

FRANCIS VEKEMAN

**Choix de véhicules et demande de
kilométrage : une approche
microéconométrique**

Mémoire présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en économie
pour l'obtention du grade de maître ès arts (M.A.)

FACULTÉ DES SCIENCES SOCIALES
UNIVERSITÉ LAVAL
QUÉBEC

NOVEMBRE 2004

©Francis Vekeman, 2004

Résumé

Cette étude utilise un modèle microéconométrique de la classe discret-continu pour étudier le choix et l'usage du parc de véhicules privés québécois. Le modèle de choix est estimé à l'aide d'un logit mixte et l'ensemble de choix est constitué de chacun des modèles de véhicules présent dans la banque de données, soit un total de 230 modèles. Une correction qui généralise celle d'Heckman pour chacun des éléments de l'ensemble de choix est intégrée à l'équation d'usage, elle-même estimée par la méthode des moindres carrés ordinaires. Les données utilisées proviennent de l'*Enquête sur le kilométrage des conducteurs et conductrices du Québec* effectuée par la Société de l'assurance automobile du Québec en 1996-97. Les résultats de différentes simulations indiquent une grande sensibilité de la composition du parc de véhicules à l'âge des conducteurs. Le sexe, le lieu d'habitation et l'âge sont parmi les principaux facteurs qui influencent l'utilisation.

Abstract

A discrete/continuous choice model is used to analyze the ownership of private motor vehicles and their use in Quebec. The estimation of the discrete choice model is based on the Mixed Logit specification and each of the 224 different vehicle models appears as specific element of the choice set. A generalization of Heckman's correction term is incorporated in the vehicle use equation for each alternative. The data come from a special survey conducted by the *Société de l'assurance automobile du Québec* in 1996-1997. The empirical results show that car choice is highly related to driver's age and fuel consumption. Prices of car do not affect the probability of owning most of vehicle classes, including SUV. Gender, household location, age, and fuel price are among the main determinants of car use.

Avant-Propos

Cette étude a bénéficié du support financier de la Chaire en économie de l'énergie électrique, du Fonds québécois de la recherche sur la société et la culture (FQRSC) ainsi que du Centre de données et d'analyse sur les transports (CDAT). Merci également à la Société de l'assurance automobile du Québec pour m'avoir autorisé à utiliser leurs données.

Je tiens à remercier d'une façon toute spéciale mon co-directeur, Denis Bolduc, qui, bien qu'il le soit devenu sans le savoir, m'a accordé plus de temps qu'un étudiant peut en espérer. Il a su me transmettre sa rigueur scientifique et sa passion pour la recherche. Je remercie également mon directeur, Jean-Thomas Bernard, pour son appui et ses conseils. Merci à Nathalie Boucher et Philippe Barla, membres du CDAT, pour leurs judicieux commentaires.

Merci à mes proches qui m'ont supporté tout au long de mes études. Finalement, merci au petit livre des grandes inventions, à travers lequel j'ai acquis la passion de découvrir.

Pangloss enseignant à Candide : **«Il est démontré [...] que les choses ne peuvent être autrement : car, tout étant fait pour une fin, tout est nécessairement pour la meilleure fin.»**

Candide ou l'optimisme

Voltaire

*«If you torture the data hard enough,
it will confess to anything.»
Un célèbre économètre*

Table des matières

Résumé	i
Abstract	ii
Avant-Propos	iii
1 Introduction	1
2 Revue de la littérature	5
3 Méthodologie	15
3.1 Modèle de choix	16
3.1.1 Probit	19
3.1.2 Logit	21
3.1.3 Logit mixte	22
3.2 Conditions d'identification	25
3.3 Fonction d'utilité indirecte	29
3.4 Correction pour l'estimation de la demande de kilométrage	31
4 Données	35
5 Résultats	42
5.1 Modèle de choix	42
5.2 Modèle d'usage	50

6 Conclusion	53
Bibliographie	56
Appendice A Résumé des études présentées au chapitre 2	60
Appendice B Définition des variables	62
Appendice C Suite des résultats du modèle de choix	65

Table des figures

3.1	Régression linéaire simple sur un choix binaire	18
5.1	Simulation 1, vieillissement de tous les conducteurs de 10 ans	45
5.2	Simulation 2, ajout d'un enfant âgé entre 7 et 15 ans	46

Liste des tableaux

5.1	Estimations du modèle de choix	43
5.2	Simulation 3, augmentation de 10 % du taux de consommation de carburant des VUS	47
5.3	Simulation 4, augmentation de 10 % du prix des grosses berlines . . .	48
5.4	Composantes d'erreur (logit mixte)	49
5.5	Estimations du modèle d'usage	51

Chapitre 1

Introduction

Les questions de choix et d'usage des véhicules automobiles ont été et sont toujours des questions de toute première importance à des fins de planification, mais aussi d'un point de vue théorique. Nous entendons par choix d'un véhicule, la décision d'un individu ou d'un ménage d'acheter telle ou telle marque, modèle et année de véhicule. Le terme usage est utilisé pour caractériser l'intensité d'utilisation du véhicule sélectionné. Ce dernier se calcule généralement en kilomètre ou en consommation de carburant. Depuis quelques années, la responsabilité grandissante des véhicules privés dans le total des émissions de gaz à effet de serre (GES) des pays industrialisés nourrit un nouvel intérêt pour ces questions.

Des conjonctures plus locales peuvent elles aussi motiver l'étude du secteur du transport privé. À ce titre, les études empiriques antérieures sont assez unanimes sur le fait que les caractéristiques sociodémographiques d'un pays, d'une région ou d'une ville influencent de façon importante non seulement la composition des parcs automobiles privés, mais aussi leur utilisation (Pas, 1984, Hensher et Smith, 1986, Schaper et Patterson, 1998, Kavalec 1999). Ainsi, les changements qui s'opèrent actuellement dans la société québécoise (vieillesse de la population, exode rural, etc.) sont

susceptibles de modifier la composition et l'intensité d'utilisation de son parc de véhicules. Pour comprendre ces changements, il est essentiel d'étudier la relation entre la propriété de véhicules spécifiques et certains facteurs sociodémographiques. À cette fin, nous proposons un modèle microéconométrique qui permet de mettre en lumière l'influence de différentes variables dont l'âge, le sexe, le revenu et le lieu d'habitation sur la composition et l'usage du parc de véhicules privés québécois. La validation empirique de notre modèle est faite à partir d'une enquête sur le kilométrage des conducteurs et conductrices du Québec effectuée en 1996-1997 par la Société de l'assurance automobile du Québec (SAAQ).

Historiquement, les écrits sur les questions de choix et d'usage ont été marqués par une dichotomie trouvant sa source dans la perspective adoptée. Certaines études considèrent la question à un niveau agrégé alors que d'autres analysent le problème en ciblant plutôt les ménages ou les individus. Cette dernière catégorie d'études permet de circonscrire au niveau du consommateur les variables susceptibles d'influencer son comportement. La mise en place au cours des dernières décennies de banques de données de plus en plus riches sur les habitudes de conduite des propriétaires de véhicules automobiles a contribué à accroître l'intérêt pour cette approche désagrégée. De plus, la publication par Train (1979) et Tardiff (1980), à la fin des années soixante-dix, de deux études rétrospectives sur la modélisation du choix des véhicules à partir d'hypothèses sur le comportement des agents et des ménages a permis de consolider les bases théoriques nécessaires au développement de ce champ de recherche en économie des transports (Hensher, 1985).

Au moment de l'achat d'un véhicule automobile, l'utilisation prévue de ce dernier est susceptible d'influencer le type de véhicule qui sera sélectionné et vice versa. Par exemple, un homme d'affaires, qui prévoit faire le trajet entre Québec et Montréal deux fois par semaine, optera pour un véhicule confortable, sans nécessairement se

préoccuper de l'espace de rangement. À l'opposé, une famille de quatre enfants, qui voyage peu, mais dont les deux plus jeunes jouent au hockey chaque samedi à l'aréna du quartier, privilégiera davantage l'espace au confort. La dynamique décisionnelle qui sous-tend l'achat d'un véhicule par un ménage est donc influencée de façon simultanée par le choix et l'usage.

En termes de modélisation, ce type de situation se traduit par une structure à deux niveaux, dite discrète-continue. La probabilité de choix de tel ou tel véhicule ou catégorie de véhicules étant estimée au premier niveau, alors que l'intensité d'utilisation, conditionnelle au véhicule sélectionné, est estimée au second niveau. Pour tenir compte de l'aspect conditionnel de l'utilisation, la méthodologie standard consiste à introduire un terme de correction qui généralise celui d'Heckman dans cette dernière équation. La présente étude suit essentiellement cette méthodologie relativement usuelle. Elle se distingue cependant par l'utilisation, pour l'estimation du modèle de choix, du logit mixte au lieu des traditionnels probit ou GEV (*generalized extreme value*). L'emploi de cet outil d'estimation permettra de révéler, s'il y a lieu, les liens entre des modalités supposées jusqu'à maintenant indépendantes étant donné les contraintes des modèles de choix utilisés¹. De plus, à l'opposé des études connues, l'ensemble de choix considéré n'est pas agrégé à des classes de véhicules, mais comprend plutôt chacun des modèles de véhicules présents dans la banque de données, soit 230. Il en découle l'inclusion d'autant de termes de correction dans l'équation d'usage. Il va sans dire que le traitement d'un ensemble de choix de cette taille donne lieu à une «gymnastique» particulière. Finalement, à notre connaissance, aucune étude n'a encore analysé de façon aussi désagrégée le choix et l'usage d'un parc de véhicules privés.

Dans le chapitre qui suit, nous présentons une brève revue des articles qui ont marqué la recherche sur le choix et l'usage des véhicules. Le chapitre 3 est consacré

¹Cette intuition est, entre autres, nourrie par les travaux de McCarthy et Tay (1998).

aux aspects méthodologiques de notre étude. Plus spécifiquement, les sections 3.1, 3.2, 3.3 et 3.4 traitent respectivement du modèle de choix, des conditions d'identification, du modèle d'usage et du terme de correction. Le chapitre 4 présente les données utilisées. Les résultats d'estimation ainsi que différentes simulations sont rapportés au chapitre 5.

Chapitre 2

Revue de la littérature

La revue de littérature qui suit présente quelques études qui traitent du choix et de l'usage des véhicules privés dans un cadre discret-continu et qui, à notre avis, ont contribué à l'avancement de ce champ de recherche¹. Les études seront d'abord présentées de façon générale en faisant ressortir ce qui les distingue les unes des autres. Cette première étape permettra de bien mettre en contexte chaque recherche. Les études seront ensuite mises en rapport à différents aspects de la modélisation².

Train et Lohrer, lors de la réunion annuelle du «Transportation Research Board»

¹De nombreux champs de spécialisation en économie ont recours aux modèles discrets-continus (économie du travail, économie de l'énergie, etc.). Il existe en conséquence une littérature très riche sur le sujet. Nous avons toutefois choisi de circonscrire notre revue à des études qui touchent directement notre thèse de recherche.

²Bien que les études choisies utilisent toutes sensiblement le même cadre méthodologique, les objectifs poursuivis par chacune d'elles diffèrent. Il serait par conséquent peu pertinent, en regard à notre problématique, de mettre en parallèle les conclusions de chaque recherche.

de 1983 à Washington, furent les premiers à présenter de façon formelle une modélisation conjointe du choix et de l'usage des véhicules automobiles (Hensher, 1985)³. Les travaux de Train et Lohrer (1983) se limitent toutefois aux ménages qui possèdent un et deux véhicules. Le modèle de Mannering et Winston (1985) représente encore à ce jour l'un des modèles les plus sophistiqués en la matière. En utilisant le cadre théorique de Train et Lohrer (1983), dans un contexte dynamique, les auteurs mettent l'accent sur les effets de préférence et de fidélité pour les marques (par exemple pour les véhicules américains ou étrangers) dans le choix de véhicules par les ménages.

Hensher et Milthorpe (1987b) n'ont pour leur part fait aucune avancée majeure au niveau de la structure même du modèle. Le point central de leur article est de vérifier l'impact sur les résultats d'estimation du modèle d'usage de différentes formules de correction pour les biais de sélection potentiels provenant de l'absence de données sur l'usage des véhicules que les ménages n'ont pas sélectionnés (les causes et les conséquences de ce problème seront exposées plus en détail à la section 3.4). Bien que l'exemple qu'ils ont choisi concerne le secteur des transports, leurs travaux sont applicables à la consommation des biens durables en général⁴. L'importance des travaux de Berkowitz *et al.* (1990) repose principalement sur leur distinction entre l'usage des véhicules pour se rendre au travail et les autres usages et ce, pour des ménages possédant zéro, un ou deux véhicules. L'utilisation d'une banque de données sur les ménages canadiens recueillie dans le but précis d'estimer un modèle de choix et d'usage des véhicules automobiles est une autre caractéristique importante. Les enquêtes utilisées par la majorité des autres études n'ayant pas été nécessairement construites à cette fin, les données nécessitent souvent des ajustements ou imputations

³Étant donné la difficulté à obtenir le document de Train et Lohrer, qui fut présenté lors d'un colloque, l'information est tirée de la revue d'Hensher (1985).

⁴À ce titre, l'un des articles les plus importants en ce qui a trait à l'estimation du choix et de l'usage de bien durable, dans le cadre d'un modèle discret-continu, est celui de Dubin et McFadden (1984) et concerne les électroménagers.

qui peuvent modifier les résultats finaux.

Au contraire de Berkowitz *et al.* (1990), Hensher, Milthorpe et Smith (1990) proposent un cadre applicable à toutes les grandeurs de parcs automobiles. Leur modèle traite ainsi chaque véhicule d'un ménage comme une entité unique, possédant des liens avec les autres véhicules à travers les niveaux d'utilisation. Finalement, Kim (2002) propose une méthode d'estimation des modèles discrets-continus basée sur l'approche bayésienne. Le cadre général qu'il préconise est similaire à celui adopté par Hensher, Milthorpe et Smith (1990), mais il fonde l'estimation du modèle de choix sur un instrument jusqu'à récemment exclu par la lourdeur des calculs qu'il nécessite.

La probabilité de sélection d'un véhicule, tout comme son utilisation, est estimée sur la base d'une fonction d'utilité. Le point important à noter ici est la nature durable et discrète du bien considéré et l'implication de ces deux attributs sur le type de fonction d'utilité préconisé. Dans une perspective générale, l'intensité d'utilisation d'un bien durable peut être vue comme un indicateur du flux de services fournis par ce dernier. Dans le présent cas, la demande de carburant ou de kilométrage⁵ est dérivée de la demande de services de transport fournis par le ou les véhicules utilisés. L'utilité retirée de la consommation de carburant ou de kilomètre doit donc être vue comme étant indirecte. Dans l'optique d'obtenir l'intensité d'usage du véhicule choisi, il est alors essentiel de travailler avec une fonction d'utilité indirecte pour le calcul de la probabilité de choix. La forme fonctionnelle de cette dernière est toutefois sujette à changement selon les auteurs.

⁵Lorsque seule l'information sur le kilométrage est disponible, la demande de carburant est obtenue en multipliant le kilométrage parcouru par l'efficacité énergétique du véhicule considéré, c'est-à-dire le nombre de litres de carburant consommés par 100 kilomètres. Il suffit d'effectuer le calcul inverse pour la demande de kilométrage lorsque seule la consommation de carburant est disponible.

Alors que Train et Lohrer (Hensher, 1985) justifient la structure de leur fonction d'utilité par ses propriétés mathématiques, c'est-à-dire la linéarité dans les paramètres, Mannering et Winston (1985) et Hensher, Milthorpe et Smith (1990) motivent la forme de la fonction d'utilité qu'ils emploient par le fait qu'elle peut intégrer la présence de coûts fixes tout en conservant la propriété de linéarité dans le revenu et les prix. Cette forme fonctionnelle, tirée par les deux groupes d'auteurs de l'article de Dubin et McFadden (1984), est basée à l'origine sur les travaux de Hausman (1981). Par l'accent qu'ils mettent sur l'usage, Berkowitz *et al.* (1990) avancent pour leur part qu'ils ne peuvent utiliser la fonction d'utilité proposée par Dubin et McFadden (1984) puisqu'elle mène à des fonctions d'usage non symétriques. Les auteurs expliquent que sans information pour distinguer le premier et le second véhicule d'un ménage, il n'y a pas de raison d'attribuer à chacun des fonctions d'usage différentes. Berkowitz *et al.* (1990) suggèrent donc une fonction d'utilité indirecte qui conduit à des fonctions d'usage symétriques.

