d'autres termes, je suppose que l'agent abandonne le véhicule lorsque l'utilité escomptée par le choix de conserver le véhicule un an de plus, qu'il l'utilise ou pas au cours de la période, est inférieure à la valeur qu'il obtiendrait par la vente de celui-ci.

Le modèle proposé appartient à la classe des processus de décision de Markov. Il est spécifié comme un modèle dynamique de choix discrets pour un agent économique tourné vers le futur. Il est défini sur 4 variables d'état

et 2 variables de décision.

Les variables d'état utilisées sont de deux types. Trois modélisent les prévisions des évolutions du prix du carburant (diesel et essence) et du revenu. L'autre, l'âge du véhicule, est déterministe. Ses évolutions sont endogènes et modélise le kilométrage cumulé.

La première variable de décision modèle le choix de possession du véhicule. Il s'agit d'une variable dichotomique qui vaut 0 lorsque le véhicule est conservé et 1 en cas d'abandon. La deuxième variable de décision modélise la demande de kilométrage, en sachant que le véhicule est conservé. Le kilométrage choisi en début de période affecte le kilométrage cumulé à la période suivante.

Le caractère stochastique du problème dérive de l'impossibilité d'observer les préférences et les attentes de l'agent. Un terme stochastique, visant à résumer l'information manquante, est donc introduit dans le modèle.

Puisque l'approche ici choisie est à horizon fini, c'est-à-dire que l'on peut observer tous les choix effectués à partir d'un moment  $t$  (2002 pour les véhicules achetés avant le début de la période d'observation cylindrée ici sélectionnée, l'année d'achat pour les autres) jusqu'à l'arrêt du processus (i.e. l'année d'abandon est observée avant 2007), on peut la résoudre par récurrence en remontant le temps. Les paramètres du modèle structural sont estimés par le biais du Nested fixed Point algorithm.

Le chapitre est structuré en 4 sections. Le modèle est présenté au cours de la section 7.1. Chaque composant du modèle est décrit, ainsi que la méthode d'estimation. L'échantillon, extrait de Parc Auto, utilisé pour l'estimation est décrit dans la section 7.2 et les résultats en 7.3. La dernière section tire les conclusions et suggère quelque évolution possible du modèle.

## 7.1 Le Modèle

La définition du modèle se fait en imposant les cinq hypothèses suivantes :

- *AS (Additive separability)* : La fonction d'utilité au moment  $t$  peut être écrite comme somme d'un composant observable et un non-observable de la façon suivante :

$$U(a, x_{it}, \varepsilon_{it}) = u(a, x_{it}) + \varepsilon_{it}(a) \quad (7.1)$$

où  $\varepsilon_{it}(a)$  est une variable aléatoire de moyenne nulle définie sur  $\mathfrak{R}$  et elle a la même dimension  $(J + 1) \times 1$  de l'espace de choix  $A$ . Le vecteur des paramètres de la fonction d'utilité  $u$  est appelé  $\theta_u$ .

- *IID (iid unobservables)* : Les variables d'état non-observables  $\varepsilon_{it}$  sont indépendantes et identiquement distribuées sur les individus et le temps avec une fonction de répartition  $G_\varepsilon(\varepsilon_{it})$  de moyenne finie et deux fois différentiable en  $\varepsilon_{it}$ .
- *CI-X (Conditional independence of futur x)* : L'information non-observable  $\varepsilon_{it}$  n'a pas d'influence sur la transition de l'état présent  $x_{it}$  au suivant  $x_{i,t+1}$  et la fonction de répartition est donc équivalente à la fonction de transition des  $x_{it}$

$$CDF(x_{i,t+1}|a_{it}, x_{it}, \varepsilon_{it}) = F_x(x_{i,t+1}|a_{it}, x_{it}) \quad (7.2)$$

Le vecteur des paramètres de  $F_x$  est appelé  $\theta_f$ .

- *CLOGIT* : L'information non-observable  $\{\varepsilon_{it}(a) : a = 0, 1, \dots, J\}$  est indépendante entre les choix et suit une distribution de Gumbel.
- *DIS (Discrete support of x)* : Le support des  $x_{it}$  est discret et fini.

Il s'agit des hypothèses qui définissent le « dynamic programming - conditional logit model » de Rust. L'estimation de ce modèle est simple mais les hypothèses imposent des restrictions sur l'interprétation des résultats.

Les choix ici modélisés sont ceux de durée de possession et d'intensité d'utilisation pour chaque ménage  $i$  présent continument de 2000 à 2007 dans le panel Parc-Auto et n'ayant jamais possédé plus d'un véhicule par année au cours de la période.

Les agents du problème sont donc des ménages *mono-équipés*. On évite de cette façon les effets de substitution.

Ni le choix du type de véhicule, ni la raison pour laquelle le ménage décide de se séparer du véhicule ne sont estimés ici.

On suppose que l'objectif du ménage  $i$  est d'établir pour son véhicule un horizon de vie optimal  $\bar{t}_i \in \{1, \dots, T\}$  et, en même temps, la séquence des kilométrages annuels à effectuer pendant la période de possession  $\mathbf{m}_{i,\cdot} = (m_{i,1}, \dots, m_{i,T})$ . On peut donc le traiter comme un problème « optimal use and stop », en considérant que pour chaque période  $t$  la décision d'abandonner ou garder ( $d$ ) et utiliser ( $m$ ) le véhicule est prise en fonction de l'état du ménage  $i$  et de son environnement au début de la période.  $\mathbf{z}_{it}$  est le vecteur des variables d'état. Les préférences du ménage peuvent alors être représentées en fonction de ces variables d'état par une fonction d'utilité time-separable actualisée  $\sum_{t=1}^T \gamma^t u(\mathbf{z}_{i,t}, d(\mathbf{z}_{i,t}), m(\mathbf{z}_{i,t}))$  où  $\gamma$  est le facteur d'actualisation et  $u(\mathbf{z}_{i,t}, d(\mathbf{z}_{i,t}), m(\mathbf{z}_{i,t}))$  est la fonction d'utilité pour la période  $t$ . Pour alléger les notations je pose  $d_{i,t} = d(\mathbf{z}_{i,t})$  et  $m_{i,t} = m(\mathbf{z}_{i,t})$ .

Au début de chaque année  $t$  le ménage décide d'abandonner ( $d_{i,t} = 1, D$ ) ou de conserver ( $d_{i,t} = 0, K$ ) son véhicule et, seulement dans le deuxième cas choisit un kilométrage  $m_{i,t}$  à parcourir au cours de l'année avec le véhicule. Lorsque le véhicule est abandonné ce kilométrage est nul ( $m_{i,t} = 0$ ) et le processus est interrompu. L'information dont il dispose pour prendre ses décisions est l'utilité résultant de son état.

Les décisions prises à la date  $t$  ont un impact sur l'évolution future des variables d'état, mais le ménage prend en compte l'incertitude présente sur les valeurs futures de son état et l'on peut modéliser ses prévisions par une fonction de distribution de transition de Markov  $G_z(\mathbf{z}_{i,t+1} | \mathbf{z}_{i,t}, d_{i,t}, k_{i,t})$ . La séquence des décisions est choisie de façon à maximiser l'utilité espérée :

$$\max_{\bar{t}_i} \left\{ \mathbb{E}_{\mathbf{z}_{i,0}} \left( \sum_{t=1}^{T_i} \gamma^t u(\mathbf{z}_{i,t}, d_{i,t}, m_{i,t}) | \mathbf{z}_{i,0} \right) \right\} \quad (7.3)$$

où  $\mathbf{z}_{i,0}$  est l'ensemble des conditions initiales.

Je fais l'hypothèse que le programme d'optimisation inter-temporel peut être défini comme un problème séquentiel dont la solution est la même en utilisant sa représentation de Belleman (Bellman (1957), Rust (1996), Bertsekas (2000)). Un seul jeu de paramètres, indépendant de la période  $t$ , est donc estimé.

Soit  $V(\mathbf{z}_{i,t})$  la fonction valeur maximale du programme dynamique (en considérant toutes les séries de décisions possibles conditionnellement aux choix en  $t$ ), définie par résolution récursive de l'équation de Bellman suivante :

$$V(\mathbf{z}_{i,t}) = \max \left\{ v_D(\mathbf{x}_{i,t}) + \epsilon_{D,i,t}, v_K(\mathbf{x}_{i,t}) + \gamma \mathbb{E}_{\mathbf{z}_{i,t+1}} [V(\mathbf{z}_{i,t+1}) | \mathbf{z}_{i,t}] + \epsilon_{K,i,t} \right\}. \quad (7.4)$$

Les valeurs qui maximisent cette fonction  $m^*(\mathbf{z}_{i,t})$  et  $d^*(\mathbf{z}_{i,t})$  représentent respectivement le kilométrage et le choix abandonner/conservé optimaux.

On peut distinguer deux type de variables d'état ( $\mathbf{z}_{i,t} = \{\mathbf{x}_{i,t}, \boldsymbol{\varepsilon}_{i,t}\}$ ) : les  $\mathbf{x}_{i,t}$  qui sont observables à la fois par le ménage  $i$  et par le chercheur, et les  $\boldsymbol{\varepsilon}_{i,t}$  observables seulement par le ménage. La partie non observable de l'information compte deux composants  $\boldsymbol{\varepsilon}_{i,t} = \{\boldsymbol{\varepsilon}_{D,i,t}, \boldsymbol{\varepsilon}_{K,i,t}\}$  que l'on suppose i.i.d. réparties selon la loi standard de Gumbel :

$$\exp(-\exp(-\varepsilon_{d,i,t})) \quad (7.5)$$

Sous l'hypothèse *AS*, l'information non observable  $\varepsilon$  est traitée comme bruit additif dans la spécification du modèle.



L'ensemble des variables d'état dans notre cas est  $\mathbf{x}_{i,t} = \{a_{i,t}, k_{i,t}, y_{i,t}, p_{i,t}\}$  où  $a_{i,t}$  et  $k_{i,t}$  sont respectivement l'âge et le kilométrage au compteur du véhicule possédé par le ménage  $i$  au début de la période  $t$ ,  $y_{i,t}$  est la classe de revenu à laquelle le ménage appartient et  $p_{i,t}$  est le prix annuel du type de carburant utilisé par le véhicule du ménage.

Toutes ces variables sont discrètes, et respectent ainsi l'hypothèse *DIS*.

On suppose que les variables d'état observables ont des influences sur l'utilité d'abandonner et celle de conserver le véhicule pour la période suivante qui sont différentes. Ces deux utilités sont définies par les fonctions suivantes :

$$\begin{cases} v_D(\mathbf{x}_{i,t}) = \alpha \frac{P_i}{1+k_{i,t}a_{i,t}} \\ v_K(\mathbf{x}_{i,t}) = \beta_1 + \sum_{q=2}^Q \beta_q \mathbb{I}(y_{i,t} = q) + \beta_p c_i p_{i,t} + \beta_k k_{i,t} \end{cases} \quad (7.6)$$

$P_i$  et  $c_i$  sont les conditions initiales et représentent le prix à l'achat du véhicule et sa consommation moyenne.  $Q$  est le nombre de classes de revenu. Je modélise l'utilité d'abandonner le véhicule comme une fonction décroissante ( $\alpha < 0$ ) de la valeur résiduelle du véhicule, ici définie comme une fonction géométrique décroissante de l'âge et du kilométrage cumulé du véhicule. L'utilité de conserver le véhicule est exprimée comme différence entre le revenu du ménage et la dépense liée à l'utilisation du véhicule au cours de l'année, où le coût d'utilisation peut être décomposé en une partie due à la maintenance du véhicule, que je suppose dépendre de son kilométrage cumulé  $k_{i,t}$ , et la dépense en carburant  $c_i * p_{i,t}$ . L'utilité est alors définie comme fonction du revenu, de la dépense  $c * p$  et du kilométrage cumulé. On peut imaginer que le coût d'utilisation ait un effet négatif sur la décision de continuer à posséder un véhicule, je m'attends donc à observer  $\beta_p < 0$  et  $\beta_k < 0$ .

Permettez-moi de m'arrêter un instant sur l'écriture des utilités.

Pour qu'un modèle logit binomiale soit identifiable on ne peut estimer que la différence entre les deux utilités. Les paramètres  $\beta = \{\beta_1, \beta_q, \beta_p, \beta_k\}$  en 7.6 correspondent alors à  $\beta^K - \beta^D$  et écrire le système présenté plus haut est équivalent à définir les utilités comme :

$$\begin{cases} v_D^1(\mathbf{x}_{i,t}) = \beta_1^D + \sum_{q=2}^Q \beta_q^D \mathbb{I}(y_{i,t} = q) + \beta_p^D c_i p_{i,t} + \beta_k^D k_{i,t} + \alpha v_D \\ v_K^1(\mathbf{x}_{i,t}) = \beta_1^K + \sum_{q=2}^Q \beta_q^K \mathbb{I}(y_{i,t} = q) + \beta_p^K c_i p_{i,t} + \beta_k^K k_{i,t} + \alpha v_K \end{cases} \quad (7.7)$$

où  $v = v_D, v_K$  est la valeur pécuniaire que l'agent obtient en faisant le choix. On a alors  $v_D = \frac{P_i}{1+k_{i,t}a_{i,t}}$  le prix de revente de son véhicule et  $v_K = 0$ . Le système d'utilités peut alors être réduit à une seule utilité, en fixant le choix d'abandonner comme référence :

$$\begin{aligned} v_K^1(\mathbf{x}_{i,t}) &= \beta_1 + \sum_{q=2}^Q \beta_q \mathbb{I}(y_{i,t} = q) + \beta_p c_i p_{i,t} + \beta_k k_{i,t} + \alpha(v_K - v_D) \\ &= v_K - \alpha \frac{P_i}{1+k_{i,t}a_{i,t}} \end{aligned} \quad (7.8)$$

Une fois les utilités définies, il faut définir le processus  $G_z$  qui guide les transitions des variables d'état  $(k_{i,t}, a_{i,t}, p_{i,t}, y_{i,t})$  et de la variable pay-off  $(m_{i,t})$ . [Aguirregabiria and Mira \(2010\)](#) montrent l'importance cruciale des hypothèses faites sur les anticipations du ménage dans la définition d'un modèle dynamique de choix discrets. En effet, ces hypothèses impactent aussi bien la forme du critère d'estimation (par exemple le maximum de vraisemblance) que la définition de la solution du problème.

La distribution jointe de Markov des transitions des variables d'état est définie comme produit des distributions marginales.

L'âge et le kilométrage cumulé au début de la période  $t$  sont des variables déterministes.

$$a_{i,t} = a_{i,t-1} + 1. \quad (7.9)$$

et

$$k_{i,t} = k_{i,t-1} + m_{i,t-1}. \quad (7.10)$$

La définition des probabilités de transitions pour ce qui concerne l'âge ne nécessite pas d'explications. Il est tout de même important de remarquer que le kilométrage cumulé dépend du kilométrage choisi à la date précédente.

Dans ce travail le kilométrage optimal, lorsque le véhicule est conservé, est défini en fonction du revenu, de la dépense en carburant et du kilométrage cumulé pour la période concernée :

$$\begin{aligned} \ln(1 + m_{i,t}^*) &= \\ \theta_0 + \sum_{j=2}^Q \theta_j \mathbb{I}(y_{i,t} = j) + \theta_p \ln(c_i p_{i,t}) + \theta_k \ln(1 + k_{i,t}) + \epsilon_{m,i,t}. \end{aligned} \quad (7.11)$$

Comme suggéré par l'analyse précédemment menée (voir chapitre 3), on peut s'attendre à avoir  $\theta_l \geq 0, l = 0, \dots, Q, \theta_p < 0$  et  $\theta_k > 0$ . Cela revient à dire que la demande de kilométrage est fonction positive du revenu, décroissante du coût d'utilisation et croissante de la demande à la période précédente (à travers le kilométrage cumulé). La décision de conserver un véhicule sans l'utiliser au cours de la période est possible dans la spécification choisie. Le kilométrage choisi en  $t$  affecte la valeur escompté en  $t + 1$  par le biais du kilométrage cumulé.

Il est important de remarquer que, malgré le caractère continu de la demande du kilométrage, l'échantillon ici utilisé ne permet pas de la traiter comme telle. En effet, en observant les réponses collectées, on dispose ici plutôt d'une représentation discrète de la variable continue. Les ménages ont tendance à déclarer une valeur arrondie pour cette variable et la taille réduite de l'échantillon amplifie l'effet de cette tendance.

Je vais alors discrétiser la demande en définissant une partition de  $\mathbf{R}^+$  qui définit l'ensemble de choix discret. Cette décision est critiquable, comme soutenu par [Small \(1987\)](#). Cependant il ne me semble pas trop restrictif, dans ce contexte, d'utiliser une représentation discrète. De plus, la spécification de Rust, choisie pour ce travail, implique de pouvoir estimer des processus discrets de transition des variables d'état.

On peut alors modéliser la demande à l'aide d'un logit polytomique ordonné. Soient  $m_{i,t} \in \{m_1, \dots, m_B\}$  les  $B$  kilométrages que le ménage peut choisir, définis comme les centres des classes définies précédemment. Les bornes des intervalles sont indiquées par  $\bar{m}_0, \dots, \bar{m}_B$  où, par définition,  $\bar{m}_0 \equiv -\infty$  and  $\bar{m}_B \equiv +\infty$ . La probabilité que le ménage  $i$  a de choisir le kilométrage  $m_b \in [\bar{m}_{b-1}; \bar{m}_b[$  pour la période  $t$ , lorsque le véhicule est conservé, est donc :

$$\Pr(m_{i,t} = m_b | k_{i,t}, y_{i,t}, p_{i,t}, c_i, d_{i,t-1} = 0; \boldsymbol{\theta}, \omega_m) = \frac{1}{1 + \exp\left(\frac{\bar{m}_b - \theta_0 - \sum_{j=2}^Q \theta_j \mathbb{I}(y_{i,t}=j) - \theta_p \ln(c_i p_{i,t}) - \theta_k \ln(1+k_{i,t})}{\omega_m}\right)} \quad (7.12)$$

$$\frac{1}{1 + \exp\left(\frac{\bar{m}_{b-1} - \theta_0 - \sum_{j=2}^Q \theta_j \mathbb{I}(y_{i,t}=j) - \theta_p \ln(c_i p_{i,t}) - \theta_k \ln(1+k_{i,t})}{\omega_m}\right)}$$

Si le véhicule est abandonné en début de période le kilométrage est nul et la probabilité vaut 1.

Puisque dans ce travail le choix des caractéristiques du véhicule n'est pas considéré, le type de carburant est une variable d'état du véhicule observé. De plus, les données concernant les prix du carburant utilisés dans ce travail sont des moyennes nationales annuelles. L'agent n'a donc pas le choix du prix, qui est défini en fonction du type du carburant du véhicule comme suit :

$$p_{i,t} = \mathbb{I}(\text{fuel type of } i \text{ is petrol}) p_{\text{petrol},t} + \mathbb{I}(\text{fuel type of } i \text{ is diesel}) p_{\text{diesel},t} \quad (7.13)$$

Les agents qui possèdent un véhicule utilisant le même carburant sont alors tous confrontés au même prix. Cette hypothèse restrictive et peu crédible est pourtant nécessaire vu le caractère des données.

Le prix du carburant est une variable continue et strictement positive. Cependant on peut supposer que les anticipations que le ménage fait pour les évolution des prix, pour chaque type de carburant, sont basées sur une discrétisation du processus de transition. En d'autres termes, l'agent considère que l'utilité escomptée en  $t$  ne change pas lorsque, en partitionnant l'ensemble en intervalles de longueur  $\psi$ , le prix actuel appartient à la même classe que celui en  $t - 1$ .

L'hypothèse que je fais pour l'évolution des prix du carburant est une simple dynamique non stationnaire d'ordre 2.  $\forall g \in \{\text{petrol, diesel}\}$ ,

$$p_{g,t} = p_{g,t-1} + \epsilon_{g,t} \quad (7.14)$$

où  $\epsilon_{g,t} \stackrel{iid}{\rightarrow} \mathcal{N}(0, \sigma_g^2)$ . Soient  $(\bar{p}_{g,0}, \dots, \bar{p}_{g,R})$  les bornes des intervalles, de pas 1 centime de €, qui modélisent les différents niveaux de prix du carburant  $g$  et  $\tilde{p}_{g,r} = (\bar{p}_{g,r} + \bar{p}_{g,r+1})/2$  leurs valeurs centrales, utilisés comme valeurs représentatives de chaque classe. La probabilité de transition est :

$$\Pr(p_{g,t} = \tilde{p}_{g,r} | p_{g,t-1} = \tilde{p}_{g,l}, d_{i,t-1} = 0; \sigma_g) = \Phi\left(\frac{\bar{p}_{g,r} - \tilde{p}_{g,l}}{\sigma_g}\right) - \Phi\left(\frac{\bar{p}_{g,r-1} - \tilde{p}_{g,l}}{\sigma_g}\right) \quad (7.15)$$

où  $\Phi$  est la fonction de répartition d'une distribution normale standardisée. Une telle approche ne permet pas de définir ni l'ampleur des intervalles ni de choisir la valeur représentative. Ces conditions sont donc à imposer à priori. J'utilise ici un pas d'un centime ( $\psi = 0.01$  €) et les centres des classes comme valeurs représentatives.

La dernière variable d'état est le revenu annuel du ménage. Les ménages participant à l'enquête doivent déclarer, chaque année, la classe

de revenu à laquelle leur ménage appartient. La variable de revenu disponible n'a donc pas un caractère continu, mais il s'agit d'une variable groupée en 13 classes. Lorsque l'on dispose d'une telle variable on pourrait être tentés d'utiliser l'estimateur des Moindres Carrés Ordinaires, en associant à chaque classe de revenu une valeur représentative, comme par exemple son point central. Un tel choix serait erroné à cause de l'inconsistance de l'estimateur MCO.

En remarquant le caractère discret ordonné d'une telle variable, comme souligné par [Stewart \(1983\)](#), [Terza \(1985\)](#) et [Bhat \(1994\)](#), on peut raisonnablement penser à utiliser l'estimateur du Maximum de Vraisemblance.

Si l'on définit  $y^*$  la variable continue (non observée),  $y$  la variable groupée en  $J$  classes (observée) et  $\mu_{j-1}$ ,  $\mu_j$   $j = 1, \dots, J$  les bornes de chaque classe  $j$  en logarithme, on en déduit

$$y = j \leftrightarrow \mu_{j-1} < \ln(y^*) < \mu_j$$

En supposant  $\ln(y^*) = \beta'x + \epsilon$ ,  $\epsilon \rightsquigarrow \text{Logistic}(0, \sigma)$ <sup>1</sup>, la probabilité d'observer une classe de revenu  $j$  en connaissant les caractéristiques  $x$  du ménage peut donc être écrite comme :

$$P(y = j | x) = \left[ F\left(\frac{\mu_j - \beta'x}{\sigma}\right) - F\left(\frac{\mu_{j-1} - \beta'x}{\sigma}\right) \right] \quad (7.16)$$

où  $F(x) = \frac{1}{1+\exp(-x)}$  est la fonction logistique.

Il s'agit donc d'un modèle *Logit Ordonnée à Bornes Connues*.

Les bornes extrêmes sont définies, par convention, comme  $\mu_0 = -\infty$  et  $\mu_J = \infty$ . On a donc

$$\begin{aligned} P(y = 1 | x) &= F\left(\frac{\mu_1 - \beta'x}{\sigma}\right) \\ &\text{et} \\ P(y = J | x) &= 1 - F\left(\frac{\mu_J - \beta'x}{\sigma}\right) \end{aligned}$$

Puisque les bornes  $\mu_j$  sont connues,  $\beta$  et  $\sigma$  sont identifiées et donc estimables.

---

1. J'avoue que l'hypothèse d'erreurs normales, proposée dans la littérature, est plus crédible pour ce problème.

Cependant, puisque les estimations obtenues avec le modèle probit sont en ligne générale proportionnelles à celles obtenues avec un modèle logit et ce deuxième est de plus simple interprétation, les erreurs sont ici considérées logistiques

La log-vraisemblance associée au modèle est calculée comme

$$LL = \sum_i \sum_j I(y_i = j) \ln(P(y_i = j | x_i)) \quad (7.17)$$

où  $I(y_i = j) = 1$  si le ménage  $i$  est observé dans la classe de revenu  $j$  et 0 autrement.

Plusieurs techniques d'optimisation de la vraisemblance ont été proposées pour ce modèle. Dans ce travail la méthode de Newton-Raphson est utilisée.

Le processus de transition supposé pour le revenu est une simple autorégression. La variable sous-jacente  $y^*$  est donc définie comme suit :

$$\ln(y_{i,t}^*) = \beta_0 + \sum_{q=2}^Q \beta_q \mathbb{I}(y_{i,t-1} = q) + \epsilon_{y,i,t} \quad (7.18)$$

où  $\epsilon_{y,i,t} \xrightarrow{iid} \mathcal{L}ogistic(0, \tau_y)$ . Au moment  $t$ , la classe de revenu du ménage  $i$  est  $y_{i,t} = j$  si et seulement si  $\bar{y}_{j-1} \leq \ln(y_{i,t}^*) < \bar{y}_j$ .  $\bar{y}_0, \dots, \bar{y}_J$  sont fixés et se suivent en définissant les  $J$  intervalles contigus constituant l'ensemble des états possibles.

La probabilité pour le ménage  $i$  d'appartenir à une classe de revenu  $j$  pour la période  $t$  s'écrit :

$$\Pr(y_{i,t} = j | y_{i,t-1}, d_{i,t-1} = 0; \boldsymbol{\beta}, \tau_y) = \frac{1}{1 + \exp\left(-\frac{\bar{y}_j - \beta_0 - \sum_{q=2}^Q \beta_q \mathbb{I}(y_{i,t-1} = q)}{\tau_y}\right)} - \frac{1}{1 + \exp\left(-\frac{\bar{y}_{j-1} - \beta_0 - \sum_{q=2}^Q \beta_q \mathbb{I}(y_{i,t-1} = q)}{\tau_y}\right)} \quad (7.19)$$

où, par convention,  $\bar{y}_0 \equiv -\infty$  et  $\bar{y}_Q \equiv +\infty$ .

Lorsque le véhicule est abandonné aucune utilité sera obtenue de celui-ci à partir de la période suivante. Le flux d'utilités futures est donc nul et l'abandon est une décision absorbante.

On veut estimer les paramètres  $\boldsymbol{\lambda} = \{\alpha, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\delta}, \omega_m, \sigma_{\text{petrol}}, \sigma_{\text{diesel}}, \tau_y\}$  et le facteur d'actualisation  $\gamma$  en observant, pour chaque ménage, la séquence de décisions prises (à la fois abandonner/conservé et kilométrage), les variables d'état  $\mathbf{x}_{i,t}$  et les conditions initiales. Seulement les prix des carburants, qui sont des moyennes nationales, sont observés lorsque le véhicule est abandonné.

