

PIERRE-OLIVIER BOUCHER

**USAGE DES THERMOSTATS ÉLECTRONIQUES
DANS LE MARCHÉ RÉSIDENTIEL QUÉBÉCOIS :
UNE ANALYSE MICROÉCONOMIQUE**

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

2005

Résumé

Récemment, les questions d'efficacité énergétique sont sur toutes les lèvres. Au Québec, comme ailleurs, il a été proposé que des gains substantiels pourraient être réalisés par l'usage résidentiel de thermostats électroniques pour le chauffage des locaux. La présente étude vise à analyser l'usage de tels thermostats par la clientèle résidentielle d'Hydro-Québec Distribution. Une modélisation faisant appel à la théorie sur les choix discrets ainsi qu'à l'analyse conditionnelle de la demande est mise de l'avant pour y arriver. Particulièrement, nous procédons à une analyse comportementale s'articulant sur un modèle discret-continu faisant intervenir un modèle multinomial logit. De cette façon, nous corrigeons pour un biais important, le biais d'auto-sélection. La dernière étape de notre modèle intègre les résultats du modèle discret-continu dans le but de faire ressortir l'impact de la composante chauffage sur la consommation totale annuelle d'électricité.

Avant-propos

Ce travail a pu être mené à terme grâce au support financier de la Chaire en économie de l'énergie électrique ainsi qu'à la disponibilité des données utilisées, assurée par Hydro-Québec, par l'entremise de la Chaire.

Je me dois de remercier mon directeur, Jean-Thomas Bernard, de m'avoir donné l'opportunité de traiter d'un sujet peu commun et de l'aborder d'une manière très novatrice, malgré les risques que cela comporte. Les embûches rencontrées ont été pour moi d'excellentes occasions de m'améliorer sur un grand nombre d'aspects essentiels à la complétion d'un tel projet. En me guidant dans cette expérience, j'ai pu bénéficier de ses conseils, de son appui et de sa compréhension. Je remercie aussi mon co-directeur, Denis Bolduc, pour m'avoir guidé et conseillé lorsque la situation l'exigeait.

Je ne pourrais passer sous silence le temps que m'a accordé un collègue, Francis Vekeman, tout au long de mon travail, alors que lui-même était à finaliser le sien. Sa grande disponibilité et ses conseils judicieux m'ont été d'une aide précieuse à des moments opportuns. Je remercie aussi tous ceux qui ont contribué, d'une façon ou d'une autre, à l'avancement de mon travail.

D'un côté plus personnel, j'ai eu et j'ai toujours la chance de compter sur un entourage extraordinaire. Je tiens à remercier ma famille, et particulièrement mon frère, pour m'avoir encouragé, épaulé et soutenu à une période où cette réalisation aurait pu être compromise. J'en suis très reconnaissant.

Table des matières

Résumé	ii
Avant-propos	iii
Table des matières	iv
Liste des tableaux	vi
Table des figures	vii
1 Introduction	1
2 Revue de la littérature	6
3 Cadre analytique et méthodologie	11
3.1 Modèles à choix discrets	11
3.2 Logit	14
3.3 Logit emboîté ou modèle GEV	16
3.4 Probit	18
3.5 Logit mixte (<i>kernel</i>)	19
3.6 Identification	20
3.6.1 Seules les différences d'utilité comptent	21
3.6.2 L'échelle absolue de l'utilité est arbitraire	22
3.7 Spécification du modèle	23
3.7.1 Analyse comportementale	23
3.7.2 Utilité indirecte	25
3.7.3 Modèle discret-continu	26
3.7.4 Méthodes de correction pour l'équation d'usage	27
3.7.5 Consommation annuelle totale	30
4 Données	35
5 Résultats	43
5.1 Analyse comportementale	43

5.1.1	Étape 1 : modèle de choix	43
5.1.2	Étape 2 : modèle d'usage	45
5.1.3	Étape 3 : modèle de consommation moyenne annuelle	46
6	Conclusion	48
	Bibliographie	50
A	Maximisation de l'utilité	52
B	Définitions des variables	54
C	Résultats du modèle de choix discrets	57
D	Résultats du modèle de consommation annuelle	62

Liste des tableaux

5.1	Résultats du modèle d'usage	45
C.1	Résultats du modèle à choix discrets	57
D.1	Résultats du modèle de consommation avec le nombre de pièces	62
D.2	Résultats du modèle de consommation avec la superficie	64

Table des figures

4.1	Répartition des réponses pour la température maintenue le jour	41
4.2	Répartition des réponses pour la température maintenue le soir	41
4.3	Répartition des réponses pour la température maintenue la nuit	42

Chapitre 1

Introduction

Les thermostats, présents dans les résidences et utilisés pour fixer le niveau de la température ambiante, ne constituent pas un sujet d'étude très répandu, surtout lorsqu'ils sont associés à des systèmes de chauffage par plinthes électriques. Néanmoins, les récentes innovations technologiques en font des instruments beaucoup plus adaptés pour des fins d'efficacité énergétique, une orientation adoptée par le Québec et bon nombre de pays industrialisés.

L'évidence des changements climatiques ne fait maintenant plus de doute et il semble que la responsabilité en incombe aux facteurs anthropiques. Les gaz à effet de serre (GES) sont principalement visés lorsqu'il est question d'expliquer les causes de ces changements du climat. La signature du Protocole de Kyoto en 1997 est la manifestation tangible à l'échelle mondiale des actions entreprises visant à réduire ces émissions.

Bien que le Québec soit présentement dans une position favorable à cet égard, il est tout à son avantage de préserver cette situation. Il est bien connu que le secteur énergétique est un grand contributeur aux émissions de GES. Le Québec a l'opportunité de compter sur une abondante ressource hydroélectrique ; cependant cette source d'énergie risque de ne pas satisfaire la demande croissante dans les années à venir, comme le montre le Plan Stratégique d'Hydro-Québec déposé à l'automne de 2003.

Dans les années soixante, suite à la nationalisation de plusieurs compagnies d'électricité, Hydro-Québec a connu une forte expansion. À cette époque, la diffusion de l'électricité affichait une importante progression, dans un contexte où l'énergie était peu dispendieuse et la ressource hydroélectrique considérée comme abondante. La crise pétrolière de 1973 intensifia davantage la commercialisation de l'électricité et entraîna une plus grande préoccupation pour les économies d'énergie qui se matérialisèrent en

d'importants projets gouvernementaux suite au second choc pétrolier. Ces projets fortement appuyés par Hydro-Québec visaient particulièrement la bi-énergie résidentielle et étaient constitués notamment du Programme canadien de remplacement du pétrole, de Énergain Québec et ÉnerGuide. Au début de la décennie 90, en réponse à la stratégie d'efficacité énergétique du gouvernement du Québec de 1992, la société d'État s'est impliquée davantage en efficacité énergétique avec l'introduction d'une quinzaine de programmes d'économie d'énergie.

Au Québec, la *Loi sur la Régie de l'énergie* établit un volume annuel maximal d'électricité patrimoniale de 165 TWh, qui doit être rendu disponible pour la consommation québécoise par Hydro-Québec Production¹. En 2002, les ventes d'Hydro-Québec Distribution ont totalisé 158,3 TWh et leur croissance annuelle est estimée à 1,5 % en moyenne sur la période 2001-2011 (Rapport annuel de 2002). Les plus récents projets de production n'entrant en opération qu'à partir de 2006, cette production risque de ne pas satisfaire à la demande, surtout de pointe, dans les années à venir. Au-delà du volume d'électricité patrimoniale, Hydro-Québec Distribution doit conclure des contrats d'approvisionnement au prix du marché pour répondre à la demande annuelle des Québécois.

L'efficacité énergétique devient alors une voie favorisée, au Québec, afin de diminuer la croissance de la demande d'énergie, d'éviter le recours aux technologies de substitution plus polluantes et de respecter les politiques de protection de l'environnement. Les seize programmes contenus dans le Plan global en efficacité énergétique (PGEÉ) de 2003 d'Hydro-Québec présentés et approuvés par la Régie de l'énergie s'inscrivent dans cet effort.

Deux de ces programmes nous intéressent particulièrement, soit la promotion du thermostat électronique dans le marché résidentiel existant ainsi que dans la nouvelle construction. Il y a peu d'écrits antérieurs concernant l'usage de thermostats électroniques non-programmables ou programmables pour contrôler le chauffage des locaux par des plinthes électriques dans le secteur résidentiel. Dans sa revue des écrits antérieurs à ce sujet, Plourde (2003) fait remarquer qu'une étude empirique de l'efficacité des thermostats programmables pour le secteur résidentiel serait tout à fait justifiée. Le présent travail vise à combler cette lacune.

¹Suite à la décision du gouvernement du Québec, en septembre 1996, d'approuver le principe du libre transit de l'électricité de gros sur le réseau de transport d'Hydro-Québec, la société d'État a été divisée en quatre filiales : Hydro-Québec Distribution, Hydro-Québec Production, Hydro-Québec TransÉnergie et Hydro-Québec Pétrole et gaz (exploration).

Selon les deux programmes cités précédemment, les thermostats électroniques (principalement non-programmables) doivent remplacer les thermostats bimétalliques (réguliers) présents dans les habitations existantes et être fortement favorisés dans les nouvelles constructions. Un certain nombre d'intervenants plaident en faveur des thermostats programmables, qui comportent, selon eux, des possibilités d'efficacité énergétique plus importantes. Nous tiendrons compte de cette différence dans notre analyse.

Le programme visant le marché existant a pour objectif d'encourager l'implantation de l'une des mesures considérée comme ayant le plus grand potentiel, soit de favoriser le remplacement de plusieurs thermostats, d'atteindre des ventes de 269 000 thermostats électroniques nettes du tendancier pour la période 2003-2006, et d'inciter la clientèle à abaisser la température de jour et\ou de nuit des thermostats.

Du côté de la nouvelle construction, la société d'État supportera financièrement l'écart de coût entre un thermostat bimétallique et un électronique². Les objectifs visés sont de favoriser l'installation des appareils électroniques dans 60 % des nouvelles constructions d'unifamiliales, de duplex et de triplex ayant un système principal de chauffage par plinthes électriques contrôlées par des thermostats muraux.

Une question se pose alors : l'expérience actuelle des clients résidentiels d'Hydro-Québec Distribution permet-elle de faire ressortir des gains pour les consommateurs liés à l'utilisation des thermostats électroniques non-programmables et programmables par rapport à l'usage des thermostats bimétalliques? Il s'agit donc de déterminer si les ménages québécois utilisant ces thermostats ont une consommation totale d'énergie différente de ceux n'en disposant pas.

Nos attentes, à ce sujet, sont plutôt positives, et les questions précédentes devraient trouver une réponse dans l'affirmative. Autrement dit, l'usage des thermostats électroniques devrait, selon nous, être associé à une plus grande efficacité de la part des ménages, autant au niveau du choix de la température ambiante intérieure que de la consommation annuelle totale qui en découle. Nous nous attendons aussi à un recours plus répandu et plus important aux abaissements de température de la part des ménages utilisant des thermostats programmables. Enfin, nous croyons que la période de la journée peut avoir une influence sur la température de chauffage intérieure fixée ainsi que sur la consommation totale d'énergie. Le jour, selon nous, par l'effet de l'ensoleille-

²Le coût moyen d'un thermostat électronique est estimé à 45 \$ et le support financier est estimé à 40 \$ par thermostat électronique.

ment, serait une période moins propice à d'importants gains énergétiques, tout comme le soir où les résidences ont un fort niveau d'occupation et d'activité. La nuit offrirait donc les meilleures possibilités d'économies en énergie reliées à l'usage de thermostats plus efficaces.

Nous verrons, cependant, que les résultats ne permettent pas de conclure aussi précisément sur ces points. Nous verrons aussi que l'utilisation d'instruments plus efficaces, si elle ne s'effectue pas de manière appropriée, ne permet pas d'atteindre des résultats intéressants. L'usage des ménages doit lui aussi être efficace, et non seulement les appareils qu'ils utilisent. Il ressort, enfin, que les données utilisées pour l'analyse manquent de profondeur et de variabilité, spécialement en ce qui à trait aux thermostats.

La modélisation élaborée en réponse à notre problématique se divise en trois étapes séquentielles : 1) estimation du choix du type de thermostat par les ménages, 2) estimation de la demande de chauffage des ménages, et 3) estimation de la consommation annuelle totale d'énergie (électrique). Cette dernière étape incorpore les résultats obtenus aux deux précédentes et repose sur les notions de l'analyse conditionnelle de la demande (Parti et Parti, 1980). Cette méthode de régression, traitée plus en détail au chapitre 4, a pour but de désagréger la consommation totale d'énergie de chaque ménage échantillonné à travers l'ensemble des équipements électriques existants, faisant ressortir les effets de la consommation de base,³ des appareils électriques et du chauffage sur cette consommation totale.

Les deux premières étapes demandent une attention particulière car nous y traitons du choix et de l'usage des thermostats, sachant que le processus de décision est simultanément influencé par le choix et l'usage. L'analyse comportementale des ménages québécois s'effectue alors en deux étapes séquentielles (niveaux) selon la formulation découlant de la théorie sur les modèles discrets-continus. Nous estimons d'abord la probabilité de choix de chacun des thermostats (alternatives) pour ensuite estimer la fonction de demande associée à ces thermostats (usage), dans le cas présent le niveau de la température intérieure fixée, conditionnel au choix.

L'aspect conditionnel de ce processus de décision se traduit par une situation dans laquelle les données utilisées pour estimer la fonction de demande d'un certain type de thermostat proviennent uniquement du sous-échantillon de ménages qui se sont eux-mêmes sélectionnés par le choix de cette alternative. Dans les travaux sur les modèles discrets-continus, un tel biais est appelé biais d'auto-sélection. D'autre part, l'aspect simultané des choix implique qu'ils soient causés, ou déterminés, par des facteurs

³La consommation de base correspond aux usages électriques communs à tous les ménages, c'est-à-dire les usages ayant les plus forts taux de pénétration.

non-observables identiques. Comme les mêmes facteurs non-observables peuvent apparaître à chacune des étapes ou niveaux, il est possible que leurs termes d'erreur respectifs soient corrélés. Afin de prendre en compte cette corrélation entre les deux niveaux et le biais d'auto-sélection, nous suivons la méthodologie traditionnelle en introduisant un terme de correction à l'équation d'usage. La section 3.7.4 examine de plus près cette question.

Nous utilisons, pour l'estimation du modèle de choix, une spécification dite logit multinomiale qui répond mieux aux contraintes de notre modèle que le probit ou encore le GEV. Toutes ces spécifications sont fortement liées et les premières sections du chapitre 3 leur sont dédiées. L'emphase est bien entendu portée au logit standard sans toutefois oublier les autres modèles qui ont des applications très importantes et largement répandues et qui méritent d'être connues.