L'intensité d'utilisation, considérée au deuxième niveau du modèle, est dérivée sur la base de la théorie de la dualité économique. L'application de l'identité de Roy⁶ à la fonction d'utilité indirecte conditionnelle au choix du véhicule, sélectionné au premier niveau, nous permet de dériver la demande Marshallienne pour l'usage de celui-ci. Cette procédure est adoptée unanimement par tous les auteurs considérés. Ces derniers ne préconisent toutefois pas tous la même décomposition de la variable d'usage.

Mannering et Winston (1985) et Hensher, Milthorpe et Smith (1990), en se basant

⁶Identité de Roy : si une fonction d'utilité indirecte $v(\mathbf{p}, y)$, où \mathbf{p} est un vecteur de prix et y le revenu, est différentiable en (\mathbf{p}^o, y^o) et $\partial v(\mathbf{p}^o, y^o)/\partial y \neq 0$ alors

$$x_i(\mathbf{p}^o, y^o) = -\frac{\partial v(\mathbf{p}^o, y^o)/\partial p_i}{\partial v(\mathbf{p}^o, y^o)/\partial y}, \quad i = 1, \dots, n.$$

sur les hypothèses adoptées par Dubin et McFadden (1984), considèrent l'utilisation totale comme étant la somme de deux composantes. Pour Mannering et Winston (1985), l'usage total est composé, d'une part de l'utilisation typique sur une période de temps donnée des véhicules de même marque (même manufacturier) et, d'autre part, de l'usage du véhicule sélectionné par le ménage (par exemple marque, modèle et année). Selon ces auteurs, une telle procédure permet de contrôler pour l'effet de fidélité aux marques. Hensher, Milthorpe et Smith (1990) considèrent quant à eux un usage typique, indépendant du choix de véhicule, et un usage conditionnel au véhicule sélectionné. Ils utilisent ensuite le fait que la première composante peut être calculée comme une moyenne agrégée des usages totaux faits par l'ensemble des ménages considérés, pour la remplacer par un ensemble d'instruments socioéconomiques. Hensher, Milthorpe et Smith (1990) espèrent ainsi améliorer la précision de leur modèle.

Du point de vue de la théorie économique, il apparaît maintenant clair que le choix d'un véhicule par un ménage résulte de la solution d'un problème de maximisation de l'utilité jointe du choix de ce dernier et de son usage. Cette approche diffère du modèle de flux et de stock néoclassique, en intégrant la nature discrète du bien durable au lieu de supposer son renouvellement de façon continue (Hensher, Milthorpe et Smith, 1990). Considérons maintenant la partie discrète du problème.

Le premier niveau d'un cadre discret-continu est constitué d'un modèle à choix discret typique. Dans le présent cas, ce choix est dit polytomique dans la mesure où les ménages doivent faire un choix parmi une multitude de véhicules ou de types de véhicules. On distingue deux classes générales de modèles de choix discrets selon que la probabilité de choix prend une forme explicite (logit, GEV) ou non (probit, logit mixte). La différence entre les deux classes provient principalement de la spécification de la loi de distribution des erreurs de la fonction d'utilité, celle-ci imposant la structure de la probabilité à estimer. Dans le cas du probit, la distribution des erreurs

est supposée normale, ce qui conduit à l'estimation d'une probabilité comportant de un à plusieurs niveaux d'intégrale (forme non explicite). D'un autre côté, la densité qui permet d'obtenir le logit est la Gumbel indépendante et identiquement distribuée (*iid*). Cette densité permet de retrouver une probabilité dont la structure globale est un ratio (forme explicite).

Le probit offre une structure qui permet de modéliser avec une plus grande flexibilité les interrelations entre les modalités considérées. Cependant, comme le nombre de niveaux d'intégrale à évaluer grandit avec le nombre d'alternatives considérées, ce dernier est souvent délaissé dans des cas où l'ensemble de choix est grand. Le développement de l'informatique a toutefois permis d'augmenter considérablement la capacité des ordinateurs et leur vitesse de calcul, offrant ainsi aux économètres de nouveaux outils pour remédier à ce problème. La simulation constitue, à ce titre, une technique puissante de plus en plus employée pour contourner le calcul des intégrales. Kim (2002) utilise d'ailleurs cette méthode pour estimer la probabilité de choix de chacune des modalités obtenue dans le cadre d'un probit polytomique. Nous verrons plus en détail au chapitre 3 en quoi consiste cette procédure. Le logit est, pour sa part, beaucoup plus restrictif quant aux interrelations qu'il permet entre les modalités. La loi de distribution des erreurs qui conduit au logit (Gumbel *iid*), implique le calcul d'une probabilité qui impose, par sa structure, l'indépendance des alternatives non pertinentes (IAN). Cette caractéristique implique que le ratio des probabilités de choix de deux alternatives dépend uniquement des attributs de ces deux alternatives. En d'autres mots, le choix relatif d'une modalité sur une autre sera toujours le même, peu importe les autres modalités disponibles et les caractéristiques de ces dernières. De ce fait, l'utilisation du logit, dans sa formulation de base, contraint le modélisateur à faire l'hypothèse d'indépendance entre les modalités qu'il considère, hypothèse parfois difficile à défendre dans le contexte de la réalité qu'il tente de saisir. À l'inverse

du probit toutefois, la probabilité de choix du logit est plus facile à estimer, même lorsque l'ensemble de choix est grand.

L'avènement du logit emboîté ou *nested logit* a permis de relâcher sensiblement l'hypothèse d'IAN entre les modalités. Cette formulation du logit repose sur une structure en arbre, l'extrémité de chaque branche étant formée d'un nid (regroupement de modalités). Ainsi, l'hypothèse d'IAN doit être respectée à l'intérieur de chaque nid alors qu'une certaine corrélation est permise entre les nids. De cette façon, le modélisateur peut regrouper entre elles les modalités qu'il estime semblables, donc peu susceptibles de voir leur probabilité de choix influencée par les autres modalités du même nid. Les modalités dont la probabilité de sélection risque de ne pas être indépendante sont placées dans des nids différents. Cette formulation du logit relâche dans une certaine mesure l'hypothèses d'IAN tout en permettant à la probabilité de conserver une forme explicite. Le logit emboîté représente donc un instrument intéressant pour estimer le modèle de choix. Il est d'ailleurs utilisé par tous les auteurs considérés dans cette revue de littérature, à l'exception de Kim (2002). Bien qu'il soit plus souple que le logit standard, le logit emboîté impose encore des contraintes quant aux interrelations possibles entre les modalités⁷. À ce niveau, un des apports du présent travail est l'utilisation, pour le calcul de la probabilité de choix des véhicules, d'un modèle qui libère de toute contrainte de cette nature. Cet instrument est le logit mixte. L'emploi de ce dernier permettra de révéler, s'il y a lieu, les liens entre des modalités supposées jusqu'à maintenant indépendantes par les contraintes des modèles de choix utilisés. Une présentation détaillée du logit mixte et de ses conditions d'utilisation sera faite à la section 3.1.

La nécessité de modéliser conjointement le choix et l'usage des véhicules est induite

⁷Le logit emboîté ne permet pas non plus de capturer plusieurs formes d'hétérogénéité incluant, par exemple, l'hétérogénéité individuelle.

par la présence potentielle de biais d'autosélection. Ce type de biais apparaît lorsque les données avec lesquelles nous travaillons proviennent uniquement d'individus qui ont fait le choix d'appartenir au groupe étudié (Maddala, 1987). Par exemple, l'information que nous avons sur l'usage des petits véhicules provient de conducteurs qui ont choisi ce type de véhicule (plutôt qu'un autre) en fonction de ses caractéristiques et de l'usage qu'ils prévoyaient en faire. Le niveau d'utilisation peut alors dépendre des caractéristiques non observées des véhicules, incluses dans le terme d'erreur de l'équation d'usage, affectant ainsi la probabilité de choisir un véhicule particulier (Berkowitz *et al.*, 1990). Par exemple, les effets non observés qui ont tendance à augmenter l'usage (le plaisir de conduire) vont influencer négativement la probabilité de sélection d'un vieux véhicule inconfortable, qui risque d'être moins plaisant à conduire. Concrètement, ce problème se traduit par la présence de corrélation entre les termes d'erreur des équations de choix et d'usage. L'estimation de l'usage par une méthode de régression standard comme les moindres carrés ordinaires (MCO) peut alors présenter des problèmes de non convergence pouvant limiter les capacités inférentielles du modèle.

De façon générale, il existe trois méthodes pour pallier les problèmes d'autosélection : la méthode des variables instrumentales, la méthode de la forme réduite et la méthode de correction des attentes conditionnelles («selectivity corrections»)⁸. Cette dernière, qui fait l'objet d'un intérêt particulier dans le présent travail, sera présentée de façon détaillée à la section 3.4. Les trois méthodes conduisent à des estimateurs convergents en présence de corrélation entre le terme d'erreur de la fonction d'usage et celui la probabilité de choix du véhicule. De plus, selon Dubin et McFadden (1984), les trois méthodes produisent des élasticités quantitativement comparables et

⁸Pour un exemple de l'utilisation des trois méthodes de correction dans le cas d'appareils ménagers voir Dubin et McFadden (1984) et pour davantage d'information sur chacune des méthodes voir Wooldridge (2001), dans le cas des deux premières, et Maddala (1987), pour la dernière.

inférieures à celles obtenues par la méthode des MCO. Dans le type de modèle que nous préconisons, la correction se basant sur les attentes conditionnelles est toutefois la plus utilisée.

Train et Lohrer (1983) sont les seuls, parmi les articles retenus, à utiliser la méthode des variables instrumentales. Ils remplacent dans l'équation d'usage les variables qui caractérisent le type de véhicule choisi par un certain nombre de variables exogènes au modèle. Mannering et Winston (1985), Berkowitz *et al.* (1990) ainsi que Hensher, Milthorpe et Smith (1990) utilisent pour leur part la méthode de correction des attentes conditionnelles. La correction qu'ils utilisent se base sur celle développée par Dubin et McFadden (1984)⁹. Cette dernière fut à l'origine développée dans le cadre de travaux sur le choix et l'usage d'appareils ménagers.

Un apport intéressant pour la méthode de correction des attentes conditionnelles est celui de Hensher et Milthorpe (1987b). Dans leur article, les auteurs identifient dans un premier temps différentes situations pouvant conduire à la présence de biais de sélection lors d'étude sur le choix et l'usage de biens durable (ils présentent un exemple empirique basé sur le choix et l'usage de véhicules automobiles) et évaluent ensuite l'impact de deux corrections : celle développée par Dubin et McFadden (1984) et une autre proposée par Lee (1983). Hensher et Milthorpe (1987b) constatent que les deux formules de correction pour l'autosélection sont statistiquement significatives (comparativement à l'utilisation des MCO sans correction) dans le modèle d'usage, pour des ménages de un et deux véhicules. Toutefois, les corrections ne semblent pas apporter de gain de prédiction sur le niveau courant d'usage des véhicules. Toujours selon les mêmes auteurs, l'effet le plus significatif de la correction est sur certaines

⁹La correction utilisée par Hensher, Milthorpe et Smith (1990) est, selon les auteurs, tirée de Hay (1980), auquel nous n'avons malheureusement pas accès. Toutefois, dans un autre article (Hensher et Milthorpe, 1987b), les deux mêmes auteurs spécifient que cette méthode de correction provient de Hay (1980) et Dubin et McFadden (1984).

variables exogènes comme le prix du carburant par kilomètre. Le fait de ne pas corriger pour l'endogénéité du choix de véhicule sur l'usage pourrait alors conduire à de mauvaises politiques dues à l'utilisation d'estimations biaisées à la baisse ou à la hausse (selon la grandeur des parcs automobiles) de l'effet moyen du prix du pétrole. Ces conclusions sont en accord avec celles obtenues par Hensher, Milthorpe et Smith (1990). Dans leur cas, seul le prix du carburant pour les ménages à un véhicule est affecté de façon non-marginale par la variable de sélection¹⁰. Berkowitz *et al.* (1990) concluent pour leur part que les biais de sélection n'ont pas d'impacts sur l'estimation du modèle d'usage, qui dans leur cas exclut les déplacements pour se rendre au travail. Toutefois, selon Hensher, Milthorpe et Smith (1990), la non-significativité des biais de sélection sur l'estimation de l'usage n'est pas un argument pour rejeter l'approche théorique choisie pour prendre en compte ces biais. Elle est simplement une évidence empirique qui suggère que la relation entre le choix des véhicules et leur usage est adéquatement modélisée par l'inclusion dans le modèle d'usage des attributs observés des véhicules.

Cette revue de littérature nous a permis non seulement d'avoir une vision d'ensemble des principaux travaux qui ont contribué jusqu'à maintenant aux modèles discrets-continus appliqués au secteur du transport privé, mais aussi de comprendre la structure théorique globale de ce type de modèle. Un résumé des différents points considérés pour chacune des études est présenté à l'appendice A. Le chapitre qui suit aborde plus en détail les aspects méthodologiques de notre étude.

¹⁰Selon leurs résultats, pour les ménages qui possèdent un seul véhicule, le fait de corriger pour la sélection diminue de 15 % l'élasticité prix du carburant.

Chapitre 3

Méthodologie

Le présent chapitre traite dans un premier temps des modèles à choix discrets en général et plus spécifiquement de celui que nous utiliserons dans cette étude, le logit mixte. Les conditions pour l'identification des différents paramètres dans un tel cadre seront ensuite discutées. Suivra la présentation de la fonction d'utilité indirecte qui lie l'usage à la probabilité de choix des véhicules. Finalement, nous discuterons du terme de correction à inclure dans l'équation de la demande de service de transport pour tenir compte de la simultanéité entre les décisions de choix et d'usage. La méthodologie présentée est principalement tirée de Dubin et McFadden (1984), Hanneman (1984), Hensher, Milthorpe et Smith (1990), Ben-Akiva et Bolduc (1991), Bernard, Bolduc et Bélanger (1996), ainsi que de Train (2002).

En termes de modélisation, le problème auquel nous sommes confrontés est le suivant. L'agent n ($n = 1, \dots, N$) fait face à un ensemble spécifique¹ de choix C_n comprenant J_n modalités (dans notre cas des véhicules) et doit choisir la modalité i ($i = 1, \dots, J_n$) associée à un usage x_{in} qui maximise son utilité. Notons par U_{in}

¹Comme tous les modèles de véhicules présents dans la banque de données ne sont pas disponibles pour toutes les années (d'une année à l'autre de nouveaux modèles de véhicules apparaissent alors que la production d'autres cesse), l'ensemble C_n est spécifique à chaque agent selon l'année où il fait son choix.

l'utilité retirée par l'agent n de l'alternative i . Ces deux décisions (choix et usage) sont reconnues pour dépendre de facteurs communs. Le maximum de vraisemblance à information complète (MVIC) serait alors la manière optimale de maximiser la probabilité jointe du couple d'événements (i, x_{in}) . La structure de corrélation sous-jacente à un tel problème rend toutefois le MVIC presque impossible à estimer, particulièrement lorsque l'ensemble de choix compte un grand nombre d'alternatives. Néanmoins, la densité jointe peut être écrite comme le produit des densités conditionnelle et marginale $P(i|x_{in}) f(x_{in})$. En ignorant les relations structurelles qui relient les paramètres apparaissant dans $P(i|x_{in})$ et dans $f(x_{in})$, il est possible de travailler à l'aide d'une structure à deux niveaux, où l'agent n choisit dans un premier temps l'alternative i et dans un second temps, conditionnellement à l'alternative choisie, le niveau d'utilisation optimal \bar{x}_{in} (Bernard, Bolduc et Bélanger, 1996).

3.1 Modèle de choix

Nous avons survolé au chapitre 2 les forces et faiblesses majeures des deux principales classes d'instruments utilisés pour l'estimation des modèles à choix discrets. Avant d'aller plus en profondeur dans l'analyse de leurs attributs respectifs, assurons-nous de bien comprendre ce qui différencie ce type de modèles des autres modèles économétriques.