La fonction de vraisemblance de l'échantillon s'écrit :

$$\sum_{i=1}^n \ln \ell(\boldsymbol{\lambda}, \gamma | \mathbf{d}_{i,\cdot}, \mathbf{m}_{i,\cdot}, \mathbf{y}_{i,\cdot}, \mathbf{p}_i) \quad (7.20)$$

où

$$\begin{aligned} \ell(\boldsymbol{\lambda}, \gamma | \mathbf{d}_{i,\cdot}, \mathbf{m}_{i,\cdot}, \mathbf{y}_{i,\cdot}, \mathbf{p}_i) = & \\ & \Pr(d_{i,\bar{t}_i} = 1 | y_{i,\bar{t}_i}, a_{i,\bar{t}_i}, k_{i,\bar{t}_i}, p_{i,\bar{t}_i}, \mathbf{x}_{i,0}; \alpha, \boldsymbol{\beta}) \\ & \prod_{t=\underline{t}_i}^{\bar{t}_i-1} \Pr(d_{i,t} = 0 | m_{i,t}, y_{i,t}, a_{i,t}, k_{i,t}, p_{i,t}, \mathbf{x}_{i,0}; \boldsymbol{\lambda}, \gamma) \\ & \prod_{t=\underline{t}_i}^{\bar{t}_i-1} \Pr(m_{i,t} | y_{i,t}, k_{i,t}, p_{i,t}, d_{i,t-1} = 0, \mathbf{x}_{i,0}; \boldsymbol{\theta}, \omega_m) \\ & \prod_{t=\underline{t}_i}^{\bar{t}_i} \Pr(p_{i,t} | p_{i,t-1}, d_{i,t-1} = 0, \mathbf{x}_{i,0}; \sigma_{\text{petrol}}, \sigma_{\text{diesel}}) \\ & \prod_{t=\underline{t}_i}^{\bar{t}_i} \Pr(y_{i,t} | y_{i,t-1}, d_{i,t-1} = 0, \mathbf{x}_{i,0}; \boldsymbol{\beta}, \tau_y) \end{aligned} \quad (7.21)$$

Puisque l'on n'observe pas le comportement d'un ménage au cours de toute la période 2001 – 2007 mais seulement sur la période de possession du véhicule, notre échantillon n'est pas cylindré. La période sur laquelle le ménage  $i$  est observé va donc de  $\underline{t}_i$ , date à laquelle le véhicule est décrit pour la première fois ( $\underline{t}_i = 2001$  si le véhicule est acheté avant le début de la période d'observation,  $\underline{t}_i = \text{année d'achat}$  autrement), jusqu'à  $\bar{t}_i$ , date d'abandon du véhicule.

Puisque l'on ne dispose pas ici d'informations sur les préférences individuelles des ménages, seulement la séquence des probabilités d'abandonner ou conserver le véhicule et, lorsque le choix de le conserver est pris, le niveau du kilométrage sont estimables. Puisque l'on suppose que les  $\varepsilon_{d,i,t}$  sont iid répartis selon la loi de Gumbel (9.4) et en considérant que les ménages ont pour objectif de maximiser l'utilité liée au choix, les probabilités d'abandonner ( $d_{i,t} = 1$ ) ou conserver ( $d_{i,t} = 0$ ) le véhicule s'écrivent respectivement :

$$\Pr(d_{i,t} = 1 | y_{i,t}, a_{i,t}, k_{i,t}, p_t, \mathbf{x}_{i,0}; \alpha, \boldsymbol{\beta}) = \frac{\exp(v_D(\mathbf{x}_{i,t}))}{\exp(v_D(\mathbf{x}_{i,t})) + \exp(v_K(\mathbf{x}_{i,t}))} \quad (7.22)$$

et

$$\Pr(d_{i,t} = 0 | y_{i,t}, a_{i,t}, k_{i,t}, p_t, \mathbf{x}_{i,0}; \boldsymbol{\lambda}, \gamma) = \frac{\exp(v_K(\mathbf{x}_{i,t}) + \gamma \mathbb{E}_{\mathbf{z}_{i,t+1}}[V(\mathbf{z}_{i,t+1}) | \mathbf{z}_{i,t}])}{\exp(v_K(\mathbf{x}_{i,t}) + \gamma \mathbb{E}_{\mathbf{z}_{i,t+1}}[V(\mathbf{z}_{i,t+1}) | \mathbf{z}_{i,t}]) + \exp(v_D(\mathbf{x}_{i,t}))} \quad (7.23)$$

Pour pouvoir évaluer la fonction de log-vraisemblance pour des valeurs précises de  $\boldsymbol{\lambda}$  et  $\gamma$ , le programme dynamique doit être résolu de façon exacte ou sa solution doit être approximée. Sous nos hypothèses on a

$$\mathbb{E}_{\mathbf{z}_{i,t}} [V(\mathbf{z}_{i,t}) | \mathbf{z}_{i,t-1}] = \sum_{q=1}^Q \sum_{r=1}^R \left\{ \begin{array}{l} \Pr(y_{i,t} | y_{i,t-1}, d_{i,t-1} = 0, \mathbf{x}_{i,0}; \boldsymbol{\beta}, \tau_y) \times \\ \Pr(p_{i,t} | p_{i,t-1}, d_{i,t-1} = 0, \mathbf{x}_{i,0}; \sigma_{\text{petrol}}, \sigma_{\text{diesel}}) \times \\ \ln(\exp(v_K(\mathbf{x}_{i,t}) + \gamma \mathbb{E}_{\mathbf{z}_{i,t+1}} [V(\mathbf{z}_{i,t+1}) | \mathbf{z}_{i,t}]) + \exp(v_D(\mathbf{x}_{i,t}))) \end{array} \right\} \quad (7.24)$$

Le problème ici étudié étant à horizon fini, les paramètres peuvent être estimés en partant de la dernière observation  $\mathbb{E}_{\mathbf{z}_{i,t+1}} [V(\mathbf{z}_{i,t+1}) | \mathbf{z}_{i,t}] = 0$  si  $d_t = 1$ . En effet, comme dit plus haut l'abandon du véhicule est une décision absorbante qui stoppe le processus. Le véhicule n'étant plus propriété de l'agent, il n'a plus aucune décision à prendre et le flux d'utilités escompté pour la décision est nul.

L'algorithme d'estimation se compose de deux étapes.

On estime avant tout les paramètres des probabilités de transition et on les utilise pour estimer, dans un deuxième temps, les paramètres du programme dynamique. Cette deuxième étape est à son tour composée de deux pas. La sous-étape "intérieure" (*inner step*) consiste à évaluer l'équation 8.21 pour les valeurs de  $\boldsymbol{\lambda}$  et  $\gamma$  obtenues à la première étape. Dans la sous-étape "extérieure" (*outer step*) on cherche des nouvelles valeurs de  $\boldsymbol{\lambda}$  et  $\gamma$  en estimant la partie concernant les probabilités de choix à l'aide d'un algorithme de maximisation non-linéaire. Les deux étapes sont répétées jusqu'à la convergence de l'algorithme.

L'estimation du facteur d'actualisation  $\gamma$  comme un paramètre du modèle entraîne des importants problèmes computationnels; pour cette raison je préfère le considérer comme un facteur fixe et estimer le modèle pour plusieurs valeurs de  $\gamma$ .

## 7.2 Les Données

Les données sont extraites du panel français "Parc Auto". L'enquête est représentative au niveau national et a vu le jour en 1983. Environ 1/3 de l'échantillon est renouvelé chaque année pour éviter les effets d'attrition. L'objectif est de récolter des informations très détaillées concernant les dynamiques de comportement de mobilité automobile.

Il s'agit de la meilleure base de données disponible pour l'étude des dynamiques de la demande automobile et du kilométrage. Une description



détaillée du ménage et ses membres est aussi disponible. Seulement une partie des informations disponibles sont utilisées dans ce travail.

Les données sont complétées par les séries des prix du carburant annuels en France au cours de la période, fournies par la base “[DIREM](#)”.

L'échantillon utilisé est composé des ménages ayant répondu sans interruption de 2000 à 2007 et n'ayant possédé qu'un véhicule par année. Ils représentent 51,6% des observations. Non seulement pour des raisons de simplification de l'approche, la proportion de cette population spécifique est assez grande pour être considérée comme intéressante.

Puisque le ménage peut abandonner son véhicule pour en acheter un autre au cours de l'observation, les véhicules ne sont pas forcément observés sur une même période. On a donc un échantillon cylindré de ménages possédant un échantillon non-cylindré de véhicules.

Afin de pouvoir estimer le modèle à horizon fini, seulement les décisions concernant les véhicules dont l'abandon est observable sont ici estimées. Il s'agit de 310 véhicules pour lesquels un total de 1151 observations sont disponibles.

Pour chaque véhicule les variables nécessaires à l'estimation du modèle présenté au cours de la section précédente (sec. 7.1) sont extraites et corrigées par la procédure d'appariement de la base Parc Auto. Ces variables renseignent les informations suivantes : âge du véhicule, kilométrage, kilométrage au compteur, type de carburant, prix d'achat, consommation moyenne de carburant et revenu du ménage le possédant.

Les caractéristiques moyennes observées lorsque la décision d'abandonner est prise sont reportées en table 8.1 pour chaque année observée et en table 8.2 par carburant pour la période entière. Puisque sont observées ici l'utilisation et l'élimination des voitures sur une période déterminée, il est naturel d'avoir une certaine attrition dans l'échantillon. Pour ce qui concerne les variables d'état considérés dans l'approche de modélisation, on observe donc une certaine variabilité dans les données.

TABLE 7.1 – Observed decisions, sample averages

Label	Frequencies and means					
	2001	2002	2003	2004	2005	2006
Car disposal <sup>a</sup>	25,99	28,71	32,30	45,31	48,05	100
Age of car <sup>b</sup>	7,33	7,22	8,40	7,29	7,26	7,63
Holding duration <sup>c</sup>	6,68	6,18	6,94	5,94	6,16	6,40
Mileage <sup>d</sup>	13414,88	13831,19	12709,40	13347,70	12206,81	15172,74
Sample size	227	202	161	128	77	46

*a.* Fraction of cars disposed during the period

*b.* in years

*c.* in years

*d.* in kilometres

TABLE 7.2 – Whole period descriptive statistics, sample averages

Label	Petrol	Diesel
Age when bought <sup>a</sup>	1,29	0,70
Holding duration <sup>b</sup>	8,39	6,48
Cumulated mileage at date of disposal <sup>c</sup>	82560,09	113347,44
Average fuel consumption <sup>d</sup>	0,75	0,65

*a.* in years

*b.* in years

*c.* in kilometres

*d.* in litre per kilometre

En distinguant par type de carburant, toutes années confondues, des différences remarquables sont à mettre en lumière. En moyenne les véhicules à essence sont achetés plus vieux mais conservés plus longtemps. L'utilisation, en termes de kilomètres parcourus est inférieure de celle observée en moyenne pour les diesel. La consommation moyenne de carburant est plus importante pour les premiers.

TABLE 7.3 – Average annual fuel prices per litre and by type

Fuel type	Year							
	2000	2001	2002	2003	2004	2005	2006	2007
Gas	1,10	1,04	1,03	1,03	1,07	1,19	1,26	1,29
Diesel	0,85	0,80	0,77	0,79	0,88	1,03	1,08	1,09

Le prix du carburant observé est un prix de marché dont l'agent est preneur. Dans cette approche, le prix utilisé est la moyenne annuelle en France métropolitaine. On remarque une forte augmentation au cours de la période considérée, comme indiqué dans le tableau 8.3. Bien que le prix nominal du diesel s'est maintenu plus bas que le prix nominal de l'essence tout au long de la période 2000 – 2007, on observe que le premier a augmenté à un taux plus élevé que ce dernier (voir section 3.2).

Le revenu annuel du ménage est supposé avoir un rôle significatif dans le processus de décision. Il est ici utilisé comme seule caractéristique résumant l'état du ménage. Seulement la classe de revenu est observable et non le niveau exact. Les bornes des classes sont par contre connues, on peut alors considérer cette variable comme une discrétisation de l'information continue de revenu.

Ces bornes, ainsi que les distributions annuelles par classe de revenu, sont reportées en table 8.4. Bien que la taille de l'échantillon diminue au fil du temps, en raison du nombre plus important de véhicules sortants qu'entrants, la relative distribution annuelle ne varie pas de façon significative.

TABLE 7.4 – Income class variable : frequencies

Class	Range in '000s of €	Year					
		2001	2002	2003	2004	2005	2006
1	< 15,2	21	20	17	15	4	4
2	15,2 – 19,1	31	23	28	15	7	5
3	19,1 – 22,9	37	31	18	20	8	7
4	22,9 – 26,7	37	29	21	16	12	3
5	26,7 – 30,5	22	17	12	14	12	8
6	30,5 – 38,1	24	22	19	9	5	2
7	38,1 – 45,7	25	30	20	22	14	10
8	45,7 – 61	16	13	13	10	7	4
9	> 61	14	17	13	7	8	3

### 7.3 Les Résultats

Les informations concernant le prix du carburant sont minces. Comme dit plus haut, on ne dispose ici que des moyennes annuelles observées en France métropolitaine. La non stationnarité des dynamiques d'évolution du prix sont testés par la méthode de la régression standard sur la série temporelle présentée en table 8.3. Malgré le nombre restreint d'observations, l'hypothèse n'est pas rejetée. Les variances des prix des carburants différenciés sont estimés de façon empirique à l'aide de la définition théorique correspondante.

TABLE 7.5 – Estimates : fuel prices

Label	Estimate
Variance $\sigma_{\text{petrol}}^2$	0,026
Variance $\sigma_{\text{diesel}}^2$	0,037

Le problème principal est celui dû à la taille de l'échantillon. Le choix d'utiliser ces estimations pour calculer les distributions discrètes des probabilités de transition entrants dans le modèle de programmation dynamique est critiquable. D'un autre côté, l'utilisation de séries chronologiques plus longues pourrait porter à incorporer d'autres problèmes, notamment dûs aux raisons différentes expliquant l'évolution des prix du carburant au cours des dernières décennies. Les changements structurels dans la formation des prix du carburant au cours des dernières décennies ne sont pas pris en compte

dans une telle hypothèse simple sur leur évolution. Il aurait alors été plus pertinent de développer une hypothèse permettant de considérer une période plus longue.

La matrice de transaction, qui n'est pas reportée aussi en raison de sa taille ( $401 \times 401$ ), est calculée en faisant varier le prix du carburant de 0€ à 4€ avec un pas de 1 centime.

Pour ce qui concerne le revenu, les résultats reportés en table 8.6 montrent, comme on pouvait s'y attendre, que plus le ménage  $i$  appartient à une classe élevée de revenu au moment  $t - 1$  plus la probabilité d'augmenter son revenu au moment  $t$  est importante.

L'hypothèse formulée pour la spécification du revenu est de transition asymétrique d'une classe de revenu à une autre. En observant la matrice de transition  $9 \times 9$ <sup>2</sup>, on peut constater cependant qu'il existe une inertie assez importante. La probabilité de rester dans une même classe de revenu varie entre 0,23 et 0,76. La probabilité de descendre ou monter en revenu est significative lorsque la différence est inférieure à deux classes, et est presque nulle lorsque cette différence augmente quelque soit son signe. Pour conclure on peut observer que la probabilité de descendre d'une classe est toujours supérieure à celle de monter.

En regardant les résultats reportés dans la table 7.7 on remarque une corrélation directe entre le kilométrage choisi et la classe de revenu d'appartenance. Plus le ménage est riche plus le kilométrage à parcourir avec le véhicule conservé est élevé. On remarque, toutefois, un ralentissement de l'augmentation du kilométrage demandé lorsque le revenu annuel dépasse les 45,7 K€. La relation entre le kilométrage choisi et le revenu n'est pas monotone croissante. Il existe plusieurs explications à ce phénomène, mais on ne peut pas les expliquer avec ce modèle puisqu'il a été défini conditionnellement au kilométrage choisi et des informations comme la zone de résidence, qui pourrait par exemple apporter un complément d'explication, n'ont pas été prises en compte dans sa formulation.

Les valeurs estimées des paramètres mettent en lumière deux autres relations intéressantes. Le kilométrage choisi est une fonction croissante du kilométrage cumulé et il a une corrélation inverse avec le prix du carburant.

---

2. Se référer au chapitre suivant 8.3

TABLE 7.6 – Estimates : income

Label	Estimate	Std. Dev. <sup>a</sup>	T-stat. <sup>b</sup>
Intercept	1,870	0,026	72,350
Previous inc. <sup>c</sup> class is [15, 2; 19, 1[ K€	0,451	0,035	12,810
Previous inc. class is [19, 1; 22, 9[ K€	0,748	0,035	21,580
Previous inc. class is [22, 9; 26, 7[ K€	0,953	0,039	24,250
Previous inc. class is [26, 7; 30, 5[ K€	1,159	0,036	32,610
Previous inc. class is [30, 5; 38, 1[ K€	1,304	0,033	39,130
Previous inc. class is [38, 1; 45, 7[ K€	1,449	0,037	39,510
Previous inc. class is [45, 7; 61[ K€	1,614	0,042	38,260
Previous inc. is $\geq 61$ K€	1,911	0,085	22,510
Variance $\tau_y^2$	0,019	0,001	13,930
Log-lik. <sup>d</sup> at convergence	1152,781		

*a.* Std. Dev. : standard deviation

*b.* T-stat. : robust Student statistic

*c.* Inc. : income, computed in thousands of €

*d.* Log-lik. : value of the partial log-likelihood function

Puisque le kilométrage cumulé à la date  $t$  dépend du kilométrage choisi à la date précédente ( $k_{i,t} = k_{i,t-1} + m_{i,t-1}$ ), la première observation signifie que plus le kilométrage choisi à la date  $t - 1$  est importante plus il le sera à la date  $t$ . L'augmentation du prix du carburant, au contraire, cause une baisse du kilométrage choisi.

De tels résultats concordent avec ceux reportés par [Goodwin \(1992\)](#), [de Jong and Gunn \(2001\)](#), [Graham and Glaister \(2002, 2004\)](#), et cela signifie que l'élasticité à long terme est supérieure, en valeur absolue, à l'élasticité à court terme<sup>3</sup>. L'agent a en effet plus d'options pour adapter ses choix aux changements de prix pour le long terme. Cependant, puisque l'objectif du modèle concerne le cycle de vie optimal pour un véhicule conditionnellement à ses caractéristiques initiales, l'effet d'une augmentation des prix du carburant est essentiellement relatif à des changements dans les comportements de mobilité.

Intéressons nous maintenant aux facteurs qui conditionnent le choix de garder le véhicule.

3. Because of the postulated functional form in equation 7.11, elasticities differ from one car to another but they are all proportional to  $\theta_p$  in the short-run and to  $\theta_p/(1 - \theta_k)$  in the long-run.

TABLE 7.7 – Estimation du choix du kilométrage

Label	Estimate	Std. Dev. <sup>a</sup>	T-stat. <sup>b</sup>
Intercept	3,774	0,304	12,420
Inc. <sup>c</sup> class is [15, 2; 19, 1[ K€	0,069	0,063	1,100
Inc. class is [19, 1; 22, 9[ K€	0,244	0,060	4,040
Inc. class is [22, 9; 26, 7[ K€	0,220	0,074	2,970
Inc. class is [26, 7; 30, 5[ K€	0,334	0,067	5,020
Inc. class is [30, 5; 38, 1[ K€	0,414	0,062	6,710
Inc. class is [38, 1; 45, 7[ K€	0,470	0,071	6,660
Inc. class is [45, 7; 61[ K€	0,418	0,086	4,850
Inc. is $\geq 61$ K€	0,261	0,153	1,700
Average fuel exp. <sup>d</sup> $\theta_p$	-0,593	0,070	-8,460
Cum. <sup>e</sup> mileage $\theta_k$	0,239	0,023	10,550
Variance $\omega_m^2$	0,088	0,005	16,300
Log-lik. <sup>f</sup> at convergence	1780,435		

<sup>a</sup>. Std. Dev. : standard deviation

<sup>b</sup>. T-stat. : robust Student statistic

<sup>c</sup>. Inc. : income, computed in thousands of €

<sup>d</sup>. exp. : expenditures. Average fuel expenditures are computed for one thousand kilometres

<sup>e</sup>. Cum. : cumulated. Cumulated mileage is computed in thousands of kilometres

<sup>f</sup>. Log-lik. : value of the partial log-likelihood function

Je suppose que ce choix est conditionné par le degré de préférence pour le présent du ménage. Le modèle est alors estimé pour trois différentes valeurs du facteur d'actualisation :

- $\gamma = 0$  signifie que le ménage n'accorde pas d'importance au futur mais il se contente de regarder au présent et il prend les décisions en fonction de sa situation actuelle. Un tel agent est dit myope.
- $\gamma \in ]0; 1[$  indique que le présent et le futur sont intéressants pour le ménage. Une valeur de  $\gamma$  inférieure à 0,5 indique un plus fort intérêt pour le présent et vice-versa. Dans l'estimation du modèle on fixe  $\gamma = 0,5$ .
- $\gamma = 1$  signifie que le ménage s'intéresse uniquement au futur. Puisque je ne considère pas plausible que le présent ne soit pas du tout pris en compte, on fixe  $\gamma = 0,99$  pour estimer ce cas extrême.

Les résultats reportés dans la table 8.10 montrent, en premier lieu, une tendance à vouloir garder le véhicule quelque soit la valeur du facteur d'actualisation.

Les rôles joués par les dépenses liées à la conservation (à travers le kilométrage cumulé) et l'utilisation du véhicule (à travers le prix du carburant), et par sa valeur résiduelle sont significatifs pour un ménage myope mais décroissent au fur et à mesure que l'intérêt pour le présent diminue, pour disparaître lorsque le ménage est intéressé quasiment que par le futur ( $\gamma = 0,99$ ). Dans ce cas, la seule information qui est considérée par le ménage lorsqu'il prend sa décision, est le revenu qu'il s'attend à avoir dans le futur. Une observation importante concerne la vitesse à laquelle les trois informations perdent de leur importance pour le choix. Le premier facteur à disparaître avec l'augmentation de l'intérêt porté au futur est le prix d'entretien du véhicule, suivi par le prix du carburant. La valeur résiduelle du véhicule est la dernière à perdre son rôle. Son impact, lorsqu'il est significatif, est positif. Cela signifie que plus la valeur obtenue en cas de revente du véhicule est importante plus le ménage a tendance à conserver le véhicule. En se rappelant que cette valeur résiduelle est calculée comme une fonction décroissante de l'âge et du kilométrage cumulé (7.6) on peut conclure qu'une utilisation plus importante dans le passé du véhicule implique une plus forte probabilité d'abandonner le véhicule. Le kilométrage cumulé est à son tour une fonction déterministe des choix passés de kilométrage que l'on a vu être inversement proportionnelle au prix du carburant. L'augmentation de ce dernier diminue alors le kilométrage choisi en augmentant par conséquent la probabilité de garder le véhicule. Cependant lorsque le facteur d'actualisation est tel que l'effet du



prix du carburant sur le choix est significatif on observe une valeur négative du paramètre associé. Cela remet en question la conclusion précédente parce qu'il signifie que l'augmentation du prix du carburant a un effet direct sur le choix de garder le véhicule contraire à celui observé.

Pour conclure on observe que le rôle joué par le revenu du ménage ne subit pas trop de changement avec le changement de la valeur d'actualisation. Les ménages ayant un revenu annuel inférieur à 30,5 K€ et ceux appartenant à l'avant dernière classe de revenu ont la même probabilité de garder le véhicule, et cette probabilité est plus importante que celle de ceux qui perçoivent entre 30,5 K€ et 45,7 K€ par an. Pour ces derniers on voit, en outre, que lorsque le revenu est supérieur à 38,1 K€ et le ménage à une nette préférence pour le futur la probabilité de garder le véhicule est encore plus faible. Les ménages ayant un revenu annuel supérieur à 61 K€ ont la plus importante probabilité de garder le véhicule, et cela indépendamment du facteur d'actualisation.

Il y a plusieurs façons d'expliquer le rôle du revenu que l'on vient d'analyser, mais puisque les facteurs qui en permettraient la compréhension n'ont pas été inclus dans ce modèle je préfère éviter toute spéculation sur le phénomène.

TABLE 7.8 – Estimation de la probabilité de conserver le véhicule

Label	Estimate	Std. Dev. <sup>a</sup>	T-stat. <sup>b</sup>	Estimate	Std. Dev.	T-stat.	Estimate	Std. Dev.	T-stat.
Intercept	1,887	0,355	5,320	1,422	0,380	3,744	0,790	0,426	1,855
Inc. <sup>c</sup> class is [15, 2; 19, 1[ K€	-0,101	0,238	-0,420	0,005	0,259	0,019	0,081	0,290	0,279
Inc. class is [19, 1; 22, 9[ K€	-0,136	0,236	-0,580	-0,059	0,257	-0,231	-0,014	0,287	-0,048
Inc. class is [22, 9; 26, 7[ K€	-0,319	0,254	-1,260	-0,168	0,276	-0,610	-0,035	0,306	-0,114
Inc. class is [26, 7; 30, 5[ K€	-0,108	0,277	-0,390	0,067	0,299	0,225	0,125	0,335	0,374
Inc. class is [30, 5; 38, 1[ K€	-0,452	0,236	-1,910	-0,551	0,263	-2,100	-0,665	0,313	-2,122
Inc. class is [38, 1; 45, 7[ K€	-0,303	0,292	-1,040	-0,557	0,352	-1,582	-0,920	0,479	-1,920
Inc. class is [45, 7; 61[ K€	0,101	0,334	0,300	-0,706	0,484	-1,459	-1,802	1,148	-1,569
Inc. is $\geq 61$ K€	1,346	0,800	1,680	10,734	0,758	14,155	14,379	0,769	18,690
Average fuel exp. <sup>d</sup> $\beta_p$	-0,010	0,004	-2,800	-0,007	0,004	-1,945	-0,004	0,004	-0,991
Cum. <sup>e</sup> mileage $\beta_k$	-0,003	0,001	-2,290	-0,002	0,001	-1,286	-0,001	0,001	-0,525
Scrap/sell-off value <sup>f</sup> $\alpha$	4,649	1,199	3,880	2,847	1,372	2,075	1,917	1,534	1,250
Discount factor $\gamma$	0			0,5			0,99		
Log-lik. <sup>g</sup> at convergence	-636,075			-500,606			-384,519		

a. Std. Dev. : standard deviation

b. T-stat. : robust Student statistic

c. Inc. : income, computed in thousands of €

d. exp. : expenditures. Average fuel expenditures are computed for one thousand kilometres

e. cum. : cumulated. Cumulated mileage is computed in thousands of kilometres

f. Scrap/sell-off value is computed in thousands of €

g. Log-lik. : value of the partial log-likelihood function

## 7.4 Conclusion

La simplicité de plusieurs des hypothèses sur lesquelles le modèle ici présenté se base laisse à penser que celui-ci ne reflète pas tout-à-fait les comportements de décision humains face à la question d'utilisation automobile.