Les écrits, peu nombreux, concernant les thermostats électroniques et présentant des études empiriques à leur sujet sont présentés brièvement au chapitre suivant. Par la suite, le chapitre 3 est consacré aux éléments théoriques et méthodologiques sous-jacents à notre modélisation. La section 3.1 présente les notions générales des modèles à choix discrets, la section 3.2 porte sur le modèle logit standard, la section 3.3 donne un aperçu des modèles GEV et particulièrement du logit emboîté, les sections 3.4 et 3.5 introduisent les modèles probit et logit mixte, la section 3.6 discute de l'identification du modèle, et enfin, la section 3.7 présente notre modélisation. Le chapitre 4 offre un survol des données utilisées pour l'analyse et finalement, le chapitre 5 présente les principaux résultats obtenus à l'aide de notre modélisation et une discussion de ceux-ci.

Chapitre 2

Revue de la littérature

Le commun des mortels n'étant pas à l'affût de toutes les distinctions entre les différents types de thermostats étudiés ici, il est important, avant de poursuivre, de présenter quelques notions pratiques les caractérisant. Les développements qui suivront seront plus aisément compréhensibles et l'intérêt relié au sujet plus appréciable.

Les thermostats programmables constituent un sujet d'étude intéressant en raison de leur capacité à réduire la consommation et les dépenses en énergie, tout en affectant au minimum le niveau de confort dont jouissent les ménages qui les utilisent. La nuit et les périodes où la résidence est inoccupée permettent une température ambiante plus basse sans toutefois s'accompagner d'une réduction du confort aussi importante que les périodes où la résidence est occupée.

D'un point de vue technique, la distinction majeure entre le thermostat régulier ou bimétallique et celui électronique se situe au niveau de la précision de l'appareil et de la consommation d'énergie électrique qui en résulte. Le thermostat régulier fonctionne selon l'expansion ou la contraction d'un ressort en métal qui déplace un petit interrupteur au mercure. Il peut permettre des écarts allant jusqu'à 2°C au-dessous ou au-dessus de la température souhaitée. De plus, il offre une faible précision de réglage et s'accompagne typiquement d'ajustements fréquents qui sont coûteux et très inefficaces.

Le thermostat électronique, quant à lui, possède généralement un afficheur numérique sur lequel la température ambiante est indiquée. Il est muni d'une sonde lisant continuellement la température et réagissant dès qu'une infime variation se fait sentir. Avec en plus la possibilité de fixer la température souhaitée beaucoup plus précisé-

ment (numériquement) qu'avec le thermostat régulier, il en résulte une consommation d'électricité plus stable. Aussi, son usage tend à éliminer les fréquents ajustements cités plus haut. De fortes variations autour d'une température imprécise font place à de très faibles variations autour d'une température précise.

Techniquement, les thermostats électroniques programmables fonctionnent sur le même principe que les non-programmables, à la différence qu'ils comportent, en plus, une minuterie électronique ajustant automatiquement la température désirée selon les habitudes de vie des occupants¹. L'intérêt de ce type de thermostat repose moins sur ses particularités techniques que sur l'usage qu'il est possible d'en faire.

Le thermostat programmable possède généralement les fonctions suivantes : il emmagasine et applique plusieurs programmes quotidiens ou hebdomadaires qui peuvent être manuellement contournés sans être altérés, il emmagasine souvent plus de six températures différentes programmées par jour et il permet de régler la température de chauffage de manière à assurer à la fois un plus grand confort et des gains énergétiques.

Il est recommandé, pour un usage plus efficace des thermostats programmables, de les programmer de façon à ce que la température ambiante soit de 17°C pendant la nuit ou lorsqu'il n'y a pas d'occupant et de 20°C le reste du temps². Ainsi, la température ambiante moyenne sera plus faible et le consommateur n'aura pas à manipuler le thermostat quotidiennement ; le tout se fera automatiquement selon les heures de la journée et les jours de la semaine.

Voilà où repose le potentiel d'économie d'énergie des thermostats programmables : choisir une température ambiante, la maintenir pour certaines heures fixées et ensuite abaisser cette température ambiante pendant un certain laps de temps. Cette période au cours de laquelle la température ambiante est plus basse est communément appelée *setback* ou *offset*. L'énergie est économisée entre le moment où la température se stabilise au plus bas niveau et le moment où le chauffage est réactivé. Plus cette période est longue, donc plus le *setback* est grand, plus le potentiel d'économie est élevé³.

¹Il existe aussi une ancienne génération de thermostats programmables dits électromécaniques, contrôlés manuellement par des mécanismes amovibles permettant de régler des températures pour la nuit et le jour.

²hydroquebec.com/residentiel

³Il faut cependant mettre en garde contre une fausse croyance voulant que plus la température fixée sur le thermostat est élevée, plus la température ambiante se réchauffera rapidement. Ce qui importe c'est la durée requise pour atteindre la température fixée.

En hiver, des économies significatives peuvent être obtenues en réduisant manuellement ou automatiquement la température de chauffage sur le thermostat. Ces économies sont attribuées à la perte de chaleur d'un bâtiment en hiver, qui dépend fortement de la différence entre les températures intérieure et extérieure. Nous tenons compte de cet élément en introduisant le degré-jour de chauffage à la seconde étape de notre modèle. Nous y reviendrons plus loin.

Pour l'instant, revenons aux écrits concernant l'usage de thermostats électroniques programmables et au concept de *setback* qui lui est directement relié. Ces travaux, en ce qui concerne le contrôle du chauffage des locaux par des plinthes électriques dans le secteur résidentiel, sont somme toute assez limités. Plourde (2003) fait remarquer qu'une étude empirique de l'efficacité des thermostats programmables dans la réduction des demandes d'énergie ne semble pas avoir été réalisée au Canada pour les secteurs résidentiel ou commercial, et serait justifiée. Notre travail serait donc une première, du moins pour le secteur résidentiel.

Les premières études effectuées dans les années soixante-dix sur l'efficacité des thermostats programmables étaient basées presque uniquement sur les systèmes de chauffage résidentiels à l'huile ou au gaz. Les conclusions étaient généralement positives et laissaient entrevoir un fort potentiel de réduction de l'utilisation d'énergie pour ces systèmes. L'article de Nelson et MacArthur (1978) représente bien l'approche méthodologique préconisée à l'époque. On leur doit notamment la règle générale, toujours valide, selon laquelle un *setback* nocturne quotidien d'une moyenne de huit heures peut entraîner une réduction de près de 1,8 % de la consommation d'énergie pour chaque degré Celsius pour les systèmes au gaz ou à l'huile.

Parallèlement à ce résultat, les auteurs ajoutent que pour un *setback* de jour identique et de même durée, les économies d'énergie sont plus faibles. On y apprend aussi que la rigueur du climat joue directement un rôle. Plus le climat est froid, plus les économies d'énergie sont importantes en utilisant des thermostats programmables. Enfin, les économies tendent à être plus élevées pour des habitations ayant une isolation plus faible, en quantité et non en épaisseur. Les similarités des systèmes de chauffage à l'huile permettent de leur attribuer des économies d'énergie du même ordre. Cependant, les auteurs entrevoient des gains moindres de l'usage de thermostats programmables dans les cas impliquant des pompes à chaleur alimentées à l'électricité.

Il est important de faire remarquer que la modélisation utilisée dans cette étude et dans les travaux à ce sujet diffèrent grandement de la nôtre. Nelson et MacArthur (1978) procèdent par la simulation de modèles de circulation d'énergie dans une habitation ayant des caractéristiques spécifiques. Ils utilisent ces modèles pour analyser le potentiel

de gains énergétiques générés par l'usage de thermostats programmables en association avec un système de chauffage au gaz naturel.

La simulation repose sur deux options : 1) choisir une température ambiante et la maintenir toute la journée, et 2) choisir la même température, la maintenir seulement pour un certain nombre d'heures, et ensuite permettre à la température ambiante d'être plus basse pour le reste de la journée⁴. Une situation de base est spécifiée et elle inclut les températures ambiantes moyennes intérieure et extérieure ainsi que la durée des *setbacks*. La simulation s'effectue en deux étapes distinctes, à l'aide de ce modèle, pour analyser l'impact des deux options précédentes.

Les études plus récentes tendent à démontrer que l'usage de thermostats programmables dans un contexte résidentiel offre des possibilités d'économie d'énergie beaucoup plus modestes, surtout lorsque le système de chauffage est à l'électricité. Déjà, Bullock (1978) en arrivait, avec un modèle similaire à Nelson et MacArthur (1978), à des conclusions indiquant des gains énergétiques beaucoup plus faibles associés aux pompes à chaleur en général.

Des résultats similaires sont obtenus par Bouchelle *et al.* (2000) quant à l'utilisation simultanée de pompes à chaleur et de thermostats programmables standards. À l'aide d'un échantillon de 200 utilisateurs de pompes à chaleur de la Floride (un climat peu rigoureux), ils concluent que les économies d'énergie sont faibles même avec un *setback* élevé. Les auteurs expliquent ces conclusions par le fait que les pompes à chaleur ont un mode de fonctionnement plus intensif, c'est-à-dire qu'elles utilisent plus d'électricité pour produire de la chaleur lorsque la période de *setback* se termine. L'usage du thermostat programmable amplifie cette particularité, ce qui mène à un plus grand besoin d'électricité à la fin de la période de *setback* et contribue à éliminer, en partie, les réductions d'énergie réalisées lors de cette période.

Des résultats intéressants, mais peu optimistes quant à l'efficacité des thermostats programmables, découlent de l'étude de Nevius et Pigg (2000) sur 299 ménages du Wisconsin. Même en présence de systèmes de chauffage au gaz, leur enquête révèle que les ménages ayant des thermostats programmables consomment une quantité d'énergie pour le chauffage identique aux ménages ayant des thermostats réguliers.

Un facteur très important souligné par les auteurs pour expliquer la piètre efficacité des thermostats programmables dans leur enquête provient du fait que les deux options citées plus haut peuvent être appliquées sans faire l'usage de tels thermostats.

⁴Plourde, André (2003) « Programmable Thermostats as Means of Generating Energy Savings : Some Pros and Cons », CBEEDAC 2003-RP-01, Edmonton, Alberta.

Les études à ce sujet recourant généralement à ces options dans leurs simulations, l'éventualité d'abaisser le thermostat manuellement aurait donc pour effet d'accaparer les économies d'énergie attribuées uniquement à l'usage du thermostat programmable dans ces simulations. La mauvaise utilisation par les ménages serait aussi un facteur non négligeable.

Nevius et Pigg (2000) font aussi état de la présence d'un biais d'auto-sélection dans leur échantillon. À titre d'exemple d'auto-sélection, un ménage ayant davantage à coeur l'efficacité énergétique optera plus vraisemblablement pour un thermostat programmable qu'un autre ne s'en préoccupant pas. L'inverse est aussi vrai. Le recours aux méthodes de choix discrets dans notre modèle permet de prendre en compte explicitement le problème posé par ce biais. Pour ce qui est de la mauvaise utilisation des thermostats programmables, nous ne pouvons pas et nous ne devons pas corriger pour ce type de problème⁵. Néanmoins, nous verrons qu'elle joue effectivement un rôle important.

On constate aisément que la vaste majorité des travaux ont porté sur les systèmes opérant au gaz naturel ou à l'huile et que les plinthes électriques n'ont pas reçu, à ce jour, un traitement sérieux. Nous proposons donc une étude considérant uniquement un système de chauffage par plinthes électriques et s'appliquant au secteur résidentiel québécois. Nous verrons au chapitre 3 que notre modélisation permet d'aborder la question dans une perspective complètement différente qui possède l'avantage de prendre en considération l'aspect comportemental des ménages étudiés. Il s'agit d'un ajout important puisque c'est aux individus, en réalité, à choisir le niveau de la température ambiante et le type de thermostat utilisé dans la résidence.

⁵Évidemment, nous devons constater ou utiliser les habitudes d'utilisation des ménages pour expliquer leur comportement et non tenter de les corriger.

Chapitre 3

Cadre analytique et méthodologie

La modélisation élaborée afin d'étudier l'efficacité des thermostats programmables sera grandement facilitée si nous présentons, d'abord, tous les éléments théoriques et méthodologiques qui la sous-tendent. Ce chapitre offrira donc au lecteur un exposé succinct, mais complet, des principales notions concernant les modèles de choix discrets-continus pertinentes à notre travail. Nous verrons ensuite comment ces notions s'appliquent à notre propre modèle. Nous débutons par les modèles à choix discrets (qualitatifs).

3.1 Modèles à choix discrets

Leur nom l'indique bien, les modèles à choix discrets sont conçus pour décrire et expliquer les situations dans lesquelles un agent¹ doit choisir parmi un nombre limité d'alternatives différentes (aussi appelées modalités). L'attrait de ces modèles tient au fait qu'il est beaucoup plus intéressant, économiquement, d'étudier le comportement à un niveau désagrégé ou encore individuel, les demandes totales n'étant que l'agrégation des actions de plusieurs agents. Les enquêtes microéconomiques fortement désagrégées étant de plus en plus répandues et de meilleure qualité, les modèles à choix discrets deviennent des instruments prisés par les chercheurs.

Pourquoi élaborer de tels modèles alors que nous disposons déjà d'une grande variété de procédures de régression standard ? La réponse est bien simple. L'expérience a démontré que les résultats de l'étude des comportements individuels ne sont pas conti-

¹Nous utilisons le terme agent pour représenter à la fois un individu, un ménage ou toute entité décisionnelle dans le but d'alléger la présentation.

nus et que les procédures de régression multiple sont inappropriées. Le problème repose dans le caractère discret de la variable dépendante. Un côté de la régression est discret alors que l'autre côté est continu. La moyenne du terme d'erreur sera nulle seulement si la probabilité d'un choix est égale à l'élément explicatif de la régression. Cette probabilité prend par définition une valeur entre zéro et un, mais rien ne contraint la partie explicative, ou son estimateur, à être comprise dans cet intervalle (Train, 2003).

Revenons aux modèles de choix. Les prémisses à la base de ces modèles concernent les alternatives formant l'ensemble de choix : 1) le nombre d'alternatives doit être fini ; 2) les alternatives doivent être mutuellement exclusives, c'est-à-dire que le choix d'une alternative exclut le choix d'une autre alternative ; et 3) l'ensemble des alternatives doit être exhaustif de manière à ce que toutes les alternatives possibles soient incluses. Si les alternatives sont correctement définies, uniquement le premier critère est restrictif. Le recours à un modèle à choix discret est donc pertinent si un agent doit faire un choix parmi un ensemble fini et exhaustif d'alternatives mutuellement exclusives.

Tous les modèles de choix ont le même objectif, soit de calculer la probabilité qu'un agent choisisse une certaine alternative parmi un ensemble fini, et cela, selon les données observées par le chercheur. Ils diffèrent cependant par la forme fonctionnelle utilisée pour relier les données observées à la probabilité de choix. Ces formes fonctionnelles sont notamment le logit, le probit, le GEV et le logit mixte, que nous verrons dans les prochaines sections.