L'objectif des modèles à choix discrets est d'expliquer une variable dépendante à données non-continues. Ces dernières sont généralement des choix effectués parmi un ensemble de modalités ou d'alternatives (systèmes de chauffage, véhicules, plan d'interurbain, etc.) par des unités décisionnelles (individus, ménages, firmes, etc.). Selon Train (2002), l'ensemble de choix doit respecter trois conditions, à savoir : les

alternatives doivent être mutuellement exclusives du point de vue de l'unité décisionnelle, l'ensemble de choix doit être exhaustif et le nombre d'alternatives doit être fini. Ces conditions peuvent être, selon le cas, plus ou moins restrictives.

Du point de vue de la régression multiple standard, le traitement d'une variable dépendante discrète pose généralement 3 problèmes². Considérons la régression suivante :

$$y_n = x_n\beta + \mu_n,$$

où n représente un individu quelconque, $y_n = 1$ avec probabilité P_n et 0 avec probabilité $1 - P_n$, x_n est un vecteur de variables explicatives, β est un vecteur de paramètres à estimer et ε_n est un terme d'erreur. Le premier problème provient du fait que le côté gauche de la régression (y_n) est discret alors que le côté droit est continu. Comme y_n est discret, théoriquement, le terme d'erreur μ_n devrait être de moyenne :

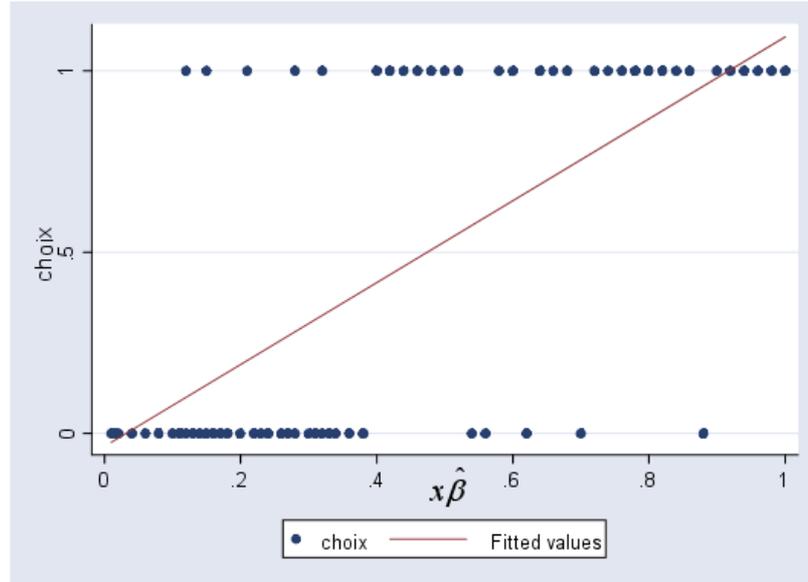
$$\begin{aligned} E(\mu_n) &= P_n(1 - x_n\beta) - (1 - P_n)x_n\beta \\ &= P_n - x_n\beta \\ &\equiv E(y_n) - x_n\beta. \end{aligned}$$

$E(\mu_n)$ sera de moyenne nulle si $P_n = x_n\beta$, condition qui sera généralement respectée si nous utilisons le bon estimateur. Le deuxième problème provient du fait que rien n'assure que $0 < x_n\beta < 1$, ce qui devrait être le cas puisque par définition P_n est entre 0 et 1. Finalement, de la même manière, si $\hat{\beta}$ est l'estimateur de β , rien n'impose que $0 < x_n\hat{\beta} < 1$. D'un point de vue graphique, l'utilisation de la régression linéaire standard pour l'estimation d'un choix binaire, basée par exemple sur $x_n\hat{\beta}$, pourrait nous mener à établir la relation fallacieuse illustrée à la figure 3.1.

Considérons maintenant le développement de la probabilité de sélection. Sous l'hypothèse de maximisation de l'utilité, l'agent n choisira la modalité i si et seulement

²Cette section est inspirée des notes du cours ECN-61694.

FIG. 3.1: Régression linéaire simple sur un choix binaire



si $U_{in} > U_{jn} \forall j \neq i$. Du point de vue du chercheur, U_{in} n'est pas connu, seul le choix ultime est observé. Considérons alors V_{in} , l'utilité estimée sur la base de X_{in} , un vecteur de caractéristiques de chacune des modalités telles que perçues par l'agent n et des caractéristiques de ce dernier. Comme V_{in} est estimé, nous aurons potentiellement $V_{in} \neq U_{in}$. La différence entre les deux quantités provient des facteurs qui influencent l'utilité, mais qui ne sont pas capturés dans V_{in} . Il est donc possible d'écrire $U_{in} = V_{in} + \mu_{in}$. Représenté ainsi, μ_{in} n'est que la différence entre U_{in} et V_{in} . Comme le chercheur ne connaît pas $\mu_{in} \forall i$, il considère ce terme comme aléatoire. Il n'y a cependant aucune raison a priori d'attribuer telle ou telle distribution à μ_{in} . Nous verrons plus loin que la spécification de μ_{in} déterminera la forme de la probabilité de choix. La probabilité

que l'individu n choisisse l'alternative i est :

$$\begin{aligned}
 P_n(i) &= P(U_{in} > U_{jn} \forall j \neq i) \\
 &= P(V_{in} + \mu_{in} > V_{jn} + \mu_{jn} \forall j \neq i) \\
 &= P(\mu_{jn} - \mu_{in} < V_{in} - V_{jn} \forall j \neq i).
 \end{aligned} \tag{3.1}$$

En spécifiant la densité jointe du vecteur aléatoire $\mu_n = (\mu_{1n}, \dots, \mu_{Jn})$, dénotée par $f(\mu_n)$, nous pouvons réécrire la probabilité (3.1) de la manière suivante :

$$P_n(i) = \int_{\mu} I(\mu_{jn} - \mu_{in} < V_{in} - V_{jn} \forall j \neq i) f(\mu_n) d\mu_n,$$

où $I(\cdot)$ est une fonction indicatrice qui vaut 1 lorsque l'expression entre parenthèse est vraie et 0 sinon. La probabilité (3.2) comporte une intégrale multi-dimensionnelle sur les valeurs admissibles du vecteur $f(\mu_n)$. Selon la loi postulée pour les μ_{in} , l'expression de cette probabilité prendra une forme explicite ou non. Considérons trois cas.

3.1.1 Probit

Si nous postulons les μ_{in} de loi normale, la probabilité (3.2) prend la forme d'un probit. Examinons un cas trichotomique où l'évaluation de l'alternative i , par l'agent n , est basée sur la fonction d'utilité :

$$U_{in} = V_{in} + \mu_{in}, \quad i = 1, 2, 3 \quad \mu_{in} \sim N(0, \sigma^2),$$

où $N(0, \sigma^2)$ dénote la distribution d'une loi normale d'espérance nulle et de variance σ^2 . Si l'utilité est linéaire dans les paramètres alors $V_{in} = X_{in}\beta$, où X_{in} a la même signification que précédemment et β est un vecteur de coefficients fixes. L'alternative 1 sera choisie si $U_{1n} > U_{2n}$ et $U_{1n} > U_{3n}$, ce qui se traduit en termes de probabilité

par :

$$\begin{aligned}
P_n(1) &= P(U_{1n} > U_{2n}, U_{1n} > U_{3n}) \\
&= P(X_{1n}\beta + \mu_{1n} > X_{2n}\beta + \mu_{2n}, X_{1n}\beta + \mu_{1n} > X_{3n}\beta + \mu_{3n}) \\
&= P(\mu_{2n} - \mu_{1n} < X_{1n}\beta - X_{2n}\beta, \mu_{3n} - \mu_{1n} < X_{1n}\beta - X_{3n}\beta). \quad (3.2)
\end{aligned}$$

Avec des termes d'erreur normaux, la probabilité de choix de l'alternative 1 prend la forme :

$$P_n(1) = \int_{-\infty}^{X_{1n}\beta - X_{2n}\beta} \int_{-\infty}^{X_{1n}\beta - X_{3n}\beta} f(v_{21,n}, v_{31,n}) dv_{21,n} dv_{31,n}, \quad (3.3)$$

où $v_{21,n} = \mu_{2n} - \mu_{1n}$ et $v_{31,n} = \mu_{3n} - \mu_{1n}$. Formulée de cette manière, la probabilité contient $J - 1$ niveaux d'intégrale, J étant le nombre d'alternatives de l'ensemble de choix. La réduction du nombre de niveaux d'intégrale de J à $J - 1$ est due à l'expression des termes d'erreur en différence.

Puisque le probit polytomique permet davantage de flexibilité dans la modélisation des interdépendances entre les modalités (voir chapitre 3), il est naturel d'opter a priori pour ce dernier. Cependant, comme la primitive associée à la loi normale n'a pas de forme explicite, l'estimation d'une probabilité comportant des intégrales doit nécessairement passer par l'évaluation numérique de celles-ci. Considérant que, pour une précision de 40 points³, le calcul numérique de trois niveaux d'intégrale demande 64 000 évaluations ($40 \times 40 \times 40$), l'évaluation numérique de J niveaux demande alors 40^J itérations. Cette procédure devient donc rapidement prohibitive en termes de temps de calcul à mesure que l'ensemble de choix grandit. Nous verrons à la section 3.1.3 qu'il est maintenant possible de contourner ce problème. Une autre contrainte des modèles probit tient au fait que les erreurs sont contraintes à suivre une distribution normale (Train, 2002).

³Il est possible d'évaluer numériquement une intégrale en effectuant une somme sur un certain nombre de points qui subdivisent l'intégrale.

3.1.2 Logit

Si nous postulons les μ_{in} de loi Gumbel *iid* (aussi appelée loi *iid* à valeurs extrêmes)⁴, la probabilité (3.2) conduit à un modèle logit de la forme :

$$P_n(i) = \int_{\mu_{in}} \left(\prod_{\substack{j \in C_n \\ j \neq i}} e^{-e^{-(\mu_{in} + V_{in} - V_{jn})}} \right) e^{-\mu_{in}} e^{-e^{-\mu_{in}}} d\mu_{in}.$$

L'expression entre parenthèses est en fait le produit des distributions cumulatives individuelles des μ_{in} , ces derniers étant *iid*. Quelques manipulations algébriques permettent de réduire cette expression à une formulation plus explicite⁵ :

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}}. \quad (3.4)$$

Ce résultat provient, entre autres, du fait que la différence de termes d'erreur Gumbel *iid* suit une distribution logistique. En reprenant le cas trichotomique de la section précédente, la probabilité (3.4) pour le choix de l'alternative 1 devient, dans le cas du logit :

$$P_n(1) = \frac{e^{X_{1n}\beta}}{e^{X_{1n}\beta} + e^{X_{2n}\beta} + e^{X_{3n}\beta}}.$$

Le principal avantage de la probabilité logistique est sa forme explicite (contrairement au probit, elle ne contient pas d'intégrale). Comme nous l'avons mentionné au chapitre 2, il faut toutefois rester conscient que le logit impose l'IAN. Rappelons que cette contrainte implique que le ratio des probabilités de choix de deux alternatives est invariant aux caractéristiques de toutes les autres alternatives. Par exemple, soit :

$$P_n(1) = \frac{e^{X_{1n}\beta}}{e^{X_{1n}\beta} + e^{X_{2n}\beta} + e^{X_{3n}\beta}} \quad \text{et} \quad P_n(2) = \frac{e^{X_{2n}\beta}}{e^{X_{1n}\beta} + e^{X_{2n}\beta} + e^{X_{3n}\beta}},$$

⁴La densité et la distribution cumulative d'une loi Gumbel *iid* pour une fonction d'utilité $U_{jn} = V_{jn} + \mu_{jn}$ sont respectivement :

$f(\mu_{jn}) = e^{-\mu_{jn}} e^{-e^{-\mu_{jn}}}$ et $F(\mu_{jn}) = e^{-e^{-\mu_{jn}}}$.

⁵Pour les détails du développement mathématique, voir Train (2002), p. 78-79.

alors

$$\frac{P_n(1)}{P_n(2)} = \frac{\frac{e^{X_{1n}\beta}}{e^{X_{1n}\beta} + e^{X_{2n}\beta} + e^{X_{3n}\beta}}}{\frac{e^{X_{2n}\beta}}{e^{X_{1n}\beta} + e^{X_{2n}\beta} + e^{X_{3n}\beta}}} = \frac{e^{X_{1n}\beta}}{e^{X_{2n}\beta}}.$$

3.1.3 Logit mixte

Le modèle préconisé dans la présente analyse est un mélange des deux modèles précédents. Le logit mixte, aussi appelé en anglais *mixed logit* ou *logit kernel*, dû à son noyau logistique, doit son caractère particulier aux deux termes d'erreur qui le composent. Le premier, de distribution Gumbel *iid*, donne la forme logistique à la probabilité centrale alors que le deuxième permet de modéliser la structure des interrelations entre les modalités. Le logit mixte comporte de nombreux avantages sur le probit et le logit standard ou même sur le logit emboîté, brièvement décrit au chapitre 2. Premièrement, il permet des variations aléatoires dans les goûts. Par exemple, des agents qui ont le même revenu peuvent accorder des valeurs différentes aux prix des véhicules. Deuxièmement, le logit mixte permet des profils de substitution non restreints. Le ratio des probabilités de sélection de deux alternatives ne dépend donc pas uniquement des caractéristiques de ces deux alternatives, mais bien des caractéristiques de l'ensemble des alternatives. Troisièmement, il permet la corrélation entre les facteurs non observés à travers le temps. Par exemple, si nous observons le choix de véhicule des individus à travers le temps, il est possible de spécifier un modèle pour lequel les facteurs non-observables, qui ont affecté le choix des véhicules aux périodes passées, influencent le choix à la période présente. Finalement, contrairement au probit, il n'est pas limité à une distribution normale des termes d'erreur. Regardons comment se construit la probabilité de choix du logit mixte dans le contexte d'un modèle où la corrélation est traitée selon une approche à erreurs composées. Pour ce

faire, considérons le cas trichotomique précédent. Soit :

$$y_{in} = \begin{cases} 1 & \text{si } U_{in} > U_{jn} \text{ pour } j = 1, 2, 3 \\ 0 & \text{sinon,} \end{cases}$$

et

$$\begin{aligned} U_{in} &= V_{in} + F_{in}\psi_n + \varepsilon_{in} \\ \varepsilon_{in} &\sim \text{Gumbel } iid, \\ \psi_n &\stackrel{iid}{\sim} NMV(0, \Sigma), \end{aligned} \tag{3.5}$$

où y_{in} représente le choix observé et F_n est une matrice constituée de 0 et de 1, qui définit la structure de corrélation entre les alternatives. Toutes les autres variables ont la même définition que précédemment. Dans un cas à trois alternatives, si nous suspectons la présence de corrélation entre la première et la troisième alternatives, correspondant respectivement à la première et la troisième ligne de la matrice F_n , celle-ci prendrait la forme suivante :

$$F_n = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}}_{(3 \times 2)}.$$

Cette méthode, aussi appelée méthode à facteurs, fut introduite par Ben-Akiva et Bolduc (1991). Nous avons choisi de poser ψ_n de loi normale, mais, dans d'autres circonstances, une autre loi de distribution aurait pu mieux convenir (log-normale, triangulaire, etc.)⁶. Avec le logit mixte, la probabilité que l'individu n choisisse l'alternative 1, telle qu'exprimée en (3.2), devient l'intégrale des probabilités logistiques standards sur tout le domaine de ψ_n , c'est-à-dire :

$$P_n(1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Lambda_n(1|\psi_n) f(\psi_n) d\psi_n,$$

⁶Pour une discussion détaillée des distributions possibles des erreurs et leurs contextes d'utilisations, voir Train (2002), chapitre 6.

où $f(\psi_n)$ est la densité multi-variée de $\psi_n = (\psi_{1n}, \psi_{2n}, \psi_{3n})$ et $\Lambda_n(1|\psi_n)$, la probabilité logistique standard :

$$\Lambda_n(1|\psi_n) = P_n(1|\psi_n) = \frac{e^{X_{1n}\beta + F_{1n}\psi_n}}{e^{X_{1n}\beta + F_{1n}\psi_n} + e^{X_{2n}\beta + F_{2n}\psi_n} + e^{X_{3n}\beta + F_{3n}\psi_n}}.$$

Cette formulation se généralise facilement à l'ensemble des J_n alternatives contenues dans C_n :

$$P_n(i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \Lambda_n(i|\psi_n) f(\psi_n) d\psi_n, \quad (3.6)$$

où

$$\Lambda_n(i|\psi_n) = \frac{e^{X_{in}\beta + F_{in}\psi_n}}{\sum_{j \in C_n} e^{X_{jn}\beta + F_{jn}\psi_n}}. \quad (3.7)$$

Notons dans la probabilité (3.6) l'intégrale non bornée de dimension égale au nombre de facteurs spécifiés dans la matrice F_n . Comme dans le cas du probit, l'évaluation numérique d'intégrales multiples devient rapidement prohibitive en termes de temps de calcul. Une solution à ce problème fut proposée par McFadden (1989) et Pakes et Pollard (1989). L'idée de base est de remplacer les probabilités (qui nécessitent le calcul de l'intégrale multi-dimensionnelle) par des simulations calculées à partir d'une variable latente du modèle. En termes concrets, cette méthode consiste premièrement à tirer des valeurs de ψ_n dans sa loi de distribution en notant par ψ_n^s le $s^{\text{ième}}$ tirage. La deuxième étape consiste à calculer les probabilités logistiques à l'aide des valeurs de ψ_n^s tirées. Il faut ensuite répéter les étapes une et deux un grand nombre de fois et effectuer une moyenne de l'ensemble des résultats recueillis. Nous obtenons ainsi la probabilité simulée :

$$\tilde{P}_n(i) = \frac{1}{S} \sum_{s=1}^S \Lambda_n(i|\psi_n^s), \quad (3.8)$$

où

$$\Lambda_n(i|\psi_n^s) = \frac{e^{X_{in}\beta + F_{in}\psi_n^s}}{\sum_{j \in C_n} e^{X_{jn}\beta + F_{jn}\psi_n^s}},$$

représente la probabilité de choix logistique pour une valeur donnée de ψ_n^s et S , le nombre total de tirages. Pour un nombre S suffisamment grand, $\tilde{P}_n(i)$ est un

estimateur convergent⁷ de $P_n(i)$.