En premier lieu j'admets que la spécification de l'utilité n'est pas vraiment proche de la réalité. En effet un grand nombre de facteurs jouant probablement un rôle dans ce processus de décision n'y sont pas inclus. Par exemple, un choc imprévisibles suite à un accident peut influencer sur la performance de la voiture en termes de consommation de carburant et / ou de dommages indésirables qui pourraient en affecter la valeur (et peut-être à l'origine de la décision de se séparer du véhicule plus tôt).

Une deuxième critique possible concerne l'hétérogénéité propre au type de données utilisées. Or, dans ce modèle ne sont traitées ni la présence d'hétérogénéité ni la corrélation entachant l'observation inobservable. Je ne considère ni la possibilité d'autocorrélation ni de corrélation entre les équations. Encore une fois, on peut penser à des chocs qui touchent le moteur et / ou provoquent des dommages physiques affectant simultanément la durée de possession et l'utilisation.

De plus la description du ménage est ici réduite à sa classe de revenu. L'environnement n'est représenté que par les évolutions des prix du carburant. L'exclusion des variables démographiques, géographiques, socio-économiques qui sont susceptibles de jouer un rôle important réduisent aussi la crédibilité du modèle.

La demande de kilométrage est fonction du prix du carburant, du revenu et de la demande de kilométrage retardée. La classe de revenu et la dépense en carburant jouent un rôle principal sur la décision de conserver ou pas le véhicule. La valeur résiduelle du véhicule joue un rôle significatif mais moindre que les deux facteurs précédents. De plus cet effet varie en fonction de la valeur d'actualisation, c'est-à-dire des hypothèses formulées sur le taux de préférence pour le futur. Cependant, la valeur d'actualisation étant fonction du kilométrage cumulé, qui est défini comme fonction déterministe du kilométrage, la décision de garder ou pas une voiture à une date  $t$  est fonction de la demande de kilométrage à cette date par le biais de l'utilité attendue maximale que le ménage obtiendrait par la possession du véhicule à la date  $t + 1$  lorsque la décision de conserver est prise. Le ménage à un comportement tourné vers le futur, c'est-à-dire qu'il balance entre le présent et l'avenir.

En tout cas on peut conclure que l'hypothèse d'agents prévoyants améliore considérablement la compréhension des processus de décision modélisés.

Comme dit plus haut, l'approche doit cependant être affinée sous plusieurs aspects.

Une première évolution possible serait d'introduire un plus large nombre de variables d'état observées, permettant une meilleure description de la structure démographique du ménage, de son état et des conditions initiales des véhicules lui appartenant. Un tel ajout d'informations pourrait permettre d'expliquer certains résultats obtenus avec la présente approche. Les caractéristiques socio-démographiques pourraient également servir à expliquer l'évolution des revenus. Dans le cas où de telles variables seraient exclues les probabilités de transitions seraient à traiter avec attention, en raison de leur probable corrélation avec le revenu.

Une autre amélioration importante serait de définir un ensemble d'hypothèses plus vraisemblables pour les variables d'état non observées, à savoir le problème de l'hétérogénéité inobservée persistante qui peut créer des corrélations non observées au cours du temps mais aussi de l'effet probable d'autres variables dans le temps sur les variables d'état non observées (comme la possibilité que la voiture soit en très mauvais état).

Un dernier point serait d'étendre le modèle en plusieurs directions. On peut par exemple modéliser la décision d'acheter une marque spécifique / modèle / type de voiture et, une fois décidé de s'en séparer, le choix de la remplacer ou non, et par quoi. En outre, il faudrait considérer le nombre de ménages qui possèdent plus d'une voiture, ainsi que la population des ménages non-motorisés.

En prenant en compte tous ces facteurs, on pourrait alors obtenir une bien meilleure description des comportements des ménages qui règlent leurs décisions concernant la possession et l'utilisation automobile. Cependant la réalisation d'un modèle complet demanderait un coût computationnel extrêmement élevé ainsi que le traitement d'un volume important d'informations.





## Chapitre 8

# A dynamic discrete\continuous choice model for forward-looking agents

Ce travail a fait l'objet d'une communication ([Cernicchiaro and de Lapparent \(2012\)](#)) à l'occasion du 66ème Meeting Européen de l'Econometric Society qui a eu lieu à Malaga du 27 au 31/08/2012.

Par absence de temps, je ne traduit pas ici ce travail et je reporte la version soumise pour cette conférence <sup>a</sup>. Le chapitre suivant est donc en anglais.

Je vais en présenter ici un résumé en français.

### Résumé

Le caractère durable des véhicules implique que l'utilisation et le temps en causent la dépréciation.

Les études portant sur les décisions de possession et d'utilisation automobile ne peuvent donc pas être dissociées. De plus, une telle caractéristique laisse à penser que l'individu soit prévoyant et adopte un comportement "tourné vers le futur" lorsqu'il est confronté à ces choix.

Ce chapitre propose un modèle dynamique "tourné vers le futur" discret/continu.

La spécification du modèle se base sur le cadre défini par [Rust \(1987\)](#). À la différence de cet auteur, le modèle ici présenté utilise un logit discret/continu, inspiré des travaux de [Bento et al. \(2009\)](#). L'utilisation d'une telle spécification pour le modèle inter-temporel permet d'éviter l'estimation a priori du choix du kilométrage. En effet, en utilisant l'identité de Roy la demande de kilométrage peut être dérivée directement de l'utilité de possession. Les deux décisions sont donc estimées simultanément.

Le modèle est estimé à horizon fini, en utilisant les données de panel issues des étapes intermédiaires de traitement de la base "Parc Auto".

Les résultats semblent montrer une amélioration des estimations lorsque l'hypothèse d'agent "tourné vers le futur" est formulée <sup>b</sup>. On observe que le "revenu conservé", c'est-à-dire la différence entre le revenu du ménage <sup>c</sup> et la valeur du véhicule a l'effet le plus important sur ses décisions. De plus, le type d'utilisateur principal semble avoir un impact, qui varie selon la préférence pour le futur supposée, sur ses décisions.

Puisque la méthode d'estimation ici utilisée permet d'obtenir des estimation seulement consistantes, et pas efficaces, et la spécification choisie est très simple (pour des raisons de temps disponible pour sa réalisation) je préfère ne pas spéculer sur les raisons qui peuvent expliquer ces résultats.

<sup>a</sup>. [http://www.eea-esem.com/files/papers/eea-esem/2012/2719/Cernicchiaro\\_How%20to%20use%20the%20car%20fleet.pdf](http://www.eea-esem.com/files/papers/eea-esem/2012/2719/Cernicchiaro_How%20to%20use%20the%20car%20fleet.pdf)

<sup>b</sup>. Le facteur d'actualisation n'est pas estimé dans ce travail, mais le modèle est estimé pour plusieurs valeurs d'actualisation possibles.

<sup>c</sup>. ici représenté par le centre de la classe de revenu que le ménage déclare au cours de l'enquête

During the last decade more than 80% of French households chose motorization, and the average holding duration of a car has increased. The modeling of car demand and use plays a central role in numerous fields, such as energy management, public transportation planning, auto industry policies, and air quality monitoring. It is thus of a great importance to identify factors driving household's behavior in making decisions on their fleet composition and use in order to be more capable of forecasting car demand.

Car demand is composed of many different dimensions. Choosing motorization, in fact, implies that households face a series of equipment choices including, to cite only a few, how many cars to own and their type, which fuel to use, and how often to use their fleet. As already highlighted by Train (1986), car ownership, fleet size and variety, fuel efficiency, and mileage are important considerations when managing transport-related policies that are in line with sustainable development.

The literature shows the important effects of car attributes, such as purchase price, maintenance and use costs, and value on this choice. Increases in automobile market and fuel prices, contractions of the household's budgetary constraint, and household socio-demographic and geographic characteristics (e.g. age and profession of the household's head, household size, urban size, habitat type, transportation supply), as highlighted by the French Automobile Constructors Committee (CCFA (2006)), are all important topics in car demand and use modelling.

Moreover, the durable character of the car can imply that agents consider a purchase as an investment. If that occur, costs and benefits of owning a car fleet and using it over a period of time have to be simultaneously considered. As shown in Julliard (2007) (figure 1), a higher initial expenditure may require a longer period of time to receive a return. At the same time, the cost of operating a car increases with its age due to physical depreciation (e.g. Baron (2002)). The agent thus has the choice enter, for example, to use a car extensively in the short-term to gain a fast return on the investment, use it less to obtain a duration of longer holding or adapt the amount to drive to the variations of the fuel price, by reducing the use of the car when it increases (Julliard (2007), Graham and Glaister (2004)). All these factors can therefore reasonably be taken into account when studying consumers' behaviors.

A large number of studies have analyzed car holding and use behaviors during the last 40 years. The existing literature is extensive and can be split into several groups depending on the level of disaggregation, data type



(describing real or hypothetical choices), temporal dynamic inclusion, use of supply\demand equilibrium and dimensions of demand considered. For our purposes one can review, for example, [Hoeherman et al. \(1983\)](#), [Mannering \(1983a\)](#), [Berkovec and Rust \(1985\)](#), [Hensher et al. \(1990\)](#), [Gilbert \(1992\)](#), [de Jong \(1990, 1996\)](#) and more recently [Bhat and Sen \(2006\)](#). These past analyses employ various data and techniques, analyzing the dynamics of car ownership, holding duration and\or use at the disaggregate level with multiple intents such as gasoline demand forecast, brand fidelity analysis, urban zone effect or tax impact on choices.

However, only a few approaches treat fleet size, vehicle type, and mileage to travel as simultaneous choices, an important aspect of car demand modeling as argued by [Train \(1986\)](#), who proposed a Generalized Extreme Value (GEV) model considering the non-ordered aspect of these choices.

Moreover, as pointed out by [de Jong and Kitamura \(2009\)](#), the majority of these studies consider the problem from a static point of view, forgetting the dynamic that very likely drives such decisions. In other words, to our knowledge and with rare exceptions (e.g. [Rust \(1987\)](#)), even when all demand dimensions are considered as simultaneous, agents' choices are considered to be myopic or, in the rare case a temporal dynamic is considered explicitly, based on past evolutions.

Even if the backward-looking agent hypothesis seems to be credible concerning the total fleet's decisions, a car is a durable good, that is a good that does not quickly wear out and that yields utility over time rather than being completely consumed in one use. For this reason, we argue that when decisions on car holding and use are taken, one must in some way consider the future decline of the vehicle.

We suppose then the agent anticipates the future evolution of his environment and bases his decisions about a single car on this knowledge, in concordance with Lucas' critique ([Lucas \(1976\)](#)). That is, households choose to scrap or keep a vehicle by anticipation of future use, fuel price evolution, future changes in household structure, and so on.

This work formulates a discrete\continuous finite-horizon optimal stopping problem under uncertainty to describe the behavior of a household as it regards car holding duration and mileage driven. A household accounts for the uncertain evolution of fuel prices and income. In the present approach, the decision to keep or dispose of a car in the fleet will be taken conditional on household and vehicle characteristics, choices made for other vehicles and principal user's characteristics.

As the approach is finite-horizon, it is therefore solved by backward induc-

tion. The parameters of the structural model are estimated using a Nested Fixed Point Algorithm.

The model is demonstrated by drawing data from the French “Parc Auto” (Car Fleet) panel survey, the best available tool for studying French household behaviors in automobile equipment terms. We focus on the population of households observed from 2002 to 2008. Data collected in 2002 are used as the initial conditions and decisions are estimated from 2003.

Although there is no innovation in terms of the estimation algorithm, we combine discrete\continuous model and Rust (1987) dynamic forward-looking specification, a proposition that, to our knowledge, was not previously implemented in the literature. Moreover, we are the first to propose an application to the topic of household car ownership and use over time, using disaggregate panel French data.

The outline of the chapter is as follows : The model is developed in section 8.1, discussing the different parts constituting the approach as well as how the parameters of the primitives are estimated. Data are then presented in section 8.2. The results are debated in section 8.3. The last section draws conclusions and defines a roadmap for future research.

## 8.1 Model

### 8.1.1 Framework

We build the specification on six assumptions that define what is known as the “dynamic programming – conditional logit” model of Rust (1987) : additive separability of utility functions, the unobserved state variables are iid with cdf  $F$ , conditional independence of future observed state variables with current unobserved state variables, conditional independence of pay-off variable with current unobserved state variables, the unobserved state variables are iid extreme value type 1 distribution, discrete support of observed state variables. This framework is the simplest for estimation : it is computationally manageable but comes at the cost of rather restrictive assumptions on the interpretation of the model.

A household  $i$ 's purpose is to simultaneously determine the optimal ownership duration  $\bar{t}_j \in \{1, \dots, T\}$  and intensity of usage  $\mathbf{m}_{j,\cdot} = (m_{j,1}, \dots, m_{j,t})$  for each of its cars  $j$ . We suppose these two decisions, car disposal or car

keeping ( $d$ ) and use ( $m$ ), to be taken at the beginning of each period  $t$  conditional on the household's state and environment. The problem is then an optimal "use and stop" problem. That is, when vehicle  $j$  is kept, the household chooses an amount of mileage  $m_{ijt}$ . If it disposes of the vehicle, the decision process stops.

For multiple vehicle households, we suppose homogeneous income repartition between cars. However, choices concerning a multiple car fleet are not simultaneously estimated in this work.

We suppose that choice of make \ model \ type of the car is made at the purchase date, and therefore it is not estimated. Due to a lack of data, we do not model the car's intended purpose and do not capture the substitution effect.

The state variables' vector is labeled  $\mathbf{z}_{ijt}$ , and it is assumed to be separable into two subsets : a first one of state variables observed by the econometrician, labeled  $\mathbf{z}_{ijt} = \{\mathbf{x}_{ijt}, \epsilon_{ijt}\}$ .  $\mathbf{x}_{ijt}$  and a vector of two unobserved state variables  $\epsilon_{D,ijt}, \epsilon_{K,ijt}$  that are assumed to be independently and identically distributed with cumulative distribution function  $\exp(-\exp(-\epsilon_{d,ijt}))$ . It is also assumed that the unobserved state variables regarding the utility functions are modelled as additive shocks.

Households' preferences over possible sequences of state variables can be represented by a time separable discounted utility function  $\sum_{t=1}^T \gamma^t u(\mathbf{z}_{ijt}, d(\mathbf{z}_{ijt}), m(\mathbf{z}_{ijt}))$  where  $\gamma$  is the discounting factor and  $u(\mathbf{z}_{ijt}, d(\mathbf{z}_{ijt}), m(\mathbf{z}_{ijt}))$  is the per-period utility function.

The sequence of decisions is taken to maximize the expected utility with respect to the distribution of the sequences of state variables :

$$\max_{\bar{t}_i} \left\{ \mathbb{E}_{\mathbf{z}_{i,\cdot}} \left( \sum_{t=1}^{T_i} \gamma^t u(\mathbf{z}_{ijt}, d_{ijt}, m_{ijt}) \mid \mathbf{z}_{ij0} \right) \right\} \quad (8.1)$$

where  $\mathbf{z}_{ij0}$  is a set of initial conditions.

The intertemporal optimization problem can be formulated as a sequential decision problem whose solution is the same when using its Bellman representation (see for instance [Bellman \(1957\)](#), [Rust \(1996\)](#), [Bertsekas \(2000\)](#)). Let  $V(\mathbf{z}_{ijt})$  denote the maximum value of all utility flows to household  $i$  knowing its state variables  $\mathbf{z}_{ijt}$ .  $V(\mathbf{z}_{ijt})$  is defined recursively by the solution to the following Bellman equation :

$$V(\mathbf{z}_{ijt}) = \max \left\{ v_D(\mathbf{x}_{ijt}) + \epsilon_{D,ijt}, v_K(\mathbf{x}_{ijt}) + \gamma \mathbb{E}_{\mathbf{z}_{ijt+1}} [V(\mathbf{z}_{ijt+1}) \mid \mathbf{z}_{ijt}] + \epsilon_{K,ijt} \right\}. \quad (8.2)$$

The optimal demand for mileage  $m^*(\mathbf{z}_{ijt})$  and the decision to keep or dispose of the car,  $d^*(\mathbf{z}_{ijt})$ , are the arguments that maximize equation 9.6. The decisions at period  $t$  affect the evolution of future values of the state variables, but the household faces uncertainty about these future values. The beliefs of household  $i$  about sequences of states are modeled by a Markov transition distribution function  $G_z(\mathbf{z}_{ijt+1}|\mathbf{z}_{ijt}, d_{ijt}, m_{ijt})$  where, by convention,  $d_{ijt} = d(\mathbf{z}_{ijt})$  and  $m_{ijt} = m(\mathbf{z}_{ijt})$ .

Considering a vehicle's value as the amount a household invests every time it decides to continue ownership one period more, as proposed by Bento et al. (2009), the utility for household  $i$  to keep vehicle  $j$  for period  $t$  can be defined as follows :

$$v_K(\mathbf{z}_{ijt}) = \beta_0 - \exp(-\beta_1(y_{ijt} - v_{jt})) + \exp\left(-\beta_2 o_{jt} + \sum_{d=1}^D \mu_d t_{djt}\right) \quad (8.3)$$

where  $y_{ijt}$  is the amount of available income used for vehicle  $j$  during period  $t$ , and  $v_{jt}$  is the vehicle's scrap value at date  $t$  modeled as a geometrically decreasing function of ownership duration  $a_{ijt}$  and cumulative use  $k_{ijt}$ .

$$v_{jt} = \frac{P_j}{1 + a_{ijt} * k_{ijt}} \quad (8.4)$$

Purchase price  $P_j$  belongs to the set of initial conditions and represents a vehicle's scrap value at the purchase date.

$o_{jt} = c_j * p_{jt}$  is vehicle operating cost, computed as a function of vehicle consumption  $c_j$  (initial condition) and annual fuel price  $p_{jt}$  defined as

$$p_{ijt} = \mathbb{I}(\text{fuel type of } i \text{ is petrol}) p_{\text{petrol},t} + \mathbb{I}(\text{fuel type of } i \text{ is diesel}) p_{\text{diesel},t}, \quad (8.5)$$

although different households owning cars with the same fuel type are faced with the same market price. This simplifying assumption is used because the available data on fuel prices are national averages by year over the considered period. These prices are actually different at several locations within the French metropolitan territory.  $t_{djt}$  are dummy variables representing age and employment status of principal user of  $j$  for period  $t$ .

- $t_{1j}$  : principal user less than 40 years old and employed
- $t_{2j}$  : principal user less than 40 years old and unemployed
- $t_{3j}$  : principal user age  $\in [40; 60[$  and employed
- $t_{4j}$  : principal user age  $\in [40; 60[$  and unemployed

As postulated by [Mannering \(1983b\)](#), principal user characteristics have an important effect on car use choice. As we focus on decisions made for each car separately, and not on the total fleet's choice, only these characteristics are included in the estimation process and a household's description is assumed to be demonstrated by its income.

Supposing that ownership decisions for other vehicles are known by household  $i$  when it makes the decision for vehicle  $j$ , we can calculate available income  $y_{ijt}$  as

$$y_{ijt} = Y_{it}/S_{K,it} \quad (8.6)$$

where  $Y_{it}$  is annual household income and  $S_{K,it}$  is fleet size obtained from keeping vehicle  $j$ . Household income is observed as a categorical variable.  $Y_{it}$  is the center value of a household's income class.

There is nothing to discuss regarding the evolution of the ownership duration variable, that is a deterministic state variable defined as :

$$a_{ijt} = a_{ijt-1} + 1. \quad (8.7)$$

The cumulative use at the beginning of a period  $t$  is a deterministic function of the chosen mileage at the beginning of period  $t - 1$ , which was optimally chosen as a function of the other state variables.

$$k_{ijt} = k_{ijt-1} + m_{ijt-1}. \quad (8.8)$$

In the present approach, vehicle  $j$  optimal use  $m_{ijt}^* = VMT_{K,u}$  is estimated using Roy's Identity :

$$VMT_K = -\frac{\delta v_K / \delta o_{jt}}{\delta v_K / \delta Y_{it}} \quad (8.9)$$

Under our definition of utility  $v_K$ , Marshallian demand for mileage is

$$m_{ijt}^* = \alpha * \exp \left( \beta_1 \left( \frac{Y_{it}}{T_{K,it}} - \frac{P_j}{1 + a_{ijt} * k_{ijt}} \right) - \beta_2 (c_j * p_{jt}) + \sum_{u=1}^4 \mu_u t_u \right) \quad (8.10)$$

The probability for vehicle  $j$  to have cumulative use  $k_{ijt+1}$  knowing  $k_{ijt}$  is the probability for household  $i$  to chose mileage  $m_{ijt}^*$  for period  $t$  :

$$P(m_{ijt}^* | \mathbf{z}_{ijt}) = \frac{1}{(2\pi)^{1/2} \sigma_i} \exp \left( -\frac{1}{2} \left( \frac{m_{ijt}^* - m_{ijt}}{\sigma_i} \right)^2 \right) \quad (8.11)$$

As stated by [Aguirregabiria and Mira \(2010\)](#), assumptions on  $G_z$  are key modeling decisions in the econometrics of dynamic discrete structural models. The form of the estimation criteria (e.g. log-likelihood function) and choice of the appropriate solution and estimation methods crucially depend on these assumptions. The joint Markov transition distribution of the state variables is defined as the product of marginal distributions.

Fleet size and principal user optimal sequences of choice are assumed to be known by household. It therefore does not face uncertainty about them, and probability transactions are not estimated.

Households are assumed to be sensitive to a fuel price evolution stage-by-stage. Expectations about this evolution are then based on a discrete representation of their respective processes. Sensitivity remains the same in between some lower and upper bound values. Our modelling point of view about the evolution of fuel prices is extremely simple. We assume that their dynamics are not stationary (at the second order) over time.  $\forall g \in \{\text{petrol, diesel}\}$ ,

$$p_{g,t} = p_{g,t-1} + \epsilon_{g,t} \quad (8.12)$$

where  $\epsilon_{g,t} \xrightarrow{iid} \mathcal{N}(0, \sigma_g^2)$ . Then let  $(\bar{p}_{g,0}, \dots, \bar{p}_{g,R})$  be a sequence of predetermined values over the real line. It serves as a bound of contiguous intervals that models the different stages of fuel price  $g$ . Let also  $\tilde{p}_{g,r} = (\bar{p}_{g,r} + \bar{p}_{g,r+1})/2$  be the centers of these intervals (up to some convention as it regards the first and the last of these intervals). They will model some “representative” values. Then :

$$\Pr(p_{g,t} = \tilde{p}_{g,r} | p_{g,t-1} = \tilde{p}_{g,l}, d_{ijt-1} = 0; \sigma_g) = \Phi\left(\frac{\bar{p}_{g,r} - \tilde{p}_{g,l}}{\sigma_g}\right) - \Phi\left(\frac{\bar{p}_{g,r-1} - \tilde{p}_{g,l}}{\sigma_g}\right) \quad (8.13)$$

where  $\Phi$  is the cumulative distribution function of a standard normal distribution. Nevertheless, such an approach neither solves the problem of choosing the right number of discrete intervals to obtain an acceptable representation of the Markov transition probability nor gives a clue about whether choosing the center of the interval as the representative value is correct. In this application, intervals have a range equal to 1 cent of €.

The last observed state variable is household income. The interviewed households have to declare, for each participation year, their annual income class. The income variable is then not continuous, but it is grouped in 13 classes.

One can be tempted to treat this kind of information as continuous using the Ordinary Least Squares estimator, choosing for each class a representative value as, for example, the central point. Such a choice will be incorrect because of the inconsistency of the OLS estimator in this type of situation. Pointing out the discrete character of the income variable, Stewart (1983), Terza (1985) and Bhat (1994) indicate that the use of Maximum Likelihood is more reasonable.

Let  $y^*$  be the latent continuous variable (not observed),  $y$  the variable grouped in  $J$  classes (observed) and  $\mu_{j-1}, \mu_j, j = 1, \dots, J$  the logarithm<sup>1</sup> of class  $j$  bounds. One can deduce the following :

$$y = j \Leftrightarrow \mu_{j-1} < \ln(y^*) < \mu_j$$

Supposing  $\ln(y^*) = \beta'x + \epsilon$ ,  $\epsilon \rightsquigarrow \text{Logistic}(0, \sigma)$ <sup>2</sup>, the probability to observe the income class  $j$  knowing the household's characteristics  $x$  can be written as :

$$P(y = j | x) = \left[ F\left(\frac{\mu_j - \beta'x}{\sigma}\right) - F\left(\frac{\mu_{j-1} - \beta'x}{\sigma}\right) \right] \quad (8.14)$$

where  $F(x) = \frac{1}{1 + \exp(-x)}$  is the logistic function.

It is then an *Ordered Logit Model with Known Bounds*.

By convention, the extreme bounds are defined as  $\mu_0 = -\infty$  and  $\mu_J = \infty$ . One has

$$\begin{aligned} P(y = 1 | x) &= F\left(\frac{\mu_1 - \beta'x}{\sigma}\right) \\ &\text{and} \\ P(y = J | x) &= 1 - F\left(\frac{\mu_J - \beta'x}{\sigma}\right) \end{aligned}$$

As bounds  $\mu_j$  are known,  $\beta$  and  $\sigma$  are identifiable and can be estimated. The log-likelihood associated to the model is computed as

$$LL = \sum_i \sum_j I(y_i = j) \ln(P(y_i = j | x_i)) \quad (8.15)$$

---

1. It is known that the normality assumption is not adapted for income.

2. We acknowledge that the hypothesis of normal errors, proposed in the literature, is more credible for the considered problem. However, as probit results are generally proportioned to the logit ones and the latter interpretation is easier than the former, the errors here are considered as logistic

where  $I(y_i = j) = 1$  if household  $i$  income is observed in class  $j$  and 0 otherwise.

Multiple optimization techniques are proposed in the literature for this model likelihood. Here we use the Newton-Raphson method<sup>3</sup>.

The hypothesis we employ in this work is that the income follows an autoregressive Markov process, meaning that the income class for period  $t$  is a function only of the income class observed for the previous period  $t - 1$ . The underlying relation is then

$$\ln(y_{i,t}^*) = \beta_0 + \sum_{j=1}^{J-1} \beta_j \mathbb{I}(y_{i,t-1} = j) + \epsilon_{y,i,t} \quad (8.16)$$

A final assumption is that disposing of a car is an absorbing decision. Thus, once the household disposes of the car, all the transition probabilities are set equal to 0.

### 8.1.2 A short discussion about Assumptions

Evidently, we acknowledge that our assumptions are restrictive and their realism can be criticized.

First of all, we treat decisions on each vehicle separately and the purchase behavior is not modeled. Thus even though the substitution effect a household is likely to apply in the modeled choice, we do not account for it.