Pour l'instant, introduisons une notation qui nous permettra de nous rendre au coeur des modèles de choix. Un agent n fait face à un ensemble d'alternatives, noté J_n , dans lequel chaque alternative est dénotée par i . Les caractéristiques observées propres à l'alternative choisie sont représentées par z_{in} alors que celles de toutes les autres modalités sont dénotées par z_{jn} . Enfin, nous représentons par s_n les caractéristiques observées de l'agent n . Selon cette notation, la probabilité que l'agent n choisisse l'alternative i est spécifiée par la forme générale suivante :

$$P_{in} = f(z_{in}, z_{jn}, s_n, \beta \ j \neq i), \quad (3.1)$$

où f est la forme fonctionnelle reliant les données observées aux probabilités de choix.

Comment, alors, déterminer quelle est l'alternative effectivement choisie par l'agent ? La réponse repose dans la théorie microéconomique du consommateur, qui est en fait une théorie de choix. Nous y apprenons que le consommateur cherche à identifier et à sélectionner une alternative préférée sur la base de ses goûts personnels et de ses contraintes. La notion de préférence est cruciale en ce sens qu'elle permet au consommateur de construire un classement en représentant ses préférences par une relation qui

reflète une habileté à faire des choix selon une certaine logique (Jehle et Reny, 2001). Dans la théorie moderne, la fonction d'utilité décrit les préférences et attribue une valeur aux différentes alternatives de façon à ce que l'alternative la plus désirable reçoive des valeurs supérieures à celles moins désirées (Varian, 1997).

Les modèles d'utilité aléatoire s'appuient sur ces concepts ordinaux et attribuent une certaine utilité U_{in} à chacune des alternatives i dans J_n . Cette utilité dépend aussi bien des caractéristiques de l'alternative i (x_{in}) que des caractéristiques de l'agent n (r_n) :

$$U_{in} = U(x_{in}, r_n) \quad \forall i \text{ dans } J_n. \quad (3.2)$$

Le choix de l'agent, en termes d'utilité, s'exprime comme suit ²:

$$\begin{aligned} U_{in} &> U_{jn} \quad \forall i \text{ dans } J_n, j \neq i \\ U(x_{in}, r_n) &> U(x_{jn}, r_n) \quad \forall i \text{ dans } J_n, j \neq i. \end{aligned} \quad (3.3)$$

L'interprétation est simple, l'agent choisit une alternative parmi celles disponibles si et seulement si cette alternative lui procure un niveau d'utilité supérieur aux autres alternatives.

Malheureusement, la fonction d'utilité est généralement inconnue, tout comme bon nombre de facteurs explicatifs. Nous séparons donc la fonction d'utilité en deux composantes, l'une déterministe et dépendant des facteurs observés et l'autre représentant tous les éléments non observés ou inconnus de l'utilité (ϵ_{in})

$$U_{in} = U(x_{in}, r_n) = V(z_{in}, s_n, \beta) + \epsilon_{in}, \quad (3.4)$$

où z_{in} est la partie observée de x_{in} et s_n la partie observée de r_n . Les termes non observés sont considérés aléatoires de densité conjointe $f(\epsilon)$.

Selon cette nouvelle spécification de la fonction d'utilité, la probabilité de choisir l'alternative i (P_{in}) devient la probabilité que l'équation (3.3) tienne, c'est-à-dire que l'utilité de l'alternative i soit supérieure à celle de n'importe quelle autre alternative (Train, 1986).

$$\begin{aligned} P_{in} &= \Pr(U_{in} > U_{jn} \quad \forall i \text{ dans } J_n, j \neq i) \\ &= \Pr(V_{in} + \epsilon_{in} > V_{jn} + \epsilon_{jn} \quad \forall i \text{ dans } J_n, j \neq i) \\ &= \Pr(\epsilon_{jn} - \epsilon_{in} < V_{in} - V_{jn} \quad \forall i \text{ dans } J_n, j \neq i). \end{aligned} \quad (3.5)$$

La différence, $V_{in} - V_{jn}$, observable, est calculée par le chercheur alors que la différence, $\epsilon_{jn} - \epsilon_{in}$, non observable, est une variable aléatoire variant parmi les agents ayant les

²La possibilité d'égalités ($U_{in} = U_{jn}$) ou ($U_{in} \geq U_{jn}$) est ignorée dans cette discussion. Nous sommes intéressés aux cas où un choix est effectivement réalisé.

mêmes composantes observées de l'utilité. Par définition, le côté droit de (3.5) est une fonction de distribution cumulative jointe : la probabilité qu'une variable aléatoire soit plus petite qu'une valeur réelle spécifiée.

Le processus comportemental peut être abordé différemment par la fonction suivante : $y = h(x, \epsilon)$, où x et ϵ ont les mêmes significations qu'auparavant. L'étape suivante est de définir une fonction indicatrice $I[h(x, \epsilon) = y]$ prenant la valeur 1 lorsque les éléments non observés incitent l'agent à choisir y et la valeur 0 sinon. Ainsi, la valeur espérée, sur toutes les valeurs possibles des facteurs non observés, représente la probabilité que l'agent choisisse y :

$$\begin{aligned} P(y|x) &= Pr(I[h(x, \epsilon) = y] = 1) \\ &= \int I[h(x, \epsilon) = y]f(\epsilon)d\epsilon. \end{aligned} \quad (3.6)$$

En utilisant la densité $f(\epsilon)$ et la fonction indicatrice, nous pouvons récrire la probabilité cumulative comme suit :

$$P_{in} = \int_{\epsilon} I(\epsilon_{jn} - \epsilon_{in} < V_{in} - V_{jn} \forall j \neq i) f(\epsilon_n) d\epsilon_n,$$

où la fonction indicatrice prend la valeur 1 si l'intérieur de la parenthèse est vrai et la valeur 0 sinon. La densité $f(\epsilon_n)$ est la distribution de la portion non observée de l'utilité parmi les agents faisant face à la même portion observée de l'utilité. La probabilité P_{in} est la fraction des agents qui choisissent l'alternative i parmi les agents qui font face à la même utilité observée, pour chaque alternative, que l'agent n (Train, 2003). Il est maintenant approprié de traiter des différentes spécifications de cette distribution, qui déterminent le type de modèle de choix. Débutons par le logit.

3.2 Logit

Le logit standard est le modèle de choix le plus simple à utiliser et il s'agit du modèle que nous utilisons pour notre analyse comportementale³. Sous l'hypothèse que chacun des ϵ_{in} est indépendamment, identiquement distribué (*iid*) valeur extrême (de loi Gumbel) pour tout i , le modèle logit peut être dérivé. Cet aspect crucial du logit, voulant que les facteurs non observés soient non corrélés et aient la même variance entre les alternatives, constitue une restriction, mais constitue aussi la source de sa simplicité en permettant une forme explicite (sans intégrale) aux probabilités de choix.

³En fait, nous utilisons un logit à plusieurs variables explicatives, appelé logit multinomial standard.

L'hypothèse de la distribution Gumbel *iid* des termes d'erreurs permet d'exprimer les probabilités de choix en fonction de la densité et de la distribution cumulative de ces termes d'erreur. La probabilité de choix logistique que l'agent n choisisse l'alternative i s'écrit⁴

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in J_n} e^{V_{jn}}} \quad \forall j \neq i. \quad (3.7)$$

La composante déterministe de l'utilité étant habituellement considérée linéaire dans les paramètres ($V_{in} = \beta w(z_{in}, s_n)$ ou $V_{in} = \beta w_{in}$), elle peut aussi s'écrire de la façon suivante :

$$P_{in} = \frac{e^{\beta w_{in}}}{\sum_j e^{\beta w_{jn}}},$$

où w est un vecteur en fonction des données observées et β est un vecteur de paramètres.

Les principales propriétés associées aux probabilités de choix logistiques sont qu'elles sont entièrement comprises entre zéro et un (comme se doit toute probabilité); elles s'additionnent à l'unité pour chaque individu, reflétant les critères d'exhaustivité et d'exclusion mutuelle des modèles de choix; et enfin, leur relation avec la composante déterministe de l'utilité pour une alternative donnée est de forme sigmoïde. Il en découle une propriété fondamentale du logit, voulant que le ratio des probabilités de choix de deux alternatives, i et k , soit indépendant de toutes les alternatives autre que i et k ⁵. Le ratio est donc indépendant par rapport aux alternatives non pertinentes (IIA, pour independent from irrelevant alternatives), c'est-à-dire autres que celles formant le ratio.

Une implication importante de la propriété IIA pour le modèle logit est que les effets de substitution se distribuent de façon proportionnelle à toutes les autres alternatives. La part relative de deux alternatives n'est fonction que des caractéristiques de ces deux alternatives. Une autre implication est qu'uniquement les différences d'utilité agissent sur les probabilités de choix. Le niveau absolu de l'utilité n'a pas d'impact sur celles-ci. Nous reviendrons sur cet élément à la section portant sur l'identification du modèle.

⁴Suivant Train (1986, 2003), cette probabilité est dérivée sachant que sous l'hypothèse de distribution de loi Gumbel *iid*, la densité et la distribution cumulative de chaque ϵ_{in} sont respectivement $f(\epsilon_{in}) = \exp(-\epsilon_{in}) \cdot \exp(-e^{-\epsilon_{in}})$ et $F(\epsilon_{in}) = \exp(-e^{-\epsilon_{in}})$. Rappelons que la probabilité du choix de i s'écrit $P_{in} = \Pr(\epsilon_{jn} < \epsilon_{in} + V_{in} - V_{jn}, \forall j \neq i)$ et que puisque les ϵ sont indépendants, nous pouvons écrire cette distribution cumulative sur tous les $j \neq i$ comme le produit des distributions cumulatives individuelles : $P_{in} | \epsilon_{in} = \prod_{j \neq i} \exp(-e^{-(\epsilon_{in} + V_{in} - V_{jn})})$. Avec les ϵ_{in} inconnus, la probabilité de choix devient $P_{in} = \int_{s=-\infty}^{\infty} (\prod_{j \neq i} \exp(-e^{-(s + V_{in} - V_{jn})})) \exp(-s) \cdot \exp(-e^{-s}) ds$ où s remplace ϵ_{in} . Puisque $V_{in} - V_{in} = 0$, nous pouvons rassembler les exposants de ϵ pour obtenir $P_{in} = \int_{s=-\infty}^{\infty} \exp(-e^{-s} \cdot \sum_j e^{-(V_{in} - V_{jn})}) e^{-s} ds$. Si nous posons $t = e^{-s}$ et $-e^{-s} ds = dt$, nous obtenons $P_{in} = \int_0^{\infty} \exp(-t \cdot \sum_j e^{-(V_{in} - V_{jn})}) dt = \frac{\exp(-t \cdot \sum_j e^{-(V_{in} - V_{jn})})}{-\sum_j e^{-(V_{in} - V_{jn})}} \Big|_0^{\infty} = \frac{e^{V_{in}}}{\sum_j e^{V_{jn}}}$ qui est bien la probabilité de choix logistique.

⁵La démonstration est simplement $\frac{P_{in}}{P_{kn}} = \frac{e^{V_{in}} / \sum_j e^{V_{jn}}}{e^{V_{kn}} / \sum_j e^{V_{jn}}} = \frac{e^{V_{in}}}{e^{V_{kn}}} = e^{V_{in} - V_{kn}}$.

L'estimation des probabilités logistiques par maximum de vraisemblance est rendue possible grâce à leur forme explicite. L'estimation du modèle logit standard s'effectue donc normalement en maximisant la vraisemblance des choix observés (échantillon) y_{in} des agents :

$$L(\beta) = \prod_{n=1}^N \prod_i P_{in}^{y_{in}},$$

où la vraisemblance d'une observation s'écrit $\prod_i P_{in}^{y_{in}}$ avec y_{in} égal à un si l'agent choisit l'alternative i et zéro sinon, et en considérant que les observations sont indépendantes entre les individus. Il est généralement plus aisé de maximiser la log-vraisemblance qui s'écrit

$$LL(\beta) = \sum_{n=1}^N \sum_i y_{in} \ln P_{in}.$$

Avec une linéarité dans les paramètres, cette fonction est globalement concave et des estimateurs peuvent être obtenus facilement à l'aide de programmes contenus dans bon nombre de logiciels statistiques.

3.3 Logit emboîté ou modèle GEV

Nous limiterons notre aperçu des modèles GEV (*generalized extreme value*) au cas du logit emboîté, la variante la plus largement utilisée de cette classe de modèles. Le modèle GEV a été dérivé du modèle d'utilité aléatoire par McFadden (1978) et se compose d'une grande famille de modèles incluant le logit multinomial et le logit emboîté (Bierlaire, 2001). Cette classe de modèles offre une variété de possibilités de substitution. Tous les modèles GEV partagent la particularité de permettre une corrélation des facteurs non observés entre les alternatives en accordant une distribution conjointe à valeur extrême généralisée à la composante non observée de l'utilité (termes d'erreur) pour toutes les alternatives. Le logit emboîté est donc une généralisation du logit standard et de sa distribution univariée à valeur extrême.

Le logit emboîté est intéressant lorsque l'indépendance par rapport aux alternatives non pertinentes rend le logit standard inapproprié, auxquels cas l'ensemble de choix d'un agent peut être divisé en sous-ensembles, appelés nids, avec l'IIA respectée à l'intérieur de chaque nid et non respectée entre les nids. Ainsi, pour deux alternatives appartenant au même nid, le ratio des probabilités est indépendant des caractéristiques des autres alternatives. L'IIA tient dans le nid. D'un autre côté, pour deux alternatives

appartenant à des nids différents, le ratio des probabilités peut dépendre des caractéristiques des autres alternatives dans les deux nids. L’IIA ne tient pas à travers les nids (Train, 2003).

Un exemple permettra de mieux saisir ces notions quelque peu abstraites. Prenons le cas d’un agent qui doit faire un choix quant au mode de transport pour se rendre au travail. Il a le choix entre prendre son auto, le taxi, l’autobus ou le métro. Supposons, dans cet exemple, que les probabilités pour l’auto et le taxi augmentent toujours de la même proportion lorsque les autres alternatives sont éliminées. Supposons aussi que les probabilités pour l’autobus et le métro augmentent de la même proportion suite à l’élimination des autres alternatives. Dans un tel cas, l’IIA tient entre l’auto et le taxi et entre l’autobus et le métro. Nous pouvons donc créer deux nids, l’un appelé « voiture » et l’autre nommé « transport en commun ». L’IIA est respectée dans chacun des nids mais pas entre les nids.