Il est possible d'utiliser $\tilde{P}_n(i)$ dans le cadre du maximum de vraisemblance (MV). Nous parlons alors de maximum de vraisemblance simulé (MVS). Une autre méthode possible serait celle des moments simulés. Lorsque l'ensemble de choix utilisé est grand, comme dans notre cas, le MVS sera toutefois privilégié, dans la mesure où il nécessite uniquement le calcul des probabilités de choix des modalités sélectionnées⁸. En substituant $P_n(i)$ par $\tilde{P}_n(i)$ dans la fonction de log-vraisemblance des modèles à choix discrets, nous obtenons la log-vraisemblance simulée (LVS) :

$$LVS = \sum_{n=1}^N \sum_{i \in C_n} d_{in} \ln \tilde{P}_n(i), \quad (3.9)$$

où $d_{in} = 1$ si le ménage n choisit la modalité i et 0 sinon. L'estimateur du MVS est alors le vecteur de paramètres qui maximise la LVS⁹. Nous dénoterons ce dernier par $\tilde{\xi}_{MVS}$. Si S croît plus rapidement que le carré du nombre d'observations, alors le MV et le MVS deviennent des problèmes asymptotiquement identiques et les propriétés de l'estimateur du MV standard s'appliquent à l'estimateur du MVS.

3.2 Conditions d'identification

Les modèles économétriques, qu'ils soient discrets ou continus, imposent généralement des restrictions sur certains paramètres aux fins d'identification. Cette contrainte

⁷Cette propriété est due au fait que les moments empiriques convergent vers les moments théoriques lorsque $S \rightarrow \infty$.

⁸Pour plus de détails, voir Bolduc et Sanga (2002).

⁹Bien que $\tilde{P}_n(i)$ soit non biaisé pour $P_n(i)$, $\ln(\tilde{P}_n(i))$ est un estimateur biaisé de $\ln(P_n(i))$ pour un nombre de tirage S fini, dans la mesure où le simulateur induit un biais dans la LVS. Ce biais décroît toutefois à mesure que S augmente et disparaît de façon asymptotique lorsque S croît plus rapidement que la racine carrée du nombre d'observations (Brownstone et Train, 1999).

provient du fait que la solution (vecteur de paramètres) provenant du problème d'optimisation des équations du modèle n'est souvent pas unique. Cela se comprend facilement lorsqu'on réalise que dans \mathfrak{R}^n ($\forall n > 2$), la fonction objectif à son optimum est un hyperplan (Walker, Ben-Akiva et Bolduc, 2003). Cette section est consacrée à deux particularités des modèles de choix qui imposent certaines conditions à l'identification des paramètres.

Comme nous l'avons vu à la section 3.1, une alternative sera sélectionnée plutôt qu'une autre si l'utilité qu'elle rapporte à l'unité décisionnelle est strictement plus grande que celle procurée par toutes les autres alternatives. Ainsi, dans notre cas trichotomique des sections précédentes, le fait que l'agent n choisisse la modalité 1 revient à dire que $U_{1n} > U_{2n}$ et $U_{1n} > U_{3n}$. La probabilité générique (3.2) illustre cette condition. Ce cas simple révèle la première propriété affectant l'identification des modèles à choix discrets : *seules les différences d'utilité importent*. Il en découle que toute transformation monotone positive laissera le classement (ordinal) des utilités inchangé et par le fait même la probabilité de sélection. La généralisation de cette caractéristique nous amène à conclure que seuls les paramètres capturant des différences entre les modalités pourront être estimés (Train, 2002).

Regardons comment cette particularité affecte les composantes d'une version simplifiée de notre modèle :

$$\begin{aligned} U_{1n} &= \alpha_{1n}^* + X_{1n}\beta_1 + S_n\theta_1 + F_{1n}\psi_n + \varepsilon_{1n} \\ U_{2n} &= \alpha_{2n}^* + X_{2n}\beta_2 + S_n\theta_2 + F_{2n}\psi_n + \varepsilon_{2n} \\ U_{3n} &= \alpha_{3n}^* + X_{3n}\beta_3 + S_n\theta_3 + F_{3n}\psi_n + \varepsilon_{3n}, \end{aligned} \tag{3.10}$$

où α_{in}^* est une constante, X_{in} un vecteur $1 \times K$ de caractéristiques propres à la modalité $i = 1, 2, 3$, telles que perçues par l'agent n , β_i un vecteur $K \times 1$ de paramètres, S_n un vecteur $1 \times M$ de caractéristiques sociodémographiques de l'agent n et θ_i un

vecteur $M \times 1$ de paramètres. Les autres variables ont la même interprétation que précédemment. Que ce soit dans un modèle de choix ou dans une régression standard (continue), les constantes servent à capturer l'effet moyen de tous les facteurs qui ne sont pas inclus dans le modèle. Dans les modèles de choix toutefois, comme seules les différences d'utilité importent, il en va de même pour les constantes. Pour qu'elles soient identifiables, il faut donc en fixer l'échelle. Pour ce faire, la procédure usuelle consiste à normaliser l'une des constantes à 0. Sous cette condition, le modèle (3.10) devient :

$$\begin{aligned} U_{1n} &= \alpha_{1n} + X_{1n}\beta_1 + S_n\theta_1 + F_{1n}\psi_n + \varepsilon_{1n} \\ U_{2n} &= \alpha_{2n} + X_{2n}\beta_2 + S_n\theta_2 + F_{2n}\psi_n + \varepsilon_{2n} \\ U_{3n} &= X_{3n}\beta_3 + S_n\theta_3 + F_{3n}\psi_n + \varepsilon_{3n}, \end{aligned} \tag{3.11}$$

où $\alpha_{1n} = \alpha_{1n}^* - \alpha_{3n}^*$ et $\alpha_{2n} = \alpha_{2n}^* - \alpha_{3n}^*$. Ainsi, la première constante est interprétée comme étant l'effet moyen des facteurs non observables de l'utilité de la première modalité relativement à la troisième. L'interprétation de la deuxième constante est similaire. Un modèle de choix comprend donc toujours $J - 1$ (ici $3 - 1$) constantes identifiables et leur interprétation se fait par rapport à celle normalisée. Les X_{in} , c'est-à-dire les caractéristiques propres à chacune des modalités, ne nécessitent pas de traitement particulier. Comme les composantes de X_{in} varient d'une alternative à l'autre, les paramètres qui y sont associés capturent les différences dans l'utilité procurée par celles-ci et peuvent, par conséquent, tous être identifiés sans avoir à en fixer l'échelle (fixer à 0 chacun des paramètres de chacun des régresseurs de X_{in} pour l'une des modalités). Les variables sociodémographiques, pour leur part, peuvent s'exprimer de deux manières. Considérées seules, elles ne varient pas entre les alternatives et doivent donc subir le même traitement que les constantes. C'est-à-dire, en reprenant

le modèle (3.11) :

$$\begin{aligned} U_{1n} &= \alpha_{1n} + X_{1n}\beta_1 + S_n\theta_1 + F_{1n}\psi_n + \varepsilon_{1n} \\ U_{2n} &= \alpha_{2n} + X_{2n}\beta_2 + S_n\theta_2 + F_{2n}\psi_n + \varepsilon_{2n} \\ U_{3n} &= X_{3n}\beta_3 + F_{3n}\psi_3 + \varepsilon_{3n}. \end{aligned}$$

Il est aussi possible d'exprimer certaines caractéristiques sociodémographiques en relation avec des attributs propres à chaque modalité. Par exemple, nous pouvons diviser le coût des véhicules par le revenu des agents ou vice versa. Ainsi, nous captions les effets du revenu sans avoir à fixer l'échelle du paramètre qui y est associé puisqu'il est maintenant en relation avec une variable dont la valeur change entre les alternatives.

Considérons maintenant la seconde propriété qui affecte l'identification des paramètres des modèles de choix : *l'échelle absolue de l'utilité est arbitraire*. Comme n'importe quelle constante qui multiplie l'utilité n'influence pas le classement ordinal des alternatives, une infinité de valeurs d'utilité mènera à la même sélection. Il est donc nécessaire d'en fixer l'échelle. Pour ce faire, la procédure usuelle consiste à normaliser la variance du terme d'erreur. Le lien entre la normalisation de l'échelle de l'utilité et la normalisation de la variance du terme d'erreur peut sembler à première vue difficile à établir, mais en y regardant de plus près, nous constatons que lorsqu'une constante τ multiplie l'utilité, la variance du terme d'erreur varie de τ^2 : $Var(\tau\varepsilon_{in}) = \tau^2 Var(\varepsilon_{in})$. Fixer l'échelle de la variance revient donc à fixer l'échelle de l'utilité. Reprenons le modèle (3.10) et supposons, pour simplifier la notation, que la matrice F est une matrice identité, c'est-à-dire qu'il y a autant de facteurs que de modalités. Dans ce cas, si ψ_{in} est *iid* et suit une loi normale d'espérance nulle et de variance σ^2 et sachant que $Var(\varepsilon_{in}) = \pi^2/6$ (par convention, la variance d'un terme de loi Gumbel *iid* est normalisée à $\pi^2/6$), alors $Var(\psi_{in} + \varepsilon_{in}) = E(\psi_{in} + \varepsilon_{in})^2 = \pi^2/6 + \sigma^2$. Comme tous les termes d'erreur ont la même variance (ψ et ε sont tous les deux *iid*), normaliser la variance de l'un revient à normaliser la variance de tous. Notre modèle

prend alors la forme :

$$U_{in} = \frac{\alpha_{in}}{\sqrt{\frac{\pi^2}{6} + \sigma^2}} + X_{in} \frac{\beta_i}{\sqrt{\frac{\pi^2}{6} + \sigma^2}} + S_n \frac{\theta_i}{\sqrt{\frac{\pi^2}{6} + \sigma^2}} + \psi_{in} + \varepsilon_{in},$$

avec $Var(\psi_{in} + \varepsilon_{in}) = 1$. Il est important de remarquer que l'interprétation des paramètres se fait alors relativement à l'écart type du terme d'erreur. Lorsque la loi de distribution des composantes non-observées est *iid*, ce type de normalisation ne pose généralement pas de problème. Toutefois, si notre modèle avait permis la présence d'hétéroscédasticité, la procédure de normalisation aurait été un peu plus complexe. En effet, dans ce cas, comme chacun des termes d'erreur a potentiellement une variance propre, le simple fait de normaliser l'échelle de l'un de ces termes n'est pas suffisant pour fixer l'échelle de l'utilité¹⁰.

3.3 Fonction d'utilité indirecte

Afin de conserver une certaine cohérence économique entre la décision de choix du véhicule et son utilisation, nous dérivons la demande de kilométrage en appliquant l'identité de Roy à la fonction d'utilité indirecte conditionnelle à chacune des alternatives considérées. Cette fonction se définit comme suit :

$$G_{in} \equiv u_{in}[B_i, P_i, Y_n, S_n] + \epsilon_{in},$$

¹⁰Voir Train (2002) pour plus de détails sur la procédure de normalisation dans un tel cas.

où :

$u_{in} \equiv$ utilité retirée du véhicule i par le ménage n ;

$B_i \equiv$ vecteur de caractéristiques du véhicule i ;

$P_i \equiv$ coût en capital du véhicule i ;

$Y_n \equiv$ revenu annuel du ménage n ;

$S_n \equiv$ vecteur de caractéristiques sociodémographiques du ménage n ;

$\epsilon_{in} \equiv$ terme d'erreur.

En ce qui a trait à la forme fonctionnelle de la fonction d'utilité indirecte, nous avons discuté au chapitre 2 des différentes raisons pour utiliser telle ou telle structure. Dans le présent travail, étant donné la taille du modèle considéré, nous préconisons une approximation linéaire de la fonction d'utilité qui nous permet d'obtenir une équation de demande linéaire dans les prix et le revenu. De plus, nous supposons que u_{in} possède toutes les propriétés standards des fonctions d'utilité indirecte. Plus précisément, u_{in} est quasiconvexe, décroissante en P_i et croissante en Y_n (Jehle et Reny, 2000). Mais, à notre égard, la propriété la plus importante de la fonction d'utilité indirecte conditionnelle est le fait qu'elle satisfait l'identité de Roy. Cette propriété nous permet de retrouver (par le biais du théorème de l'enveloppe) la demande Marshallienne (usage) pour chaque véhicule sélectionné, c'est-à-dire :

$$x_{in}(B_i, P_i, Y_n, S_n) = -\frac{\partial u_{in}(B_i, P_i, Y_n, S_n)/\partial P_i}{\partial u_{in}(B_i, P_i, Y_n, S_n)/\partial Y_n}. \quad (3.12)$$

Cette expression nous donne la demande conditionnelle de service de transport du véhicule i par le ménage n . En termes opérationnels, cette équation d'usage s'écrit :

$$g_{in} = Z_{in}\gamma + \omega_{in}, \quad (3.13)$$

où g_{in} représente la demande de kilométrage de l'individu n pour la modalité i , Z_{in} un vecteur ($1 \times T$) de caractéristiques de chacune des modalités telles que perçues par

l'agent n et des caractéristiques de ce dernier, γ un vecteur $(T \times 1)$ de paramètres et ω un terme d'erreur. Ce dernier terme est introduit pour prendre en compte le fait que les préférences des ménages sont observées de façon incomplète par le modélisateur. Considérons maintenant la correction nécessaire pour tenir compte de la simultanéité des décisions de choix et d'usage.

3.4 Correction pour l'estimation de la demande de kilométrage

Au chapitre 2, nous avons introduit les problèmes de non convergence pouvant résulter de l'estimation de l'usage d'un bien durable à l'aide de méthodes économétriques standards, telles que les MCO. Ces problèmes proviennent de la simultanéité des décisions de choix et d'usage du bien en question. Ainsi, dans un modèle discret-continu, comme les mêmes facteurs non-observables peuvent apparaître à chacun des niveaux, il est possible que leurs termes d'erreur respectifs soient corrélés (ε_{in} dans (3.5) et ω_{in} dans (3.13)). Dans la famille des méthodes de correction des attentes conditionnelles, méthodes privilégiées dans la présente étude, il existe deux approches pour pallier ce type de problème. La première, dite à information complète, consiste à estimer en une seule étape à la fois le choix i_n et l'usage g_{in} . Comme cette méthode est généralement complexe et peut conduire à d'importants biais en cas de mauvaise spécification, nous laisserons aux lecteurs intéressés le soin d'approfondir la question¹¹. La seconde approche, dite en deux étapes, est plus couramment utilisée étant donné sa relative simplicité. Elle fut d'abord proposée par Heckman (1979) dans le cas d'un probit dichotomique et reprise quelques années plus tard par Dubin et McFadden (1984) dans le cas d'un logit polytomique. Cette méthode consiste dans un premier temps à calculer la probabilité de choix de chacune des alternatives

¹¹Voir, par exemple, Sanga (1999).