We also consider only principal user characteristics, assuming that income represents the household's status. Needs related to no-driver members are then left out of the estimation.

Moreover, a remaining issue is the credibility of the forward-looking behavior assumption. A recent study ([Turrentine and Kurani \(2007\)](#)) conducted on 56 American household shows, in fact, that households do not explicitly think of a car use in terms of investment.

However, as previously argued, the durable character of a car implies a decline of its value and efficiency that is well-known by the household. Assuming that an agent takes into account the future impact of his current car usage can, however, be considered correct because, even if he will not consider the car as a way of refunding his initial investment in the future, such thinking might be considered irrational.

Overall, this work intends to serve as a starting point for a more comprehensive way to study automobile equipment choices. We consider then that the previous issues are limits to explore in future research.

---

3. The trust regions algorithm was originally chosen, but the model did not converge



### 8.1.3 Estimation

The estimation of the structural parameters in which we are interested  $\boldsymbol{\lambda} = \{\alpha, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\delta}, \omega_m, \sigma_{\text{petrol}}, \sigma_{\text{diesel}}, \tau_y\}$  is based on the sequences of choices (keep \ dispose and mileages), state variables and initial conditions observed for each vehicle  $j$  of sample's households.

According to [Magnac and Thesmar \(2002\)](#), the discount factor  $\gamma$  is "poorly" identified. As a result, even though econometric theory would allow estimating it as an additional parameter, it often leads to serious numerical problems when utility functions are linear in parameters. To this end, it is considered as fixed, and we estimate the model for different values.

The estimation process is composed of two steps.

In the first step, parameters of the transition probability distributions are estimated.

The probabilities for mileage, fuel price and income to have the observed values are obtained by equations 8.11, 8.13, 8.14 in the former subsection.

In the second step, we estimate the parameters of the dynamic programming problem given the transition probability distributions.

Let  $\underline{t}_{ij}$  be the first date that vehicle  $j$  is observed in household  $i$  during the considered period and  $\bar{t}_{ij}$  be the date household  $i$  disposed of the car. As the panel is not cylindrical, it is necessary to define it.

The likelihood function of the observed sample may be written as

$$\sum_{i=1}^n \ln \ell(\boldsymbol{\lambda}, \gamma | \mathbf{d}_{i,\cdot}, \mathbf{m}_{i,\cdot}, \mathbf{y}_{i,\cdot}, \mathbf{p}_{i,\cdot}) \quad (8.17)$$

where

$$\begin{aligned} \ell(\boldsymbol{\lambda}, \gamma | \mathbf{d}_{i,\cdot}, \mathbf{m}_{i,\cdot}, \mathbf{y}_{i,\cdot}, \mathbf{p}_{i,\cdot}) = & \\ & \Pr(d_{i,\bar{t}_i} = 1 | y_{i,\bar{t}_i}, a_{i,\bar{t}_i}, k_{i,\bar{t}_i}, p_{i,\bar{t}_i}, \mathbf{x}_{ij0}; \alpha, \boldsymbol{\beta}) \\ & \prod_{t=\underline{t}_i}^{\bar{t}_i-1} \Pr(d_{ijt} = 0 | m_{ijt}, y_{ijt}, a_{ijt}, k_{ijt}, p_{ijt}, \mathbf{x}_{ij0}; \boldsymbol{\lambda}, \gamma) \\ & \prod_{t=\underline{t}_i}^{\bar{t}_i-1} \Pr(m_{ijt} | y_{ijt}, k_{ijt}, p_{ijt}, d_{ijt-1} = 0, \mathbf{x}_{ij0}; \boldsymbol{\theta}, \omega_m) \\ & \prod_{t=\underline{t}_i}^{\bar{t}_i} \Pr(p_{ijt} | p_{ijt-1}, d_{ijt-1} = 0, \mathbf{x}_{ij0}; \sigma_{\text{petrol}}, \sigma_{\text{diesel}}) \\ & \prod_{t=\underline{t}_i}^{\bar{t}_i} \Pr(y_{ijt} | y_{ijt-1}, d_{ijt-1} = 0, \mathbf{x}_{ij0}; \boldsymbol{\beta}, \tau_y) \end{aligned} \quad (8.18)$$

Under our assumptions one obtains as choice probabilities, assuming that households are utility maximizers,

$$\Pr(d_{ijt} = 1 | y_{ijt}, a_{ijt}, k_{ijt}, p_t, \mathbf{x}_{ij0}; \alpha, \boldsymbol{\beta}) = \frac{1}{1 + \exp(v_K(\mathbf{x}_{ijt}))} \quad (8.19)$$

and

$$\Pr(d_{ijt} = 0 | y_{ijt}, a_{ijt}, k_{ijt}, p_t, \mathbf{x}_{ij0}; \boldsymbol{\lambda}, \gamma) = \frac{\exp(v_K(\mathbf{x}_{ijt}) + \gamma \mathbb{E}_{\mathbf{z}_{ijt+1}}[V(\mathbf{z}_{ijt+1}) | \mathbf{z}_{ijt}])}{\exp(v_K(\mathbf{x}_{ijt}) + \gamma \mathbb{E}_{\mathbf{z}_{ijt+1}}[V(\mathbf{z}_{ijt+1}) | \mathbf{z}_{ijt}]) + \exp(v_D(\mathbf{x}_{ijt}))}, \quad (8.20)$$

In order to evaluate the log-likelihood function for particular values of  $\boldsymbol{\lambda}$  and  $\gamma$ , the dynamic programming problem needs to be solved exactly or its solution approximated in some way. Under the assumptions, one obtains

$$\mathbb{E}_{\mathbf{z}_{ijt}} [V(\mathbf{z}_{ijt}) | \mathbf{z}_{ijt-1}] = \sum_{q=1}^Q \sum_{r=1}^R \left\{ \begin{array}{l} \Pr(y_{ijt} | y_{ijt-1}, d_{ijt-1} = 0, \mathbf{x}_{ij0}; \boldsymbol{\beta}, \tau_y) \times \\ \Pr(p_{ijt} | p_{ijt-1}, d_{ijt-1} = 0, \mathbf{x}_{ij0}; \sigma_{\text{petrol}}, \sigma_{\text{diesel}}) \times \\ \ln(\exp(v_K(\mathbf{x}_{ijt}) + \gamma \mathbb{E}_{\mathbf{z}_{ijt+1}}[V(\mathbf{z}_{ijt+1}) | \mathbf{z}_{ijt}]) + \exp(v_D(\mathbf{x}_{ijt}))) \end{array} \right\} \quad (8.21)$$

The next step can be broken into two parts. The inner step consists of evaluating equation 8.21 for the current value of  $\boldsymbol{\lambda}, \gamma$ . The outer step consists of finding a new value of  $\boldsymbol{\lambda}, \gamma$  by iterating over the partial log-likelihood function that regards choice probabilities using a non-linear maximization algorithm. These two steps are repeated until convergence.

As the problem is a finite horizon, backward induction is used. Estimation starts from the last date of observation and the terminal condition  $\mathbb{E}_{\mathbf{z}_{ijt+1}} [V(\mathbf{z}_{ijt+1}) | \mathbf{z}_{ijt}]$  if  $d_t = 1$  (See Rust (1987), Rust (1996), Aguirregabiria and Mira (2010), for details about the procedure).

## 8.2 Data

### 8.2.1 Statistical sources

Data are drawn from the French ‘‘Parc Auto’’ panel survey. The survey is nationally representative, beginning in 1983. It is the best available database to study the dynamics of the demand for car equipment and use.

Annual rounds of a self-reported survey are designed by SOFRES, with IFSTTAR scientific council. Data are collected at the beginning of each year that describe the previous year.

The database is organized into a two-level hierarchy. The first level describes household information (i.e. socio-economic level, demographic structure, zone of residence, fleet size). At the second level, vehicles’ characteristics are detailed. Only three vehicles are described when the car fleet size is larger than

three. In the present approach, we restrict our analysis to a simple subset of variables.

The subsample of the population of motorized households that were observed over the period 2002 – 2008 is used for estimation. It is composed of 693 households, 437 of which own 637 petrol-powered or diesel-powered vehicles and scrap (or sell) it during the observation period.

Data are completed by drawing a time series of fuel prices in France over the time period in the “DIREM” database.

### 8.2.2 Descriptive statistics

We consider, first of all, sample means of observed decisions, reported in table 8.1 for each year.

As we observe use and disposal of cars on a fixed period, it is natural to have attrition in our sample. However, in this case, we consider that sample attrition occurs randomly such that the underlying causes of attrition are independent of the survey responses being studied. We know that it is rarely the case in panel data, but we argue that the reasons causing a household to exit the panel are not directly linked to the car’s disposal choice<sup>4</sup> that we study here. Therefore, sample attrition would not pose any challenges in estimation.

A distinction is made here between general vehicle characteristics (Age of car and Mileage) and those related only to the household’s use of the vehicle (Holding duration and Use).

The sample means of observed decisions are reported in table 8.2 for the whole period, distinguishing by fuel type.

Some important differences in observed decisions can be highlighted by separately analyzing petrol-powered and diesel-powered vehicles. On average,

---

4. We suggest that it is not completely true. When a single-vehicle household chooses to scrap its sole vehicle, it effectively chooses to end its motorization. Thus, the household can consider that its inclusion in the panel is no longer useful for the survey’s purpose and exit, thereby causing non-random attrition.

TABLE 8.1 – Observed decisions, sample averages

Label	Frequencies and means					
	2003	2004	2005	2006	2007	2008
Car disposal <sup>a</sup>	18,41	36,04	55,69	71,61	87,36	100
Age of car <sup>b</sup>	8,36	8,04	8,26	8,01	8,95	8,83
Holding duration <sup>c</sup>	6,03	5,98	6,39	5,98	6,44	6,70
Mileage <sup>d</sup>	101295	90764	106677	104747	101193	110439
Use <sup>e</sup>	74187	67580	83219	75171	74322	79729
Sample size	1012	1006	1014	983	991	962

DATA SOURCE : PARC AUTO

SOFTWARE : SAS

- 
- a.* Fraction of cars disposed during the period  
*b.* in years  
*c.* in years  
*d.* cumulative mileage of the car in kilometres  
*e.* cumulative mileage driven with the car by current household in kilometres

TABLE 8.2 – Whole period descriptive statistics, sample averages

Label	Petrol	Diesel
Age when bought <sup>a</sup>	2,42	1,81
Holding duration <sup>b</sup>	7,06	5,20
Cumulated mileage at date of disposal <sup>c</sup>	84782	115289
Cumulated use at date of disposal <sup>d</sup>	68866	84739
Average fuel consumption <sup>e</sup>	7,6	6,6

DATA SOURCE : PARC AUTO

SOFTWARE : SAS

- 
- a.* in years  
*b.* in years  
*c.* in kilometres  
*d.* in kilometres  
*e.* in litre per 100 kilometres

the former are bought at an older age, held longer, driven fewer kilometres and consume less fuel on average than the latter.

As mentioned previously, we use average annual fuel prices in metropolitan France. Fuel price is imposed to the household by market. He can't thus operate on its value but he can try to anticipate it.

Looking at the available values over the considered period (table 8.3), one observes a sharp rise both in petrol and diesel's prices. The latter rose at a higher rate than the former since 2000, remaining all the same always lower.

TABLE 8.3 – Average annual fuel prices per litre and by type

Fuel type	Year								
	2000	2001	2002	2003	2004	2005	2006	2007	2008
Gas	1,10	1,04	1,03	1,03	1,07	1,19	1,26	1,29	1,37
Diesel	0,85	0,80	0,77	0,79	0,88	1,03	1,08	1,09	1,27

DATA SOURCE : DIREM

SOFTWARE : SAS

As stated earlier, the income was collected in 13 classes. Though we observe only income classes, not actual levels, information is available on how these classes are defined.

Due to the inequality of the sampled households' distribution in these classes, we group the variable into 9 classes. Even though the size of the observed population of households diminishes from year to year, the relative proportions do not differ significantly (table 8.4).

TABLE 8.4 – Income class variable : frequencies

Class	Range in '000s of €	Year					
		2003	2004	2005	2006	2007	2008
1	< 15,2	55	81	58	50	49	49
2	15,2 – 19,1	73	80	70	74	56	62
3	19,1 – 22,9	109	105	96	94	77	71
4	22,9 – 26,7	92	93	90	87	100	91
5	26,7 – 30,5	67	70	80	80	89	78
6	30,5 – 38,1	71	45	66	70	55	54
7	38,1 – 45,7	93	85	89	98	112	104
8	45,7 – 61	63	65	62	56	54	73
9	> 61	69	66	76	76	92	102

DATA SOURCE : PARC AUTO

SOFTWARE : SAS

### 8.3 Results

We first estimate fuel price, performing the standard regression technique, based on the time series presented in table 8.3, to test for non-stationarity of fuel price dynamics. Though we have little information on fuel prices and a small samples, we do not reject the baseline assumption (table 8.5). We compute transition matrices for fuel prices varying from 0€ to 4€ at the cent level. To conserve space, we do not report the two  $401 \times 401$  matrices<sup>5</sup>.

TABLE 8.5 – Estimates : fuel prices

Label	Estimate
Variance $\sigma_{\text{petrol}}^2$	0,026
Variance $\sigma_{\text{diesel}}^2$	0,037

DATA SOURCE : DIREM

SOFTWARE : SAS

We acknowledge that choosing to use such a small sample (2000 \ 2008 observations) is not a negligible issue and that, due to this fact, the estimates used to compute the discrete transition probability distributions may be criticized. However, employing a longer annual time series for fuel prices would instead create additional problems as the evolution of fuel prices results from

5. The reader may contact the authors whenever he/she desires to obtain any of the estimated transition matrices regarding fuel prices.

various causes in recent history. Therefore, doing so may likely be irrelevant because of structural changes in formation of fuel prices over the past few decades that are not taken into account with such a simple assumption on their evolution.

Results concerning income are reported in table 8.6.

One can expect that the probability of belonging to a higher class of income at the observation date  $t$  would be greater the richer household  $i$  was at date  $t - 1$ .

The fact that the effect of the previous income class on the current one increases with the higher classes confirms our expectation.

TABLE 8.6 – Estimates : income

Label	Estimate	Std. Dev. <sup>a</sup>	T-stat. <sup>b</sup>
Intercept	21,9575	0,01012	259,65
Previous inc. <sup>c</sup> class is [15, 2; 19, 1[ K€	2,0909	0,01340	18,67
Previous inc. class is [19, 1; 22, 9[ K€	3,7148	0,01335	33,32
Previous inc. class is [22, 9; 26, 7[ K€	4,9304	0,01399	42,17
Previous inc. class is [26, 7; 30, 5[ K€	6,0385	0,01467	49,26
Previous inc. class is [30, 5; 38, 1[ K€	7,4954	0,01332	67,35
Previous inc. class is [38, 1; 45, 7[ K€	9,1370	0,01484	73,71
Previous inc. class is [45, 7; 61[ K€	10,8784	0,01622	80,25
Previous inc. is $\geq 61$ K€	13,1990	0,02522	62,64
$\sigma$	8,3549	0,000450	31,82
Log-Lik <sup>d</sup> at convergence	-5761		
Number of Observations <sup>e</sup>	4122		
Log-lik. H0 <sup>f</sup>	-18868		

DATA SOURCE : PARC AUTO

SOFTWARE : SAS

<sup>a</sup>. Std. Dev. : robust standard deviation

<sup>b</sup>. T-stat. : robust Student statistic

<sup>c</sup>. Inc. : income, computed in thousands of €

<sup>d</sup>. Log-lik. : value of the partial log-likelihood function

<sup>e</sup>. Used Data : each available observation for considered households (see section 8.2.1)

<sup>f</sup>. Trivial log-likelihood

As mentioned in section 8.1 (see equations 8.14 and 8.16), the probability for a household in income class  $j$  at date  $t - 1$  to be in class  $i$  for the following period  $t$  is computed as

$$P(i_t|j_{t-1}) = \frac{1}{1 + \exp(-\sigma\mu_i + \beta_0 + \beta_j)} - \frac{1}{1 + \exp(-\sigma\mu_{i-1} + \beta_0 + \beta_j)} \quad (8.22)$$

where  $\beta_1 = 0$  (i.e. reference class).

Looking at the  $9 \times 9$  estimated transition matrix (table 8.7), one observes that the greatest probability is for a household to remain in the same class from a year to the following, whatever the previous class. This probability ranges from 0,27 (corresponding to the central class 5) to 0,69 (observed for the two extremes classes).

The probability of moving to the lower or upper income classes is always inferior to 0,3 and remains considerable only when a household transitions through less than two income classes. Moreover, households characterized by an income lower than 38,1 K€ in the previous year have a higher probability of moving down, while one observes the opposite for the others.

TABLE 8.7 – Income : transition probability

Actual Income	Previous Income								
	1	2	3	4	5	6	7	8	9
1	<b>0,6854</b>	0,2121	0,0504	0,0155	0,0052	0,0012	0,0002	4,1E-5	4,0E-6
2	0,2509	<b>0,4326</b>	0,2131	0,0804	0,0287	0,0069	0,0013	0,0002	2,3E-5
3	0,0490	0,2473	<b>0,3562</b>	0,2298	0,1037	0,0277	0,0056	0,0010	0,0001
4	0,0106	0,0755	0,2349	<b>0,3096</b>	0,2276	0,0823	0,0182	0,0033	0,0003
5	0,0028	0,0215	0,0924	0,2058	<b>0,2710</b>	0,1713	0,0478	0,0091	0,0009
6	0,0011	0,0092	0,0444	0,1303	0,2820	<b>0,4338</b>	0,2629	0,0679	0,0073
7	0,0002	0,0013	0,0067	0,0222	0,0627	0,1995	<b>0,3621</b>	0,2070	0,0297
8	4,2E-5	0,0003	0,0017	0,0058	0,0174	0,0698	0,2645	<b>0,5306</b>	0,2695
9	4,2E-6	3,4E-5	0,0002	0,0006	0,0017	0,0074	0,0373	0,1809	<b>0,6922</b>

DATA SOURCE : PARC AUTO

SOFTWARE : SAS

We first estimate the model in a static framework, aiming to test our hypothesis on the aforementioned factors (table 8.8).

The effects of unemployed principal user are not significant. We can then test the hypothesis that  $\mu_2 = \mu_4$ .

Let the principal user typology now be defined now as follows :



- $t_{1j}$  : principal user less than 40 years old and employed
- $t_{2j}$  : principal user age  $\in [40; 60[$  and employed
- $t_{3j}$  : principal user less than 60 years old and unemployed

Thus, we estimate the model again.

The estimation results, reported in table 8.9, fail to reject the hypothesis of independence of principal user age for unemployed users.

TABLE 8.8 – Estimates : static model

Label	Estimate	Std. Dev. <sup>a</sup>	T-stat. <sup>b</sup>
Keep Utility Intercept ( $\beta_0$ )	0,2963	0,06820	4,35
VMT Intercept ( $\alpha$ )	0,6985	0,1630	4,29
Vehicle value ( $\beta_1$ )	0,02234	0,002393	9,33
Operating cost ( $\beta_2$ )	0,1893	0,03529	5,36
Principal user type 1 <sup>c</sup> ( $\mu_1$ )	0,4005	0,08359	4,79
Principal user type 2 ( $\mu_2$ )	-0,4491	0,2726	-1,65
Principal user type 3 ( $\mu_3$ )	0,3113	0,07605	4,09
Principal user type 4 ( $\mu_4$ )	-0,6457	0,3341	-1,93
$\sigma$	0,9167	0,01784	51,39
Log-Lik <sup>d</sup> at convergence	-2913,05		
AIC <sup>e</sup>	5844,1		
Number of Observations <sup>f</sup> : 1944			

DATA SOURCE : PARC AUTO  
SOFTWARE : SAS

- <sup>a</sup>. Std. Dev. : robust standard deviation
- <sup>b</sup>. T-stat. : robust Student statistic
- <sup>c</sup>. definition in 8.1.1
- <sup>d</sup>. Log-lik. : value of the partial log-likelihood function
- <sup>e</sup>. AIC : Log Akaike Information Criterion (small si better)
- <sup>f</sup>. Used Data : select panel's observation (see section 8.2.1)

We then estimate the latter in dynamic framework.

As previously argued, multiple hypotheses can be constructed about a household's behavior facing uncertainty. The discount rate  $\gamma$  represents this concept.  $\gamma = 0$  indicates that the household is myopic, e.g. it has an infinite preference for the present. When  $\gamma = 1$ , the household has no preference for the present : only the future is taken into account. When  $\gamma$  takes any value in between 0 and 1, the household trades off both the present and future.

TABLE 8.9 – Estimates : unemployed users' effect test

Label	Estimate	Std. Dev. <sup>a</sup>	T-stat. <sup>b</sup>
Keep Utility Intercept ( $\beta_0$ )	0,2981	0,06808	4,38
VMT Intercept ( $\alpha$ )	0,7065	0,1635	4,32
Vehicle value ( $\beta_1$ )	0,02227	0,002385	9,34
Operating cost ( $\beta_2$ )	0,1904	0,03517	5,41
Principal user type 1 ( $\mu_1$ )	0,3995	0,08357	4,78
Principal user type 2 ( $\mu_2$ )	0,3104	0,07604	4,08
Principal user type 3 ( $\mu_3$ )	-0,5416	0,2202	-2,46
$\sigma$	0,9164	0,01781	51,46
Log-Lik <sup>c</sup> at convergence	-2913,15		
AIC <sup>d</sup>	5842,3		
Number of Observations <sup>e</sup> : 1944			

DATA SOURCE : PARC AUTO  
SOFTWARE : SAS

*a.* Std. Dev. : robust standard deviation

*b.* T-stat. : robust Student statistic

*c.* Log-lik, : value of the partial log-likelihood function

*d.* AIC : Log Akaike Information Criterion (small is better)

*e.* Used Data : select panel's observation (see section 8.2.1)

We estimate the model using three different discount factors. Since it is not plausible that a household considers only the future, the model is estimated for  $\gamma = 0,99$  as a limiting case. For both present and future preferences, we define  $\gamma = 0,5$ , while the myopic behavior is tested with  $\gamma = 0$ . Results, reported in table 8.10, are in line with our expectation.

When we suppose agents to be myopic, we observe that a decrease in the vehicle scrap value has a double effect. For the same available income, the car's decline implies a decrease in the utility to keep it, but it also causes an increase in the demand for mileage. The same trend can be highlighted for a household income effect when the scrap value remains the same.

We previously defined, in equation 8.4, the scrap value of a car as a geometrically decreasing function of its ownership duration and cumulative use. That is, when the cumulative mileage increases, scrap value decreases. The demand for mileage of a household is then an increasing function of the past mileage, through the cumulative mileage variable.

Considering income and scrap value simultaneously, we can note that a greater vehicle value (i.e. a large positive difference between income and scrap value) implies a lower utility from owning it. Such a result can be interpreted as the desire to own a "better" vehicle if the available income is large enough.

As the latter is defined with regard to fleet size, this effect can be caused by a high household wealth level as well as a small fleet size. In other words, a household possessing a small number of cars is more tempted to scrap one for which the value is low.

When the vehicle is owned, a greater difference causes an increase in demand for mileage. This result indicates a tendency to use a vehicle more intensively when its scrap value is low relative to income, perhaps due to the high probability of scrapping it early.

According to results reported by [Goodwin \(1992\)](#), [de Jong and Gunn \(2001\)](#), [Graham and Glaister \(2002, 2004\)](#), the demand for mileage significantly decreases with average fuel expenditure. Moreover, the utility to keep a vehicle also decreases when operating costs increase.

Observing the principal user effect one can highlight that, *ceteris paribus*, an unemployed user possesses a decreased demand for mileage, while demand increases when the principal user is employed, whatever his or her age. This result can derive from the fact that the presence of a regular activity implies a greater quantity of trips than for unemployed users.

Moreover, looking at employed principal users, the results show that the utility of keeping the car, as well as demand for mileage, decreases conditional on age. Such an effect can be linked to an unobserved factor. We observe that, in fact, a household's head and spouse are more frequently declared as the principal users of a car fleet as opposed to adult sons (or daughters) and, furthermore, a younger head of household is more likely to include young children in the car use structure. This finding implies, for example, a higher travel need for these types of households. However, as we left household characteristics out of our approach, we cannot analyze this idea too deeply.

Estimation results for forward-looking agents show how the underlying behaviour that drives the choice to keep and use or to dispose of the car changes when considering a different discount rate  $\gamma$ .

Vehicle value is the only factor remaining almost the same when testing for different degrees of forward-looking behavior. That is, the effect of saved income is nearly stable whatever the discount factor. Results from unemployed principal users do not significantly impact owning decision and demand for mileage.

All of the other effects and their significance increase when accounting for a larger discount factor. Moreover, the gap between type 1 and 2 principal user effects decreases when the preference for the future increases.

There are potentially many underlying reasons behind these results, but we argue that deeper studies should be conducted and thus prefer to avoid speculation in discussing the results.



TABLE 8.10 – Estimates : probability to keep

	Estimate	Std. Dev. <sup>a</sup>	T-stat. <sup>b</sup>	Estimate	Std. Dev.	T-stat.	Estimate	Std. Dev.	T-stat.
Keep Utility Intercept ( $\beta_0$ )	0,8997	0,0879	10,24	0,9409	0,0692	13,59	0,7244	0,0690	10,49
VMT Intercept ( $\alpha$ )	0,5039	0,0787	6,40	0,7521	0,1147	6,55	0,8946	0,1326	6,74
Vehicle value ( $\beta_1$ )	0,0208	0,0024	8,56	0,0212	0,0027	7,97	0,0208	0,0028	7,34
Operating cost ( $\beta_2$ ) <sup>c</sup>	0,1287	0,0234	5,49	0,2198	0,0274	8,02	0,2581	0,0288	8,95
Principal user type 1 <sup>d</sup> ( $\mu_1$ )	0,6437	0,1014	6,35	0,7796	0,1163	6,70	0,8199	0,1231	6,66
Principal user type 2 <sup>e</sup> ( $\mu_2$ )	0,5111	0,1035	4,95	0,7045	0,1183	5,95	0,7578	0,1241	6,11
Principal user type 3 <sup>f</sup> ( $\mu_3$ )	-0,4497	0,2731	-1,65	-0,1099	0,2523	-0,43	0,1046	0,2230	0,47
$\sigma$	0,8252	0,0058	141,94	0,8386	0,0059	140,99	0,8459	0,0060	139,82
Discount Factor $\gamma$	0			0,5			0,99		
Log-lik. <sup>g</sup> at convergence	-2818,56			-2573,66			-2373,76		
Number of Observations <sup>h</sup> : 1944									

DATA SOURCE : PARC AUTO  
SOFTWARE : SAS

- a. Std. Dev. : standard deviation  
 b. T-stat. : robust Student statistic  
 c. REMEMBER :  $\beta_2$  enter negatively in 8.3 and 8.10  
 d. less than 40 years old and employed  
 e. age  $\in [40; 60[$  and employed  
 f. less than 60 years old and unemployed  
 g. Log-lik. : value of the partial log-likelihood function  
 h. Used Data : select panel's observation (see section 8.2.1)

## 8.4 Conclusions

The demand for mileage is function of saved income, computed by difference between household's income and vehicle's scrap value, operating cost and principal user's characteristics,

Saved income effect principally drive household's decisions on car's holding duration and use. Vehicle's scrap value is function of the cumulated mileage, defined as a deterministic function of mileage. When future is considered by the household in taking, at date  $t$ , the decision to own or not a car (i.e. the agent is forward-looking), that then depend on the mileage chosen for this period through the computation of the expected maximum utility obtained from the car at the next date whenever it is kept one year more.