Ce qui distingue le logit emboîté du logit standard, outre l’introduction de K nids notés B_1, \dots, B_K au modèle, provient de l’hypothèse que le vecteur conjoint des termes d’erreur $\epsilon_{in} = (\epsilon_{1n}, \dots, \epsilon_{Jn})$ est supposé de distribution cumulative $\exp[-\sum_{k=1}^K (\sum_{i \in B_k} e^{-\epsilon_{in}/\lambda_k})^{\lambda_k}]$. Pour le logit, chacun des ϵ_{in} est indépendant de distribution Gumbel *iid*. Pour le logit emboîté, la distribution marginale de chacun des ϵ_{in} est univariée à valeur extrême. Cependant, les ϵ_{in} sont corrélés à l’intérieur des nids. Pour le reste, l’utilité obtenue de l’alternative i par l’agent n dans le nid B_k est dénotée de la même manière, soit $U_{in} = V_{in} + \epsilon_{in}$, où V_{in} est observée et ϵ_{in} n’est pas observé et aléatoire.

Au sein du nid k , le degré d’indépendance dans la part non observée de l’utilité (termes d’erreur) entre les alternatives est exprimé par le paramètre λ_k . Plus la valeur de ce paramètre est élevée, plus l’indépendance est forte et plus la corrélation est faible. Dans le cas où λ_k est égal à l’unité pour tous les k , il y a indépendance entre les alternatives dans tous les nids et la distribution GEV devient le produit de termes indépendants de distribution $F(\epsilon_{in}) = \exp(-e^{-\epsilon_{in}})$. Les modèles logit standard et emboîté sont identiques.

La probabilité de choix de l’alternative i dans le nid B_k devient

$$P_{in} = \frac{e^{V_{in}/\lambda_k} (\sum_{j \in B_k} e^{V_{jn}/\lambda_k})^{\lambda_k - 1}}{\sum_{l=1}^K (\sum_{j \in B_l} e^{V_{jn}/\lambda_l})^{\lambda_l}}. \quad (3.8)$$

Dans le cas où λ_k est égal à l’unité, l’indépendance est parfaite dans le nid k et nous retrouvons, en tant que cas particulier, les probabilités de choix du modèle logit standard.

Si nous considérons que le dénominateur de (3.8) est identique pour toutes les alternatives, alors le ratio des probabilités est le ratio des numérateurs. Ainsi, nous pouvons démontrer que l’IIA est respectée au sein de chaque nid mais pas entre les nids. Supposons deux alternatives, i et m , qui appartiennent respectivement aux nids B_k et B_l . Le ratio des probabilités s’écrit comme suit :

$$\frac{P_{in}}{P_{mn}} = \frac{e^{V_{in}/\lambda_k} (\sum_{j \in B_k} e^{V_{jn}/\lambda_k})^{\lambda_k - 1}}{e^{V_{mn}/\lambda_l} (\sum_{j \in B_l} e^{V_{jn}/\lambda_l})^{\lambda_l - 1}}.$$

Avec $l = k$, i et m sont dans le même nid et les termes entre parenthèses s’annulent pour donner

$$\frac{P_{in}}{P_{mn}} = \frac{e^{V_{in}/\lambda_k}}{e^{V_{mn}/\lambda_l}}.$$

Nous constatons que ce ratio est indépendant de toutes autres alternatives. Une dernière remarque concerne le fait que le ratio ne dépend pas des caractéristiques des alternatives se trouvant dans les nids autres que ceux contenant i et m (Train, 2003).

Tout comme le logit, les modèles GEV possèdent des expressions explicites et leurs paramètres peuvent être estimés par les procédures de maximum de vraisemblance standards. Le logit emboîté peut aussi être estimé, avec certaines complications, de façon séquentielle. Cette approche est fondée sur le fait que les probabilités de choix peuvent se décomposer en des probabilités logistiques marginales et conditionnelles. Il s’agit d’estimer en premier le modèle final consacré au choix de l’alternative au sein d’un nid. Ensuite, à l’aide des coefficients estimés, une valeur incluse est calculée pour chacun des modèles finaux. Enfin, le modèle primal, pour le choix du nid, est estimé avec la valeur incluse s’ajoutant aux variables explicatives. Pour en savoir davantage sur les complications liées à cette méthode ainsi que pour une discussion plus approfondie des modèles GEV, voir Train (1986, 2003), Ben-Akiva (1973) et Lee (1999).

3.4 Probit

Le modèle probit vise à contourner les restrictions posées par le logit, principalement en ce qui a trait à l’indépendance par rapport aux alternatives non pertinentes découlant de l’hypothèse que les facteurs non observés de l’utilité sont *iid*. Le probit relâche cette hypothèse en supposant que la densité conjointe de ces facteurs est de loi normale multivariée. Il permet à chaque ϵ_{in} pour tout i dans J_n de posséder une variance différente et d’être corrélé avec les autres ϵ_{jn} .

Plus explicitement, nous supposons que le vecteur des termes d'erreur est de distribution normale avec une espérance nulle et une matrice variance-covariance Ω : $\tilde{\epsilon}_n = (\epsilon_{1n}, \dots, \epsilon_{Jn}) \text{ NMV}(0, \Omega)$. La densité de $\tilde{\epsilon}_n$ est

$$\phi(\tilde{\epsilon}_n) = \frac{1}{(2\pi)^{J/2} |\Omega|^{1/2}} e^{-1/2 \tilde{\epsilon}_n \Omega^{-1} \tilde{\epsilon}_n},$$

où $|\Omega|$ est le déterminant de Ω . Rappelons que la probabilité du choix i est la probabilité que l'utilité associée au choix i soit supérieure à celle de tout autre choix : $P_{in} = Pr(\epsilon_{jn} < \epsilon_{in} + V_{in} - V_{jn} \forall j \neq i)$, où le côté droit de l'équation est une distribution cumulative. Rappelons aussi que l'intégrale d'une fonction de densité est une distribution cumulative. Sachant que la valeur de ϵ_{in} n'est pas connue, nous pouvons récrire la probabilité du choix i comme étant la probabilité de le choisir pour toute valeur donnée de ϵ_{in} intégrée sur toutes les valeurs possibles de ϵ_{in} , $P_{in} = \int P_{in}(\epsilon_{in}) \phi(\tilde{\epsilon}_n) d\epsilon_{in}$, où $P_{in}(\epsilon_{in})$ est la probabilité du choix i étant donnée une quelconque valeur possible de ϵ_{in} (la densité du vecteur aléatoire $\tilde{\epsilon}_n$ intégrée de $-\infty$ à $\epsilon_{in} + V_{in} - V_{jn}$).

Cette intégrale ne possède pas de forme explicite et doit être évaluée numériquement. Évidemment, une telle complexité de la forme des probabilités de choix constitue un important désavantage du modèle probit. Un grand nombre d'alternatives et de variables explicatives le rendent très coûteux à utiliser. Pour une discussion beaucoup plus approfondie, voir Train (2003).

3.5 Logit mixte (*kernel*)

Le logit mixte est le dernier modèle de choix traité car il est une généralisation des modèles précédents. Il suppose que la composante non observée de l'utilité consiste en une partie qui suit une distribution spécifiée par le chercheur, englobant toute corrélation et hétéroscédasticité, et une partie qui est Gumbel *iid*. Le logit mixte surmonte les limitations du logit en permettant la variation aléatoire des préférences, tous les scénarios de substitution et la corrélation temporelle des facteurs non observés, tout en évitant les restrictions d'une distribution normale.

Rappelons que l'utilité, pour l'agent n , de choisir l'alternative i s'exprime par $U_{in} = \beta_n X_{in} + \epsilon_{in}$ où X_{in} est un vecteur de variables explicatives observées et β_n et ϵ_{in} ne sont pas observés et sont traités comme ayant des effets stochastiques. Le logit mixte est spécifié en partitionnant la composante stochastique en deux parties non corrélées. L'une est hétéroscédastique et corrélée sur les alternatives, alors que l'autre est *iid* sur les alternatives et les agents : $U_{in} = \beta' x_{in} + [\eta_{in} + \epsilon_{in}]$ où η_{in} est un terme aléatoire

d'espérance nulle dont la distribution sur les agents et les alternatives dépend en général de paramètres implicites et de données observées en lien avec l'alternative i et l'agent n . ϵ_{in} est un terme aléatoire d'espérance nulle *iid* sur les alternatives et qui ne dépend pas de paramètres implicites.

Le terme η peut suivre une distribution normale, log-normale, triangulaire, etc. Sa densité est notée $f(\eta|\theta)$ où les θ sont les paramètres fixes de la distribution. Si la valeur de η est donnée, la probabilité de choix conditionnelle est le logit, car le terme d'erreur restant est Gumbel *iid* :

$$L_i = \frac{e^{\beta'x_i + \eta_i}}{\sum_j e^{\beta'x_j + \eta_j}}.$$

Puisque η n'est pas donné, la probabilité de choix est l'intégration de cette formule logit, sur toutes les valeurs de η , pondérée par la densité de η : $P_i = \int L_i(\eta) f(\eta|\theta) d\eta$. Les probabilités ne répondent pas à l'indépendance par rapport aux alternatives non pertinentes et différents scénarios de substitution sont obtenus par une spécification appropriée de $f(\cdot)$ (Greene et Hensher (2001)).

Le logit standard est simplement un cas particulier où la fonction de densité $f(\eta)$ est dégénérée à des paramètres fixes b : $f(\eta) = 1$ pour $\eta = b$ et 0 pour $\eta \neq b$ pour une utilité linéaire dans les paramètres (Train, 2003). Le logit mixte est un outil beaucoup plus puissant que le laisse entrevoir le bref traitement qui lui est dévoué dans cette section. Le lecteur intéressé devrait se tourner vers les travaux de Greene et Hensher (2001) et Train (2003).

3.6 Identification

Avant de procéder aux estimations, il est essentiel de s'assurer que le modèle de choix utilisé, le logit multinomial standard dans notre cas, est correctement spécifié ou identifié au niveau des paramètres. L'identification du modèle est une étape cruciale qui s'effectue selon deux propriétés fondamentales des modèles de choix, soit que seules les différences d'utilité importent et que l'échelle de l'utilité est arbitraire. Arrêtons-nous un moment pour discuter des implications de ces conditions sur l'identification.

3.6.1 Seules les différences d'utilité comptent

Cette condition revient simplement à dire que le choix de l'agent n est identique que l'utilité soit $U_{in} \forall i$ ou $U_{in} + c \forall i, c$. L'utilité étant un concept ordinal, une transformation monotone positive ne l'affecte pas. Rappelons que la probabilité de choix s'exprime par $P_{in} = \Pr(U_{in} > U_{jn} \forall j \neq i) = \Pr(\epsilon_{jn} - \epsilon_{in} < V_{in} - V_{jn} \forall j \neq i)$. Elle repose clairement sur des différences. Il en découle que seuls les paramètres capturant des différences entre les alternatives peuvent être identifiés et estimés.

Dans la plupart des cas, il est intéressant d'introduire des constantes spécifiques aux alternatives, puisque ces constantes capturent l'effet moyen sur l'utilité de tous les facteurs qui ne sont pas inclus dans le modèle. Notamment, pour le modèle logit standard, en plus d'engendrer une espérance nulle pour la composante non observée de l'utilité, l'inclusion de constantes spécifiques peut atténuer, voire éliminer, les inexactitudes causées par l'indépendance par rapport aux alternatives non pertinentes. À l'instar de l'utilité, seules les constantes d'une forme en différence peuvent être estimées. Cependant, il est impossible d'estimer une constante pour chaque alternative car une infinité de combinaisons de constantes ont les mêmes différences et procurent les mêmes probabilités de choix.

La façon de procéder est de normaliser les niveaux absolus des constantes en normalisant l'une des constantes à zéro (le choix est arbitraire). L'interprétation des autres constantes se fait par rapport à celle normalisée. La même procédure s'applique à toute variable qui ne varie pas avec les alternatives. Les variables sociodémographiques, comme le revenu, doivent donc être introduites dans le modèle seulement si elles sont spécifiées de façon à créer une différence d'utilité. Leurs coefficients sont eux aussi normalisés à zéro, sauf si elles interagissent avec les caractéristiques des alternatives. Cette procédure de normalisation est simplement démontrée par un exemple trichotomique ($i = 1, 2, 3$). Supposons

$$U_{1n} = C_{1n} + \beta_{1n}X_{1n} + \gamma_{1n}S_n + \epsilon_{1n} = V_{1n} + \epsilon_{1n}$$

$$U_{2n} = C_{2n} + \beta_{2n}X_{2n} + \gamma_{2n}S_n + \epsilon_{2n} = V_{2n} + \epsilon_{2n}$$

$$U_{3n} = C_{3n} + \beta_{3n}X_{3n} + \gamma_{3n}S_n + \epsilon_{3n} = V_{3n} + \epsilon_{3n}$$

où C_{in} est une constante spécifique à l'alternative i , β_{in} et γ_{in} sont des vecteurs de paramètres, X_{in} est un vecteur de caractéristiques propres à chaque alternative et S_n est un vecteur de caractéristiques sociodémographiques de l'agent n . Les γ_{in} capturent les effets des variations des variables sociodémographiques sur l'utilité de l'alternative i .

Puisque seulement les différences d'utilité comptent, les niveaux absolus des γ_{in} ne peuvent pas être estimés. Nous normalisons pour obtenir

$$\begin{aligned} U_{1n} &= \alpha_{1n} + \beta_{1n}X_{1n} + \omega_{1n}S_n + \epsilon_{1n} \\ U_{2n} &= \alpha_{2n} + \beta_{2n}X_{2n} + \omega_{2n}S_n + \epsilon_{2n} \\ U_{3n} &= \beta_{3n}X_{3n} + \epsilon_{3n} \end{aligned}$$

où $\alpha_{1n} = C_{1n} - C_{3n}$ et $\alpha_{2n} = C_{2n} - C_{3n}$ représentent respectivement les niveaux moyens de l'impact des facteurs non inclus dans le modèle sur l'utilité de l'alternative 1 par rapport à l'alternative 3 et de l'alternative 2 par rapport à 3. Les ω_{in} s'interprètent de la même façon. ω_{1n} mesure l'effet différentiel de S_n sur l'utilité de 1 par rapport à 3. Donc, si $\omega_{1n} > 0$, cela signifie que $\gamma_{1n} > \gamma_{3n}$.

3.6.2 L'échelle absolue de l'utilité est arbitraire

L'alternative avec l'utilité la plus élevée reste inchangée, peu importe l'ordre de grandeur de l'utilité. Selon notre notation, $U_{in} = V_{in} + \epsilon_{in} \forall i$ est équivalent à $U_{in}^* = cV_{in} + c\epsilon_{in} \forall i, c > 0$. L'échelle de l'utilité est normalisée par l'entremise de la variance des termes d'erreur puisque par définition, les deux sont reliées⁶. Normaliser la variance des termes d'erreur revient donc à normaliser l'échelle de l'utilité.