(composante discrète) et dans un second temps à estimer l'équation d'usage de l'alternative sélectionnée (composante continue), en conditionnant sur sa probabilité de choix. Il est ainsi possible de prendre en compte la corrélation potentielle entre le terme d'erreur de chacune des équations.

Si notre modèle de choix était un logit standard, c'est-à-dire si ψ_n était omis dans (3.6), l'équation d'usage (3.13) augmentée de la correction s'écrirait :

$$g_{in} = Z_{in}\gamma + E(\omega_n | i_n) + \varsigma_{in}. \quad (3.14)$$

Par la loi des espérances itératives, nous avons que $E(\omega_n | i_n) = E_{\varepsilon_n | i_n} E(\omega_n | \varepsilon_{jn})$, ce qui revient à écrire

$$E(\omega_n | i_n) = \frac{1}{\Lambda_n(i_n)} \int_{A_{in}} E(\omega_n | \varepsilon_{jn}) f(\varepsilon) d\varepsilon, \quad (3.15)$$

où A_{in} représente l'ensemble des valeurs possibles de ε cohérentes avec le choix i . Rappelons que $\Lambda_n(i_n)$ est la probabilité logistique (3.7) pour un ψ_n connu, que ε_{jn} est de loi Gumbel *iid* et ω_n de loi normale. En se basant sur la démarche de Dubin et McFadden (1984), posons la distribution de ω_n conditionnelle aux $(\varepsilon_{1n}, \dots, \varepsilon_{Jn})$ de moyenne $(\sigma\sqrt{2}/\lambda) \sum_{j \in C_n} R_j \varepsilon_{jn}$ et de variance $\sigma^2(1 - \sum_{j \in C_n} R_j^2)$, où $\sum_{j \in C_n} R_j = 0$ et $\sum_{j \in C_n} R_j^2 < 1$. R_j est alors la corrélation entre ω_n et ε_{jn} . En substituant $E(\omega_n | \varepsilon_{jn})$ dans (3.15), nous obtenons :

$$\begin{aligned} E(\omega_n | i_n) &= \frac{1}{\Lambda_n(i_n)} \int_{A_{in}} \frac{\sigma\sqrt{2}}{\lambda} \sum_{j \in C_n} R_j \varepsilon_{jn} f(\varepsilon) d\varepsilon \\ &= \frac{\sigma\sqrt{2}}{\lambda} \sum_{j \in C_n} R_j E(\varepsilon_{jn} | i_n). \end{aligned} \quad (3.16)$$

La valeur espérée de ε_{jn} , sachant que l'alternative i a été sélectionnée par l'agent n , doit satisfaire

$$E(\varepsilon_{jn} | i_n) = \begin{cases} -\ln \Lambda_n(i_n) \cdot \frac{\lambda\sqrt{3}}{\pi} & \text{si } j = i_n \\ \frac{\Lambda_n(j)}{1-\Lambda_n(j)} \ln \Lambda_n(j) \cdot \frac{\lambda\sqrt{3}}{\pi} & \text{si } j \neq i_n, \end{cases}$$

ce qui, une fois substituée dans (3.16), nous donne

$$\begin{aligned} E(\omega_n | i_n) &= \frac{\sigma\sqrt{6}}{\pi} \left[\sum_{j \in C_n} R_j \frac{\Lambda_n(j)}{1 - \Lambda_n(j)} \ln \Lambda_n(j) - R_{i_n} \frac{\ln \Lambda_n(i_n)}{1 - \Lambda_n(i_n)} \right] \\ &= \sum_{j \in C_n} \frac{\sigma\sqrt{6}R_j}{\pi} \left[\frac{\ln \Lambda_n(j)}{1 - \Lambda_n(j)} \right] \cdot [\Lambda_n(j) - \delta_{ji_n}], \end{aligned}$$

où $\delta_{ji_n} = 1$ si $j = i_n$. Sachant que $\sum_{\substack{j \in C_n \\ j \neq i_n}} R_j + R_{i_n} = 0$, nous obtenons, après quelques manipulations, la forme finale de notre correction :

$$E(\omega_n | i_n) = \sum_{\substack{j \in C_n \\ j \neq i_n}} \left[\frac{\sigma\sqrt{6}R_j}{\pi} \right] \left[\frac{\Lambda_n(j) \ln \Lambda_n(j)}{1 - \Lambda_n(j)} + \ln \Lambda_n(i_n) \right].$$

Dans le cas du logit mixte, en conditionnant explicitement sur une valeur de ψ_n donnée, ce terme de correction s'écrit :

$$E(\omega_n | i_n, \psi_n) = \sum_{\substack{j \in C_n \\ j \neq i_n}} \left[\frac{\sigma\sqrt{6}R_j}{\pi} \right] \left[\frac{\Lambda_n(j|\psi_n) \ln \Lambda_n(j|\psi_n)}{1 - \Lambda_n(j|\psi_n)} + \ln \Lambda_n(i_n|\psi_n) \right].$$

Comme ψ_n est inconnu du point de vue du modélisateur, pour avoir l'espérance inconditionnelle de ω_n , il faut intégrer sur tout le domaine de ψ_n , c'est-à-dire¹² :

$$E(\omega_n | i_n) = \int_{\psi} E(\omega_n | i_n, \psi_n) f(\psi) d(\psi). \quad (3.17)$$

Posons

$$\sum_{\substack{j \in C_n \\ j \neq i_n}} (\cdot) \equiv \sum_{j=1}^J \kappa_{jn}(\cdot),$$

où

$$\kappa_{jn} = \begin{cases} 1 & \text{si } j \in C_n \wedge j \neq i_n \\ 0 & \text{sinon.} \end{cases}$$

¹²Cette expression fut initialement proposée par Moyneur (2002).

Un simulateur convergent de (3.17) est alors :

$$\begin{aligned}
\frac{1}{S} \sum_{s=1}^S E(\omega_n | i_n, \psi_n^s) &= \frac{1}{S} \sum_{s=1}^S \sum_{j=1}^J \kappa_{jn} \left[\frac{\Lambda_n(j|\psi_n^s) \ln \Lambda_n(j|\psi_n^s)}{1 - \Lambda_n(j|\psi_n^s)} + \ln \Lambda_n(i_n|\psi_n^s) \right] \left[\frac{\sigma \sqrt{6} R_j}{\pi} \right] \\
&= \sum_{j=1}^J \kappa_{jn} \left[\frac{1}{S} \sum_{s=1}^S \left[\frac{\Lambda_n(j|\psi_n^s) \ln \Lambda_n(j|\psi_n^s)}{1 - \Lambda_n(j|\psi_n^s)} + \ln \Lambda_n(i_n|\psi_n^s) \right] \frac{\sigma \sqrt{6}}{\pi} \right] R_j \\
&= \sum_{j=1}^J W_{jn} R_j \\
&= W_n R,
\end{aligned}$$

où l'indice «s» indique le $s^{\text{ième}}$ tirage de ψ_n . Rappelons que les R_j sont des termes estimables qui représentent les corrélations entre les termes d'erreur du modèle de choix et celui du modèle d'usage. Le correcteur est estimé en remplaçant β dans $\Lambda_n(j|\psi_n^s)$ par $\tilde{\xi}_{MVS}$, le vecteur de paramètres qui maximise la log-vraisemblance simulée des probabilités du logit mixte. Dénotons par $E(\omega_n | i_n, \widehat{\tilde{\xi}_{MVS}}) = \hat{W}_n R$ l'expression ainsi obtenue. Substituée dans l'équation (3.14), cette correction nous permet d'obtenir un modèle pour l'usage estimable par des méthodes simples telles que les MCO :

$$g_{in} = Z_{in} \gamma + \hat{W}_n R + \zeta_{in}. \quad (3.18)$$

Sous cette forme, ζ est indépendant des régresseurs et le modèle satisfait les hypothèses du modèle linéaire classique. Précisons en terminant que le vecteur \hat{R} provenant de l'estimation MCO de (3.18) est un estimateur convergent de R .

Chapitre 4

Données

Les données utilisées proviennent de l'*Enquête sur le kilométrage des conducteurs et conductrices du Québec*, effectuée par la Société de l'assurance automobile du Québec (SAAQ) en 1996-1997. L'objectif de cette enquête est de fournir les renseignements nécessaires à l'analyse du phénomène des accidents de la route. Il nous est toutefois possible d'y puiser des informations utiles à notre étude. Au total, des 19304 questionnaires initialement envoyés, 7225 ont été retournés et sont valides aux fins d'analyse, pour un taux de réponse global de 38,4 %. De plus, une enquête téléphonique complémentaire a été effectuée pour le sous-échantillon des non répondants.

L'enquête postale contenait deux parties. La première interrogeait le répondant sur un certain nombre de ses caractéristiques sociodémographiques dont l'âge, le revenu annuel du ménage, le niveau de scolarité, le lieu d'habitation, le nombre d'enfants dans le ménage, etc. La seconde partie prenait la forme d'un carnet de déplacements que le conducteur visé par l'enquête devait remplir pendant deux jours¹. Parmi les renseignements demandés sur la base de chaque déplacement figurent : le kilométrage indiqué sur l'odomètre avant et après, le but principal, la marque, le modèle et l'année

¹Certains répondants ont toutefois rempli le questionnaire pendant une seule journée.

du véhicule utilisé, la date complète, l'heure, etc.

La première étape de préparation de la banque de données fut la validation des modèles de véhicules. En se basant sur le *Guide de l'auto* ainsi que sur le *Guide d'évaluation Hebdo* des voitures usagées, la correspondance entre les marques et les modèles a pu être établie. La validité des années de fabrication a aussi été vérifiée. Par exemple, les véhicules achetés à la fin d'une année sont souvent des modèles de l'année suivante. Par conséquent, il arrive à l'occasion que des propriétaires de véhicules achetés neufs rapportent des modèles une année en avance.

Afin de considérer le plus grand nombre de facteurs susceptibles d'influencer le choix et l'usage des véhicules, des informations supplémentaires ont été ajoutées à la banque de données originale de la SAAQ. Parmi celles-ci figurent le prix et le taux de consommation moyen de carburant des véhicules sélectionnés. Les prix proviennent du *Guide de l'auto* pour les années de véhicules considérées, soit de 1986 à 1997. Ils sont obtenus, à l'exception de 1986², à partir d'une moyenne des bornes inférieure et supérieure d'un intervalle de prix fourni par le guide pour chaque marque, modèle et année de véhicule. Les taux de consommation estimés en laboratoire proviennent du *Guide de consommation de carburant* (GCC), publié tous les ans par Ressources naturelles Canada. Étant donné le niveau de désagrégation élevé du classement des véhicules de ce guide, il est impossible d'effectuer une fusion directe de son contenu avec les données de la SAAQ, beaucoup moins détaillées.

Pour des raisons de commodité, la fusion a été établie sur la base de trois caractéristiques principales du véhicule. Précisément, les valeurs de taux de consommation de carburant que nous avons utilisées sont des moyennes pour chaque véhicule de même marque, modèle et année de fabrication contenu dans le GCC. De plus, le taux

²En 1986, un seul prix est fourni par le *Guide de l'auto*.

de consommation de carburant utilisé est un taux combiné constitué de 55 % du taux de consommation en ville et de 45 % du taux de consommation sur la grande route.

Dans toutes les études que nous avons recensées, l'estimation des modèles de choix de véhicules est faite sur la base de classes de véhicules plutôt que sur la base de chaque marque et modèle. Cette situation découle historiquement de la présence de deux contraintes. Premièrement, l'estimation d'un modèle de choix avec un très grand nombre de modalités (dans le présent cas 230) représentait un fardeau relativement important en termes de temps de calcul. Deuxièmement, le manque de raffinement des instruments d'estimation eux-mêmes ont souvent forcé les chercheurs à agréger les modalités en classes, soit pour assurer le respect de l'IAN (logit, GEV) ou pour augmenter la variabilité de certains attributs entre les modalités (classes). Cette limitation était particulièrement critique lorsque les choix étudiés présentaient des structures de corrélation complexes et des caractéristiques similaires. Ces deux contraintes sont aujourd'hui grandement relâchées d'une part, par l'augmentation constante de la puissance de calcul des ordinateurs, qui permet d'obtenir de bonnes estimations en un laps de temps raisonnable et d'autre part, par le développement de modèles plus flexibles tels que le logit mixte. Mannering et Winston (1985) ont testé la présence possible de violations de l'IAN dans le cadre d'un modèle logit pour 10 classes de véhicules. À l'époque, les résultats de leur test n'ont pas permis de rejeter l'hypothèse nulle de la structure du logit multinomial. Toutefois, l'évolution incessante de la gamme de véhicules offerte aux consommateurs nous porte à croire qu'une modélisation libérée de toutes contraintes sur la structure de corrélation entre les modalités permettrait de révéler des liens jusqu'à maintenant ignorés.

Nous avons tout de même regroupé les modèles en classes, et ce pour la raison suivante. Bien que nous estimons la probabilité de sélection de chacun des 230 modèles de véhicules, il n'est pas d'un grand intérêt, dans notre cas, d'utiliser des paramètres

spécifiques à chacun d'eux afin d'estimer l'impact des variables sociodémographiques. Par exemple, il est peu pertinent de savoir si l'âge influence différemment la probabilité de sélection d'une Nissan Maxima et d'une Honda Accord. Les paramètres associés aux variables sociodémographiques sont donc spécifiques à des classes de véhicules plutôt qu'à chaque véhicule.

Il n'existe pas de méthodes formelles pour effectuer des regroupements de véhicules (Hensher, 1985). Selon l'objectif poursuivi, différents classements sont possibles. Par exemple, Ressources naturelles Canada regroupe les voitures selon un indice d'espace intérieur³. Cette catégorisation ne correspond toutefois pas exactement à nos besoins. En effet, nous voulons que des véhicules possédant des caractéristiques similaires de consommation de carburant et de prix se retrouvent dans une même classe. Cette volonté est nourrie par l'intuition que le prix et le taux de consommation de carburant sont des facteurs prépondérants dans la sélection d'un véhicule. La formation des classes doit donc refléter une certaine homogénéité des caractéristiques de prix et de consommation (spécialement pour les voitures, par opposition aux véhicules à usages spéciaux, qui sont classés dans leur catégorie respective). S'il existe une certaine correspondance entre l'espace intérieur et le taux de consommation de carburant (plus un véhicule possède d'espace intérieur, plus il est susceptible d'être volumineux et plus sa consommation de carburant aura tendance à être élevée), le lien est beaucoup moins clair entre l'espace intérieur et le prix. Par exemple, le Toyota Camry et la BMW 318i, qui pourraient potentiellement se retrouver dans la même classe sur la base de leur espace intérieur respectif, sont loin d'appartenir à la même catégorie de prix. Pour cette raison, nous avons utilisé pour regrouper les véhicules la classification établie par le *Guide de l'auto 1997*, qui considère aussi le prix comme facteur distinctif. Cette

³Cet indice correspond au volume total de l'habitacle et du coffre ou de l'espace utilitaire. De plus, les camionnettes, fourgonnettes et véhicules à usage spéciaux forment des catégories à part entière. Pour plus de renseignement sur la classification de Ressources naturelles Canada, voir le site <http://oee.nrcan.gc.ca/vehicules/econergetiques/econergetiques.cfm>.

classification comporte seize catégories de véhicules. Pour que le modèle admette un nombre suffisant de degrés de liberté, nous avons réduit ce nombre à neuf : minis voitures, sous-compactes, compactes, intermédiaires, grosses berlines (comprenant aussi les familiales et les berlines de luxe), voitures sport, fourgonnettes, véhicules utilitaires sport et camionnettes.