Our estimation results confirm the hypothesis that accounting for forward-looking behaviour greatly improves the understanding of the modelled decision processes.

The approach, however, needs to be refined in several aspects.

As stated, the optimal use of a car depends on its principal user characteristics. This choice of the principal user for a household's car is an important topic in mileage demand estimation. By using GEV as an inter-temporal choice model, one can circumvent this problem.

Another limit of this model is that the optimal use for a vehicle  $j$  in a multi-vehicle household depends on other vehicles' use but these are not estimated simultaneously. [Mannering \(1983b\)](#) proposes a simultaneous equation model to bypass this issue. Defining available income for a vehicle  $j$  as the difference between household income and other car use expenditure, as in the previous model, one can calculate mileage utility as being dependent on other cars' optimal mileage, in accordance with [Mannering \(1983b\)](#). Moreover, it can be of a great importance to combine keep and use decisions with purchase choices.

More observed state variables, especially those related to household demographics and descriptions of the initial conditions, could be introduced in order to account for other possibles effects.

One can think that a more appropriate hypothesis for income transition probabilities estimation should be that the household's income class depend on his characteristics and that the latter follow a autoregressive process. Moreover, it's plausible to think that the household's income for a period  $t$  play a role on some event, as for example a new birth or the chief retirement. One have then the following equations system :

$$\begin{aligned} x_{i,t} &= \beta x_{i,t-1} + \gamma y_{i,t-1} \\ y_{i,t} &= \theta x_{i,t} \end{aligned} \tag{8.23}$$

where  $x_{i,t}$  is a vector of household  $i$  characteristics for period  $t$  and  $z_{i,t}$  his income class.

Because of simultaneity, the income class is to be firstly estimated and the fitted class enter in the household's characteristics estimation.

Another significant improvement would involve a more realistic specification of the unobserved state variables, i.e. the problem of persistent unobserved heterogeneity may create unobserved correlations over time as well as the likely effect of additional time-varying unobserved state variables (like the possibility that the car is in a dilapidated state).

A last, but not least, possible improvement to this model would involve estimating different car usage. Our "Parc Auto" data enables us, for example, to distinguish between long travel mileage, work-travel mileage, and so on. This information would potentially allow for a joint estimation of car choice and trip behaviour in the future.

Combining all of these aspects would provide a better description of the underlying behaviours that drive choices of a household as it regards its demand for cars.

# Chapitre 9

## Les Décisions d'Achat de Véhicules

L'importance que la voiture a dans notre société, la place qu'elle prend dans notre vie quotidienne, ainsi que l'impact environnemental qu'elle implique, font que l'étude des comportements automobiles des ménages est l'un des principaux sujets traités dans la littérature du consommateur.

Au cours des 40 dernières années, un grand nombre d'études ont analysé les comportements de possession et utilisation des véhicules : Berkovec (1985a), Berkovec and Rust (1985), de Jong (1990), de Jong (1996), Bunch and Golob (1996), Choo and Mokhtarian (2004), Bhat and Sen (2006), Fang (2008), Goldberg (1996), Golob and Brownstone (1997), Golob and Ren (1995), Hensher and Le Plastrier (1985), Hocherman et al. (1983), Kitamura (2009), Lave and Train (1979), Lave and Bradley (1980), Mannering (1983a), Mannering (1983b), Mannering and Winston (1985), Mohammadian and Miller (2003), Mueller and de Haan (2009), Train and Winston (2007), Turrentine and Kurani (2007), van Wissen and Golob (1992). Cependant, comme souligné par de Jong and Kitamura (2009), le sujet est loin d'avoir été étudié dans tous ses aspects.

*“[...] many questions remain unanswered. For example, why do some households purchase brand-new vehicles and hold them until their usable life has depleted, while others replace their vehicles every two years or so? When and how does a household decide to replace a car? Why do some households purchase brand-new vehicles, while others buy used vehicles that are equally expensive? How does a household decide whether to purchase a durable vehicle and hold it for a long period, or to purchase a few less-durable vehicles within the same period? In particular, how much of vehicle transaction behavior is pre-planned by the household and what is the planning horizon?”*



Une limite importante rencontrée dans la littérature vient du caractère dynamique de ces décisions.

En effet, la plupart des études considèrent le problème comme un problème statique, sans prendre en compte les dynamiques temporelles qui vraisemblablement impactent ce genre de décisions. Lorsque le problème est traité comme dynamique, celle-ci est supposée être tournée vers le passé. [Kitamura \(2009\)](#) suppose par exemple une dynamique auto-régressive pour le choix du nombre de véhicules composant le parc automobile.

Le choix de considérer ce problème comme statique amène les travaux existant à confondre sous le nom de choix de possession les choix de continuer à posséder un véhicule précédemment acheté et d'en acheter un nouveau (mis à part [Berkovec \(1985a\)](#)). Dans la littérature, lorsque la décision d'achat est étudiée, ce n'est pas le fait d'acheter qui est modélisé mais seulement le type d'achat.

Si je suis en accord avec l'idée que la satisfaction obtenue par la possession et l'utilisation d'un véhicule dépend effectivement des services que, de par ses caractéristiques, il peut fournir et donc avec l'idée d'une utilité unique, je crois par contre que le raisonnement sous-jacent aux travaux existant ne considère pas un facteur à mon sens important dans ce processus. En effet, dire qu'à chaque moment  $t$  l'agent considère l'ensemble des choix possibles (i.e. les types de véhicules disponibles sur le marché), en partie ou dans sa totalité, pour décider quel type de parc automobile il veut posséder revient à dire qu'il lui est égal de devoir l'acheter ou pas (i.e. il décide qu'il veut une petite essence, s'il l'a tant mieux, sinon il l'achète). Train soulève le problème et propose comme solution de restreindre l'ensemble de choix pour chaque agent à un ensemble de taille finie composé des types de véhicule dernièrement possédés par l'agent et une sélection randomisée (i.e. au hasard) des autres.

L'achat d'un véhicule comporte une rupture dans l'habitude. De plus, comme le souligne [Byun \(2001\)](#), il s'agit de l'un des plus importants investissements, ou souvent le plus important après les biens immobiliers, auquel les individus sont confrontés au cours de leur vie. On pourrait donc supposer que le coût lié à l'achat n'a pas seulement une valeur pécuniaire mais est caractérisé aussi par un "prix ressenti", représentant l'effort requis pour effectuer un achat qui dépend de l'agent et de ses besoins.

L'utilité escomptée par l'achat d'un véhicule est alors décomposable en une partie due aux services qu'il peut fournir, l'utilité de possession, et une due exclusivement à l'acte d'acheter <sup>1</sup>.

Une dernière observation est, comme dit plus haut, que les rares études considérant l'aspect dynamique du problème font l'hypothèse d'un processus tourné vers le passé.

Or, comme le suggèrent [de Jong and Kitamura \(2009\)](#), il semble important de s'interroger sur l'existence d'un processus de planification à long terme de ce type de choix. On pourrait alors supposer que les agents ont un comportement tourné vers le futur vis-à-vis des décisions d'achat automobile.

Dans ce travail je propose de considérer la question de l'achat comme un problème régénérateur de planification optimale.

L'agent anticipe les évolutions de ses ressources ainsi que de ses besoins de déplacement aux fins d'établir un temps d'attente optimal entre un achat et le suivant. Le problème est régénérateur puisque lorsque le ménage décide d'acheter le processus s'arrête et un nouveau commence, dans le but d'estimer la date d'achat optimale suivante. La décision n'est donc pas absorbante.

J'avoue que l'hypothèse est très simpliste. Je ne considère en effet ni l'état du parc automobile ni les caractéristiques du véhicule acheté. De plus, l'effet de l'environnement et ses évolutions, ici représenté par les prix nationaux du carburant, rentre indirectement dans le modèle, par le biais des transitions du kilométrage (i.e. la probabilité de changer de classe de kilométrage est ici estimée en fonction de ce prix). Cependant, l'idée que je présente par l'étude d'un tel modèle est que le processus de décision d'achat automobile est affecté par les dynamiques dictées par l'habitude. D'un côté l'habitude est plus simple à affronter que la nouveauté et en même temps pourrait causer un déclin de la satisfaction obtenue.

Le modèle est estimé à horizon infini, en utilisant l'algorithme *Nested Fixed Point* proposé par [Rust \(1987\)](#), qui sera présenté en conclusion de la section [9.1](#).

L'échantillon utilisé pour la résolution du modèle est tiré du panel Parc Auto, l'outil le plus complet permettant d'étudier les comportements d'équipement et utilisation automobile des ménages Français. Les agents étudiés sont les

---

1. Par souci de simplicité ainsi que pour les besoins de la méthode d'estimation choisie (expliqué par la suite) je fais le choix de n'étudier ici que cette deuxième.

ménages ayant participé à l'enquête au moins deux ans de suite entre 2000 et 2008.

Au cours de ce chapitre je vais avant tout présenter la spécification du modèle ainsi que la méthode d'estimation (Sec. 9.1) et les données utilisées (Sec. 9.2). Les résultats, pour plusieurs intensités de préférence pour le futur, sont reportés et commentés en section 9.3. Les conclusions et les évolutions possibles d'un tel modèle sont discutées à la fin (Sec. 9.4).

## 9.1 Le Modèle

Le problème étudié porte sur la décision d'achat automobile. Il est traité ici comme un problème dynamiques de choix discret, où le temps  $t = \{2002, \dots, T\}$  est discret, représenté en années, et son horizon  $T$  est infini. Le choix de considérer le problème à horizon infini permet de contourner le besoin d'observer la suite complète des décisions en évitant par conséquent de réduire de façon considérable la taille de l'échantillon.

Ce type de problème se base sur le Principe d'Optimalité formulé par Bellman (1957).

**Principle of Optimality :** *An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.*

(9.1)

Le cadre choisi pour la spécification du modèle est défini par les 5 hypothèses formulées par Rust (1987) : séparabilité additive des utilités, variables non-observables iid, indépendance conditionnelle des  $x$  futures, information non-observable indépendante sur les choix et distribuée selon une loi de Gumbel, support des  $x$  discret.

Bien que restrictif, cet ensemble d'hypothèses rend le modèle plus maniable.

L'utilité d'acheter peut être écrite comme une fonction temps-séparable actualisée

$$\sum_{t=1}^T \gamma^t u(\mathbf{z}_{i,t}, d(\mathbf{z}_{i,t})) \quad (9.2)$$

où  $\gamma$  est le facteur d'actualisation et  $u(\mathbf{z}_{i,t}, d(\mathbf{z}_{i,t}))$  est la fonction d'utilité pour la période  $t$ . Afin d'alléger les notations, je représenterai par la suite la fonction de décision  $d(\mathbf{z}_{i,t})$  par  $d_{it}$ .

En accord avec le principe d'optimalité de Bellman 9.1, le programme d'optimisation inter-temporel est défini comme un problème séquentiel dont la solution est la même pour chaque  $t$  en utilisant sa représentation de Bellman.

Chaque période  $t$  l'agent  $i$  décide si acheter ( $d_{it} = 1$ ) ou pas ( $d_{it} = 0$ ) un nouveau véhicule, en connaissant son état résumé par le vecteur  $\mathbf{z}_{i,t}$ . Le vecteur des variables d'état peut être décomposé en deux parties :

$$\mathbf{z}_{i,t} = \{\mathbf{x}_{i,t}, \boldsymbol{\varepsilon}_{i,t}\} \quad (9.3)$$

$\mathbf{x}_{i,t} = \{r_{i,t}, k_{i,t}, a_{i,t}\}$  résume l'information observable à la fois par le ménage  $i$  et par le chercheur. Les 3 variables observables définissant l'état de l'agent sont le revenu annuel réel du ménage  $r_{i,t}$ , sa demande de kilométrage total  $k_{i,t}$  et le nombre d'années écoulées depuis le dernier achat  $a_{i,t}$ . En accord avec les hypothèses imposées, l'information est discrétisée pour les 3 variables.

$\boldsymbol{\varepsilon}_{i,t} = \{\boldsymbol{\varepsilon}_{0,i,t}, \boldsymbol{\varepsilon}_{1,i,t}\}$  récolte la part de l'information non observable par le chercheur. Dans le cadre de travail choisi, ces erreurs sont supposées i.i.d. réparties selon la loi standard de Gumbel :

$$\exp(-\exp(-\varepsilon_{i,t})) \quad (9.4)$$

L'utilité d'acheter dépend de la satisfaction résiduelle du dernier achat, représentée par le temps écoulé, de l'état du ménage, représenté par son revenu annuel et sa demande totale de kilométrage en voiture, et des conséquences que ce choix aura dans le futur.

Le ménage est donc censé faire face aux évolutions de son état afin d'établir la date optimale à laquelle effectuer un nouvel achat, et on peut modéliser ses prévisions par une fonction de distribution de transition de Markov  $G_z(\mathbf{z}_{i,t+1} | \mathbf{z}_{i,t}, d_{i,t})$ .

La séquence des décisions est alors choisie de façon à maximiser l'utilité espérée :

$$\max_{\bar{t}_i} \left\{ \mathbb{E}_{\mathbf{z}_{i,\cdot}} \left( \sum_{t=1}^{T_i} \gamma^t u(\mathbf{z}_{i,t}, d_{i,t}) \mid \mathbf{z}_{i,0} \right) \right\} \quad (9.5)$$

où  $\mathbf{z}_{i,0}$  est l'ensemble des conditions initiales.

La fonction valeur du programme dynamique  $V(\mathbf{z}_{i,t})$  est définie par résolution récursive de sa représentation de Bellman

$$V(\mathbf{z}_{i,t}) = \max_{d \in (0,1)} \left\{ v_d(\mathbf{x}_{i,t}) + \epsilon_{d,i,t} + \gamma \int V(\mathbf{z}_{i,t+1}) G_z(\mathbf{z}_{i,t+1} \mid \mathbf{z}_{i,t}, d) \right\}. \quad (9.6)$$

La décision optimale inter-temporelle d'achat  $d^*(\mathbf{z}_{i,t})$  est la valeur de  $d$  qui maximise cette fonction pour la période  $t$ .

Je définis les utilités d'acheter ou pas par le système suivant :

$$\begin{cases} v_0(\mathbf{x}_{i,t}) = \sum_{q=2}^Q \beta_q \mathbb{I}(r_{i,t} = q) + \sum_{m=2}^M \beta_m \mathbb{I}(k_{i,t} = m) \\ \quad + [\gamma_d + \sum_{cm=2}^{cM} \gamma_{dm} \mathbb{I}(k_{i,t} \in cm)] \mathbb{I}(a_{0,i,t} \leq \psi) a_{0,i,t} \\ \quad + [\gamma_{d_\psi} + \sum_{cm=2}^{cM} \gamma_{d_\psi m} \mathbb{I}(k_{i,t} \in cm)] \mathbb{I}(a_{0,i,t} > \psi) \\ v_1(\mathbf{x}_{i,t}) = \alpha + [\gamma_d + \sum_{cm=2}^{cM} \gamma_{dm} \mathbb{I}(k_{i,t} \in cm)] a_{1,i,t} \end{cases} \quad (9.7)$$

où  $Q$  et  $M$  sont les nombres de classes de revenu et kilométrage, respectivement.

$cM$  est une généralisation de la partition définie par l'ensemble  $M$ . Les classes de kilométrage  $y$  sont regroupées en trois tranches : bas, moyen et haut.

$\psi$  est une durée d'attente seuil, telle que si l'agent a déjà attendu  $\psi$  années l'effet reste le même  $\forall a_{0,i,t} > \psi$ .

$\alpha = \alpha^1 - \alpha^0$  peut être considéré comme l'effet du changement dans l'habitude dû à l'achat d'un nouveau véhicule. Je m'attends à observer un effet négatif ( $\alpha < 0$ ).

Les paramètres  $\beta$  correspondent aux variables d'état indépendantes du choix.  $\beta_q = \beta_q^0 - \beta_q^1$  et  $\beta_m = \beta_m^0 - \beta_m^1$  sont les deux vecteurs des effets dûs respectivement au niveau des ressources disponibles pour l'agent et à ses besoins de déplacements en voiture sur la décision d'achat. On pourrait s'attendre à observer un effet sur l'utilité de ne pas acheter décroissant avec le

revenu ( $\beta_q < \beta_p \forall q > p$ ). Aucune hypothèse à priori n'est formulée sur l'effet de la demande de kilométrage.

Les paramètres  $\gamma$  représentent les effet du temps écoulé depuis le dernier achat qui varie selon que l'achat ait lieu ou pas. Effectivement, si l'agent achète,  $t$  devient la nouvelle dernière date d'achat et la durée d'attente est réinitialisée à 0. Si aucun achat n'a lieu, l'agent décide d'attendre un an de plus avant d'acheter et la variable est incrémentée d'une unité

$$a_{d,i,t} = \begin{cases} a_{i,t-1} + 1 & \text{if } d = 0 \\ 0 & \text{if } d = 1 \end{cases} \quad (9.8)$$

En considérant l'achat d'un véhicule comme choix de référence, le système des utilités peut être récrit en une seule fonction :

$$\begin{aligned} v_0^*(\mathbf{x}_{i,t}) = & -\alpha + \sum_{q=2}^Q \beta_q \mathbb{I}(r_{i,t} = q) + \sum_{m=2}^M \beta_m \mathbb{I}(k_{i,t} = m) \\ & + [\gamma_d + \sum_{cm=2}^{cM} \gamma_{dm} \mathbb{I}(k_{i,t} \in cm)] \mathbb{I}(a_{i,t} \leq \psi) a_{i,t} \\ & + [\gamma_{d_\psi} + \sum_{cm=2}^{cM} \gamma_{d_\psi m} \mathbb{I}(k_{i,t} \in cm)] \mathbb{I}(a_{i,t} > \psi) \end{aligned} \quad (9.9)$$

où  $a_{i,t}$  est le temps écoulé depuis le dernier achat, dans l'hypothèse de ne pas acheter en  $t$ .

Le processus de décision d'achat automobile est donc supposé ici être régénérateur.

Lorsque le choix de l'agent est d'effectuer un nouvel achat on obtient  $a_{i,t} = a_{i,0}$ . L'attente se réinitialise et un nouveau processus de décision commence.

L'effet du temps attendu entre un achat et le suivant est ici étudié par rapport à la demande de kilométrage. Je fais donc l'hypothèse que l'effet de l'habitude n'est pas le même selon le niveau d'utilisation automobile du ménage. Cette hypothèse est motivée par les deux idées suivantes.

D'un côté le caractère durable des véhicules implique qu'une utilisation plus intense cause un déclin plus rapide. Par conséquent un agent caractérisé par un besoin de déplacement en voiture élevé pourrait user son parc plus rapidement et être donc confronté à un "rythme de remplacement" différent d'un autre moins mobile.

De plus, comme dit plus haut, je suppose que l'effet de l'habitude ait un caractère dichotomique. L'effet de l'habitude pourrait alors avoir un effet inversé selon la "préférence pour la voiture" de l'agent (i.e. le niveau de plaisir

tiré, ou dépense, du fait de conduire). Les effets croisés de la durée d'attente et de la demande de kilométrage sont alors inclus dans la spécification, en supposant que le niveau d'utilisation automobile reflète en partie cette préférence.

L'état de l'environnement du ménage n'apparaît pas dans la spécification du modèle. Le prix du carburant, qui pourrait être considéré comme une information résumant cet état, est en effet une variable indispensable dans ce type de problème et son absence pourrait compromettre les résultats. Cependant en incluant une telle variable, la taille de la matrice des probabilités de transitions (25263), et par conséquent le coût d'estimation, seraient trop importants. Une autre idée possible est d'utiliser le revenu réel. Puisque ici le revenu n'est pas observé comme variable continue mais en classes à bornes connues, on devrait alors redéfinir les bornes chaque année comme le rapport de ces revenus nominaux par l'indice des prix du carburant annuel. Les classes glisseraient alors d'une année sur l'autre et les probabilités de transition du revenu dépendraient du temps. La fonction valeur escomptée étant définie en fonction de ces probabilités, ici aussi le coût d'estimation serait élevé.

La présence du kilométrage total en voiture prévu pour l'année pourrait pourtant pallier, au moins en partie, à ce problème. En effet, le processus sous-jacent à la demande de kilométrage du ménage dépend aussi bien de ses besoins que de sa "préférence pour la voiture". Or, les prix du carburant jouent vraisemblablement aussi un rôle important dans sa définition. La demande de kilométrage pourrait alors résumer les besoins du ménage et l'état de son environnement à la fois.

Avant d'introduire la méthode d'estimation il faut encore définir la distribution jointe de Markov  $G_z(\mathbf{z}_{i,t+1} \mid \mathbf{z}_{i,t}, d_{i,t})$ .

Une des conditions définissant le cadre de travail étant l'indépendance conditionnelle de  $x$  et  $\epsilon$ , on peut décomposer la distribution comme suit :

$$G_z(\mathbf{z}_{i,t+1} \mid \mathbf{z}_{i,t}, d) = G_x(\mathbf{x}_{i,t+1} \mid \mathbf{x}_{i,t}, d)G_\epsilon(\epsilon_{i,t} \mid \mathbf{x}_{i,t}) \quad (9.10)$$

où la densité  $G_\epsilon$  est supposée suivre une loi de Gumbel 9.4 et  $G_x$  est la distribution jointe de Markov des transitions de l'état du ménage  $(r_{i,t}, k_{i,t})$ , les transitions de la durée d'attente étant déterministes.

Il ne me semble pas approprié de supposer que les transitions du revenu

et de la demande de kilométrage soient indépendants, la définition de la distribution jointe  $G_x$  mérite donc d'être traitée avec attention.

Le revenu et la demande de kilométrage du ménage sont deux quantités continues représentables par le système d'équations linéaires suivant :

$$\begin{cases} r_{i,t}^* = f_r(c_{i,t}, \epsilon_i^r, \epsilon_t^r) \\ k_{i,t}^* = f_k(c_{i,t}, r_{i,t}, \epsilon_i^k, \epsilon_t^k) \end{cases} \quad (9.11)$$

où  $c_{i,t}$  est un vecteur de caractéristiques de l'agent.

Les  $\epsilon$  sont des termes d'erreur dépendant d'une part des préférences du ménage  $i$  et de l'autre du temps  $t$ . N'ayant pas d'informations suffisantes pour l'estimation des transitions démographiques, dans ce travail les caractéristiques du ménage sont traitées comme part de l'information non observable  $\epsilon_i$ .

Pour les propriétés de la probabilité de l'intersection de deux événements indépendants, la probabilité de transition d'un état  $(r_{i,t-1}, k_{i,t-1})$  à un autre  $(r_{i,t}, k_{i,t})$  est décomposable en :

$$\Pr(r_{i,t} \cap k_{i,t} | r_{i,t-1}, k_{i,t-1}) = \Pr(k_{i,t} | r_{i,t-1}, k_{i,t-1}, r_{i,t}) \Pr(r_{i,t} | r_{i,t-1}) \quad (9.12)$$

Les probabilités de transition  $P(k_{i,t} | k_{i,t-1}, r_{i,t-1}, r_{i,t})$  et  $P(r_{i,t} | r_{i,t-1})$ , du kilométrage et du revenu respectivement, sont estimées séparément par deux modèles logit ordonnés à bornes connues.



**Logit Ordonné à Bornes Connues**

Si l'on définit  $y^*$  la variable continue (non observée),  $y$  la variable groupée en  $J$  classes (observée) et  $\mu_{j-1}, \mu_j, j = 1, \dots, J$  les bornes de chaque classe  $j$ , on en déduit

$$y = j \Leftrightarrow \mu_{j-1} < y^* < \mu_j$$

La probabilité d'observer  $y^*$  appartenant à la classe  $j$  peut donc être écrite comme :

$$P(y = j | x) = \left[ F\left(\frac{\mu_j - y^*}{\sigma}\right) - F\left(\frac{\mu_{j-1} - y^*}{\sigma}\right) \right] \quad (9.13)$$

où  $F(x) = \frac{1}{1 + \exp(-x)}$  est la fonction logistique.

En récrivant les équations composant le système 9.11 en différence on obtient le système de fonction de régression linéaire 9.14.

$$\left\{ \begin{array}{l} r_{i,t}^* = f_r(r_{i,t-1}, \Delta_{t-1}^r \epsilon^r) \\ \quad = \delta_0^r + \sum_{q=2}^Q \delta_q \mathbb{I}(r_{i,t-1} = q) + \Delta \epsilon^r \\ k_{i,t}^* = f_k(\Delta_{t-1}^r, k_{i,t-1}, \Delta_{t-1}^k \epsilon^k) \\ \quad = \delta_0^k + \sum_{m=2}^M \delta_m \mathbb{I}(k_{i,t-1} = m) + \sum_{j=1}^3 \delta_j \mathbb{I}(\Delta r = j) + \Delta \epsilon^k \end{array} \right. , \Delta \epsilon^r, \Delta \epsilon^k \text{ i.i.d.} \quad (9.14)$$

où le changement de revenu est classifié de la façon suivante :

$$\Delta r = \begin{cases} 0 & \text{if } (r_{i,t} - r_{i,t-1}) \leq 0 \\ r_{i,t} - r_{i,t-1} & \text{if } (r_{i,t} - r_{i,t-1}) \in (1, 2) \\ 3 & \text{if } (r_{i,t} - r_{i,t-1}) \geq 3 \end{cases} \quad (9.15)$$

Comme dit plus haut, le prix du carburant est supposé jouer un rôle dans le processus de transition du kilométrage. En indiquant avec  $pc_t$  le prix moyen annuel observé en France au moment  $t$  on peut alors récrire la fonction du kilométrage subjacent comme

$$k_{i,t}^* = \delta_0^k + \sum_{m=2}^M \delta_m \mathbb{I}(k_{i,t-1} = m) + \sum_{j=1}^3 \delta_j \mathbb{I}(\Delta r = j) + \Delta pc + \Delta \epsilon^k \quad (9.16)$$

Cependant on ne peut pas utiliser la matrice  $(401 \times 401)$  de transition du prix de carburant, estimée comme résultat d'un processus dynamique non stationnaire d'ordre 2<sup>2</sup>, dans l'estimation du modèle dynamique.