Pour le modèle logit, dans lequel les termes d'erreur sont supposés *iid*, la normalisation de la variance, à un nombre déterminé, de l'un de ces termes fixe la variance pour tous les autres. Par convention, les variances sont normalisées à $\pi^2/6$ dans un logit standard. Il est à noter que le problème d'échelle de l'utilité intervient lorsqu'il y a comparaison de résultats entre différents modèles ou différentes bases de données. Dans ces situations, la normalisation doit être considérée. D'autres situations, que nous ne discutons pas ici, où les erreurs sont hétéroscédastiques ou corrélées, font intervenir des procédures de normalisation beaucoup plus complexes. Nous laisserons le soin au lecteur intéressé à en savoir davantage sur le sujet à consulter, entre autres, le document de Train (2003).

⁶ $Var(c\epsilon_{in}) = c^2Var(\epsilon_{in})$.

3.7 Spécification du modèle

Nous avons introduit, dans les sections précédentes, des notions sur les modèles à choix discrets qui font partie intégrante de notre modélisation. Il est maintenant opportun de présenter plus en détail cette modélisation qui vise, rappelons-le, à analyser les effets sur la consommation totale annuelle d'électricité des ménages québécois de l'utilisation des thermostats réguliers, électroniques programmables et non-programmables. Nous incorporerons, par la même occasion, les notions concernant les méthodes dites discrètes-continues qui permettent une composante continue aux situations de choix.

Notre modèle se compose de trois étapes séquentielles. La première étape vise à estimer la probabilité de choix des thermostats utilisés sur un échantillon aléatoire de ménages québécois. La seconde étape estime l'usage, ou le niveau de la température ambiante fixée par les ménages, conditionnel au choix effectué à la première étape. Il s'agit d'un choix parmi un ensemble continu, et c'est pourquoi nous avons recours à une méthode discrète-continue afin de tenir compte de cette particularité. Ensemble, ces deux étapes considèrent l'aspect comportemental de notre analyse. L'étape finale consiste à régresser la consommation annuelle totale d'électricité sur trois composantes représentant la base, la possession d'appareils ménagers et les niveaux de chauffage obtenus à la seconde étape. Cette technique est dérivée de l'analyse conditionnelle de la demande introduite par Parti et Parti (1980).

3.7.1 Analyse comportementale

Chacun des ménages fait face à deux choix, choisir une alternative parmi un ensemble fini et exhaustif d'alternatives mutuellement exclusives (le type de thermostat) ; et déterminer l'usage relatif à une alternative particulière (le niveau de la température de chauffage fixée), avec l'usage représenté par une variable continue. Notre modèle doit donc décrire une telle situation de choix discret-continu. Afin de l'estimer, nous aurons recours à des méthodes qui ont été développées par Heckman (1978, 1979) et Dubin et McFadden (1984).

Notre objectif est de décrire la situation de choix d'un ménage en spécifiant la probabilité qu'il choisisse une alternative et la fonction d'usage qui en découle. Chaque ménage fait face à un ensemble de choix J ($i = 1, \dots, 7$). Les sept alternatives consistent à posséder des thermostats 1) réguliers seulement ; 2) électroniques non-programmables seulement ; 3) électroniques programmables seulement ; 4) électroniques non-programmables et programmables ; 5) réguliers et électroniques non-programmables ; 6) réguliers et

électroniques programmables; et 7) réguliers et électroniques non-programmables et programmables. Cette spécification fournit bien un ensemble fini et exhaustif d'alternatives mutuellement exclusives.

Notre modèle de choix est linéaire dans les paramètres et s'exprime comme suit pour l'agent n :

$$\begin{aligned}
 U_i &= \alpha_i + \beta_{i1} \cdot ql1 + \beta_{i2} \cdot enfant + \beta_{i3} \cdot hab1971_80 \\
 &+ \beta_{i4} \cdot hab1981_90 + \beta_{i5} \cdot hab1991_02 + \beta_{i6} \cdot bsl_gas \\
 &+ \beta_{i7} \cdot cotenord + \beta_{i8} \cdot sag_lac + \beta_{i9} \cdot qcmaurichau \\
 &+ \beta_{i10} \cdot regionmtl1 + \beta_{i11} \cdot vha100_200k + \beta_{i12} \cdot vha200k_ \\
 &+ \beta_{i13} \cdot revenu + \epsilon_i, \quad i = 2, \dots, 7 \text{ dans } J,
 \end{aligned}$$

où

- $ql1$: nombre de personnes dans le ménage
- $enfant$: présence d'au moins un enfant dans le ménage
- $hab1971_80$: l'habitation a été construite dans la décennie 70
- $hab1981_90$: l'habitation a été construite dans la décennie 80
- $hab1991_02$: l'habitation a été construite entre 1991 et l'année du sondage
- bsl_gas : Bas-Saint-Laurent-Gaspésie
- $cotenord$: Côte-Nord
- sag_lac : Saguenay-Lac-Saint-Jean
- $qcmaurichau$: Québec-Mauricie-Chaudière-Appalaches
- $regionmtl1$: Canton-de-l'Est-Outaouais-Lanaudière-Laurentides
- $vha100_200k$: valeur de l'habitation entre 100 000 \$ et 200 000 \$
- $vha200k_$: valeur de l'habitation 200 000 \$ et plus
- $revenu$: revenu total du ménage.

Les variables explicatives⁷ concernant l'année de construction et la valeur de l'habitation sont des *dummies*, c'est pourquoi nous n'avons pas inclus les variables sur les habitations construites avant 1971 et d'une valeur inférieure à 100 000 \$. Il est aussi important de noter que ce modèle ne possède pas d'équation pour l'alternative 1. Cela est dû au respect de la condition d'identification discutée précédemment. La base de données utilisée ne fournissant aucune information sur les caractéristiques des thermostats utilisés par les ménages, il nous est impossible de construire des variables

⁷Une liste complète des définitions de toutes les variables explicatives de notre modèle se retrouve à l'annexe B.

explicatives propres aux différentes alternatives⁸. Le modèle repose uniquement sur des variables socioéconomiques propres aux agents et puisque seulement les différences d'utilité comptent, nous normalisons pour obtenir un modèle à six équations.

Les deux niveaux de choix dépendent généralement de facteurs sous-jacents similaires qui les rendent interreliés. Cependant, si nous les considérons indépendants, une modélisation conjointe est inutile et chaque niveau peut être abordé séparément, en tenant compte de leur corrélation. Nous suivons cette approche et nous utilisons le modèle logit multinomial au premier niveau. Ce modèle, rappelons-le, exprime la probabilité de choix en fonction d'un classement ordinal des utilités (directes) associées à chaque alternative. Sachant que le deuxième niveau repose lui aussi sur des notions de maximisation de l'utilité, nous devons recourir à une forme d'utilité simple à manipuler et qui intègre les deux niveaux de choix. La fonction d'utilité indirecte fournit cela.

3.7.2 Utilité indirecte

La théorie microéconomique stipule que les préférences d'un agent peuvent être représentées de façon équivalente par une fonction d'utilité directe ou indirecte. Par contre, il est beaucoup plus facile de travailler avec l'utilité indirecte pour dériver des fonctions de demande. Cela est dû à un résultat, l'identité de Roy, qui établit que la demande marshallienne pour le bien i est le ratio des dérivées partielles de l'utilité indirecte par rapport au prix du bien i et au revenu, suite à un changement de signe.

$$-\frac{\partial v(\mathbf{p}, y)/\partial p_i}{\partial v(\mathbf{p}, y)/\partial y} = x_i(\mathbf{p}, y) \quad (3.9)$$

où $x_i(\mathbf{p}, y)$ est la fonction de demande marshallienne ou ordinaire et $v(\mathbf{p}, y)$ est la fonction d'utilité indirecte. Elle représente le niveau d'utilité de la courbe d'indifférence la plus élevée que l'agent peut atteindre, étant donné les prix \mathbf{p} et le revenu y ⁹.

Essentiellement, l'utilité directe dépend uniquement des quantités de chaque alternative alors que l'utilité indirecte dépend des prix des alternatives et du revenu. Nous pouvons appliquer l'utilité indirecte à notre situation par

$$v_i = v_i(p_i, y, z_i, s, \mu_i) \quad (3.10)$$

⁸Il aurait été possible d'inclure au modèle un prix moyen propre à chacune des alternatives. Cependant, il nous est apparu rapidement que cette option manquait de précision et qu'elle ne possédait pas de fondement économique valable. En effet, les prix des thermostats en cause étant minimes par rapport au revenu des ménages ainsi qu'à leurs dépenses les plus importantes, le pouvoir explicatif de telles variables était lui-même minime et discréditait leur utilisation.

⁹Voir l'annexe mathématique pour une discussion des notions pertinentes de la maximisation de l'utilité et de la relation entre utilité directe et indirecte.

qui est la fonction d'utilité indirecte conditionnelle pour l'alternative i . Elle donne l'utilité maximum qu'un ménage peut obtenir s'il choisit une certaine alternative. Les caractéristiques observées des alternatives sont notées par z_i , le niveau de température de chauffage par x , le revenu du ménage par y , les caractéristiques observées du ménage par s , le prix pour chaque alternative par p_i et tous les facteurs non observés par μ_i .

3.7.3 Modèle discret-continu

Si nous adaptons cette nouvelle forme d'utilité aux modèles à choix discrets, un ménage choisira l'alternative i si et seulement si l'utilité indirecte conditionnelle est plus élevée pour i que pour toute autre alternative :

$$v_i(p_i, y, z_i, s, \mu_i) > v_j(p_j, y, z_j, s, \mu_j) \quad \forall j \neq i.$$

La probabilité de choix devient

$$P_i = \Pr[v_i(p_i, y, z_i, s, \mu_i) > v_j(p_j, y, z_j, s, \mu_j)] \quad \forall j \neq i. \quad (3.11)$$

Étant donné que les μ_i présents dans l'utilité indirecte sont non observés, nous la séparons en deux composantes, l'une observée et l'autre non observée,

$$v_i(p_i, y, z_i, s, \mu_i) = V_i(p_i, y, z_i, s) + \epsilon_i,$$

et nous retrouvons la même notation qu'aux sections précédentes. Si nous supposons chacun des ϵ_i de loi Gumbel *iid*, nous obtenons le modèle logit standard, et les probabilités de choix s'écrivent

$$P_i = \frac{e^{V_i(p_i, y, z_i, s)}}{\sum_j e^{V_j(p_j, y, z_j, s)}}.$$

Chaque variable entrant dans l'utilité déterministe est exogène et les paramètres des probabilités de choix peuvent être estimés comme si la composante continue était inexistante.

L'usage relatif au choix i , la deuxième étape, est déterminé grâce à la fonction d'utilité indirecte conditionnelle et à l'identité de Roy. Ainsi, l'usage conditionnel au choix de l'alternative i est

$$x_i = x_i(p_i, y, z_i, s, \mu_i).$$

La deuxième étape consiste donc à estimer le niveau de la température de chauffage fixée par les ménages québécois. Pour les fins de l'analyse et afin d'être en mesure de tenir compte du concept de *setback* discuté précédemment, la journée est séparée en

trois périodes : le jour (07 h 00 à 17 h 00), le soir (17 h 00 à 23 h 00) et la nuit (23 h 00 à 07 h 00). Selon nos attentes, un *setback* devrait être appliqué par les ménages le jour, alors que la résidence est habituellement inoccupée, et la nuit alors que les occupants devraient moins se soucier de leur confort. Il s'agit aussi des périodes de la journée généralement favorisées pour des fins d'efficacité énergétique reliées au chauffage des locaux. Autrement dit, nous estimons le niveau de température intérieure fixée pour le jour, le soir et la nuit et nous croyons qu'il devrait être plus faible pour le jour et la nuit, reflétant l'application d'un *setback* par les ménages échantillonnés.

Nous avons adapté notre modélisation à ces particularités. Nous procédons à l'estimation de trois équations d'usage distinctes représentant les trois périodes de la journée. Aussi, afin de prendre en considération l'impact de la température extérieure dans la décision d'usage, nous avons incorporé une variable indiquant le nombre total de degrés-heures de chauffage par période, pour les mois de septembre à mai, pour huit grandes régions du territoire d'Hydro-Québec. Les autres variables présentes dans l'équation d'usage portent sur le nombre de personnes formant le ménage ; la présence d'enfant(s) ; le revenu total du ménage ; le niveau de scolarité du répondant ; et les corrections pour le biais d'auto-sélection (sujet de la prochaine section). Les trois équations d'usage utilisent les mêmes variables explicatives dans le but évident de garder une certaine logique entre elles.

Il est important de remarquer qu'un ménage choisirait une alternative différente si, en raison d'une variation dans un facteur sous-jacent, l'usage avait changé, et qu'un ménage choisirait un usage différent si, en raison d'une variation dans un facteur sous-jacent, le ménage optait pour une alternative différente (Train, 1986). Il ne s'agit pas d'une causation proprement dite entre les deux niveaux de choix, mais plutôt d'une corrélation entre les facteurs sous-jacents. L'estimation simultanée des paramètres des choix discrets et continus, bien que possible, pose des difficultés, et il est usuel de procéder séquentiellement. Cependant, nous devons tenir compte du fait que certaines des variables explicatives dans l'équation d'usage sont corrélées avec le terme d'erreur (ϵ_i corrélés avec les μ_i), engendrant des estimateurs par moindres carrés ordinaires biaisés. Nous devons corriger pour ce biais.

3.7.4 Méthodes de correction pour l'équation d'usage

Nous limiterons la discussion à l'approche par correction de sélectivité, pour laquelle nous avons opté. Il existe une autre approche, par instrumentation, qui peut être appliquée selon trois méthodes différentes. Ces méthodes reposent sur un remplacement des variables explicatives endogènes par les estimateurs de fonctions de variables exo-

gènes. Elles dépendent, ainsi, de la fonction de variables exogènes utilisée. Pour revenir à l'approche par correction de sélectivité, utilisons une version dichotomique simplifiée de notre modèle pour la présenter.

$$x_1 = \alpha_1 + \beta_1 y + \epsilon_1 \quad (3.12)$$

$$x_2 = \alpha_2 + \beta_2 y + \epsilon_2 \quad (3.13)$$

Cette spécification comporte des paramètres qui diffèrent entre équations et qui contient seulement le revenu du ménage comme variable explicative.

Si nous procédions à l'estimation de la première équation sur les ménages ayant uniquement des thermostats réguliers (1) et la deuxième équation sur ceux ayant uniquement des thermostats électroniques non-programmables (2), les estimateurs seraient biaisés. Par exemple, un ménage à faible revenu aurait une faible probabilité de choisir un thermostat électronique qui est relativement plus cher. Si un ménage choisit un tel thermostat, alors des facteurs non observés tels qu'un plus grand souci pour l'efficacité, doivent l'avoir incité à agir ainsi. Ces mêmes facteurs inciteraient aussi le ménage à choisir une température de chauffage plus basse (plus faible consommation d'électricité). Lorsque le revenu est faible pour un ménage qui a choisi un thermostat électronique, ce ménage doit choisir une température de chauffage plus basse que prévue afin de l'inciter à se munir d'un thermostat électronique. Un tel biais est appelé biais d'auto-sélection car l'estimation est effectuée sur un sous-échantillon de ménages qui se sont eux-mêmes sélectionnés pour l'estimation par leur choix de l'alternative.