Pour leur part, les répondants ont été regroupés selon la population de leur municipalité de résidence. Plus la population d'une ville est importante, plus il est facile d'offrir un service de transport en commun de qualité. Par conséquent, le fait de connaître la population d'une région donnée permet d'avoir une idée de son accessibilité au transport en commun. Par exemple, les résidents de la ville de Montréal jouissent d'une très grande disponibilité des transports en commun, disponibilité sans doute nulle ou très faible à St-Nicéphore. Bien que nous n'étudions pas le choix de posséder ou non un véhicule (l'accessibilité au transport en commun étant un déterminant important de cette décision), nous croyons que la disponibilité des transports en commun peut influencer l'achat de certains types de véhicules comme les mini-fourgonnettes. De plus, la taille d'une population étant souvent positivement corrélée à sa densité, nous croyons qu'une population élevée peut affecter négativement la décision d'opter pour de gros véhicules plus encombrants. Finalement, l'importance de la population peut aussi refléter, dans une certaine mesure, le niveau d'entretien des routes (particulièrement durant la période hivernale) et ainsi influencer la probabilité d'acheter des véhicules utilitaires sport ou des camionnettes.

Le premier classement des lieux d'habitation comportait 7 regroupement : Montréal, ceinture de Montréal, très grandes villes (100 000 habitants et plus), grandes villes (entre 60 000 et 99 999 habitants), moyennes villes (entre 20 000 et 59 999 habitants), petites villes (entre 5000 et 19 999 habitants) et zones rurales (entre 1 et 4999 habitants). Après quelques tests toutefois, cette catégorisation s'est révélée

trop désagrégée. Nous avons donc opté pour les regroupements plus généraux suivants : Montréal, ceinture de Montréal, autres villes et zones rurales. L'information sur la population de chaque municipalité en 1997 a été obtenue grâce au répertoire des municipalités du Québec.

La variable du revenu annuel des ménages, du nombre d'enfants pour chaque catégories d'âge (6 ans et moins, 7 à 15 ans et 16 ans et plus), du nombre total d'individus au sein du ménage, du nombre de détenteurs de permis de conduire et du nombre de véhicules que possède un ménage comportent des valeurs manquantes qui sont imputées. Les revenus manquants sont estimés par la méthode des MCO à partir d'une spécification inspirée de Lyssiotou, Pashardes et Stengos (2001). Parmi les variables explicatives utilisées figurent entre autres : le niveau de scolarité, l'âge et l'occupation du répondant (temps plein, temps partiel, temporaire, etc.), ainsi que le nombre d'enfants et le nombre de véhicules du ménage. Seuls les revenus manquants sont imputés par les revenus estimés⁴. Le nombre d'enfants par catégorie d'âge est imputé sur la base du nombre d'enfants moyen par catégorie d'âge du répondant (16-19, 20-24, 25-44, 45-64, 65 et plus), de même pour le nombre total de membres dans le ménage. Le nombre de détenteurs de permis de conduire ainsi que le nombre de véhicules possédés par le ménage sont quant à eux imputés sur la base d'une moyenne par nombre total de membres du ménage.

Deux derniers aspects des données restent à considérer. Premièrement, les informations contenues dans la banque de données ne nous permettent pas de différencier les véhicules achetés neufs de ceux achetés usagés. Pour pallier ce manque d'information, nous avons fait l'hypothèse que tous les véhicules ont été achetés neufs. L'année

⁴Il n'existe pas de véritable consensus à savoir s'il est préférable de remplacer toutes les observations d'une variable par les valeurs estimées correspondantes (les valeurs imputées étant estimées avec erreur, le modélisateur s'assure ainsi que toutes les valeurs sont affectées de la même erreur) ou simplement remplacer les valeurs manquantes. Notre choix est basé sur la meilleure performance de la variable revenu lorsqu'uniquement les valeurs manquantes sont imputées.

de fabrication et d'achat des véhicules sont donc considérées équivalentes. Deuxièmement, comme l'enquête de la SAAQ nous informe sur le profil sociodémographique des individus en 1996-97, il y a fort à parier que leur situation a changé entre la date d'achat du véhicule et la date de l'enquête (nombre et âge des enfants, revenu annuel du ménage, dernier niveau de scolarité, etc.). L'impact de certaines variables sociodémographiques sur la probabilité de choix et l'intensité d'utilisation est donc à considérer avec prudence, spécialement pour les véhicules plus âgés. Nous avons envisagé d'autres alternatives à l'hypothèse de l'achat des véhicules neufs. Par exemple, supposer que les ménages doivent racheter en 1997 les mêmes véhicules que ceux rapportés dans la banque de données. Le fait que de nombreux modèles fabriqués entre 1986 et 1996 ne sont plus disponibles en 1997 exclut toutefois cette solution. Afin d'améliorer la concordance entre la date d'achat des véhicules et le profil sociodémographique des répondants, il sera éventuellement possible d'appliquer des «facteurs de changement» sur certaines variables comme le nombre d'enfants par groupe d'âge, le revenu annuel du ménage et le dernier niveau de scolarité atteint par le répondant.

Chapitre 5

Résultats

5.1 Modèle de choix

En raison des contraintes que les modèles à choix discrets imposent sur l'identification de leurs paramètres (voir section 3.2), l'interprétation de ceux-ci est souvent moins directe que dans le cas des modèles économétriques conventionnels. Afin de contourner ce problème, nous avons procédé par simulation. La procédure consiste à modifier de façon marginale la base de données utilisée pour estimer le modèle, par exemple en augmentant le prix d'une classe de véhicules en particulier, et à utiliser les estimations obtenues pour inférer l'effet de ce changement sur la probabilité de sélection de chaque classe de véhicules. Il est ainsi plus facile d'exploiter toute l'information contenue dans les paramètres pour isoler l'impact d'un changement dans l'une des variables explicatives du modèle. Les variables dont les valeurs changent entre les alternatives n'étant pas soumises aux contraintes d'identification, il est possible d'interpréter directement leur paramètres, sans passer par la simulation. Ainsi, les paramètres significatifs associés aux variables du log du revenu divisé par le prix des véhicules (Log rev/Prix véhic), rapportés au tableau 5.1, ont presque tous le signe escompté, c'est-à-dire un signe positif pour les gros véhicules et un signe négatif pour les petits véhicules. Ainsi,

TAB. 5.1: Estimations du modèle de choix

Variables	Estimations logit	Statistiques t	Estimations logit mixte	Statistiques t
Log rev/Prix véhic				
Minis Voitures	-0.185	-8.578	-0.382	-7.801
Sous-compactes	-0.132	-6.607	-0.034	-1.134
Compactes	0.091	3.551	0.083	2.020
Intermédiaires	0.125	3.311	0.081	1.527
Grosses Berlines	0.565	11.569	0.566	9.243
Voitures sport	0.171	2.494	0.138	1.358
Mini-fourgonnettes	0.021	0.266	-0.113	-1.031
VUS	0.037	0.674	-0.012	-0.189
Camionnettes	-0.082	-1.170	-0.170	-2.361
Conso carburant				
Minis Voitures	-0.008	-0.479	-0.111	-2.902
Sous-compactes	-0.049	-2.739	-0.022	-1.011
Compactes	-0.108	-3.988	-0.070	-2.292
Intermédiaires	-0.148	-5.572	-0.184	-3.540
Grosses Berlines	-0.141	-3.578	-0.145	-3.436
Voitures sport	-0.103	-1.808	-0.118	-2.011
Mini-fourgonnettes	-0.098	-2.094	-0.128	-1.998
VUS	-0.086	-2.301	-0.109	-2.539
Camionnettes	-0.056	-0.886	-0.065	-0.842

les petits véhicules jouent le rôle de biens inférieurs, dans la mesure où, toutes choses étant égales par ailleurs, un accroissement du revenu annuel du ménage entraîne une perte d'utilité pour les détenteurs de tels véhicules. Pour les autres classes de véhicules, une augmentation du revenu conduit à un gain d'utilité. Les paramètres associés aux prix des véhicules s'interprètent de manière inverse. Le signe négatif associé à la classe des camionnettes est toutefois à considérer avec plus de prudence. À notre avis, les camionnettes ont potentiellement une dimension utilitaire plus importante que toutes les autres classes de véhicules considérées (capacité de chargement de matériaux en vrac, puissance, etc.). Il est donc possible que leur demande soit affectée de manière

différente par le rapport entre le revenu annuel du ménage et le prix comparativement aux autres classes. Fait à noter, le paramètre associé à la variable de prix n'est pas significatif pour les VUS. Des politiques visant à enrayer l'engouement pour ce type de véhicules en augmentant son prix pourraient donc ne pas avoir l'effet souhaité.

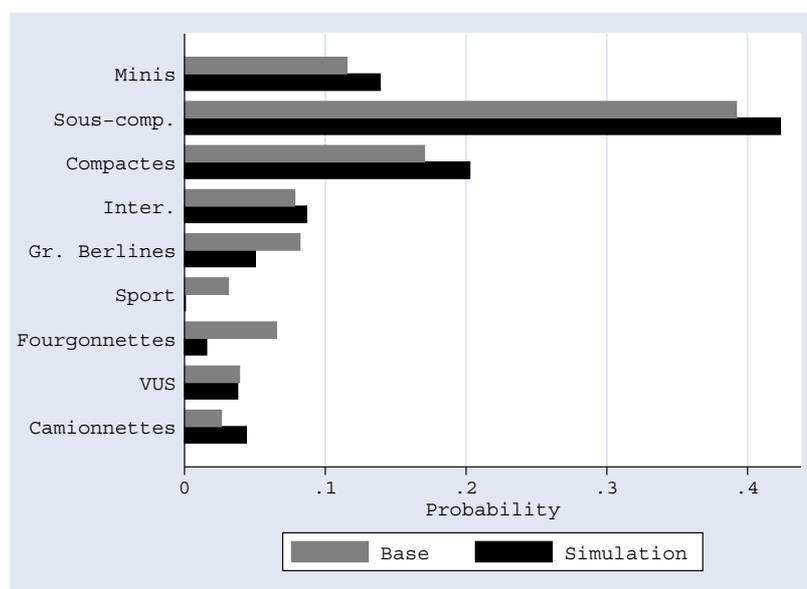
En ce qui à trait au taux de consommation de carburant (Conso carb), tous les paramètres ont le signe anticipé. Une augmentation du taux de consommation de carburant entraîne une baisse de l'utilité et ce, peu importe la classe de véhicules. Un détail qui renforce notre intuition sur l'aspect utilitaire des camionnettes est la faiblesse (et le fait qu'il n'est pas significatif) du paramètre associé au taux de consommation de carburant de ces derniers. Comme les camionnettes sont généralement achetées pour des usages spécifiques, des modifications de leur taux de consommation de carburant n'influencera pas leur probabilité de sélection. La définition des autres variables explicatives du modèle de choix et leur résultat d'estimation sont respectivement donnés aux appendices B et C.

Les quatre simulations considérées sont : le vieillissement de tous les conducteurs de 10 ans, l'ajout d'un enfant âgé entre 7 et 15 ans dans tous les ménages, une augmentation de 10 % de la consommation de carburant des VUS ainsi qu'une augmentation de 10 % du prix des grosses berlines.

La figure 5.1 nous renseigne sur la sensibilité de la composition du parc automobile québécois à l'âge des conducteurs. On doit noter que cette simulation n'est pas basée sur des prévisions démographiques officielles, mais représente simplement une application du modèle à des fins de prédiction. Un vieillissement des conducteurs d'une décennie entraîne une augmentation de la probabilité de sélection des minis voitures, des sous-compactes, des compactes, des intermédiaires ainsi que des camionnettes.

Ces changements se font principalement au détriment des voitures sports, des fourgonnettes et des grosses berlines, alors que la probabilité de choix des VUS reste pratiquement inchangée. Dans la perspective d'un vieillissement généralisé de la population québécoise, il est donc à prévoir qu'une substitution s'opérera vers de plus petits véhicules et des camionnettes. Ce résultat présuppose toutefois que les autres caractéristiques du ménage, y compris les préférences de ses membres, demeurent inchangées.

FIG. 5.1: Simulation 1, vieillissement de tous les conducteurs de 10 ans

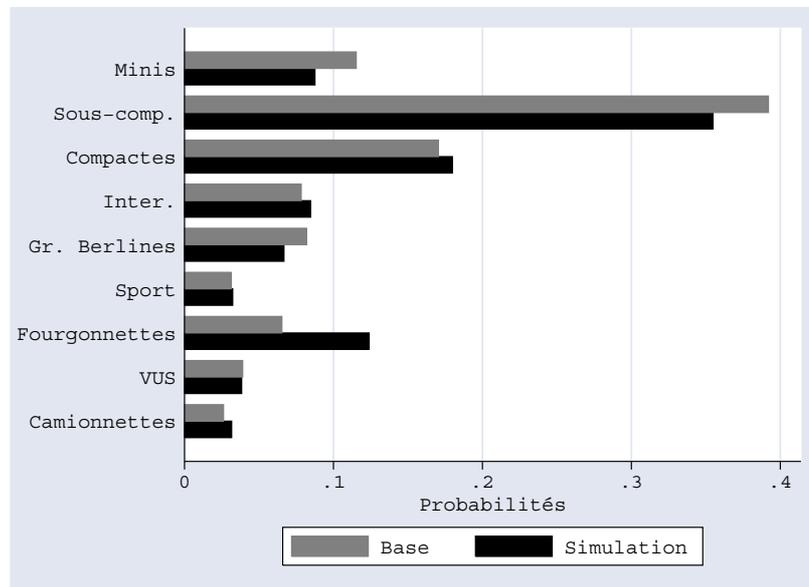


La figure 5.2, pour sa part, est assez éloquent quant à l'impact des enfants sur la probabilité de sélection des fourgonnettes. D'après nos estimations (voir appendice C), ces résultats sont toutefois valides uniquement pour les enfants âgés entre 7 et 15 ans, le paramètre associé aux fourgonnettes pour les enfants plus jeunes n'étant pas significatif. Ce résultat, de prime abord contre intuitif, pourrait s'expliquer par le fait suivant : comme les ménages possèdent généralement un véhicule plusieurs années¹,

¹Aujourd'hui, cette tendance est de moins en moins vraie avec les possibilités de location à court

il est possible que ce dernier soit conservé après l'arrivée d'un enfant et encore plus s'il est le premier. Il est donc normal que la présence d'enfants en bas âges influence moins le choix des véhicules puisque bon nombre de ceux-ci ont potentiellement été sélectionnés avant le début de la grossesse. Par contre, la présence d'enfants âgés entre 7 et 15 ans implique une période de temps suffisamment longue (au moins 7 ans) pour que le ménage ait pu ajuster son véhicule à sa nouvelle situation familiale.

FIG. 5.2: Simulation 2, ajout d'un enfant âgé entre 7 et 15 ans



Le tableau 5.2 rapporte les variations en pourcentage de la probabilité de sélection des classes de véhicules suite à une augmentation de 10 % du taux de consommation de carburant des VUS. Cette simulation vise à représenter la tendance dans ce secteur à produire des véhicules toujours plus gros et plus énergivores. Avec une diminution de la probabilité de sélection de plus de 63 %, l'effet sur la classe des VUS est sans ambiguïté. La substitution s'opère principalement à l'avantage des camionnettes, des fourgonnettes et des voitures compactes, ce dernier résultat étant toutefois contre terme (2 ou 3 ans).

intuitif. Nous nous serions attendus intuitivement à ce qu'une personne qui transfère de la classe des VUS vers la catégorie des voitures choisisse une grosse berline ou, à la limite, une intermédiaire. Les transferts de probabilité vers ces deux classes de voitures sont toutefois inférieurs à celui des compactes. Le tableau 5.3 rapporte la même mesure

TAB. 5.2: Simulation 3, augmentation de 10 % du taux de consommation de carburant des VUS

Classes de véhicules	Élasticités
Minis voitures	0.025
Sous-compactes	0.014
Compactes	0.185
Intermédiaires	0.088
Grosses berlines	0.024
Voitures sport	0.058
Fourgonnettes	0.123
VUS	-0.634
Camionnettes	0.085

que le tableau 5.2, mais pour une augmentation de 10 % du prix des grosses berlines. Sans grande surprise, ces dernières voient leur probabilité de sélection diminuer alors que les fourgonnettes, les intermédiaires et les compactes sont les classes privilégiées suite à ce changement.