Si l'on suppose indépendantes les évolutions du prix du carburant et du revenu<sup>3</sup> pour le théorème de la probabilité totale, on peut récrire la probabilité pour le ménage  $i$  de passer de  $k_{i,t-1}$  à la classe de kilométrage  $m$ , conditionnée à la transition de revenu, en 9.12 comme

$$\Pr(k_{i,t}|r_{i,t-1}, k_{i,t-1}, r_{i,t}) = \sum_{p=0}^4 \Pr(k_{i,t}|r_{i,t-1}, k_{i,t-1}, r_{i,t}, \Delta pc) \Pr(\Delta pc) \quad (9.17)$$

où

$$\Pr(k_{i,t} = m | k_{i,t-1}, r_{i,t-1}, r_{i,t}, \Delta pc; \boldsymbol{\beta}, \tau_k) = \frac{1}{1 + \exp\left(-\frac{\bar{k}_m - k_{i,t}^*}{\tau_k}\right)} - \frac{1}{1 + \exp\left(-\frac{\bar{k}_{m-1} - k_{i,t}^*}{\tau_k}\right)} \quad (9.18)$$

La probabilité pour le ménage  $i$  d'appartenir à la classe  $q$  de revenu l'année  $t$ , en sachant sa classe de revenu retardée  $r_{i,t-1}$ , s'écrit comme

$$\Pr(r_{i,t} = q | r_{i,t-1}; \boldsymbol{\beta}, \tau_r) = \frac{1}{1 + \exp\left(-\frac{\bar{r}_q - r_{i,t}^*}{\tau_r}\right)} - \frac{1}{1 + \exp\left(-\frac{\bar{r}_{q-1} - r_{i,t}^*}{\tau_r}\right)} \quad (9.19)$$

En supposant  $\tau_r$  et  $\tau_k$  indépendants, ces probabilités sont indépendantes et l'on peut calculer la probabilité de transition d'état par leur produit. Je reconnais que l'hypothèse d'indépendance entre les variances est peu réaliste. Cependant l'utilisation d'un outil plus adapté, le nested logit, impliquerait de considérer un ensemble comptant 63 choix possibles auquel est attaché un coût d'estimation trop important.

Je peux maintenant décrire la méthode d'estimation, l'algorithme *Nested Fixed Point (NFXP)*.

---

2. Voir les chapitres précédents pour l'estimation de cette probabilité.

3. Pour que cette hypothèse soit réaliste il faudrait utiliser le pouvoir d'achat du ménage plutôt que le revenu nominal comme bornes dans l'estimation des probabilités de transition de classe de revenu. Pourquoi cela n'est pas fait ici est expliqué plus haut.

Il s'agit de l'algorithme proposé par Rust (1988) pour l'estimation structurale maximum likelihood des paramètres de processus stochastiques visant à résoudre, à horizon infini, les problèmes de décision dynamiques de Markov à variables discrètes. L'algorithme est dit imbriqué car il se compose de deux sous-algorithmes : un algorithme du point fixe (**inner**) emboîté dans un algorithme d'optimisation non-linéaire standard du maximum de vraisemblance (**outer**).

Pour comprendre l'idée qui motive un tel algorithme reprenons l'équation 9.6. Le dernier terme peut être considéré comme représentant la fonction valeur escomptée pour l'agent  $i$  s'il prend la décision  $d$  au moment  $t$  :

$$\mathbb{E}V(\mathbf{z}_{i,t}, d) = \int V(\mathbf{z}_{i,t+1}) G_z(\mathbf{z}_{i,t+1} | \mathbf{z}_{i,t}, d) \quad (9.20)$$

Comme on l'a vu plus haut, l'hypothèse d'indépendance conditionnelle des  $x$  permet de récrire  $G_z(\mathbf{z}_{i,t+1} | \mathbf{z}_{i,t}, d)$  comme le produit 9.10. La participation de l'agent  $i$  à la fonction de vraisemblance de l'échantillon attachée aux paramètres  $\theta$

$$\sum_{i=1}^n \ln \ell(\theta, \gamma | \mathbf{d}_{i,\cdot}, \mathbf{r}_{i,\cdot}, \mathbf{k}_{i,\cdot}, \mathbf{a}_{i,\cdot}) \quad (9.21)$$

s'écrit alors comme

$$\ell(\theta) \equiv \ell(x_1, \dots, x_T, d_1, \dots, d_T) = \prod_{t=2}^T P(d_t | x_t, \theta) p(x_t | x_{t-1}, d_{t-1}, \theta) \quad (9.22)$$

où la probabilité conditionnelle de choix est définie par la fonction logit suivante :

$$P(d_t | x_t, \theta) = \begin{cases} \frac{\exp\{v_0^*(x_t) + \gamma[\mathbb{E}V(x_t, 0) - \mathbb{E}V(x_t, 1)]\}}{1 + \exp\{v_0^*(x_t) + \gamma[\mathbb{E}V(x_t, 0) - \mathbb{E}V(x_t, 1)]\}} & \text{if } d_t = 0 \\ \frac{1}{1 + \exp\{v_0^*(x_t) + \gamma[\mathbb{E}V(x_t, 0) - \mathbb{E}V(x_t, 1)]\}} & \text{if } d_t = 1 \end{cases} \quad (9.23)$$

où  $v_0^*$  est l'utilité unique définie en 9.9. Rust montre que l'on obtient des estimations consistantes de  $\theta$  même en ne maximisant que la fonction de vraisemblance partielle :

$$\ell_p(\theta) \equiv \prod_{t=1}^T P(d_t | x_t, \theta) \quad (9.24)$$

Pour pouvoir estimer le modèle il faut donc calculer, ou néanmoins approximer, la valeur de  $\mathbb{E}V$ .

Puisque l'on travaille ici à horizon infini cette valeur ne peut qu'être approximée, d'où la raison d'un algorithme en deux étapes emboîtées.

Rust propose de définir un opérateur non-linéaire  $T_\theta : B \rightarrow B$  ( $B$  espace de Banach des fonctions mesurables sur  $[0, \infty)$ )

$$T_\theta(W)(x, d) = \int \log [\exp \{\gamma W(x, 1)\} + \exp \{v_0^*(x_t) + \gamma W(x, 0)\}] p(dx, d, \theta) \quad (9.25)$$

Il peut être montré que  $\|T_\theta(W) - T_\theta(V)\| \leq \beta \|W - V\|$ ,  $\beta \in [0, 1)$ . Le théorème du point fixe garantit alors l'existence d'une solution unique  $\mathbb{E}V_\theta = T_\theta(\mathbb{E}V_\theta)$

La solution du problème du point fixe peut être obtenue par le "polyalgorithme" **INNER**.

En suivant la terminologie utilisé par Rust (1994), je l'appelle polyalgorithme puisqu'il s'agit de la combinaison de deux méthodes itératives.

**L'algorithme itératif de contraction** approxime dans un premier temps  $\mathbb{E}V_\theta$ . La fonction valeur escomptée est obtenue par itérations successives  $\mathbb{E}V_k = T_\theta(\mathbb{E}V_{k-1})$  à partir d'une valeur initiale  $\mathbb{E}V_0 = 0$ .

Cet algorithme, bien que de très simple utilisation, lorsque  $\beta$  approche l'unité la méthode converge très lentement et le nombre d'itérations nécessaires à atteindre une erreur acceptable ( $\simeq 10^{-6}$ , devient trop important. Cet algorithme est donc efficace seulement pour les 20 – 30 premières itération mais ne suffit pas à résoudre le problème.

Comme proposé par Rust, on peut alors compléter l'approximation en utilisant l'**algorithme Newton-Kantorovich**. Cet algorithme présente en effet le problème inverse : il converge très lentement pour des valeurs loin de la solution mais à vitesse quadratique une fois dans le *domaine d'attraction* du point fixe.

À partir de l'approximation obtenue par le premier algorithme, les itérations successives sont cette fois définies comme

$$\mathbb{E}V_k = \mathbb{E}V_{k-1} - [1 - T'_\theta]^{-1} (1 - T_\theta) (\mathbb{E}V_{k-1}) \quad (9.26)$$

Pour conclure, l'objectif de l'algorithme d'optimisation **OUTER** est de trouver les paramètres  $\theta$  qui maximisent 9.22 ou 9.24, où  $\mathbb{E}V$  est l'unique point fixe de 9.25. La vraisemblance est ici maximisé par l'**algorithme à régions de confiance**

## 9.2 Les Données

Les données ici utilisées sont issues de l'enquête "Parc Auto" ([TNS-SOFRES](#)), le meilleur outil disponible pour l'étude de la motorisation des ménages Français et de leur usage de l'automobile.

TABLE 9.1 – Mouvements Annuels des Cohortes

	2003	2004	2005	2006	2007	2008	Total
2002	1822	1300	988	736	620	564	6030
2003	0	365	242	154	122	112	995
2004	0	0	376	198	138	115	827
2005	0	0	0	307	215	169	691
2006	0	0	0	0	448	309	757
2007	0	0	0	0	0	794	794
Total	1822	1665	1606	1395	1543	2063	10094

### Achats

	2003	2004	2005	2006	2007	2008	Total
2002	264	196	135	83	65	56	799
2003	0	65	43	17	14	13	152
2004	0	0	44	28	23	16	111
2005	0	0	0	44	34	24	102
2006	0	0	0	0	70	43	113
2007	0	0	0	0	0	112	112
Total	264	261	222	172	206	264	1389

TABLE 9.2 – Distributions Annuelles

Classe de Revenu							<i>Durée d'Observation</i>					
	2003	2004	2005	2006	2007	2008	1	2	3	4	5	6
< 15,2	19,21	19,82	17,06	18,21	14,97	14,25	17,66	19,13	16,15	17,47	13,11	13,12
15,2 – 19,1	15,42	13,51	12,89	14,19	12,64	11,10	13,33	13,07	13,60	13,98	12,43	11,88
19,1 – 22,9	13,34	12,97	15,19	13,62	13,93	14,11	13,01	13,78	14,91	13,67	15,85	15,43
22,9 – 26,7	11,42	13,81	13,01	13,55	13,67	12,75	12,09	13,21	14,08	13,46	14,07	13,48
26,7 – 30,5	10,15	8,47	8,97	9,75	8,88	9,79	9,75	8,66	8,49	10,17	9,70	9,75
30,5 – 38,1	14,76	15,14	16,00	14,27	14,97	16,04	15,73	14,97	14,98	14,59	15,71	14,01
38,1 – 45,7	7,14	7,69	8,34	8,03	9,20	8,97	7,86	8,13	9,25	8,32	8,33	8,51
45,7 – 61	5,93	5,95	5,60	5,66	7,71	8,05	6,98	6,49	5,24	5,24	6,97	8,69
> 61	2,63	2,64	2,93	2,72	4,02	4,94	3,60	2,56	3,31	3,08	3,83	5,14
Total	1822	1665	1606	1395	1543	2063	4112	2264	1449	973	732	564

Classe de Kilométrage							<i>Durée d'Observation</i>					
	2003	2004	2005	2006	2007	2008	1	2	3	4	5	6
< 2500	7,35	8,95	7,97	10,68	9,14	9,65	7,15	8,61	10,35	11,51	12,30	10,46
[2500, 5000[	7,35	9,01	8,66	9,39	9,07	8,48	6,69	8,66	9,94	11,20	11,07	11,35
[5000, 10000[	22,50	21,68	23,16	21,72	21,84	21,91	19,82	22,70	23,60	23,64	23,91	28,19
[10000, 15000[	20,42	19,82	20,86	21,72	20,74	21,86	20,48	19,79	20,84	22,20	22,40	24,65
[15000, 20000[	15,70	15,62	13,89	14,05	14,00	12,85	15,37	15,37	13,46	12,33	12,70	10,28
[20000, 35000[	20,20	19,10	19,24	16,92	17,56	18,32	21,94	17,98	17,05	15,62	14,21	12,06
>= 35000	6,48	5,83	6,23	5,52	7,65	6,93	8,56	6,89	4,76	3,49	3,42	3,01
Total	1822	1665	1606	1395	1543	2063	4112	2264	1449	973	732	564

Chaque vague du panel décrit en effet le parc automobile et son utilisation pour un échantillon représentatif de la population française comptant 10000 ménages. Le taux de réponse annuel moyen varie entre 65% et 70% au fil du temps. L'échantillon est renouvelé au taux d'environ 1/3 par an afin de gérer les problèmes d'attrition.

Afin de construire un échantillon exploitable pour ce travail, à partir des 6133 ménages observés au moins deux fois entre 2002 et 2008 dont le parc automobile ne compte jamais plus de 3 véhicules et les informations sont disponibles<sup>5</sup>, deux conditions sont imposées.

En premier lieu, puisque la décision étudiée porte sur la durée d'attente entre deux achats, le dernier achat du ménage doit être observable. Or, lorsqu'un ménage non-équipé est observé dans la base, seulement la date du dernier abandon est collecté. Les ménages non-équipés au moment de la première observation ne sont inclus que s'ils achètent au cours de la période d'observation et s'ils répondent encore au moins un an après.

Pour conclure, les ménages ne sont pas sélectionnés s'ils ont effectué plus d'un achat la même année au moins une fois au cours de la période d'observation. Bien que ces achats soient identifiables<sup>6</sup>, cela concerne 3,6% des observations. Je ne dispose donc pas d'une quantité d'information suffisante pour étudier ce cas particulier.

L'échantillon sélectionné compte alors 4226 ménages. Il s'agit d'un échantillon non cylindré dont les cohortes sont composées par les ménages entrant dans le panel la même année. L'achat n'est pas une condition de sortie pour le ménage, les mouvements annuels (i.e. les entrées et sorties du panel) sont dictés seulement par la participation à l'enquête. En table 9.1 sont reportées les entrées et sorties des ménages, ainsi que le nombre annuel d'achats par cohorte.

Dans ce travail 3 variables sont utilisées :

- **Revenu Annuel** : Collecté comme variable discrète en 13 classes, regroupées ici sur 9 classes.

---

5. Voir sec. 4.2 pour plus d'informations

6. Voir la procédure décrite au cours du chapitre ref.

- **Demande de Kilométrage Annuel** : Calculé comme somme des kilométrages déclarés par un même ménage pour les véhicules utilisés au cours de l'année. Il s'agit donc d'une variable continue ici discrétisée en 7 classes.
- **Durée d'Attente** : Nombre d'années écoulées entre deux achats automobiles consécutifs.

Le panel n'étant pas cylindré, les distributions par classe de revenu et demande de kilométrage sont reportées en table 9.2 aussi bien annuellement que par année d'observation.

Pour ce qui concerne la classe de revenu, mis à part une légère sous-représentation des revenus  $< 15,2$  k€ (i.e. moins de 1300€ par mois) on n'observe rien de particulier. Les distributions annuelles, ainsi que leurs évolutions au fil du temps, sont proches de celles observées sur la population représentative. Si l'on regarde par année d'observation des ménages on remarque aussi une certaine stabilité des distributions au cours de l'enquête, malgré une inflation certes modérée.

La demande de kilométrage annuel des ménages semble aussi en ligne avec celle observée dans l'échantillon Parc Auto. Une précision est tout de même nécessaire. Il a été dit plus haut que les ménages sont inclus dans l'échantillon sélectionné seulement à partir d'une date qui permet de définir le temps écoulé depuis le dernier achat. Les ménages non-équipés en début d'observation ne sont pas utilisés. Cependant, si un ménage retenu se démotorise au cours de l'observation cela ne cause pas sa sortie de l'échantillon (Les répartitions annuelles par nombre de véhicules composant le parc automobile du ménage sont reportées en table 9.3). Les ménages caractérisés par une demande nulle de kilométrage sont par conséquent très faiblement représentés. De plus je considère ici le kilométrage comme une approximation des besoins de déplacement et de la préférence pour la voiture, que le ménage évalue pour décider de sa motorisation. En supposant que si le ménage décide de ne pas se motoriser signifie que la résultante de ces deux facteurs est faible, les demandes nulles de kilométrage sont incorporées à la classe inférieure ( $< 2500$  km).



TABLE 9.3 – Répartition par Type de Motorisation

	2003	2004	2005	2006	2007	2008	Total
0	8	19	19	20	17	27	110
1	1348	1215	1153	1017	1099	1406	7238
2	449	413	408	339	409	608	2626
3	17	18	26	19	18	22	120

## Achat

	2003	2004	2005	2006	2007	2008	Total
0	0	0	2	0	2	0	4
1	178	172	145	106	122	160	883
2	82	86	69	62	78	100	477
3	4	3	6	4	4	4	25
Total	264	261	222	172	206	264	1389

La dernière information ici considérée concerne le nombre d'années écoulées depuis le dernier achat. Il s'agit d'une variable qui dépend du choix. En effet, si le ménage décide de ne pas acheter, sa durée d'attente est incrémentée d'une unité, alors que s'il achète, celle-ci est réinitialisée à 0. En table 9.4 sont reportées les répartitions et les distributions par durée d'attente au moment de l'entrée du ménage dans le panel ainsi que lorsqu'un achat a lieu.

On observe en entrée un pourcentage faible de “longue attente” (plus de 15 ans). Ce pourcentage se fait encore plus faible lorsque l'on observe le nombre d'années effectivement attendues entre un achat et le suivant. On peut alors s'interroger sur le fait que l'effet de la durée d'attente ne soit pas le même si le ménage a déjà attendu longtemps depuis le dernier achat ou pas. A cette fin je construis deux variables, une première continue qui contient la durée d'attente obtenue sans acheter si celle-ci est inférieure à 15 ans et une indicatrice qui vaut 1 si et seulement si cette attente dépasse les 15 ans.

Pour conclure, on peut s'interroger sur l'existence d'un effet croisé de la durée d'attente et de l'état du ménage. En effet, comme montré en figure 9.1, le temps attendu entre deux achats semble décroître lorsque le revenu et, de façon plus marquée, le kilométrage. Le croisement de la classe de kilométrage, répartie en Basse (Classes 1,2 et 3), Moyenne (Classes 4 et 5)

TABLE 9.4 – Durée d’Attente Entre 2 Achats Automobiles

	ENTREE		ACHAT	
	Freq.	%	Freq.	%
0	824	20,04	195	14,04
1	678	16,49	227	16,34
2	562	13,67	174	12,53
3	435	10,58	152	10,94
4	386	9,39	134	9,65
5	258	6,27	107	7,70
6	224	5,45	88	6,34
7	188	4,57	75	5,40
8	142	3,45	51	3,67
9	96	2,33	48	3,46
10	81	1,97	34	2,45
11	62	1,51	27	1,94
12	48	1,17	26	1,87
13	35	0,85	19	1,37
14	27	0,66	15	1,08
15	19	0,46	6	0,43
16	22	0,54	7	0,50
17	6	0,15	2	0,14
18	6	0,15	2	0,14
19	5	0,12		
20	1	0,02		
21	3	0,07		
22	1	0,02		
23	1	0,02		
28	1	0,02		
33	1	0,02		

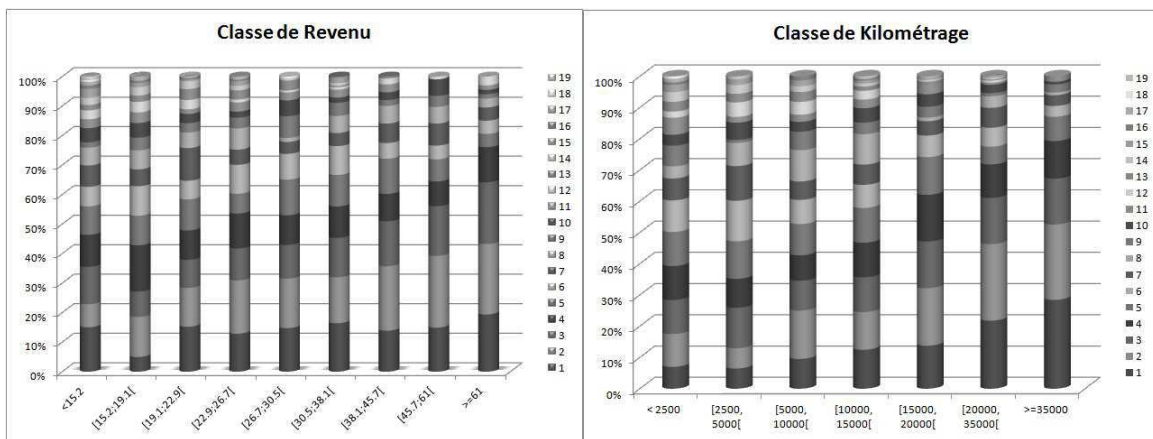


FIGURE 9.1 – Distributions par Durée d’Attente

et Haute (Classes 6 et 7), et de la durée d'attente est alors incluse au modèle.

## 9.3 Les Résultats

Dans cette section je vais présenter les résultats du travail.

Comme expliqué en introduction, l'agent étudié par ce modèle est supposé avoir un comportement tourné vers le futur lorsqu'il se confronte aux décisions d'équipement automobile. Il est donc en mesure de prévoir les évolutions de son état afin d'évaluer les conséquences de ses décisions et il les optimise sur cette base. Pour pouvoir estimer le modèle d'achat il faut donc connaître les probabilités de transition, c'est à dire les probabilités pour un ménage de passer d'un état en  $t$  à un autre en  $t + 1$ .

Dans un premier temps les estimations des probabilités de transition d'état sont présentées (sec. 9.3.1). Les estimations du modèle sont commentées ensuite (sec. 9.3.2).

### 9.3.1 Les transitions

L'état d'un ménage est défini simultanément par sa classe de revenu et sa demande de kilométrage en voiture pour l'année. Puisque la classe de revenu joue vraisemblablement un rôle sur la demande automobile du ménage, on ne peut pas supposer la probabilité de transition totale être le produit des résultant de deux processus auto-régressifs indépendants.

Pour pallier à cela, j'estime en premier lieu la classe de revenu comme issue d'un processus auto-régressif. La probabilité du kilométrage est ensuite estimée en fonction de l'information retardée et de la variation du revenu observée. La probabilité jointe est alors obtenue par produit de ces deux probabilités indépendantes (eq. 9.12).

Pour ces estimations le plus large échantillon disponible (i.e. tout ménage répondant à l'enquête au moins deux années de suite au cours la période considérée) a été utilisé. La seule condition imposée sur l'échantillon non

TABLE 9.5 – Répartition et Distribution par Classe de Revenu selon le Revenu Retardé

Revenu	Revenu Retardé									Total
	0 15, 2	15, 2 19, 1	19, 1 22, 9	22, 9 26, 7	26, 7 30, 5	30, 5 38, 1	38, 1 45, 7	45, 7 61	61	
< 15, 2	4380 (73,84)	662 (20,92)	253 (8,79)	107 (4,07)	47 (2,25)	62 (2,01)	18 (1,01)	17 (1,13)	28 (3,55)	5574
15, 2 – 19, 1	855 (14,41)	1404 (44,37)	460 (15,98)	197 (7,49)	64 (3,07)	50 (1,62)	5 (0,28)	5 (0,33)	10 (1,27)	3050
19, 1 – 22, 9	311 (5,24)	677 (21,40)	1188 (41,26)	443 (16,85)	153 (7,33)	102 (3,30)	20 (1,13)	12 (0,80)	5 (0,63)	2911
22, 9 – 26, 7	148 (2,49)	216 (6,83)	603 (20,94)	1001 (38,08)	379 (18,16)	223 (7,21)	54 (3,04)	17 (1,13)	5 (0,63)	2646
26, 7 – 30, 5	93 (1,57)	98 (3,10)	203 (7,05)	454 (17,27)	694 (33,25)	379 (12,26)	80 (4,51)	27 (1,80)	10 (1,27)	2038
30, 5 – 38, 1	76 (1,28)	73 (2,31)	120 (4,17)	319 (12,13)	596 (28,56)	1600 (51,75)	361 (20,34)	113 (7,52)	13 (1,65)	3271
38, 1 – 45, 7	25 (0,42)	14 (0,44)	28 (0,97)	68 (2,59)	116 (5,56)	531 (17,17)	822 (46,31)	242 (16,11)	25 (3,17)	1871
45, 7 – 61	20 (0,34)	13 (0,41)	14 (0,49)	26 (0,99)	32 (1,53)	124 (4,01)	369 (20,79)	848 (56,46)	128 (16,22)	1574
> 61	24 (0,40)	7 (0,22)	10 (0,35)	14 (0,53)	6 (0,29)	21 (0,68)	46 (2,59)	221 (14,71)	565 (71,61)	914
Total	5932	3164	2879	2629	2087	3092	1775	1502	789	23849

cylindré 2002 – 2008 Parc Auto pour la construction de l'échantillon pour l'estimation du revenu est que la classe de revenu, actuelle et retardée, est déclarée. Pour l'estimation du kilométrage, la condition que les kilométrages sont déclarés de façon cohérente chaque année pour tous les véhicules que le ménage possède au cours de l'observation. Les répartitions, annuelles pour le revenu et par classe pour le kilométrage, sont reportées respectivement en table 9.5 et 9.6.

Dans les deux cas, une seule observation par ménage est retenue afin d'éviter d'inclure des effets de panel.

TABLE 9.6 – Classes de Kilométrage

Classe	N	Moyenne	Ecart-type	Minimum	Maximum
< 2500	4142	289,51	640,89	0	2490
[2500, 5000[	1132	3748,67	701,27	2500	4998
[5000, 10000[	2863	7386,41	1476,29	5000	9995
[10000, 15000[	2789	12177,68	1488,37	10000	14998
[15000, 20000[	1987	17166,35	1469,71	15000	19999
[20000, 35000[	2864	25524,28	4115,89	20000	34990
≥ 35000	1228	56335,25	36170,88	35000	333443

Pour ce qui concerne le **REVENU** aucune remarque particulière n'est à faire aux résultats de l'estimation, reportés en table 9.7. Les probabilités de transition ainsi obtenues sont présentées en table 9.8.

La probabilité de recevoir un revenu du même ordre de grandeur que celui perçu l'année précédente est la plus importante, quelque soit la classe de revenu retardée.

La probabilité de connaître un changement du volume de ressources reste significative seulement si cette variation ne compte plus de 2 et est systématiquement inférieure à 5% pour des variations plus importantes.

Pour conclure, la probabilité de connaître une baisse de revenu est toujours plus importante que celle de monter en classe, sauf lorsque le revenu retardé se place dans une classe centrale ( $26, 7 \leq r < 45, 7$ )<sup>7</sup>.

---

7. Ces résultats pourraient sembler paradoxaux. Cependant cela s'explique par le fait qu'il n'est pas utilisé ici le revenu réel, corrigé par rapport aux prix.