Le revenu est seulement un exemple. Le véritable problème concerne les probabilités de choix. Une variable qui affecte la probabilité de choix P_2 est corrélée, dans le sous-échantillon ayant choisi le thermostat électronique, avec le terme d'erreur ϵ_2 dans l'équation d'usage. Plus explicitement, l'espérance de ϵ_2 n'est pas nulle pour chaque observation, comme l'exige les moindres carrés ordinaires. Elle est plutôt une fonction de la probabilité de choix P_2 . La manière de procéder pour régler ce problème est de décomposer le terme d'erreur en son espérance et une déviation de son espérance :

$$\epsilon_i = E(\epsilon_i) + \eta,$$

où $E(\epsilon_i)$ est en fonction de la probabilité de choix de i . Le terme de déviation η n'est pas relié au choix entre les alternatives et est indépendant de P_i . Nous pouvons récrire l'équation d'usage par

$$x_i = \alpha_i + \beta_i y + E(\epsilon_i) + \eta. \quad (3.14)$$

$E(\eta) = 0$ et y et $E(\epsilon_i)$ sont indépendants de η , donc cette équation peut être estimée par moindres carrés ordinaires. Le terme $E(\epsilon_i)$ est appelé correction de sélectivité puisqu'il corrige pour le biais de sélection lorsqu'il est inclus dans l'équation d'usage.

Les travaux de Heckman, et par la suite de Dubin et McFadden (1984), ont démontré que si les probabilités de choix sont logistiques et que les ϵ_i sont normalement distribués, alors la correction de sélectivité s'exprime par

$$\begin{aligned} E(\epsilon_i) &= \frac{\sqrt{6\sigma^2}}{\pi} \left[\sum_{j=1}^J \frac{R_j \ln P_j}{(1 - P_j)} - \frac{R_i \ln P_i}{(1 - P_i)} \right] \\ &= \sum_{j \neq i}^J (\sqrt{6\sigma^2} R_j) \left[\frac{P_j \ln P_j}{1 - P_j} + \ln P_i \right] \end{aligned} \quad (3.15)$$

où σ^2 est la variance de ϵ dans la population entière et R_i est la corrélation de η et ϵ_i . La distribution de η conditionnelle aux $(\epsilon_1, \dots, \epsilon_J)$ est supposée d'espérance $(\sqrt{2\sigma^2}/\lambda) \sum_{i=1}^J R_i \epsilon_i$ et de variance $\sigma^2(1 - \sum_{i=1}^J R_i^2)$ où $\sum_{i=1}^J R_i = 0$ et $\sum_{i=1}^J R_i^2 < 1$. Donc, avec N alternatives dans l'ensemble de choix J , il y a $N - 1$ termes de correction de sélectivité à introduire à la régression d'usage, avec les coefficients $(\sqrt{6\sigma^2}/\pi)R_j$ pour chaque $j \neq i$.

Nous disposons d'une estimation des probabilités de choix et nous pouvons donc calculer le terme entre crochets. L'équation d'usage s'exprime maintenant par

$$x_i = \alpha_i + \beta_i y + \gamma_i C_i + \eta, \quad (3.16)$$

où C_i est le terme de correction de sélectivité pour l'alternative i et γ_i est le coefficient du terme de correction, prenant la valeur $(\sqrt{6\sigma^2}/\pi)R_j$.

Plus spécifiquement, notre modèle d'usage, identique pour les trois périodes de la journée, consiste à estimer la température intérieure maintenue le jour, la nuit et le soir en régressant les variables appropriées de la base de données sur différentes variables explicatives et sur les corrections pour le biais d'auto-sélection. Le modèle s'exprime de la façon suivante

$$\begin{aligned} temperature_{j,s,n} &= \alpha + \beta_1 \cdot total(J, S, N)reel + \beta_2 \cdot enfant + \beta_3 \cdot ql1 + \beta_4 \cdot revenu \\ &+ \beta_5 \cdot primaire + \beta_6 \cdot collegial + \beta_7 \cdot universite + \gamma_2 \cdot corr2 \\ &+ \gamma_3 \cdot corr3 + \gamma_4 \cdot corr4 + \gamma_5 \cdot corr5 + \gamma_6 \cdot corr6 + \gamma_7 \cdot corr7 + \eta. \end{aligned}$$

Note : dans ce qui précède, nous faisons l'hypothèse que les agents choisissent le niveau de la température ambiante parmi un ensemble continu d'alternatives. Cela nous permet d'avoir recours au modèle de choix discret-continu présenté plus haut. Théoriquement, cette hypothèse n'est pas fautive : les agents pourraient choisir n'importe quelle température entre 15°C et 25°C et l'ensemble serait bel et bien continu. Or, dans les faits, la situation est différente. À l'exception du thermostat régulier, sur lequel la

température est fixée sans grande précision et pourrait alors être exprimée par un ensemble continu, les autres thermostats (électroniques) offrent la possibilité de fixer la température à l'unité seulement. Cela fait en sorte de restreindre l'ensemble d'alternatives, qui n'est plus continu. De plus, nous avons observé que les ménages interrogés sur la température *moyenne* maintenue pendant les mois de chauffage, restreignent leurs réponses à une demie unité près (par exemple, 15, 15,5 ou 16°C)¹⁰.

Comment pourrions-nous estimer notre modèle sachant que les deux situations de choix sont discrètes ? Une façon de procéder serait de considérer une seule situation de choix où le choix de la température intérieure fixée serait conditionnel au choix du type de thermostat. Si nous nous rapportons à ce qui a été dit au chapitre 3, le modèle logit emboîté offre un tel cadre d'analyse. Le *cross nested* logit, un cas particulier du logit emboîté, serait lui aussi un modèle intéressant puisqu'il permettrait de tenir compte explicitement de la présence d'un même type de thermostat dans plusieurs alternatives différentes. Ces analyses ne sont pas possibles dans le cadre du présent travail, mais il serait très intéressant de vérifier, éventuellement, leur pouvoir explicatif.

3.7.5 Consommation annuelle totale

La troisième et dernière étape de notre modèle vise à faire ressortir l'effet sur la consommation annuelle totale d'électricité de l'usage de différents thermostats et de variables associées à la base et à divers appareils électriques. C'est à cette étape que l'efficacité énergétique due aux thermostats électroniques doit ressortir, s'il y a lieu. Nous ne procédons pas à une véritable analyse conditionnelle de la demande (ACD), mais nous nous inspirons fortement de l'approche développée par Parti et Parti (1980) à ce sujet.

Voici une brève description de cette méthode. L'ACD consiste en une méthode de régression ayant pour but de désagréger la consommation totale d'énergie (électrique) de chaque ménage échantillonné à travers l'ensemble des équipements électriques existants et des différentes caractéristiques des ménages et de leur résidence pouvant influencer la demande en électricité. La consommation totale d'électricité d'un ménage se traduit ainsi par une fonction linéaire des équipements électriques qu'il possède et d'une interaction de certaines caractéristiques exogènes des ménages (Lacroix, 2004).

¹⁰Nous verrons que les résultats obtenus à l'aide du modèle discret-continu ne sont pas très convaincants. Il y a fort à parier que la nature (véritable) discrète de l'ensemble de choix que nous tentons d'estimer de façon continue contribue à ce phénomène.

De façon plus explicite, cela se traduit par

$$E_i = \sum_{j=0}^M b_{ij} \cdot H_j + \epsilon_i = b_{i0} + \sum_{j=1}^M b_{ij} \cdot H_j + \epsilon_i, \quad i = 1, \dots, N, \quad (3.17)$$

où E_i représente les consommations unitaires, A_i est une variable indicatrice, H_j est une variable exogène caractérisant le ménage ou son habitation et b_{ij} est un paramètre associé à l'interaction entre A_i et H_j , sur le nombre d'usages électriques N et le nombre de variables exogènes M .

Les consommations unitaires E_i sont inconnues et, sachant que la consommation totale d'électricité des ménages est la seule variable connue, l'ACD permet d'estimer les paramètres de l'équation (3.17) en supposant que la consommation totale est égale à la somme des consommations unitaires :

$$E = E_0 + \sum_{i=1}^N E_i + \epsilon. \quad (3.18)$$

E_0 est la consommation d'électricité des usages non définis, c'est-à-dire des appareils ayant des taux de pénétration trop élevés ou ayant des consommations relativement faibles (Lacroix, 2004).

Selon cette approche, la consommation totale se représente donc par une fonction de demande qui est conditionnelle à la possession de l'appareil :

$$E_i = A_i \cdot [f_i(V) + \epsilon_i], \quad i = 1, \dots, N, \quad (3.19)$$

où A_i est une variable dichotomique représentant chacun des usages définis et prenant la valeur 1 pour les ménages possédant l'appareil i et 0 sinon. L'hypothèse selon laquelle l'équation (3.19) est linéaire permet de récrire l'équation (3.18) sous la forme suivante

$$E = \sum_{i=0}^N \sum_{j=0}^M b_{ij} (A_i V_j) + \epsilon. \quad (3.20)$$

La demande pour le chauffage des locaux n'est pas directement observable, c'est pourquoi nous utilisons les éléments théoriques de l'ACD afin de faire ressortir les effets du chauffage des locaux, des appareils électriques et des caractéristiques des ménages et de leur habitation sur la consommation annuelle totale. De cette façon, nous supposons que la consommation totale est simplement la somme des trois composantes, avec les résultats du modèle discret-continu (deuxième étape) entrant dans la composante sur le chauffage.

Les variables explicatives formant la composante sur les caractéristiques des ménages et de leur habitation vont du nombre de personnes et du revenu total du ménage à l'âge et à la valeur de l'habitation. Du côté des appareils électriques, les différentes conditions imposées à notre base de données font en sorte de limiter le nombre de variables disponibles. Nous considérons l'usage d'un chauffe piscine électrique, d'une cuisinière électrique avec auto-nettoyant, le nombre de réfrigérateurs à dégivrage manuel, l'usage d'une pompe à eau électrique (puits artésien ou sub-pompe), et l'usage d'un système d'appoint au bois, à l'huile et au gaz. Les autres variables disponibles ont soit un taux de pénétration trop élevé, soit une consommation trop faible parmi les ménages échantillonnés restants.

La composante pour le chauffage repose sur des variables construites à l'aide des résultats des estimations du modèle discret-continu. À cette étape, nous avons obtenu des estimations pour la température intérieure moyenne fixée par les agents pour le jour, le soir et la nuit, le tout pour les mois de chauffage de septembre à mai. À eux seuls, ces résultats ne sont pas suffisants pour expliquer le chauffage des locaux. Nous devons aussi tenir compte, à cet effet, de la superficie à chauffer ainsi que de la température extérieure pour les mêmes mois de chauffage. Tous ces éléments sont nécessaires à la représentation du besoin en chauffage des locaux.

Le chauffage, pour chaque période de la journée, se représente donc par la différence entre la température intérieure calculée à la seconde étape (niveau de référence) et la température (horaire) réelle extérieure, multipliée par la superficie de l'habitation. La mesure exprimée par cette différence de températures représente le degré-heure, que nous présentons dans le chapitre qui suit. La variable est complétée en multipliant la superficie aux degrés-heures totaux de jour, de soir et de nuit.

De plus, nous suspectons cette variable d'être corrélée avec le terme d'erreur dans la régression de la consommation totale. Dans ce cas, si les moindres carrés ordinaires sont utilisés, tous les estimateurs seront biaisés. C'est pourquoi nous procédons à une estimation par variables instrumentales du modèle de consommation totale. Nous terminerons cette section en discutant brièvement de cette méthode.

Supposons un modèle linéaire standard à deux variables explicatives ¹¹ :

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \mu_1. \quad (3.21)$$

La variable dépendante y_1 est endogène (corrélée avec μ_1). Nous supposons que $E(\mu_1) = 0$, que la variable z_1 est exogène dans (3.21) (z_1 est non corrélée avec μ_1) et que la variable y_2 est probablement corrélée avec μ_1 . En général, il est suffisant de considérer que μ_1 contient des variables omises corrélées avec y_2 .

Si nous estimons (3.21) par moindres carrés ordinaires (MCO), nous obtenons des estimateurs biaisés. Il nous faut donc d'autres variables exogènes (z_2 et z_3) qui n'entrent pas dans (3.21), qui ne sont pas corrélées avec μ_1 et qui sont corrélées avec les variables pour lesquelles elles sont des instruments. Puisque z_1 , z_2 et z_3 sont non corrélées avec μ_1 , une combinaison linéaire des variables exogènes est elle aussi non corrélée avec μ_1 et donc une variable instrumentale valide.

Nous pouvons écrire

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2, \quad (3.22)$$

où

$$E(v_2) = 0, \text{ Cov}(z_1, v_2) = 0, \text{ Cov}(z_2, v_2) = 0 \text{ and } \text{Cov}(z_3, v_2) = 0. \quad (3.23)$$

La variable instrumentale pour y_2 est la combinaison linéaire des z_j dans (3.22) que nous notons y_2^* :

$$y_2^* = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3. \quad (3.24)$$

Pour que cette variable instrumentale ne soit pas parfaitement corrélée avec z_1 , π_2 et/ou π_3 ne doit pas être nul.

Ainsi, l'équation (3.22) peut être vue comme un moyen de séparer y_2 en deux parties. La première est y_2^* (non corrélée avec le terme d'erreur) et la seconde est v_2 (peut-être corrélée avec le terme d'erreur). La première étape, la forme réduite, peut être estimée par MCO afin d'obtenir les valeurs prédites :

$$\hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3, \quad (3.25)$$

qui sont utilisées comme variable instrumentale pour y_2 . Avec plusieurs instruments, déduits d'un modèle complet, l'estimateur par variable instrumentale est aussi appelé estimateur 2SLS (*two stage least squares*).

¹¹Les propos qui suivent s'inspirent de Woolridge (2000), chap. 15

Finalement, pour que les écarts-type et les statistiques t soient asymptotiquement valides, nous supposons que la variance de l'erreur μ_1 ne peut dépendre d'aucune des variables exogènes (homoscédasticité). Lorsqu'il y a plus d'une variable explicative endogène dans le modèle, une condition d'identification nécessaire, appelée condition d'ordre, stipule qu'il faut au moins autant de variables exogènes exclues qu'il y a de variables explicatives endogènes incluses dans l'équation structurelle.