Nous pouvons remarquer au tableau 5.1 et à l'appendice C que la valeur des paramètres estimés à l'aide du logit mixte est généralement plus élevée que celle obtenue à l'aide du logit standard. Cette particularité est due à la normalisation du terme Gumbel *iid* (ε), qui fixe l'échelle de l'utilité. Dans un logit standard, toute la composante stochastique est absorbée par ce terme d'erreur. Sa variance a donc tendance à être plus élevée que pour le logit mixte, où une partie de cette composante est capturée par le deuxième terme (ψ), de loi normale dans notre cas. Comme l'échelle

de l'utilité est normalisée dans les deux cas à l'aide de la variance du terme Gumbel *iid*, il en découle alors que l'utilité, et par le fait même l'amplitude des paramètres, est normalisée à la baisse avec le logit standard comparativement au logit mixte.

TAB. 5.3: Simulation 4, augmentation de 10 % du prix des grosses berlines

Classes de véhicules	Élasticités
Minis voitures	0.048
Sous-compactes	0.038
Compactes	0.431
Intermédiaires	0.246
Grosses berlines	- 0.512
Voitures sport	0.057
Fourgonnettes	0.207
VUS	0.097
Camionnettes	0.094

L'ajout des composantes d'erreur dans le logit mixte comparativement au logit standard est significative selon un test du ratio de vraisemblance au niveau de confiance de 95 %. Leur valeur, présentée au tableau 5.4, nous informe sur la variabilité des utilités à l'intérieur de chacune des classes de véhicules postulées. Nous apprenons ainsi que les compactes, les mini-fourgonnettes, les VUS, les camionnettes, les véhicules de 40000 dollars et plus, les petites voitures (regroupement des minis voitures, des sous-compactes et des compactes), les grosses voitures (regroupement des intermédiaires et des grosses berlines) et les véhicules offrant beaucoup d'espace (regroupement des mini-fourgonnettes et des VUS) sont des classes de véhicules relativement homogènes. En effet, la variance des utilités à l'intérieur de ces classes n'est pas significativement différente de zéro. À l'opposé, les minis voitures, les sous-compactes, les intermédiaires, les grosses berlines et les voitures sport présentent une

certaine variance dans l'utilité qu'ils procurent. Trois raisons principales peuvent expliquer cette différence.

TAB. 5.4: Composantes d'erreur (logit mixte)

Variables	Estimations	Statistiques t
Composantes d'erreur		
Minis	12.600	3.613
Sous-compactes	28.200	2.835
Compactes	-0.319	-0.466
Intermédiaires	1.950	2.451
Grosses berlines	-2.160	-2.034
Voitures sport	1.980	2.016
Mini-fourgonnettes	-0.371	-0.459
VUS	0.970	1.168
Camionnettes	-0.244	-0.244
40 000\$ + ^a	-0.149	-0.327
Petites voitures ^b	-0.102	-0.220
Grosses voitures ^c	-0.434	-0.872
Bcp d'espace ^d	-0.073	-0.158
Log-vraisemblance		
logit	-16625.9498	
logit mixte	-16556.3584	

^aRegroupement de tous les véhicules de 40 000 dollars et plus

^bRegroupement des minis voitures, des sous-compactes et des compactes

^cRegroupement des intermédiaires et des grosses berlines

^dRegroupement des mini-fourgonnettes et des VUS

Premièrement, il est possible que les véhicules que nous croyions homogènes tant dans leurs attributs que dans leur fonction ne le soient pas vraiment. Par exemple, à l'intérieur des intermédiaires, il existe peut-être une distinction entre les voitures étrangères et les voitures américaines, les premières étant possiblement plus luxueuses. Dans ce contexte, il serait normal que l'utilité retirée des véhicules varie en fonction

de leur provenance. Deuxièmement, il se peut que des véhicules soient effectivement homogènes en ce qui a trait à leurs caractéristiques, à l'intérieur de leur classe respective, mais que la perception de celles-ci ne soit pas la même entre les individus. Par exemple, une sous-compacte, qui constitue le second véhicule d'un ménage possédant aussi une grosse berline, ne sera pas utilisée de la même manière et pour les mêmes raisons qu'une sous-compacte qui représente le seul véhicule à la disposition d'un ménage. Troisièmement, il se peut que la perception des individus face à certaines catégories de véhicules change à travers le temps. Par exemple, sur les 11 années conservées aux fins de cette étude, les minis voitures ont connu un essor important. De nombreux modèles sont apparus, ceux qui existaient déjà ont connu des modifications majeures et la clientèle cible s'est élargie. Tous ces changements contribuent à modifier l'appréciation des consommateurs à l'égard de cette catégorie de véhicules à travers le temps, produisant ainsi de la variabilité dans les utilités perçues.

5.2 Modèle d'usage

Les résultats du modèle d'usage sont globalement conformes à nos attentes. Les principales variables ayant un pouvoir explicatif significatif sont l'âge, le sexe, et le lieu de résidence (voir tableau 5.5). Dans ce dernier cas, les ménages de Montréal et de sa ceinture ne semblent pas avoir un usage significativement différent de ceux des autres villes, alors que les ménages des régions ont clairement une tendance à utiliser davantage leurs véhicules. L'âge a l'effet négatif escompté sur l'utilisation. Selon nos résultats, le vieillissement de la population combiné à l'exode des régions rurales, deux phénomènes qui affectent le Québec de façon particulièrement sévère², induiraient

²Pour plus de détails sur les perspectives démographiques du Québec, voir le site de l'Institut de la statistique du Québec.

donc, toutes choses étant égales par ailleurs, une diminution de l'utilisation globale du parc de véhicules privés québécois. Notons que le prix du carburant, comme prévu, a un effet négatif sur le kilométrage parcouru. Finalement, soulignons que l'ajout des termes de correction est globalement significatif selon un test de Fisher au niveau de confiance de 95 % et que les paramètres les plus affectés par ces derniers sont ceux de la constante, de l'âge du véhicule ainsi que de la consommation de carburant. Les paramètres associés aux termes de correction provenant de la Mazda Miata, du Toyota 4Runner et de l'Infiniti Q45 sont les plus significatifs.

TAB. 5.5: Estimations du modèle d'usage

Variables	Estimations sans corr.	Statistiques t	Estimations avec corr.	Statistiques t
Constante	28.647	1.38	30.704	1.43
Semaine	-	-	-	-
FDS	3.561	1.39	4.909	2.11
Français	-	-	-	-
Anglais	7.028	1.72	5.136	1.37
Log rev/Prix carb	3.185	2.78	3.043	2.85
Été	-	-	-	-
Automne	0.227	0.07	2.646	0.92
Hiver	-3.229	-1.00	-1.328	-0.45
Printemps	3.564	1.13	4.993	1.73
Homme	-	-	-	-
Femme	-15.984	-6.78	-15.886	-7.26
Plein	-	-	-	-
Partiel	-4.643	-1.05	-3.585	-0.89
Temporaire	16.142	1.31	14.194	1.27
Retraite	-4.035	-0.92	-2.848	-0.71
Sans	-9.609	-1.52	-8.841	-1.53
Foyer	-12.827	-2.33	-10.490	-2.07
Étudiant	-9.845	-2.55	-8.096	-2.29
Autre Occ	16.281	2.57	12.024	2.05
Nbr personnes (6-)	-6.064	-1.99	-4.573	-1.63
Nbr personnes (7-15)	-1.874	-0.80	-0.512	-0.24
Nbr personnes (16+)	-5.090	-2.80	-3.840	-2.30

Estimation du modèle d'usage (suite)

Variables	Estimations sans corr.	Statistiques t	Estimations avec corr.	Statistiques t
Taille ménage (2-4)	-	-	-	-
Taille ménage (1)	-3.960	-0.93	0.160	0.04
Taille ménage (5+)	4.897	0.95	1.778	0.38
Âge véhic (3-5)	-	-	-	-
Âge véhic (0-2)	-3.540	-1.01	-7.459	-1.00
Âge véhic (6-8)	-4.528	-1.57	-18.524	-2.71
Âge véhic (9+)	-10.591	-3.38	-10.075	-1.36
Conso carb	0.547	0.88	-0.116	-0.14
Autres villes	-	-	-	-
Montreal	-2.198	-0.56	-1.313	-0.36
Ceinture	2.202	0.68	3.658	1.24
Rural	17.161	5.23	17.190	5.71
Un véhic	-	-	-	-
Plus d'un véhic	3.958	1.44	3.151	1.25
Âge	-0.342	-3.31	-0.299	-3.06
SCR	15 900 116		11 604 489	
Statistique F (230, 3378)	5.44			
Valeur critique	1.30			

Chapitre 6

Conclusion

La présente étude se distingue sur de nombreux aspects. Premièrement, d'un point de vue méthodologique, l'utilisation dans un cadre discret-continu du logit mixte comme outil d'estimation du modèle de choix permet de prendre en compte avec une grande flexibilité les interactions possibles entre les alternatives considérées. Toujours d'un point de vue méthodologique, l'ensemble de choix que nous utilisons inclut chacun des modèles de véhicules présent dans la banque de données. Cette particularité implique l'estimation de la probabilité de sélection de 230 alternatives et l'inclusion d'autant de termes de correction dans l'équation d'usage. Finalement, à notre connaissance, le cadre analytique préconisé est le premier du genre à être utilisé pour étudier le secteur du transport privé québécois.

Les résultats de la présente étude ont une grande résonance et ce, à plusieurs égards. Premièrement, le respect de l'Accord de Kyoto, récemment ratifié par le Canada, devra passer par l'élaboration, au niveau des provinces, de programmes de réduction des émissions de GES. Dans un souci d'efficacité, de telles politiques devront être construites à la lumière de données suffisamment désagrégées pour révéler les groupes à cibler et leur sensibilité aux mesures préconisées. À ce titre, nos résultats

offrent un point de départ intéressant. Par exemple, les mesures fiscales que le gouvernement du Québec projette de mettre en place afin de ralentir l'engouement pour les véhicules trop gourmands en carburant pourraient ne pas avoir l'effet escompté dans la mesure où les consommateurs qui privilégient ce type de véhicules présentent une faible sensibilité au prix. D'un autre côté, nous avons vu que l'âge et le fait de vivre en région urbaine plutôt que rurale avaient un impact négatif sur l'utilisation des véhicules. Les tendances démographiques que connaît le Québec à l'heure actuelle pourraient donc mener à une réduction de l'utilisation globale de son parc de véhicules privés. Finalement, dans une perspective plus large, nous croyons que nos recherches permettront aux autorités concernées de mieux comprendre l'évolution du choix et de l'usage des véhicules privés.

Notre modèle contient toutefois certaines limites. Premièrement, comme Manning et Winston (1985) le font remarquer, le fait de modéliser le choix et l'usage des véhicules dans un cadre statique évacue très certainement de nombreux aspects. Parmi ceux-ci, il y a les effets de substitution pour des véhicules plus efficaces suite à des mesures fiscales ou une augmentation du prix du carburant. Le fait de considérer le modèle dans un cadre dynamique permettrait aussi de prendre en compte les effets de préférence et de fidélité aux marques. Deuxièmement, la pauvreté de nos données en ce qui a trait à la description des véhicules nous oblige à faire des classes très générales. Il est donc possible que des résultats plus spécifiques et potentiellement contre intuitifs restent cachés derrière des tendances plus générales. Troisièmement, une autre lacune de nos données est le manque d'information sur l'ensemble du parc automobile des ménages, lacune qui nous contraint à considérer uniquement le ou les véhicules utilisés lors de l'enquête. Dans la mesure où les véhicules possédés par un ménage ont potentiellement des influences réciproques sur leur probabilité de sélection et leur usage, un modèle plus parcimonieux devrait considérer l'ensemble du parc

de véhicules de chacun des ménages.

Finalement, une extension intéressante au modèle serait de simuler, à l'aide d'effets croisés, de véritables scénarios de projection sociodémographique. Pour se faire toutefois, il nous faudrait pondérer les données de l'enquête utilisée à l'aide de poids statistiques appropriés pour obtenir des estimations représentatives de l'ensemble des propriétaires de véhicules québécois.

Bibliographie

- [1] Ben-Akiva, Moshe et Denis Bolduc (1991), «Multinomial Probit with Autoregressive Error Structure», *Groupe de Recherche en Économie de l'Énergie et des Ressources Naturelles*, Cahier 91-22, pp. 29.
- [2] Berkowitz, Micheal K. *et al.* (1990), «Disaggregate Analysis of the Demand for Gasoline», *The Canadian Journal of Economics*, 23 (2), 253-275.
- [3] Bernard, Jean-Thomas, Denis Bolduc et Donald Bélanger (1996), «Quebec residential electricity demand : a microeconomic approach», *Canadian Journal of Economics*, 24 (1), 92-113.
- [4] Bolduc, Denis et Dimitri Sanga (2002), «Two-Step Estimation of Discrete/Continuous Econometric Models with Interdependent Multinomial Choices», document de travail, Université Laval.
- [5] Brownstone, David et Kenneth Train (1999), «Forecasting New Product Penetration With Flexible Substitution Patterns», *Journal of Econometrics*, 89, 109-129.
- [6] Dubin Jeffrey A. et Daniel L. McFadden (1984), «An Econometric Analysis of Residential Electric Appliance Holdings and Consumption», *Econometrica*, 52 (2), 345-362.
- [7] Duval, Jacques et Denis Duquet (1986-1997), *Le guide de l'auto*, Montréal, Éditions de l'Homme.

- [8] Hanemann Michael W. (1984), «Discrete/Continuous Models of Consumer Demand», *Econometrica*, 52 (3), 541-562.
- [9] Hausman, Jerry A. (1981), «Exact Consumer's Surplus and Deadweight Loss», *The American Economic Review*, 71 (4), 662-676.
- [10] Heckman J. (1979), «Sample Selection Bias as a Specification Error», *Econometrica*, 47, 153-161.
- [11] Hensher, David A. (1985), «Empirical Vehicle Choice and Usage Models in the Household Sector : A Review», *International Journal of Transport Economics*, 12 (3), 231-251.
- [12] Hensher, David A. et Frank W. Milthorpe (1987b), «Selectivity Correction in Discrete-Continuous Choice Analysis», *Regional Science and Urban Economics*, 17, 123-150.
- [13] Hensher, David A., Frank W. Milthorpe et Nariida C. Smith (1990), «The Demand for Vehicle Use in the Urban Household Sector», *Journal of Transport Economics and Policy*, 24 (2), 119-137.
- [14] Hensher, David A. et Nariida C. Smith (1986), «A Structural Model of the Use of Automobiles by Households : A Case Study of Urban Australia», *Transport Reviews*, 6 (1), 87-111.
- [15] Jehle, Geoffrey A. et Philip J. Reny (2000), *Advanced Microeconomic Theory*, second edition, Boston, Addison Wesley, pp. 543.
- [16] Kavalec, Chris (1999), «Vehicle Choice in an Aging Population : Some Insights from a Stated Preference Survey for California», *Energy Journal*, 20 (3), 123-138.
- [17] Kim Yeonbae (2002), «Estimation of Discrete/Continuous Choice Model : Application of Bayesian Approach Using Gibbs Sampling», *Applied Economics Letters*, 9, 305-309.

- [18] Lee, L. F. (1983), «Generalised Econometric Models With Selectivity», *Econometrica*, 51 (2), 507-512.
- [19] Lyssiotou, Pashardes et Stengos (2001), «Estimates of the Black Economy based on Consumer Demand Approaches», document de travail, pp. 28.
- [20] Maddala, G. S. (1987), *Limited-Dependent and Qualitative Variables in Econometrics*, New York, Cambridge University Press, pp. 401.
- [21] Mannering, Fred et Clifford Winston (1985), «A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization», *The Rand Journal of Economics*, 16 (2), 215-236.
- [22] McCarthy, Patrick S. et Richard S. Tay (1998), «New Vehicle Consumption and Fuel Efficiency : A Nested Logit Approach», *Transportation Research : Part E : Logistics and Transportation Review*, 34 (1), 39-51.
- [23] McFadden, Daniel (1989), «A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration», *Econometrica*, 57 (5), 995-1026.
- [24] Moyneur, Érick (2002), «Méthodes d'inférence dans les modèles discrets/continus en présence d'instrumentation faible», mémoire de maîtrise, Université Laval.
- [25] Pakes, Ariel et David Pollard (1989), «Simulation and the Asymptotics of Optimization Estimators», *Econometrica*, 57 (5), 1027-1057.
- [26] Pas, E I (1984), «The Effect of Selected Sociodemographic Characteristics on Daily Travel-Activity Behavior», *Environment and Planning A*, 16, 571-581.
- [27] Sanga, D.M. (1999), «Estimations des modèles économétriques discret/continu : une approche par simulation», Thèse de doctorat, Université Laval.
- [28] Schaper, Vincent et Philip Patterson (1998), «Factors that Affect VMT Growth», <http://www.ott.doe.gov/pdfs/vmtwhite.pdf>.