TABLE 9.7 – Fonction de Transition du Revenu Estimée

Factor	Est.	Std.Dev.	T-Stat
Intercep	2,587	0,0043	599,23
Previous inc. $\in [15, 2; 19, 1[$ k€	0,294	0,0062	47,42
Previous inc. $\in [19, 1; 22, 9[$ k€	0,474	0,0063	75,74
Previous inc. $\in [22, 9; 26, 7[$ k€	0,632	0,0064	98,78
Previous inc. $\in [26, 7; 30, 5[$ k€	0,773	0,0068	113,78
Previous inc. $\in [30, 5; 38, 1[$ k€	0,922	0,0061	150,31
Previous inc. $\in [38, 1; 45, 7[$ k€	1,121	0,0072	156,15
Previous inc. $\in [45, 7; 61[$ k€	1,327	0,0079	168,29
Previous inc. $\geq 61$ k€	1,637	0,0122	134,29
Variance $\tau_r^2$	0,020	0,0003	73,66

TABLE 9.8 – Probabilités de Transition du Revenu Estimées

	< 15, 2	-19, 1	-22, 9	-26, 7	-30, 5	-38, 1	-45, 7	-61	$\geq 61$
< 15, 2 k€	<b>72,17</b>	24,49	8,31	2,89	1,08	0,38	0,09	0,02	2,4E-03
[15, 2; 19, 1[ k€	20,71	<b>37,51</b>	23,01	10,12	4,14	1,50	0,37	0,09	0,01
[19, 1; 22, 9[ k€	5,04	23,49	<b>30,88</b>	22,03	11,37	4,59	1,20	0,28	0,03
[22, 9; 26, 7[ k€	1,37	9,10	20,77	<b>26,47</b>	20,48	10,55	3,12	0,76	0,09
[26, 7; 30, 5[ k€	0,43	3,24	9,61	18,87	<b>23,09</b>	17,44	6,63	1,75	0,20
[30, 5; 38, 1[ k€	0,22	1,73	5,78	14,81	27,77	<b>37,26</b>	26,93	9,73	1,25
[38, 1; 45, 7[ k€	0,04	0,33	1,18	3,44	8,42	18,45	<b>30,90</b>	21,73	3,92
[45, 7; 61[ k€	0,01	0,11	0,40	1,20	3,16	8,43	25,31	<b>45,78</b>	25,50
$\geq 61$ k€	2,1E-03	0,02	0,06	0,18	0,49	1,39	5,45	19,85	<b>68,99</b>

On a vu dans la section 9.1 que le kilométrage n'est pas estimé directement en fonction du revenu mais du  $\Delta Revenu$ . Pour calculer les probabilités du kilométrage total il nous faut alors connaître les probabilités du type de transition conditionnées à la classe retardée ( $\Pr(\Delta r|r_{i,t-1})$ ).

Pour chaque classe de revenu on obtient la probabilité de ne pas monter en classe comme union des probabilités conditionnées que le ménage ait un revenu inférieur à la borne supérieure de cette classe de revenu retardée et de la probabilité de monter plus de 2 classes comme union des probabilités conditionnées que le ménage ait un revenu supérieur à la borne inférieure de la troisième classe successive à la classe retardée. Puisque l'appartenance à une classe  $j$  est un événement incompatible avec l'appartenance à une autre classe  $i$  ( $\forall i \neq j$ ), la probabilité de l'union est la somme des probabilités et l'on peut écrire :

$$\Pr(\Delta r|r_{t-1}) = \begin{cases} \sum_{j=1}^{r_{t-1}} \Pr(r_t = j|r_{t-1}) & \text{si } \Delta r = 0 \\ \Pr(r_t = r_{t-1} + j|r_{t-1}) & \text{si } \Delta r = j, \quad r_{t-1} \leq 9 - j \\ 0, & j \in (1, 2) \quad r_{t-1} > 9 - j \\ \sum_{j=r_{t-1}+3}^9 \Pr(r_t = j|r_{t-1}) & \text{si } \Delta r \geq 3, \quad r_{t-1} \leq 6 \\ 0, & r_{t-1} > 6 \end{cases} \quad (9.27)$$

Les évolutions des probabilités du  $\Delta Revenu$  conditionné à la classe retardée ( $\Pr(\Delta r|r_{i,t-1})$ ) sont représentées en figure 9.2.

En analysant le graphique on confirme que quelque soit le revenu initial, la probabilité de ne pas monter en classe est la plus importante, ainsi que toutes les observations précédentes.

Le **KILOMETRAGE** est estimé comme fonction de l'information retardée et des transitions du revenu et du prix du carburant. La table 9.9 reporte ces estimations.

Les probabilités de transitions sont reportées en figure 9.3. Comme expliqué à la section 9.1 les probabilités totales, sur la gauche, sont calculées comme :

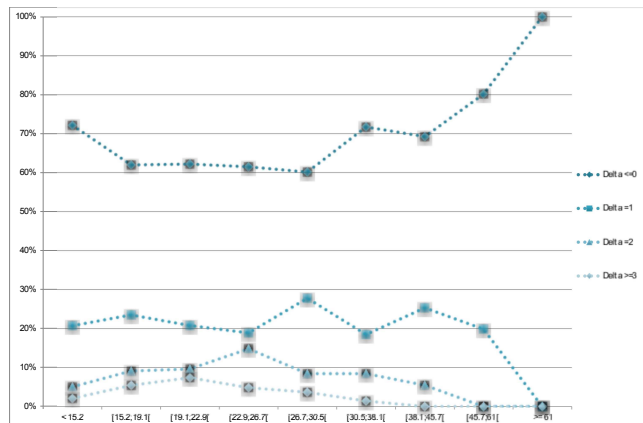
FIGURE 9.2 – Probabilités du  $\Delta Revenu$ 

TABLE 9.9 – Fonction de Transition du Kilométrage Estimée

Factor	Est.	Std.Dev.	T-Stat
Intercept	2,991	0,0999	29,95
Previous km < 2500	-3,630	0,0641	-56,63
Previous km $\in [2500, 5000[$	-2,463	0,0707	-34,82
Previous km $\in [5000, 10000[$	-2,268	0,0580	-39,09
Previous km $\in [10000, 15000[$	-1,876	0,0573	-32,76
Previous km $\in [15000, 20000[$	-1,481	0,0595	-24,89
Previous km $\in [20000, 35000[$	-0,818	0,0581	-14,08
$\Delta r = 1$	0,111	0,0327	3,39
$\Delta r = 2$	0,167	0,0560	2,99
$\Delta r \geq 3$	0,287	0,0645	4,46
$\Delta pc$	0,122	0,0794	1,54
Variance $\tau_k^2$	0,309	0,0081	38,06



$$\Pr(k_{i,t} | r_{i,t-1}, k_{i,t-1}, r_{i,t}) = \sum_{d=0}^3 \left[ \sum_{p=0}^4 \Pr(k_{i,t} | r_{i,t-1}, k_{i,t-1}, r_{i,t}, pc_t) \Pr(pc_t = pc | pc_{t-1}) \right] \Pr(\Delta r = d) \tag{9.28}$$

Ici aussi la probabilité de rester dans la même classe deux ans de suite est la plus importante, exception faite pour les classes extrêmes<sup>8</sup>.

Si l'on confronte ces probabilités à celles conditionnées au changement de revenu, reportées sur la droite de l'image, on s'aperçoit de différences marquées en fonction de ce changement. Effectivement, la probabilité d'augmenter devient de plus en plus importante au fur et à mesure que le ménage monte en classe de revenu.

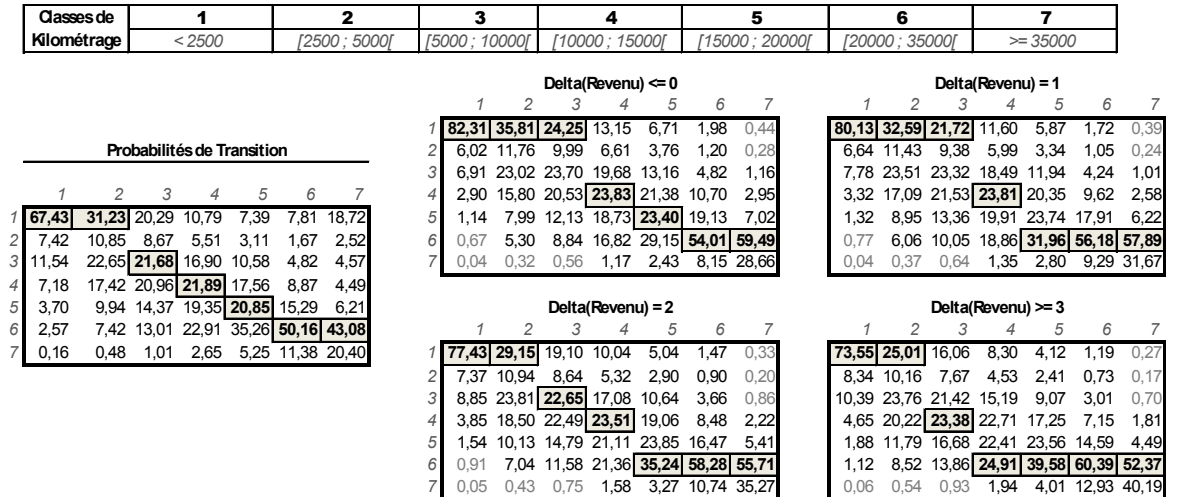


FIGURE 9.3 – Probabilités de Transition du Kilométrage Estimées

### 9.3.2 Le modèle

On a maintenant toutes les clés pour analyser les résultats des estimations.

8. On observe en effet deux classes (2 et 6) qui n'ont jamais probabilité maximale, mais cela pourrait venir d'une mauvaise classification de la variable. Ce sont effectivement ces deux classes qui sont le moins représentées.

Le problème inter-temporel est résolu en statique dans un premier temps afin de produire des valeurs initiales pour l'estimation dynamique. Ces résultats, reportés en table 9.10, donnent déjà une première idée des effets interagissant dans ce problème<sup>9</sup> et sont cohérents avec ce à quoi on pouvait s'attendre.

Comme dit en section 9.1 au moment de définir les utilités 9.7, le premier paramètre ( $\alpha$ ) représente l'utilité d'acheter.

Le fait d'effectuer un achat, toute chose égale par ailleurs, a donc un effet négatif. Cela pourrait être la conséquence à la fois de l'attachement à l'habitude et du coût découlant du choix pour le ménage.

On observe un effet négatif de la classe de revenu sur l'utilité attachée à la décision de ne pas acheter qui se fait de plus en plus important au fur et à mesure que le revenu augmente. En d'autres termes, plus le revenu du ménage est élevé, moins l'effet négatif de l'achat a d'impact sur la décision.

Si l'on analyse l'effet de la demande de kilométrage, sans prendre en compte la durée d'attente, on observe un effet positif sur la décision de ne pas acheter qui s'accroît au fur et à mesure que la demande augmente, pour atteindre un maximum lorsque la demande de kilométrage se trouve en classe 4 et décroître après. Cela signifierait que plus la demande de kilométrage du ménage est importante, moins il est tenté par l'achat. Il pourrait venir du fait qu'une demande de kilométrage élevé implique déjà pour le ménage un coût significatif auquel ce n'est pas optimal de rajouter un achat.

Cet effet est tout même mieux compréhensible en considérant ses interactions avec la durée d'achat.

Lorsque l'on regarde la durée d'attente, celle-ci semble avoir un effet positif, toute chose égale par ailleurs. C'est à dire que plus le nombre d'années écoulées depuis le dernier achat est important et plus le ménage tend à ne pas acheter.

Or, si l'on considère l'interaction entre les deux effets, l'interprétation de ces résultats se fait plus claire. En effet, les paramètres attachés au croisement de ces deux variables sont négatifs. En d'autres termes, l'effet de la durée d'attente change de signe lorsque la demande de kilométrage est importante.

---

9. Cependant ces résultats ne sont pas comparables avec ceux issus de l'estimation dynamique.

TABLE 9.10 – Estimation Statique des Utilités d'Acheter ou Pas

Factor		Estimate	Std.Dev. <sup>a</sup>	T-stat. <sup>b</sup>
$\alpha$ : Effet Achat		-0,936	0,2099	-4,46
Revenu	$\in [15, 2; 19, 1[$ k€	-0,209	0,1837	-1,14
Revenu	$\in [19, 1; 22, 9[$ k€	-0,470	0,1772	-2,65
Revenu	$\in [22, 9; 26, 7[$ k€	-0,543	0,1809	-3
Revenu	$\in [26, 7; 30, 5[$ k€	-0,582	0,1899	-3,06
Revenu	$\in [30, 5; 38, 1[$ k€	-0,833	0,1649	-5,05
Revenu	$\in [38, 1; 45, 7[$ k€	-0,893	0,1926	-4,64
Revenu	$\in [45, 7; 61[$ k€	-0,911	0,1996	-4,56
Revenu	$\geq 61$ k€	-1,101	0,2377	-4,63
Kilométrage	[2500, 5000[	0,991	0,2128	4,66
Kilométrage	[5000, 10000[	1,337	0,1709	7,82
Kilométrage	[10000, 15000[	1,987	0,2476	8,03
Kilométrage	[15000, 20000[	1,800	0,2435	7,39
Kilométrage	[20000, 35000[	1,492	0,2298	6,49
Kilométrage	$\geq 35000$	1,494	0,2497	5,99
Attente <sup>c</sup>		0,038	0,0218	1,75
–	$km_M$ <sup>d</sup>	-0,097	0,0335	-2,91
–	$km_H$ <sup>e</sup>	-0,078	0,0371	-2,12
At, $> 15$ <sup>f</sup>		1,600	0,5511	2,9
–	$km_M$	-4,801	1,0484	-4,58
–	$km_H$	-3,894	1,1590	-3,36
Log-lik. <sup>g</sup>	-1691,65			

<sup>a</sup>. Std.Dev. : standard deviation

<sup>b</sup>. T-stat. : robust Student statistic

<sup>c</sup>. Attente : Nombre d'années qui s'écouleraient du dernier achat si l'agent décidait de ne pas acheter ( $\leq 15$ ).

<sup>d</sup>. Effet Attente pour les ménages ayant une demande de Kilométrage Moyenne

<sup>e</sup>. Effet Attente pour les ménages ayant une demande de Kilométrage Haute

<sup>f</sup>. At.  $> 15$  : Indique que l'attente dépasserait 15 ans si l'agent décidait de ne pas acheter.

<sup>g</sup>. Log-lik. :value of the partial log-likelihood function at convergence.

On pourrait y voir d'une part l'effet du caractère durable des véhicules. En effet, la valeur et l'efficacité d'un véhicule déclinent au cours du temps et en fonction de l'utilisation que l'on en fait. Cela implique que pour un ménage caractérisé par une utilisation automobile importante, le besoin de s'équiper à nouveau pourrait se présenter plus rapidement que chez un ménage moins mobile.

D'un autre côté cela pourrait aussi découler de ce que j'ai précédemment appelé "l'effet négatif de l'habitude", c'est à dire le besoin de nouveauté qui pourrait pousser le ménage à acheter. Je postule que ce besoin de nouveauté dépend de la "préférence pour la voiture" de l'individu. Lorsque cette préférence est faible, l'agent pourrait alors ressentir comme très négatif l'effet d'acheter et préférer continuer à attendre tant qu'il peut. En faisant l'hypothèse que cette préférence joue un rôle dans la définition de ses besoins de kilométrage automobile, les résultats pourraient alors refléter cet effet.

Une dernière observation est à faire sur le choix de la classe absorbante de kilométrage. Si l'effet de la durée était linéaire, c'est à dire si une année d'attente supplémentaire augmente l'utilité de 0,02 indépendamment du nombre d'années déjà attendu, on observerait la valeur estimée pour l'indicateur de durée  $> 15$  seulement une fois que l'attente dépasse les 74 années ( $1,48/0,02 = 74$ ). Cela pourrait signifier alors qu'il existe un point de rupture dans cet effet.

Les résultats de l'estimation dynamique sont reportés en table 9.11 pour 3 facteurs d'actualisation fixés. Puisque les hypothèses de préférence absolue pour le présent ( $\gamma = 0$ ) ou pour le futur ( $\gamma = 1$ ) ne me semblent pas crédibles, les deux cas extrêmes sont  $\gamma = 0,1$  et  $\gamma = 0,8$ <sup>10</sup>. Le cas d'agents qui balancent de façon équilibrée entre présent et futur est représenté par  $\gamma = 0,5$ .

Une première observation est que, comme prévu, la vraisemblance augmente lorsque l'on suppose une préférence pour le futur de plus en plus importante. Cependant, cette augmentation semble moins nette au-delà de 0,5. Cela pourrait signifier que l'hypothèse d'agent ayant un comportement tourné vers le futur vis-à-vis de ce type de décision est correcte mais que les conséquences de court terme ont plus d'impact.

Bien que ces résultats ne soient pas vraiment comparables aux résultats de

---

10. L'algorithme ne converge pas pour  $\gamma = 0,9$

l'estimation statique, on peut quand même remarquer que les paramètres estimés en dynamique sont toujours plus significatifs.

L'utilité d'acheter est toujours négative mais diminue en supposant une préférence pour le futur plus élevée. Cela pourrait signifier que plus l'agent est prévoyant à ce sujet, moins il "subit" l'achat. En effet, plus il anticipe ses décisions et plus il a le temps de les ajuster.

En comparant les 3 estimations, on remarque que l'effet du revenu reste stable pour toute valeur de  $\gamma$ . Cependant cet effet n'est pas tout-à-fait le même que celui suggéré par l'estimation statique. En effet, bien qu'ici aussi il diminue au fur et à mesure que la classe de revenu augmente, cette hausse n'est pas linéaire pour les classes les plus élevées.

L'effet du kilométrage devient de plus en plus important, tout en gardant la même allure que celle commentée plus haut, au fur et à mesure que la préférence pour le futur supposée augmente. Cela pourrait signifier que plus l'agent est tourné vers le futur plus il anticipe ses besoins et son équipement. Il se pourrait alors qu'il ait moins de chance de se trouver démuné et donc forcé à l'achat.

Pour conclure, on peut regarder l'effet de la durée d'attente. Les résultats des estimations dynamiques reflètent les mêmes effets remarqués dans l'estimation initiale.

L'effet reste stable pour tout facteur d'actualisation lorsque l'attente est inférieure ou égale à 15 ans. L'effet des très longues attentes s'estompe si le comportement est supposé de plus en plus tourné vers le futur.

TABLE 9.11 – Estimation Dynamique de la Probabilité d'Acheter

Factor	Estimate	Std.Dev. <sup>a</sup>	T-stat. <sup>b</sup>	Estimate	Std.Dev.	T-stat.	Estimate	Std.Dev.	T-stat.
$\alpha$ : Effet Achat	-1,06	0,13	-8,13	-1,01	0,13	-7,47	-0,95	0,14	-6,75
Rev. classe 2	-0,14	0,12	-1,19	-0,14	0,12	-1,16	-0,13	0,12	-1,13
Rev. classe 3	-0,46	0,11	-4,16	-0,46	0,11	-4,14	-0,46	0,11	-4,11
Rev. classe 4	-0,61	0,11	-5,46	-0,61	0,11	-5,44	-0,60	0,11	-5,42
Rev. classe 5	-0,71	0,12	-5,83	-0,70	0,12	-5,80	-0,70	0,12	-5,77
Rev. classe 6	-0,64	0,11	-5,85	-0,64	0,11	-5,83	-0,63	0,11	-5,79
Rev. classe 7	-0,91	0,12	-7,53	-0,91	0,12	-7,51	-0,91	0,12	-7,47
Rev. classe 8	-0,74	0,13	-5,52	-0,75	0,14	-5,51	-0,74	0,14	-5,49
Rev. classe 9	-0,95	0,16	-5,96	-0,95	0,16	-5,93	-0,95	0,16	-5,90
Km. classe 2	1,06	0,13	8,50	1,13	0,13	8,99	1,19	0,13	9,45
Km. classe 3	1,49	0,10	14,31	1,58	0,11	14,89	1,66	0,11	15,25
Km. classe 4	1,96	0,16	12,26	2,02	0,17	12,22	2,08	0,17	12,18
Km. classe 5	1,76	0,16	11,23	1,84	0,16	11,21	1,90	0,17	11,18
Km. classe 6	1,31	0,15	8,81	1,37	0,15	9,00	1,43	0,16	9,14
Km. classe 7	1,15	0,16	7,04	1,21	0,17	7,13	1,26	0,18	7,21
Attente <sup>c</sup>	0,02	0,01	1,87	0,02	0,01	2,34	0,02	0,01	2,73
– $km_M$ <sup>d</sup>	-0,09	0,02	-4,21	-0,08	0,02	-4,16	-0,07	0,02	-4,02
– $km_H$ <sup>e</sup>	-0,05	0,02	-2,30	-0,04	0,02	-2,22	-0,03	0,01	-2,02
At. > 15 <sup>f</sup>	1,48	0,30	5,00	1,03	0,19	5,42	0,73	0,12	6,24
– $km_M$	-3,00	0,65	-4,61	-2,58	0,57	-4,55	-2,20	0,50	-4,39
– $km_H$	-2,20	0,81	-2,72	-1,42	0,62	-2,30	-0,87	0,46	-1,88
$\gamma$ <sup>g</sup>	0,1			0,5			0,8		
Log-lik. <sup>h</sup>	-4033,20			-4032,65			-4032,33		

a. Std.Dev. : standard deviation

b. T-stat. : robust Student statistic

c. Attente : Nombre d'années qui s'écouleraient du dernier achat si l'agent décidait de ne pas acheter (si  $\leq 15$ ).

d. Effet Attente pour les ménages ayant une demande de Kilométrage Moyen

e. Effet Attente pour les ménages ayant une demande de Kilométrage Haut

f. At. > 15 : Indique que l'attente dépasserait les 15 ans si l'agent décide de ne pas acheter.

g. Facteur d'Actualisation

h. Log-lik. :value of the partial log-likelihood function at convergence.

## 9.4 Conclusion

Le travail ici présenté a pour objectif d'étudier les dynamiques qui régissent les comportements d'achat des ménages.

Le caractère durable des véhicules d'une part, et le rôle que l'habitude joue sur ce type de décisions de l'autre, motivent l'hypothèse d'agent "tourné vers le futur" face à ce type de décisions.

Pour cette raison, il est supposé dans ce travail que la suite des décisions annuelles sur le fait d'acheter ou pas est la solution d'un programme dynamique de Rust.

L'hypothèse ici formulée est que la décision du type de véhicule à acheter n'est pas une question statique que le ménage se pose annuellement mais plutôt une question à laquelle il décide chaque année de se confronter ou pas. Il affrontera ce choix seulement si la décision d'acheter est prise. Ce qui est modélisé ici est le processus de décision en amont, c'est à dire la décision d'effectuer un achat.

Les hypothèses formulées pour la spécification du modèle sont simples et peuvent être critiquées. Notamment le modèle est peu réaliste pour les deux raisons suivantes.

Avant tout, le cadre d'estimation choisi ne relaxe aucune des hypothèses restrictives nécessaires à l'estimation du modèle de Rust (erreurs i.i.d., distribués selon F Gumbel, utilité séparable additive et support discret des variables explicatives).

Bien que cela implique probablement des problèmes d'hétérogénéité non traités, le modèle est plus maniable en imposant ces conditions et peut être traité comme un simple problème de point fixe.

En deuxième lieu, seulement l'acte d'acheter et non le type d'achat est étudié. En effet, je ne considère ni le type de véhicule acheté (ex. véhicule à essence ou diesel), ni le type de transaction choisie (ex. avec un particulier ou un privé). La décision de simplifier ainsi le problème dérive du temps que l'écriture d'un programme d'estimation pour un problème dynamique logit emboîté aurait demandé. Ne disposant pas d'autant de temps, j'ai préféré n'étudier ici que le choix d'acheter, sans m'intéresser aux décisions qui en découlent.

Pour conclure, très peu de facteurs ont été inclus dans la définition des

utilités d'acheter ou pas.

Tout ce qui concerne les autres choix d'équipement et mobilité automobile (ex. quels véhicules "hérités" du parc précédent et comment les utiliser) est considéré comme subjacent au modèle et n'est pas étudié ici. Le ménage d'ailleurs n'est représenté que par sa classe de revenu et le volume de sa demande de kilométrage annuel en voiture. Et l'état de son environnement, résumé par le prix du carburant, entre dans le modèle indirectement par le biais de la demande de kilométrage.

Ces choix sont motivés principalement pour des raisons de coût d'estimation.





# Conclusion Générale

L'objectif de ce travail est l'étude des comportements d'équipement automobile des ménages français.

En me basant sur les travaux de Rust, je propose de considérer ce type de décisions comme résultant de processus dynamiques.

Un modèle complet pour cette classe de choix a été proposé au cours de la première partie de ce travail. L'idée présentée est à considérer comme un projet de travail futur.

Cette proposition est à développer sous plusieurs aspects, que ce soit dans la méthode d'estimation que dans les hypothèses sur les termes d'erreur et les paramètres du modèle.

La deuxième partie présente le traitement effectué sur la base "Parc Auto", fourni par l'IFSTTAR, permettant de traiter les tables des vagues annuelles comme un panel de véhicules.

Ce travail a permis d'observer les difficultés causées par la façon dans laquelle les données sont récoltées, lorsque l'on veut utiliser le panel de véhicules décrit dans la base.

Plusieurs applications pour l'étude de ce sujet sont présentées au cours de la partie concluant le document.

En premier lieu, une étude hédoniste des prix des véhicules est présentée. En raison de l'importance que la consommation du véhicule me semble avoir aux yeux des agents, je propose l'estimation d'un système d'équations visant à définir simultanément la consommation de carburant et le prix du véhicule. Ces deux variables sont transformées au préalable avec la méthode Box-Cox, dans le but d'utiliser une forme fonctionnelle plus adaptée que celle linéaire simple pour ces problèmes.

Les prix d'occasion sont estimés en les supposant fonction décroissante du prix neuf, décroissante selon l'âge et le kilométrage.

La méthode des prix hédonistes a été appliquée à l'étude d'un grand nombre de biens. Cependant très peu d'applications pour l'étude des véhicules sont disponibles. De plus, la structure en deux équations, traitant la consommation comme variable endogène au modèle de prix, ne me semble pas avoir été proposée dans la littérature.

Les 3 chapitres qui suivent cette première application formulent l'hypothèse d'agent prévoyant qui anticipe les conséquences de ses choix ainsi que les évolutions de son état et son environnement, lorsqu'il est confronté aux décisions d'utilisation, possession et achat des véhicules.

Bien que ce type de modèle a été souvent utilisé, rares sont les applications pour le problème de l'équipement automobile. Mon travail se veut donc de reprendre la proposition de Rust, appliquée dans son travail aux décisions de remplacement d'une flotte de bus, et la reporter à l'étude des comportements du ménage. Le chapitre 7 n'apporte aucune amélioration à la spécification de Rust, alors que le chapitre 8 représente le problème inter-temporel par un modèle logit discret/continu. Pour conclure, très rares sont les études portant sur les décisions d'achat automobile du ménage. Lorsque ce problème fait l'objet d'une étude dans la littérature, seulement le type d'achat et pas le fait même d'acheter est modélisé. Pour cette raison le modèle de planification des achats est, à mon avis, innovant pour le sujet.