Revenons à notre modèle pour la consommation totale annuelle. L'équation structurelle s'exprime comme suit :

$$cons_an = \beta_0 + \beta_1 Y + \beta_2 Z + \mu, \quad (3.26)$$

où Y est le vecteur des variables explicatives endogènes et X est le vecteur des variables explicatives exogènes. Les variables endogènes sont au nombre de trois. Elles représentent les variables sur le chauffage, formées de la différence entre la température intérieure calculée à la seconde étape et la température réelle extérieure, multipliée par la superficie de l'habitation, pour le jour, le soir et la nuit. La condition nécessaire d'ordre est respectée si nous utilisons trois variables exogènes absentes de (3.26) mais corrélées avec celles-ci. Nos variables instrumentales sont identiques aux variables endogènes, à l'exception qu'elles sont formées de la différence entre une température de référence (18°C, exogène) et la température réelle extérieure, multipliée par la superficie de l'habitation, pour le jour, le soir et la nuit. Nous estimons notre équation structurelle par 2SLS.

Avant de discuter des résultats obtenus à chacune des étapes de notre modélisation, présentées dans le présent chapitre, nous procédons à un survol des données utilisées, des modifications apportées à celles-ci ainsi que des diverses limitations qui découlent de leur utilisation.

Chapitre 4

Données

Les données utilisées aux fins de l'analyse proviennent essentiellement de la banque de données associée à l'étude d'Hydro-Québec, *Utilisation de l'électricité dans le marché résidentiel, édition 2002*. Toutes les informations concernant les ménages, soit les caractéristiques de l'habitation, l'utilisation des équipements électriques, le chauffage des locaux, les thermostats et les caractéristiques du ménage sont tirées de cette étude, garantissant une certaine homogénéité. Nous avons aussi eu recours à des données climatiques rendues disponibles via le site internet d'Environnement Canada. Ces données, exogènes, ont été traitées et fusionnées à l'étude d'Hydro-Québec afin de disposer d'une mesure de la température extérieure.

La cueillette de la version 2002 de l'étude d'Hydro-Québec a été réalisée, du 14 février au 18 mars 2002, à l'aide d'un sondage téléphonique. Au total, 10 047 entrevues ont été complétées. La sélection des répondants a été effectuée de façon aléatoire selon une liste d'abonnés correspondant à la distribution réelle de l'ensemble des inscriptions au fichier du Système de Gestion des Abonnements. C'est donc dire que nous disposons d'une base de données obtenues par un échantillonnage aléatoire de la population. Cette propriété nous permet donc de recourir aux méthodes d'estimation basées sur les échantillons aléatoires.

Le sondage a été effectué auprès de clients résidentiels à mesurage individuel assujettis aux tarifs D, DT et DH, propriétaires ou locataires. Cela signifie que l'ensemble des ménages sondés ont un compteur propre à leur unité de logement et paient directement leur électricité à Hydro-Québec. Ces ménages se regroupent en cinq grands territoires : Île-de-Montréal (27 %) ; Laurentides (24 %) ; Montmorency (18 %) ; Nord-Est (9 %) ; et Richelieu (22 %).

Les ménages échantillonnés sont propriétaires de la résidence dans laquelle ils habitent dans 60 % des cas. De ces propriétaires, 55 % habitent dans une résidence individuelle et moins de la moitié (46 %) possèdent une habitation de moins de 100 000 \$. Les ménages logeant dans des maisons individuelles sont les plus grands consommateurs d'électricité avec une consommation moyenne de 20 417 kWh par année par rapport aux 17 104 kWh de l'ensemble de l'échantillon. Cela s'explique grandement par le nombre de pièces dans les habitations et par la diffusion des appareils électriques. La consommation moyenne annuelle d'électricité est aussi plus élevée chez les propriétaires (19 775 kWh) que chez les locataires (12 106 kWh).

L'âge des résidences (34 ans en moyenne) affecte aussi la consommation annuelle moyenne. En général, cette dernière est plus élevée pour les résidences construites après 1970. Ceci s'explique par le fait que les résidences construites après 1970 utilisent l'électricité comme source principale d'énergie pour le chauffage des locaux dans 77 % des cas alors que celles construites avant 1970 l'utilisent dans 57 % des cas (Hydro-Québec, 2002). Les consommations moyennes correspondent à la moyenne des consommations facturées pour l'année 2001, ajustée à 365 jours.

Plus généralement, l'échantillon partage les caractéristiques démographiques de la population québécoise. Un peu moins de la moitié des répondants (47 %) se situent dans le groupe d'âge 35 à 54 ans, alors qu'un peu plus du quart (28 %) ont 55 ans et plus. En moyenne, les ménages se composent de 2,6 personnes. Sur la base des réponses obtenues (26 % des répondants n'ont pas répondu), il semble que la plus forte proportion des ménages se situe dans la strate comprise entre 30 000 et 60 000 \$ de revenu familial (29 %).

Nous y apprenons, notamment, que la source principale d'énergie la plus courante pour le chauffage des locaux est l'électricité (67 %), suivie par le mazout (12 %), le bois (seul ou combiné avec une autre source d'énergie) (10 %), le gaz naturel (5 %) et la bi-énergie (4 %). Suite à une hausse entre 1990 et 1994, le taux de diffusion de l'électricité (excluant la bi-énergie) est demeuré stable de 1994 à 2002 (62 % en 1990, 67 % en 1994, 66 % en 1999 et 67 % en 2002). Dans les maisons individuelles, l'électricité est moins présente qu'elle ne l'était au milieu des années '90 (1994 : 64 %, 1999 : 58 %, et 2002 : 60 %) alors qu'elle est en hausse dans les maisons en rangée (1994 : 71 %, 1999 : 76 %, et 2002 : 78 %) et dans les duplex et triplex (1994 ; 1999 ; 2002 : duplex : 66 %, 69 % et 73 % ; triplex : 75 %, 77 % et 79 %). Le gaz naturel, comme le mazout, continue de décliner en 2002 (5 %).

Les plinthes électriques constituent le système de chauffage principal des locaux le plus répandu (56 %). Il est suivi par les systèmes centraux à air chaud (13 %) et à eau chaude (10 %) ainsi que par les fournaies centrales ou individuelles au bois et les poêles à bois (9 %) et enfin les thermopompes (5 %). Ces parts de marché ont peu bougé au cours de la dernière décennie. En 2002, pour contrôler la température dans leur résidence, la majorité des ménages (63 %) utilisent des thermostats muraux (par pièce-étage). Enfin, pour l'ensemble des ménages, la température moyenne ambiante fixée varie entre 19°C et 20°C selon la période de la journée pour les jours de semaine. Au total, 45 % des ménages affirment abaisser la température lorsqu'ils s'absentent de leur résidence.

Du côté du chauffage de l'eau à l'aide de chauffe-eau individuels, la situation est relativement stable depuis 1990, avec l'électricité dominant fortement les sources concurrentes avec une diffusion de l'ordre de 90 % en 2002. Enfin, en ce qui concerne les électroménagers, il ressort que les fours à micro-ondes poursuivent leur progression, si bien qu'en 2002, 92 % des ménages en possèdent un. Tous les ménages disposent d'au moins un réfrigérateur, alors qu'un ménage sur deux possède au moins un congélateur indépendant du réfrigérateur (52 %).

Les différentes caractéristiques de l'échantillon, surtout en ce qui concerne la source d'énergie et le système principalement utilisé pour le chauffage des locaux, justifient la pertinence d'étudier l'usage des thermostats électroniques pour contrôler le chauffage des locaux par des plinthes électriques dans le secteur résidentiel. De plus, ces caractéristiques sont d'une grande utilité lors de la troisième étape de notre modèle qui vise à estimer la consommation moyenne annuelle d'électricité. Nous pouvons, à l'aide des taux de diffusion, déterminer les variables à inclure ou à exclure du modèle et vérifier, dans certains cas, notre intuition sur les effets attendus de diverses variables explicatives.

Toutes les observations n'étant pas pertinentes aux objectifs d'estimation visés, nous avons procédé à un tri et à l'exclusion de plusieurs variables. Seulement les observations associées aux résidences principales unifamiliales, plex ou en rangée, à l'électricité seule comme principale source d'énergie pour le chauffage des locaux ainsi qu'avec des plinthes électriques comme système de chauffage principal à l'électricité ont été retenues. De plus, nous avons supprimé les observations (inutiles) reliées aux ménages utilisant des thermostats centraux ou intégrés à la plinthe ainsi que celles associées aux répondants qui ne savent pas ou ne répondent pas aux questions sur l'utilisation des thermostats électroniques et réguliers. Autrement dit, des 10 047 observations initiales, nous conservons uniquement celles qui possèdent toutes les caractéristiques précédentes, ce qui réduit notre échantillon à 3 045 observations. Nous les appellerons « ménages types ».

Par la suite, nous avons classé les observations selon les types de thermostats utilisés. De tous les ménages types utilisant au moins un thermostat mural, nous distinguons ceux ayant des thermostats réguliers de ceux ayant des thermostats électroniques (la possibilité existe qu'un ménage possède les deux à la fois). La dernière étape consiste à subdiviser à leur tour les ménages possédant des thermostats électroniques en deux sous-groupes, ceux ayant au moins un thermostat électronique programmable et ceux ayant au moins un thermostat électronique non-programmable. Cette démarche mène directement à notre ensemble de choix discrets constitué de sept alternatives.

La réduction éventuelle de notre échantillon à 2 170 observations résulte simplement de l'élimination de données manquantes ainsi que de réponses inutiles pour des variables essentielles qui étaient difficilement imputables : nombre de personnes habitant le foyer, âge du répondant, scolarité du répondant, âge et valeur de l'habitation. Néanmoins, d'autres modifications s'imposèrent afin de procéder aux estimations. Les plus importantes concernent le revenu du ménage, le lieu de résidence et la température extérieure par région.

L'obstacle le plus important posé par notre base de données concerne le revenu. Typiquement, les répondants sont peu enclins à divulguer leur revenu annuel (dans le sondage d'Hydro-Québec, ils sont 26 % à ne pas répondre à cette question). Néanmoins, cette variable est cruciale à toute analyse microéconomique et il est essentiel de disposer de valeurs à son sujet. Nous avons adopté deux méthodes différentes afin de combler les réponses manquantes ou encore les abstentions des ménages échantillonnés. Une manière de procéder fut d'imputer les valeurs manquantes en régressant les valeurs disponibles du revenu sur différentes variables socio-économiques des ménages avant de remplacer ces valeurs manquantes. Nous avons aussi appliqué le revenu moyen, associé aux différentes régions créées à cette fin.

La représentation des ménages selon les cinq territoires présentés plus haut n'était pas appropriée à nos besoins. Nous avons procédé à un nouveau regroupement des ménages qui tient compte d'un découpage géographique davantage désagrégé et réaliste. Ces nouvelles régions permettent de mieux capturer l'effet du lieu de résidence des ménages sur leurs habitudes de consommation et sur leur comportement. Il est maintenant aussi possible d'attribuer certaines données météorologiques externes propres à ces régions, données indispensables lorsque l'objectif est d'expliquer les habitudes de chauffage. Les huit grandes régions sont : Bas-Saint-Laurent-Gaspésie, Côte-Nord, Saguenay-Lac-Saint-Jean, Nord-du-Québec, Québec-Mauricie-Chaudière-Appalaches, Canton-de-l'Est-Outaouais-Lanaudière-Laurentides, Montréal-Laval-Montérégie-Haut-Saint-Laurent-Centre-du-Québec et Abitibi-Témiscamingue.

Nous l'avons mentionné plus haut, le chauffage est directement relié à la température intérieure et extérieure ainsi qu'à la superficie de l'habitation. Notre base de données ne contient aucune information sur la température extérieure et les données à ce sujet doivent être importées. Nous utilisons les données climatiques du Service météorologique du Canada (SMC) rendues disponibles sur le site internet d'Environnement Canada. Nous nous intéressons particulièrement aux degrés-jours et aux degrés-heures calculés pour les stations météorologiques associées aux régions d'intérêt. Les degrés-jours entrent dans l'estimation des probabilités de choix en tant que variable reflétant la sévérité du climat. Les degrés-heures interviennent, quant à eux, dans l'équation d'usage et dans la régression de la consommation totale d'électricité. Ils servent eux aussi d'indicateurs de la rigueur du climat, mais à la différence qu'ils permettent d'introduire une précision supplémentaire selon la période de la journée. Nous espérons ainsi être en mesure de faire ressortir l'impact du climat sur la demande en chauffage le jour, le soir et la nuit.

Le degré-jour se définit par le degré d'écart entre la température moyenne d'une journée et un seuil de référence donné. Les degrés-jours de chauffage sont la somme des écarts journaliers entre la température extérieure moyenne et la température intérieure de référence (18°C) calculée tout au long des jours de chauffage. Si la température est égale ou supérieure à 18°C, alors leur nombre est nul. C'est donc dire qu'un nombre élevé de degrés-jours indique une moyenne de température extérieure basse et un climat plus rigoureux et, selon nous, un usage différent des thermostats par les ménages. Plus la température extérieure est basse, plus la consommation d'énergie nécessaire pour les besoins en chaleur est élevée.

Le degré-heure se définit et s'interprète de façon similaire. Par contre, l'utilisation que nous en faisons est différente. D'abord, dans le but de tenir compte des périodes de *setback*, nous construisons des mesures de degrés-heures de jour, de soir et de nuit pour les mois de chauffage (septembre à mai) pour chaque station représentative. Ensuite, nous additionnons ces degrés-heures afin d'obtenir des degrés-heures totaux de jour, de soir et de nuit. À cette dernière étape, nous spécifions trois variables différentes, l'une basée sur le seuil théorique de 18°C; une autre fondée sur la température intérieure de chauffage fournie par les répondants du sondage; et une dernière reposant sur les résultats de l'estimation d'usage (l'estimation de la réponse précédente) comme seuil de référence. Cette façon de procéder nous permet de régresser la consommation totale d'électricité à l'aide d'une méthode par variables instrumentales tel que mentionné au chapitre précédent.

Nous terminerons notre discussion sur les données en exposant un problème important qui s'est révélé à nous lors de l'estimation de l'équation d'usage. Nous avons constaté, à cette étape, que la question du sondage d'Hydro-Québec demandant « De façon spécifique, quelle température maintenez-vous habituellement dans votre habitation durant la SAISON DE CHAUFFAGE. . . a) Le jour (7 a.m. à 5 p.m.) pendant les JOURS DE SEMAINE ? d) Le soir (5 p.m. à 11 p.m.) pendant les JOURS DE SEMAINE ? g) La nuit (11 p.m. à 7 a.m.) pendant les JOURS DE SEMAINE ? »¹, comportait plusieurs irrégularités et un grand manque de variabilité parmi les réponses. Malheureusement, cette question était cruciale et centrale à notre analyse, étant la variable dépendante de l'équation d'usage.