- [29] Train, K. (1979), «Consumers Responses to Fuel-Efficient Vehicles : a Critical Review of Econometric Studies», *Transportation*, 8 (3), 237-258.
- [30] Train, K. (2002), *Discrete Choice Methods with Simulation*, Cambridge, Cambridge University Press, pp. 334.
- [31] Tardiff, T.J. (1980), «Vehicle Choice Models : Review of Previous Studies and Direction for Further Research», *Transportation Research*, 14(A) (5-6), 327-335.
- [32] Walker, Joan, Moshe Ben-Akiva et Denis Bolduc, «Identification of the Logit Kernel (or mixed Logit) Model», document de travail, juillet 2003.
- [33] Wooldridge, Jeffrey M. (2001), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, The MIT Press, pp. 752.

Appendice A Résumé des études présentées au chapitre 2

Auteurs	Fonction d'utilité indirecte	Choix des véhicules	Usage des véhicules	Modèle de choix	Base de données
Train et Lohrer (1983)	Fonction d'utilité indirecte permettant la linéarité dans les paramètres.	Parcs automobiles de 0, 1, 2 véhicules.	Kilométrage annuel pour des ménages possédant 1 et 2 véhicules. MCO avec variables instrumentales pour corriger les problèmes de sélection.	Logit emboîté	Echantillon aléatoire stratifié de 1095 ménages choisis sur l'ensemble du territoire américain. Données de Cambridge sur les caractéristiques systématiques des véhicules de 1978.
Manning et Winston (1985)	Fonction d'utilité indirecte pouvant intégrer des coûts fixes, tout en permettant une équation de demande linéaire dans les paramètres.	Choix entre 1 et 2 véhicules parmi un ensemble de 10 véhicules non classés, pour chaque période et pour des périodes combinées.	Une équation pour chaque grandeur de parcs automobiles. MCO avec correction des attentes rationnelles.	Logit emboîté	Données du «National Interim Energy Consumption Survey» et du «Household Transportation Panel Groups» permettant de connaître le profil de choix et d'usage d'un échantillon de ménages américains entre décembre 1978 et juin 1980. C'est-à-dire avant, pendant et après le choc pétrolier de 1979.
Hensher et Milthorpe (1987b)	-	-	MCO pour les ménages possédant 1 véhicule et MCO en trois étapes pour les ménages possédant 2 véhicules et plus. Évaluation de 2 formules pour la correction des attentes rationnelles.	Logit emboîté	Données sur 1434 ménages tirés d'une enquête effectuée dans la région métropolitaine de Sydney entre 1981 et 1982.

(suite)

Berkowitz <i>et al.</i> (1990)	Fonction d'utilité indirecte qui conduit à des usages symétriques. Une fonction par grandeur de parc automobile (0, 1, 2).	Choix entre 0, 1 ou 2 voitures.	Moindres carrés généralisés avec correction des attentes rationnelles. Distinction entre usages discrectionnaires et non discrectionnaires.	Logit emboîté	Enquête longitudinale conduite auprès de 2400 ménages canadiens à l'hiver 1982. Les données sur les coûts d'utilisation sont tirées, selon l'année, du Ward's Canadian Black Book ainsi que du MacLean-Hunter's Red book.
Hensher, Milthorpe et Smith (1990)	Fonction d'utilité indirecte non linéaire dans les paramètres et pouvant intégrer des coûts fixes.	Modèle permettant le traitement de toutes les grandeurs de parcs automobiles.	MCO en trois étapes avec correction des attentes rationnelles.	Logit emboîté	Données sur 1434 ménages tirés d'une enquête effectuée dans la région métropolitaine de Sydney entre 1981 et 1982.
Kim (2002)	-	Distingue uniquement entre les ménages possédant un véhicule et ceux possédant plus d'un véhicule.	-	Probit polytomique avec approche bayésienne.	Enquête conduite auprès de 293 ménages choisis aléatoirement dans la ville de Séoul en octobre 1998. Les prix et les caractéristiques des véhicules sont tirés du Hyundai Motors 1998.

Appendice B

Définition des variables

Modèle de choix

Variables	Définitions
Log rev/Prix véhic	Log du revenu annuel du ménage divisé par le prix du véhicule
Conso carb	Taux de consommation de carburant
Femme	Variable binaire indiquant que la répondante est un Femme
Homme	Variable binaire indiquant que le répondant est un homme
Nbr véhic	Nombre de véhicules dans le ménage
Log nbr membres	Log du nombre de membres dans le ménage
Nbr enfants (6-)	Nombre d'enfants âgés de 6 ans et moins
Nbr enfants (7-15)	Nombre d'enfants âgés entre 7 et 15 ans
Age ²	Âge du répondant au carré
Gr (16-19)	Variable binaire indiquant que le répondant appartient au groupe d'âge des 16-19 ans
Gr (20-24)	Variable binaire indiquant que le répondant appartient au groupe d'âge des 20-24 ans
Gr (25-44)	Variable binaire indiquant que le répondant appartient au groupe d'âge des 25-44 ans
Gr (45-64)	Variable binaire indiquant que le répondant appartient au groupe d'âge des 45-64 ans
Gr (65+)	Variable binaire indiquant que le répondant appartient au groupe d'âge des 65 ans et plus
Autres villes	Variable binaire indiquant que le ménage habite une région autre que les trois mentionnées ci-dessous
Montréal	Variable binaire indiquant que le ménage habite la ville de Montréal
Ceinture	Variable binaire indiquant que le ménage habite la ceinture de Montréal
Rurale	Variable binaire indiquant que le ménage habite un région rurale (agglomération de moins de 5000 habitants)

Modèle d'usage

Variables	Définitions
Semaine	Variable binaire indiquant que le déplacement a été effectué pendant la semaine
FDS	Variable binaire indiquant que le déplacement a été effectué pendant la fin de semaine
Français	Variable binaire indiquant que la langue de correspondance avec le conducteur sélectionné est le français
Anglais	Variable binaire indiquant que la langue de correspondance avec le conducteur sélectionné est l'anglais
Log rev/Prix carb	Log du revenu annuel du ménage divisé par le prix du carburant
Été	Variable binaire indiquant que le déplacement a été effectué pendant l'été
Automne	Variable binaire indiquant que le déplacement a été effectué pendant l'automne
Hiver	Variable binaire indiquant que le déplacement a été effectué pendant l'hiver
Printemps	Variable binaire indiquant que le déplacement a été effectué pendant le printemps
Homme	Variable binaire indiquant que le répondant est un homme
Femme	Variable binaire indiquant que le répondant est une femme
Plein	Variable binaire indiquant que le répondant est employé à temps plein
Partiel	Variable binaire indiquant que le répondant est employé à temps partiel
Temporaire	Variable binaire indiquant que le répondant est mise à pied temporairement (grève)
Retraite	Variable binaire indiquant que le répondant est à la retraite
Sans	Variable binaire indiquant que le répondant est sans emploi
Foyer	Variable binaire indiquant que le répondant est une personne au foyer
Étudiant	Variable binaire indiquant que le répondant est étudiant à temps plein
Autre occ	Variable binaire indiquant que le répondant occupe une autre occupation que celles mentionnées ci-dessus

Modèle d'usage (suite)

Variables	Définitions
Nbr personne (6-)	Nombre de personnes âgées de 6 ans et moins
Nbr personne (7-15)	Nombre de personnes âgées entre 7 et 15 ans
Nbr personne (16+)	Nombre de personne âgées de 16 ans et plus
Taille ménage (2-4)	Variable binaire indiquant un ménage comprenant entre 2 et 4 membres
Taille ménage (1)	Variable binaire indiquant un ménage comprenant 1 membre
Taille ménage (5+)	Variable binaire indiquant un ménage comprenant 5 membres et plus
Âge véhic (3-5)	Variable binaire indiquant un véhicule âgé entre 3 et 5 ans
Âge véhic (0-2)	Variable binaire indiquant un véhicule âgé entre 0 et 2 ans
Âge véhic (6-8)	Variable binaire indiquant un véhicule âgé entre 6 et 8 ans
Âge véhic (9+)	Variable binaire indiquant un véhicule âgé de 9 ans et plus
Conso carb	Taux de consommation de carburant du véhicule
Autres villes	Variable binaire indiquant que le ménage habite une région autre que les trois mentionnées ci-dessous
Montréal	Variable binaire indiquant que le ménage habite la ville de Montréal
Ceinture	Variable binaire indiquant que le ménage habite la ceinture de Montréal
Rurale	Variable binaire indiquant que le ménage habite un région rurale (agglomération de moins de 5000 habitants)
Un véhic	Variable binaire indiquant que le ménage possède un véhicule
Plus d'un véhic	Variable binaire indiquant que le ménage possède plus d'un véhicule
Âge	Âge du répondant

Appendice C

Suites des résultats du modèle de choix

Variables	Estimations logit	Statistiques t	Estimations logit mixte	Statistiques t
Constantes				
Petites voitures ^a	-	-	-	-
Moyennes voitures ^b	-2.720	-5.693	-4.200	-1.622
Grosses berlines	-5.350	-8.106	-8.130	-2.752
Voitures sport	-1.050	-0.799	-1.580	-0.520
Mini-fourgonnettes	-2.610	-3.015	-2.600	-0.959
VUS	-3.080	-3.385	-3.910	-1.416
Camionnettes	-4.390	-3.363	-5.010	-1.710
Homme (Femme)				
Petites voitures	-	-	-	-
Moyennes voitures	0.237	2.807	4.900	2.957
Grosses berlines	0.619	4.596	5.300	3.182
Voitures sport	0.957	4.361	5.880	3.492
Mini-fourgonnettes	0.693	4.745	5.530	3.327
VUS	0.834	4.743	5.650	3.387
Camionnettes	2.150	7.027	6.970	4.108
Nbr véhic				
Petites voitures	-	-	-	-
Moyennes voitures	-0.224	-3.617	-1.860	-2.712
Grosses berlines	-0.130	-1.238	-1.730	-2.454
Voitures sport	0.230	1.876	-1.280	-1.820
Mini-fourgonnettes	-0.431	-3.865	-2.080	-3.010
VUS	-0.058	-0.489	-1.690	-2.422
Camionnettes	0.353	2.741	-1.240	-1.777

^aRegroupement des minis voitures et des sous-compactes

^bRegroupement des compactes et des intermédiaires

(suite)

Variables	Estimations logit	Statistiques t	Estimations logit mixte	Statistiques t
Log nbr membres				
Petites voitures	-	-	-	-
Moyennes voitures	0.765	6.157	5.760	3.096
Grosses berlines	0.318	1.661	5.210	2.781
Voitures sport	-0.422	-1.525	4.210	2.210
Mini-fourgonnettes	0.930	4.076	5.870	3.135
VUS	0.208	0.870	5.170	2.741
Camionnettes	-0.392	-1.329	4.600	2.424
Nbr enfants (6-)				
Petites voitures	-	-	-	-
Moyennes voitures	-0.005	-0.037	0.697	0.702
Grosses berlines	0.022	0.090	0.778	0.734
Voitures sport	-0.381	-1.096	0.234	0.215
Mini-fourgonnettes	0.426	2.795	1.140	1.146
VUS	0.069	0.312	0.752	0.740
Camionnettes	0.102	0.303	0.808	0.761
Nbr enfants (7-15)				
Petites voitures	-	-	-	-
Moyennes voitures	0.006	0.066	0.838	1.239
Grosses berlines	-0.192	-0.977	0.584	0.778
Voitures sport	0.365	1.828	1.260	1.699
Mini-fourgonnettes	0.551	4.668	1.510	2.214
VUS	0.119	0.687	0.972	1.391
Camionnettes	0.520	2.653	1.420	2.032
Age²				
Petites voitures	-	-	-	-
Moyennes voitures	≈ 0	4.967	0.004	2.770
Grosses berlines	0.001	4.914	0.004	2.994
Voitures sport	-0.001	-3.206	0.001	0.820
Mini-fourgonnettes	≈ 0	1.995	0.003	2.582
VUS	≈ 0	1.308	0.003	2.538
Camionnettes	≈ 0	1.279	0.004	2.578

(suite)

Variables	Estimations logit	Statistiques t	Estimations logit mixte	Statistiques t
Gr (16-19) (Gr (45-64))				
Petites voitures	-	-	-	-
Moyennes voitures	-0.014	-0.055	-1.850	-0.819
Grosses berlines	-0.679	-1.627	-2.490	-1.071
Voitures sport	-2.490	-2.689	-5.490	-1.989
Mini-fourgonnettes	-0.364	-0.853	-2.390	-1.038
VUS	-0.236	-0.426	-2.180	-0.934
Camionnettes	0.240	0.351	-1.480	-0.619
Gr (20-24) (Gr (45-64))				
Petites voitures	-	-	-	-
Moyennes voitures	-0.218	-0.858	-3.410	-1.478
Grosses berlines	-0.740	-1.796	-3.940	-1.670
Voitures sport	-1.960	-2.287	-6.120	-2.276
Mini-fourgonnettes	-1.080	-2.329	-4.480	-1.899
VUS	0.119	0.231	-3.130	-1.325
Camionnettes	0.003	0.005	-3.080	-1.262
Gr (25-44) (Gr (45-64))				
Petites voitures	-	-	-	-
Moyennes voitures	-0.100	-0.504	-1.590	-0.983
Grosses berlines	-0.416	-1.295	-1.910	-1.126
Voitures sport	-1.170	-1.834	-3.260	-1.681
Mini-fourgonnettes	-0.245	-0.819	-1.840	-1.125
VUS	0.154	0.395	-1.310	-0.787
Camionnettes	-0.145	-0.284	-1.530	-0.889
Gr (65+) (Gr (45-64))				
Petites voitures	-	-	-	-
Moyennes voitures	-0.534	-2.486	-4.820	-1.896
Grosses berlines	-0.509	-1.825	-4.890	-1.911
Voitures sport	1.650	1.622	-1.790	-0.592
Mini-fourgonnettes	-0.697	-1.872	-4.910	-1.919
VUS	-1.300	-2.650	-5.520	-2.130
Camionnettes	-1.350	-2.197	-5.550	-2.116

(suite)

Variables	Estimations logit	Statistiques t	Estimations logit mixte	Statistiques t
Montréal (Autres villes)				
Petites voitures	-	-	-	-
Moyennes voitures	0.064	0.513	0.865	0.801
Grosses berlines	0.153	0.826	1.170	1.039
Voitures sport	-0.093	-0.269	0.654	0.552
Mini-fourgonnettes	-0.287	-1.242	0.497	0.450
VUS	-0.367	-1.339	0.322	0.289
Camionnettes	-0.195	-2.664	-1.230	-0.890
Ceinture (Autres villes)				
Petites voitures	-	-	-	-
Moyennes voitures	-0.206	-1.919	-1.890	-1.844
Grosses berlines	-0.185	-1.057	-1.810	-1.713
Voitures sport	-0.025	-0.099	-1.600	-1.417
Mini-fourgonnettes	-0.252	-1.391	-1.980	-1.905
VUS	-0.680	-2.926	-2.410	-2.290
Camionnettes	-1.280	-3.331	-3.070	-2.815
Rural (Autres villes)				
Petites voitures	-	-	-	-
Moyennes voitures	0.200	1.718	3.270	2.474
Grosses berlines	0.462	2.617	3.590	2.681
Voitures sport	0.054	0.190	3.100	2.259
Mini-fourgonnettes	0.488	2.662	3.540	2.647
VUS	0.301	1.449	3.440	2.565
Camionnettes	1.020	4.348	4.160	3.065