L'hypothèse d'agent "tourné vers le futur", comme attendu, améliore les résultats obtenus aussi bien pour les modèles de possession et utilisation que pour le modèle d'achat. Les observations mises en lumière dans la littérature sont confirmées par ces travaux.

Plusieurs développements futurs sont envisageables pour ces études. En premier lieu, les modèles estimés pourraient être redéfinis en relâchant les hypothèses simplistes et restrictives adoptées dans la spécification. Les effets de la localisation géographique, qui n'est pas utilisée ici, ainsi que des caractéristiques de l'environnement entourant l'agent, sont vraisemblablement des facteurs importants dans ce processus et devraient être capturés par le modèle.

Le processus suggéré au cours de la première partie pourrait alors être estimé.

# Bibliographie

- Aguirregabiria, V. and Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, 7(1) :1 – 53.
- Aguirregabiria, V. and Mira, P. (2010). Dynamic discrete choice structural models : A survey. *Journal of Econometrics*, 156(1) :38 – 67.
- Altonji, J. (1986). Intertemporal substitution in labor supply : Evidence from micro data. *The Journal of Political Economy*, 94 :176–215.
- Altug, S. and Miller, R. (1990). Household choices in equilibrium. *Econometrica*, 58(3) :543 – 570.
- Amemyia, T. (1981). Qualitative response models : a survey. *Journal of Economic Literature*, 19 :1483–1536.
- Ayres, R., Landen, J., Humpstone, C., Dussec, D., and Doggett, R. (1976). *Automobile forecasting models*.
- Barbusse, S. and Gagnepain, L. (2003). La climatisation automobile, impact énergétique et environnemental (automobile air conditioning : energy and environmental impact). *Données et références, Ademe report, Paris*.
- Baron, J.-F. (2002). La consommation automobile depuis quarante ans. *INSEE Première*, 844 :1–4.
- Becker, G. (1965). A theory of the allocation of time. *The Economic Journal*, 75(299) :493 – 517.
- Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.
- Ben-Akiva, M. and Bierlaire, M. (1999). Discrete choice methods and their applications to short-term travel decisions. In Hall, R., editor, *Handbook of Transportation Science*, pages 5–34. Kluwer.
- Ben-Akiva, M. and Bolduc, D. (1991). Multinomial probit with autoregressive error structure. Technical report, Laval-Recherche en Energie.
- Ben-Akiva, M. and Lerman, S. (1974). Some estimation results of a simultaneous model of auto ownership and mode choice to work. *Transportation*, 3 :357–376.
- Ben-Akiva, M. and Lerman, S. (1985). *Discrete Choice Analysis : Theory and Application to Travel Demand*. MIT Press, Cambridge, MA.

- Bento, A. M., Goulder, L. H., Jacobsen, M. P., and von Haefen, R. H. (2009). Distributional and efficiency impacts of increased us gasoline taxes. *American Economic Review*, 99(3) :667–699.
- Berkovec, J. (1985a). Forecasting automobile demand using disaggregate choice models. *Transportation Research Part B*, 19(4) :315–329.
- Berkovec, J. (1985b). New car sales and used car stocks : A model of the automobile market. *The Rand Journal of Economics*, 16(2) :195–214.
- Berkovec, J. and Rust, J. (1985). A nested logit model of automobile holdings for one vehicle households. *Transportation Research Part B*, 19(4) :275–285.
- Berkovec, J. and Stern, S. (1991). Job exit behavior of older men. *Econometrica*, 59(1) :189 – 210.
- Berkowitz, M. K. et al. (1990). Disaggregate analysis of the demand for gasoline. *The Canadian Journal of Economics*, 23(2) :253–275.
- Berkson, J. (1951). Why i prefer logits to probits. *Biometrics*, 7(4) :327–339.
- Bernard, J., Bolduc, D., and Bélanger, D. (1996). Quebec residential electricity demand : a microeconomic approach. *Canadian Journal of Economics*, 24(1) :92–113.
- Berry, S. (1992). Estimation of a model of entry in the airline industry. *Econometrica : Journal of the Econometric Society*, 60 :889–917.
- Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica : Journal of the Econometric Society*, 63(4) :841–890.
- Bertsekas, D. P. (2000). *Dynamic Programming and Optimal Control (2<sup>nd</sup> edition)*. Athena Scientific.
- Bhat, C. (1994). Imputing a continuous income variable from grouped and missing income observations. *Economics Letters*, 46(4) :311–320.
- Bhat, C. and Sen, S. (2006). Household vehicle type holdings and usage : An application of the multiple discrete–continuous extreme value (MDCEV) model. *Transportation Research Part B*, 40(1) :35–53.
- Bierlaire, M. (1995). A robust algorithm for the simultaneous estimation of hierarchical logit models. GRT Report 95/3, Department of Mathematics, Facultés Universitaires ND de la Paix, Namur, Belgium.
- Birchen, J. and Richard, E. (2008). Politique de la ville et déplacements – fiche n° 10 la mobilité pour tous, un enjeu de cohésion sociale.
- Birkeland, M. and Jordal-Jorgensen, J. (2001). Energy efficiency of passenger cars. In *Paper presented at the European Transport Conference 2001, PTRC, Cambridge, UK*.
- Boursin, P. [http ://philippe.boursin.perso.sfr.fr/auto0.htm](http://philippe.boursin.perso.sfr.fr/auto0.htm).
- Bowman, K. and Ethridge, D. (1992). Characteristics supplies and demands in a hedonic framework : U.s. market for cotton fiber attributes. *American Journal of Agricultural Economics*, 74 :991 – 1002.

- Box, G. and Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Serie B*, 26(2) :211 – 252.
- Bresnahan, T. and Reiss, P. (1990). Entry in monopoly market. *The Review of Economic Studies*, 57(4) :531–553.
- Bresnahan, T. and Reiss, P. (1991). Entry and competition in concentrated markets. *Journal of Political Economy*, 99 :977–1009.
- Brock, W. and Durlauf, S. (2001). Discrete choice with social interactions. *The Review of Economic Studies*, 68(2) :235 – 260.
- Brown, J. and Rosen, S. (1982). On the estimation of structural hedonic price models. *Econometrica*, 50(1) :765 – 768.
- Browning, M., Deaton, A., and Irish, M. (1985). A profitable approach to labor supply and commodity demands over the life-cycle. *Econometrica : Journal of the Econometric Society*, 53 :503–543.
- Brownstone, D., Bunch, D., and Train, K. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B : Methodological*, 34(5) :315–338.
- Bunch, D.S. Brownstone, D. and Golob, T. (1996). A dynamic forecasting system for vehicle markets with clean–fuel vehicles. *World Transport Research*, 1 :189 – 203.
- Byun, D. (2001). The AHP approach for selecting an automobile purchase model. *Information and Management*, 38 :289 – 297.
- Cassel, E. and Mendelsohn, R. (1985). The choice of functional forms for hedonic price equations : Comment. *Journal of Urban Economics*, 18(2) :135 – 142.
- Cavailhès, J. (2005). Le prix des attributs du logement. *Economie et Statistiques*, 381 - 382.
- CCFA (2006). *Analyse and statistiques*. Comité des Constructeurs Français Automobiles.
- Cernicchiaro, G. <http://giuliacernicchiaro.altervista.org/>.
- Cernicchiaro, G. and de Lapparent, M. (2012). How to use the car fleet? a dynamic discrete\ continuous choice model for forward-looking agents. In *66th European Meeting of the Econometric Society*.
- Choo, S. and Mokhtarian, P. (2004). What type of vehicle do people drive? the role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A : Policy and Practice*, 38(3) :201 – 222.
- Court, L. (1941). Entrepreneurial and consumer theories for commodity spectra. *Econometrica*, 9(3) :135 – 162.
- Cowling, K. and Cubbin, J. (1972). Hedonic price indexes for united kingdom cars. *The Economic Journal*, 82(327) :963 – 978.
- Cox, D. R. (1978). Some procedures connected with the logistic qualitative response curve. In Wiley, F. N. D. J., editor, *Research Papers in Statistics*, pages 55–71. London.

- Cragg, J. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39(5) :829 – 844.
- Cropper, M.L. Deck, L. and McConnell, K. (1988). On the choice of functional form for hedonic price functions. *The Review of Economics and Statistics*, 70(4) :668 – 675.
- Das, M. and van Soest, A. (1999). A panel data model for subjective information on household income growth. *Journal of Economic Behavior and Organization*, 40 :409 – 426.
- de Jong, G. (1990). An indirect utility model of car ownership and private car use. *European Economic Review*, 34 :971–985.
- de Jong, G. (1996). A disaggregate model system of vehicle holding duration, type choice and use. *Transportation Research Part B*, 30(4) :263–276.
- de Jong, G., Fox, J., Daly, A., Pieters, M., and Smit, R. (2004). Comparison of car ownership models. *Transport Reviews*, 24(4) :379–408.
- de Jong, G. and Gunn, H. (2001). Recent evidence on car cost and time elasticities of travel demand in Europe. *Journal of Transport Economics and Policy*, 35 :137–160.
- de Jong, G. and Kitamura, R. (1992). A review of household dynamic vehicle ownership models : holdings models versus transactions models. In *PTRC Summer Annual Meeting, 20th, 1992, University of Manchester, United Kingdom*.
- de Jong, G. and Kitamura, R. (2009). A review of household dynamic vehicle ownership models : holdings models versus transactions models. *Transportation*, 36 :733–743.
- De Lapparent, M. (2005). Déplacements domicile–travail en Île–de–France et choix individuels du mode de transport. *L'Actualité économique*, 81(3) :485 – 520.
- de Lapparent, M. and Cernicchiaro, G. (2012). How long to own and how much to use a car? a dynamic discrete choice model to explain holding duration and driven mileage. *Economic Modelling*, 29(5) :1737–1744.
- Deaton, A. (1988). Life-cycle models of consumption : Is the evidence consistent with the theory?
- Deaton, A. and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review*, 70(3) :312 – 326.
- Delucchi, M.A. Murphy, J. and McCubbinc, D. (2002). The health and visibility cost of air pollution : a comparison of estimation methods. *Journal of Environmental Management*, 64(2) :139 – 152.
- DIREM. <http://www.developpement-durable.gouv.fr/Prix-de-vente-moyens-des,10724.html>.
- Dubin, J. A. and McFadden, D. L. (1984). An econometric analysis of residential electric appliance holding and consumption. *Econometrica*, 52(2) :345–362.
- Eckstein, Z. and Wolpin, K. (1989). The specification and estimation of dynamic stochastic discrete choice models. *Journal of Human Resources*, 24(4) :562 – 598.

- Engers, M. and Stern, S. (2008). Long-term care and family bargaining. *International Economic Review*, 43(1) :73–114.
- Epple, D. (1987). Hedonic prices and implicit markets : Estimating demand and supply functions for differentiated products. *Journal of Political Economy*, 95(1) :59 – 80.
- Fang, H. (2008). A discrete/continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B : Methodological*, 42(9) :736 – 758.
- Finney, D. (1971). *Probit Analysis : 3d Ed.* Cambridge University Press.
- Gardes, F. and Cardoso, N. (1996). Estimations de lois de consommation sur un pseudo–panel d'enquêtes de l'insee (1979,1984,1989). *Économie et prévision*, 126(126) :111 – 125.
- Gaudry, M. and Wills, M. (1978). Estimating the functional form of travel demand models. *Transportation Research*, 12(4) :257 – 289.
- Gilbert, C. (1992). A duration model of automobile ownership. *Transportation Research Part B*, 26 :97–114.
- Goldberg, P. (1996). Dealer price discrimination in new car purchases : Evidence from the consumer expenditure survey. *Journal of Political Economy*, 104(3) :622 – 654.
- Goldberg, P. (1998). The effects of the corporate average fuel efficiency standards in the us. *The Journal of Industrial Economics*, 46(1) :1–33.
- Golob, T.F. Bunch, D. and Brownstone, D. (1997). A vehicle use forecasting model based on revealed and stated vehicle type choice and utilisation data. *Journal of Transport Economics and Policy*, 31(1) :69 – 92.
- Golob, T.F. Kim, S. and Ren, W. (1995). How households use different types of vehicles : A structural driver allocation and usage model. *Transportation Research*, 30 :103 – 118.
- Golob, T. (1990). The dynamics of household travel time expenditures and car ownership decisions. *Transportation Research Part A : General*, 24(6) :443 – 463.
- Goodman, A. C. (1983). Willingness to pay for car efficiency : A hedonic price approach. *Journal of Transport Economics and Policy*, 17(3) :247 – 266.
- Goodwin, P. (1986). A panel analysis of changes in car ownership and bus use. *Traffic engineering & control*, 27(10) :519–525.
- Goodwin, P. (1989). Family changes and public transport use 1984–1987. *Transportation*, 16(2) :121–154.
- Goodwin, P. (1992). A review of new demand elasticities with special reference to short and long run effects of price changes. *Journal of Transport Economics and Policy*, 26 :155–163.
- Goodwin, P. and Mogridge, M. (1981). Hypotheses for a fully dynamic model of car ownership. *International Journal of Transport Economics*, 8(3) :313–26.



- Graham, D. and Glaister, S. (2002). The demand for automobile fuel : a survey of elasticities. *Journal of Transport Economics and Policy*, 36 :1–26.
- Graham, D. and Glaister, S. (2004). Road traffic demand elasticity estimates : A review. *Transport Reviews*, 24(3) :261–274.
- Griliches, Z. and Ohta, M. (1976). Automobile prices revisited : Extensions of the hedonic hypothesis. In Terleckyj, N. E., editor, *Household Production and Consumption*, pages 325 – 398. Elsevier.
- Guidez, J. (2007). La mobilité urbaine des années 2000. – fiche n° 3 vers un “découplage” entre la possession et l’usage de la voiture? In CERTU, editor, *Mobilités : faits et chiffres*.
- Hall, R. (1978). Stochastic implications of the life cycle-permanent income hypothesis : Theory and evidence. *Journal of Political Economy*, 86 :971–988.
- Hanemann, W. (1984). Discrete/continuous models of consumer demand. *Econometrica*, 52(3) :541 – 562.
- Hansen, L. and Singleton, K. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica : Journal of the Econometric Society*, 50 :1269–1286.
- Hansen, L. and Singleton, K. (1984). Errate. *Econometrica*, 52 :267–268.
- Harrison, D. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5 :81 – 102.
- Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition. In Bresnahan, T. F. and Gordon, R. J., editors, *The Economics of New Goods*, volume 2, chapter 5, pages 209 – 247. University of Chicago Press.
- Heckman, J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In Manski, C. F. and McFadden, D. L., editors, *Structural Analysis of Discrete Data with Econometric Applications*, chapter 4, pages 179–195. MIT Press, Cambridge, USA.
- Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2) :271 – 320.
- Hensher, D. and Le Plastrier, V. (1985). Towards a dynamic discrete choice model of household automobile fleet size and composition. *Transportation Research Part B : Methodological*, 19(6) :481 – 495.
- Hensher, D. and Milthorpe, F. (1987). Selectivity correction in discrete-continuous choice analysis. *Regional Science and Urban Economics*, 17 :123–150.
- Hensher, D. A. (1985). Empirical vehicle choice and usage models in the household sector : A review. *International Journal of Transport Economics*, 12(3) :231–251.
- Hensher, D. A., Milthorpe, F. W., and Smith, N. C. (1990). The demand for vehicle use in the urban household sector. *Journal of Transport Economics and Policy*, 24(2) :119–137.

- Hess, A. (1977). A comparison of automobile demand equations. *Econometrica*, 45(3) :683 – 702.
- Hivert, L., Wingert, J., et al. (2010). Automobile et automobilité : quelles évolutions de comportements face aux variations du prix des carburants de 2000 à 2008 ? *Pétrole, mobilité, CO2 : Les politiques publiques et l'automobilité face à la variation des prix du pétrole. Rapport Final : mai 2010*.
- Hocherman, I., Prashker, J. N., and Ben-Akiva, M. (1983). Estimation and use of dynamic transaction models of automobile ownership. *Transportation Research Records*, 944 :134–141.
- Hotz, V. and al. (1994). A simulation estimator for dynamic models of discrete choice. *The Review of Economic Studies*, 61(2) :265 – 289.
- Hotz, V., Kydland, F., and G., S. (1988). Intertemporal preferences and labor supply. *Econometrica : Journal of the Econometric Society*, 5(335-360) :6.
- Hotz, V. J. and Miller, R. (1988). An empirical analysis of life cycle fertility and female labor supply. *Econometrica*, 56(1) :91 – 118.
- Hotz, V. J. and Miller, R. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3) :497 – 529.
- INED. [http://www.ined.fr/fr/ressources\\_documentation/focus\\_sur/age\\_pacs\\_mariage/](http://www.ined.fr/fr/ressources_documentation/focus_sur/age_pacs_mariage/).
- INSEE. <http://www.insee.fr/fr/themes/theme.asp>.
- INSEE (2004). La situation démographique en 2004 - mouvement de la population. [http://www.insee.fr/fr/themes/document.asp?ref\\_id=sd2004](http://www.insee.fr/fr/themes/document.asp?ref_id=sd2004).
- INSEE (2010). Bilan démographique 2010. [http://www.insee.fr/fr/themes/document.asp?ref\\_id=ip1332](http://www.insee.fr/fr/themes/document.asp?ref_id=ip1332).
- Julliard, M. (2007). Le budget automobile des ménages s'adapte aux prix des carburants. *INSEE Première*, 1159 :1–4.
- Keane, M. P. and Wolpin, K. I. (1994). The solution and estimation of discrete choice dynamic programming models by simulation and interpolation : Monte carlo evidence. Technical Report 181, Federal Reserve Bank of Minneapolis.
- Kim, C.W. Phippsb, T. and Anselinc, L. (2003). Measuring the benefits of air quality improvement : a spatial hedonic approach. *Journal of Environmental Economics and Management*, 45(1) :24 – 39.
- Kim, Y. (2002). Estimation of discrete/continuous choice model : Application of bayesian approach using gibbs sampling. *Applied Economics Letters*, 9 :305–309.
- Kitamura, R. (1987). A panel analysis of household car ownership and mobility. In *Proceedings of the Japan Society of Civil Engineers*, volume 383, pages 13–27.
- Kitamura, R. (1988). An evaluation of activity-based travel analysis. *Transportation*, 15(1 - 2) :9 – 34.
- Kitamura, R. (1990). Panel analysis in transportation planning : An overview. *Transportation Research Part A : General*, 24(6) :401 – 415.

- Kitamura, R. (2009). A dynamic model system of household car ownership, trip generation, and modal split : model development and simulation experiment. *Transportation*, 36(6) :711 – 732.
- Kitamura, R., Golob, T., Yamamoto, T., and Wu, G. (2000). Accessibility and auto use in a motorized metropolis. In *Paper presented at the 79th Transportation Research Board Annual Meeting, Washington, DC*.
- Kitamura, R. and van der Hoorn, T. (1987). Regularity and irreversibility of weekly travel behavior. *Transportation*, 14 :227–51.
- Kooreman, P. (2006). Estimation of econometric models of some discrete games. *Journal of Applied Econometrics*, 9(3) :255–268.
- Koppelman, F. and Wen, C. (1998). Alternative nested logitmodels : structure, properties and estimation. *Transportation Research Part B : Methodological*, 32(5) :289 – 298.
- Koppelman, F. and Wen, C. (2000). The paired combinatorial logit model : properties, estimation and application. *Transportation Research Part B : Methodological*, 34(2) :75 – 89.
- Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74 :132 – 156.
- Lave, C. and Train, K. (1979). A disaggregate model of auto-type choice. *Transportation Research Part A : General*, 13(1) :1 – 9.
- Lave, C. A. and Bradley, J. (1980). Market share of imported cars : A model of geographic and demographic determinants. *Transportation Research Part A : General*, 14(5-6) :379 – 387.
- Lucas, R. J. (1976). Econometric policy evaluation : A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(1) :19–46.
- MaCurdy, T. (1994). A framework for relating microeconomic and macroeconomic evidence on intertemporal substitution. In *Advances in Econometrics—Fifth World Congress of the Econometric Society Vol. II*, pages 149–176.
- Magnac, T. and Thesmar, D. (2002). Identifying dynamic discrete decision processes. *Econometrica*, 70 :801–816.
- Mankiw, N., Rotemberg, J., and Summers, L. (1985). Intertemporal substitution in macroeconomics. *Quarterly Journal of Economics*, 100 :225–251.
- Mannering, F. (1983a). *Dynamic Econometric Models of Car Ownership and Utilization*. PhD thesis, Massachusetts Institute of Technology.
- Mannering, F. (1983b). An econometric analysis of vehicle use in multivehicle households. *Transportation Research Part A*, 17(3) :183–189.
- Mannering, F. and Hensher, D. (1987). Discrete/continuous econometric models and their application to transport analysis. *Transport Reviews : A Transnational Transdisciplinary Journal*, 7(3) :227 – 244.
- Mannering, F. and Winston, C. (1985). A dynamic empirical analysis of household vehicle ownership and utilisation. *The Rand Journal of Economics*, 16(2) :215–236.

- Manning, F., Winston, C., and Starkey, W. (2002). An exploratory analysis of automobile leasing by us households. *Journal of Urban Economics*, 52(1) :154–176.
- Manski, C. (1977). A dynamic empirical analysis of household vehicle ownership and utilisation. *Theory and Decision*, 8(3) :229 – 254.
- Manski, C. (1980). *Short run equilibrium in the automobile market*. Maurice Falk Institute for Economic Research in Israel.
- Manski, C. and Sherman, L. (1980). An empirical analysis of household choice among motor vehicles. *Transportation Research Part A*, 14(6) :349–366.
- Marsh, L. (2012). Mindscapes and landscapes : Hayek and simon on cognitive extension.
- Massot, M. and Orfeuill, J. (2005). La mobilité au quotidien, entre choix individuel et production sociale. *Cahiers internationaux de sociologie*, (1) :81–100.
- Mazuy, M. Prioux, F. and Barbieri, M. (2011). L'évolution démographique récente en france. quelques différences entre les départements d'outre-mer et la france métropolitaine. *Population - F*, 66(3-4) :503 – 504.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, chapter 4, pages 105–142. New York : Academic Press.
- McFadden, D. (1977). Modelling the choice of residential location. Cowles Foundation Discussion Papers 477, Cowles Foundation, Yale University.
- McFadden, D. (1980). Econometric models for probabilistic choice among products. *The Journal of Business*, 53(3) :S13 – S29.
- McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15 :447 – 470.
- Mohammadian, A. and Miller, E. (2003). Empirical investigation of household vehicle type choice decisions. *Transportation Research Record : Journal of the Transportation Research Board*, 1854 :99 – 106.
- Mueller, M. and de Haan, P. (2009). How much do incentives affect car purchase? agent-based microsimulation of consumer choice of new cars. part i : Model structure, simulation of bounded rationality, and model validation. *Energy Policy*, 37(3) :1072 – 1082.
- Page, M., Whelan, G., and Daly, A. (2000). Modelling the factors which influence new car purchasing. In *Paper presented at the European Transport Conference 2000, PTRC, Cambridge, UK*.
- Patalano, R. and Rizzello, S. (2002). Il concetto di image nel pensiero di kenneth boulding e le implicazioni per la teoria economica contemporanea. Technical report, CESMEP Working Paper 08/2002.
- Pla, A. and Beaumel, C. (2011). Bilan démographique 2010. la population française atteint 65 millions d'habitants. *INSEE Première*, 1332 :1–4.
- Rosen, S. (1974). Hedonic prices and implicit markets : Product differentiation in pure competition. *Journal of Political Economy*, 82 :34 – 35.

- Rust, J. (1987). Optimal replacement of GMC bus engines : An empirical model of Harold Zurcher. *Econometrica*, 55(5) :999–1033.
- Rust, J. (1988). Maximum likelihood estimation of discrete control processes. *SIAM Journal on Control and Optimization*, 26(5) :1006–1023.
- Rust, J. (1994). Structural estimation of markov decision process. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Computational Economics*, volume 4, chapter 51, pages 3081 – 3143. Elsevier.
- Rust, J. (1996). Numerical dynamic programming in economics. In Amman, H. M., Kendrick, D. A., and Rust, J., editors, *Handbook of Computational Economics*, volume 1, chapter 14, pages 619–729. Elsevier.
- Rust, J. (1997). Using randomization to break the curse of dimensionality. *Econometrica*, 65(3) :487–516.
- Sheppard, S. (1999). Hedonic analysis of housing markets. In Mills, E. and (eds), P. C., editors, *Handbook of Regional and Urban Economics*, volume 3, Applied Urban Economics, pages 1595 – 1635.
- Small, K. A. (1987). A discrete choice model for ordered alternatives. *Econometrica*, 55(2) :409–24.
- Smith, N., Hensher, D., and Wrigley, N. (1991). A dynamic discrete choice sequence model : Method and an illustrative application to automobile transactions. *International Journal of Transport Economics*, 18(2) :123–150.
- Stewart, M. (1983). On least squares estimation when the dependent variable is grouped. *Review of Economic Studies*, 50 :141–149.
- Terza, J. (1985). Ordered probit : A generalization. *Communications in Statistics ? A. Theory and Methods*, 14 :1–11.
- Terza, J. (1987). Estimating linear models with ordinal qualitative regressors. *Journal of Econometrics*, 34(3) :275 – 291.
- TNS-SOFRES. <http://www.tns-sofres.com/notre-offre/etudes-collectives/BF2991386A804D6BBFD23E50473CF89D.aspx>.
- Toivanen, O. and Waterson, M. (2000). Empirical research on discrete choice game theory models of entry : An illustration. *European Economic Review*, 44(4) :985–992.
- Train, K. (1986). *Qualitative Choice Analysis : Theory, Econometrics, and an Application to Automobile Demand*. MIT Press.
- Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Train, K. and Lohrer, M. (1982). *Vehicle ownership and usage : an integrated system of disaggregate demand models*. Cambridge Systematics, Inc.
- Train, K. and Winston, C. (2007). Vehicle choice behavior and the declining market share of u.s. automakers. *International Economic Review*, 48(4) :1469–1496.
- Turrentine, T. S. and Kurani, K. S. (2007). Car buyers and fuel economy. *Energy Policy*, 35 :1213–1223.

- van Wissen, L. and Golob, T. (1992). A dynamic model of car fuel-type choice and mobility. *Transportation Research Part B : Methodological*, 26(1) :77 – 96.
- Wen, C. and Koppelmanb, F. (2001). The generalized nested logit model. *Transportation Research Part B : Methodological*, 35(7) :627 – 641.
- West, S. (2004). Distributional effects of alternative vehicle pollution control policies. *Journal of public Economics*, 88(3) :735–757.
- Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20(1) :39–54.