Suite à une revue attentive des réponses fournies par les ménages échantillonnés, il ressort que dans une large proportion, les gens interrogés n'ont pas répondu avec une grande rigueur ou encore avec beaucoup de sérieux. Les observations manquent souvent de logique et sont parfois totalement insensées. De plus, la nature même de la question entraîne une invariabilité symptomatique des réponses. En effet, les répondants sont amenés à fournir une réponse « spécifique », à une question très générale (maintenez-vous habituellement), cependant très détaillée, couvrant une vaste période (la saison de chauffage). À une telle question, il est évident que ces répondants opteront intuitivement pour la réponse la plus facile et pas nécessairement la plus exacte. Les réponses prendront la forme de moyennes manquant de variabilité et de profondeur. Cette question aurait avantage à faire place à plusieurs questions, plus précises, dont les réponses seraient directes et aisément disponibles aux répondants.

Après suppression des ménages ne répondant pas à la question, voici une représentation graphique de la distribution des réponses données par nos ménages types :

¹Hydro-Québec (2002) « Utilisation de l'électricité dans le marché résidentiel », Édition 2002 (emphasis dans l'original).

FIG. 4.1 – Répartition des réponses pour la température maintenue le jour

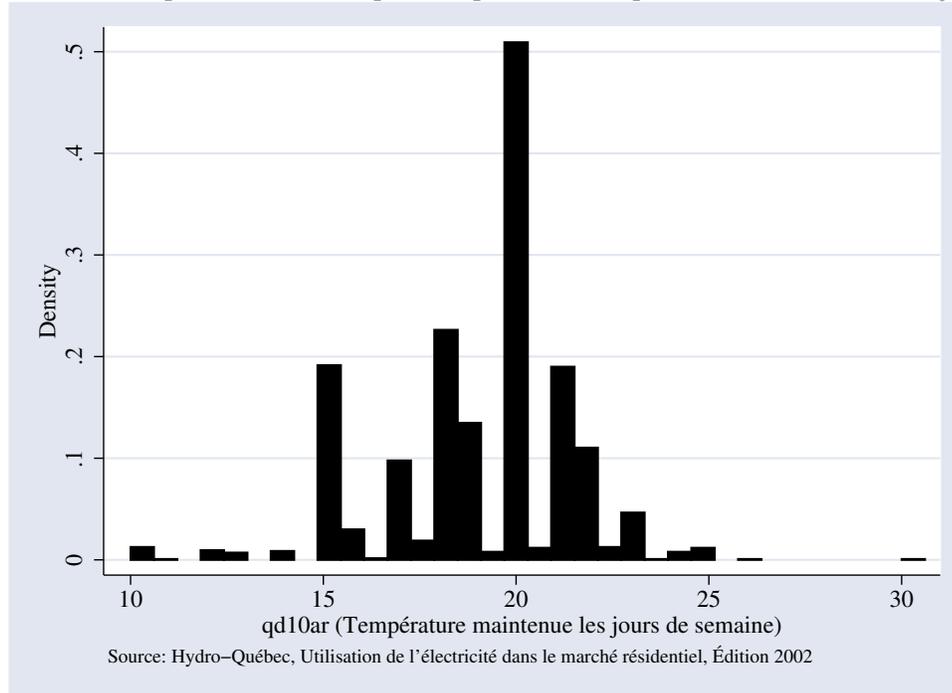


FIG. 4.2 – Répartition des réponses pour la température maintenue le soir

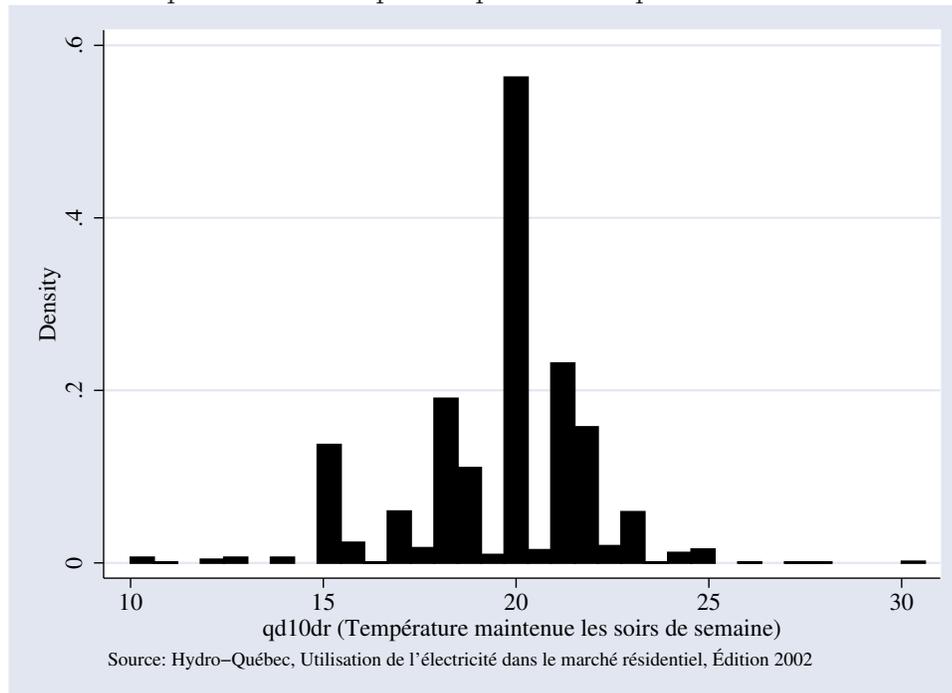
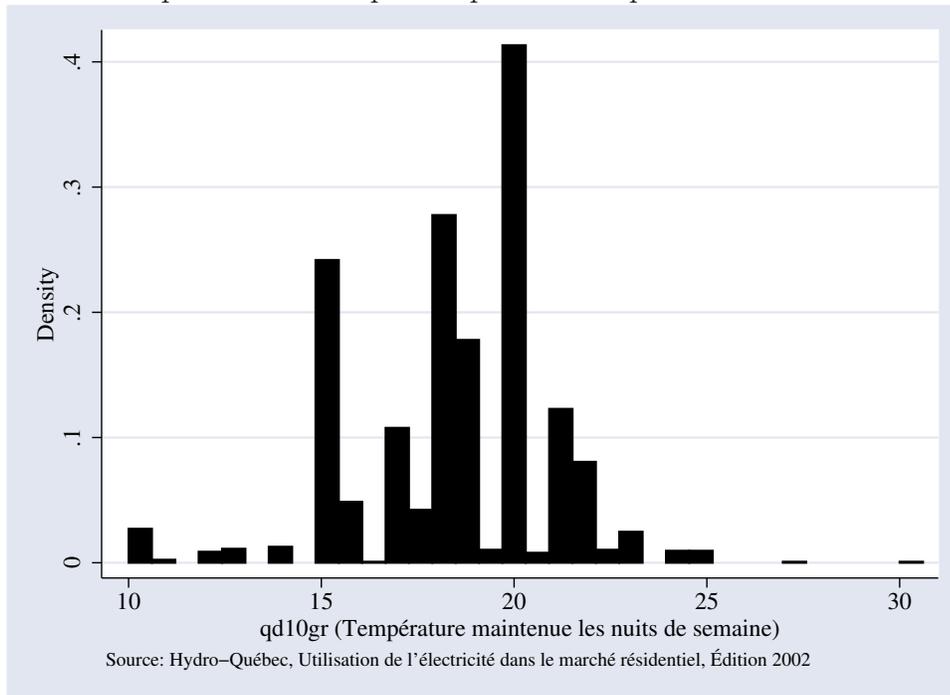


FIG. 4.3 – Répartition des réponses pour la température maintenue la nuit



Il apparaît clairement, comme nous l'avons indiqué plus haut, que la majorité des réponses se situent entre 19°C et 20°C. Il y a aussi un nombre important de valeurs « extrêmes » parmi les réponses. Lors de la régression d'usage, nous avons éliminé les observations qui étaient selon nous insensées, comme de maintenir une température supérieure à 26°C et inférieure à 15°C pour le jour, le soir ou la nuit. Cependant, ce que nous avons gagné en cohérence, nous l'avons perdu en variabilité. Malgré le recours à une transformation élevant les données de la variable dépendante au carré, dans le but d'étirer l'échelle des valeurs, la régression n'en a pas bénéficié. Ces résultats, ainsi que ceux obtenus à la première et dernière étape de notre modèle sont exposés dans le chapitre qui suit.

Chapitre 5

Résultats

L'objectif de la première étape de notre modélisation, l'estimation des probabilités de choix par un modèle de choix discrets, était moins d'expliquer le choix fait par les ménages québécois que de tenir compte de ce choix dans notre analyse de l'usage des différents thermostats disponibles. Pour cette raison, nous discuterons que brièvement des résultats obtenus à cette étape et de leur interprétation. En retour, le modèle d'usage et le modèle de consommation moyenne totale, formant la deuxième et troisième étapes, occuperont une part plus importante de notre discussion.

5.1 Analyse comportementale

5.1.1 Étape 1 : modèle de choix

L'élaboration de notre modèle de choix discrets multinomial logit s'est butée à deux limitations provenant à la fois de la question étudiée et de la banque de données utilisée. Ces limitations découlaient de l'impossibilité d'introduire une variable pour les prix des thermostats qui ait un sens économique et un certain degré explicatif et de l'impossibilité de disposer de variables de caractéristiques propres aux alternatives dans le sondage d'Hydro-Québec. Généralement, les prix reliés aux diverses alternatives sont un déterminant important dans la prise de décision d'un agent du choix de celles-ci. Dans notre contexte, cependant, l'impact des prix est beaucoup plus difficile à cerner. Cela devient évident si nous examinons la question en terme du revenu total du ménage. Nous savons

que les prix des thermostats étudiés varient de 10 \$ à 40 \$¹. Le coût des thermostats représente donc une part minime du revenu des ménages, surtout que cette dépense est de toute évidence ponctuelle et survient lors d'une nouvelle construction ou de rénovations effectuées sur la résidence, où elle s'insère dans une dépense beaucoup plus importante. Néanmoins, nous avons tenté d'introduire les prix dans notre modèle, mais les résultats obtenus n'avaient ni logique ni pouvoir explicatif. Nous avons finalement ignoré les prix par la suite.

L'absence de questions portant sur les caractéristiques des thermostats utilisés par les ménages dans le sondage d'Hydro-Québec était un autre handicap pour lequel nous ne pouvions rien. Peut-être encore davantage que l'absence des prix dans le modèle, l'impossibilité d'introduire des caractéristiques propres aux modalités ont sans aucun doute affaibli le pouvoir explicatif du modèle. À ce titre, il serait intéressant, dans le futur et dans l'éventualité d'une analyse de l'usage des thermostats semblable à la nôtre, de disposer de données plus complètes, précises et diversifiées en ce qui concerne les thermostats détenus par les ménages québécois².

Notre modèle repose donc entièrement sur des variables de caractéristiques propres aux ménages. Étant donné que nous n'avons pas procédé à des simulations et que l'interprétation des paramètres de telles variables est moins directe que dans les modèles économétriques standards, en raison des contraintes d'identification des modèles à choix discrets, nous nous limiterons à présenter les variables qui semblent les plus significatives dans le choix des thermostats. L'ensemble des résultats se retrouve à l'annexe C.

Deux éléments ressortent de l'estimation de notre modèle, les constantes spécifiques sont toutes significatives et le revenu total du ménage est la variable sociodémographique la plus déterminante. Nous avons vu que les constantes spécifiques capturent l'effet moyen sur l'utilité de tous les facteurs qui ne sont pas inclus dans le modèle. C'est un résultat auquel nous nous attendions étant donné notre obligation d'ignorer les variables importantes que sont les prix et les caractéristiques des alternatives. Le revenu des ménages, quant à lui, semble hausser l'utilité des thermostats électroniques, principalement programmables, par rapport aux thermostats réguliers. Bien que cette relation ne ressorte pas nettement, néanmoins, selon toute logique, un revenu plus élevé semble favoriser le choix de thermostats électroniques.

¹Ces prix sont le reflet des modèles les plus populaires selon un magasin à grande surface de la région de Québec : thermostat régulier (10 \$), thermostat électronique non-programmable (30 \$) et thermostat électronique programmable (37 \$).

²Nous pouvons penser au prix, à la marque du fabricant, aux fonctions disponibles, à la date d'acquisition, etc.

Les variables telles que le nombre de personnes dans le ménage, la présence d'enfant(s) et le niveau de scolarité du ménage n'ont pas d'effet marqué sur le choix. Les variables portant sur le lieu de résidence des ménages sont généralement significatives. Cependant, étant donné qu'il s'agit de variables dummies, et que leur interprétation doit s'effectuer par rapport à la fois à la variable omise et à l'alternative normalisée, cet exercice est complexe et ne correspond pas aux objectifs visés dans cette étape. Nous tournerons donc notre attention immédiatement sur les résultats du modèle d'usage.

5.1.2 Étape 2 : modèle d'usage

L'étape 1 de notre modèle discret-continu nous ayant permis de calculer les corrections pour le biais d'auto-sélection associées aux estimations des probabilités de choix, nous les avons intégrées dans notre modèle d'usage qui comptait déjà sur des variables portant sur les caractéristiques des ménages, la température extérieure et le revenu. Rappelons que nous effectuons une estimation de la température intérieure moyenne maintenue le jour, le soir et la nuit. Le tableau suivant présente les résultats de ces estimations :

TABLEAU 5.1 – Résultats du modèle d'usage

Variables	Coefficients			Statistiques t		
	Jour	Soir	Nuit	Jour	Soir	Nuit
totalreel	0,002	0,001	-0,0001	2,90	0,55	-0,14
enfant	-12,081	0,735	4,755	-2,15	0,14	0,85
q11	3,476	-0,319	0,902	1,51	-0,15	0,39
millirevenu	-0,259	-0,076	-0,048	-3,27	-1,02	-0,61
primaire	5,004	-15,123	0,118	0,64	-2,07	0,02
collegial	-7,975	-3,119	-4,475	-1,65	-0,69	-0,93
universite	-11,370	-4,595	-7,115	-2,20	-0,95	-1,39
constante	307,383	355,960	340,602	15,76	20,11	15,62

Estimations par moindres carrés ordinaires au seuil de 5%.

$R^2 = 0,068; 0,082$ et $0,023$ pour le jour, le soir et la nuit. $n = 1955$.

Selon ces résultats, il semble évident que l'usage des thermostats est beaucoup plus difficile à estimer pour le soir et la nuit que pour le jour. Le jour, les ménages types maintiendraient une température intérieure plus élevée lorsque la température extérieure est plus rigoureuse. Ce résultat va dans le sens de nos attentes et reflète le fait que les ménages tentent de contre-balancer une température extérieure plus basse en fixant une température intérieure plus élevée. À l'inverse, le revenu et un niveau élevé de scolarité tendent à diminuer le niveau de la température intérieure choisi. Encore là, ces résultats sont en accord avec nos hypothèses initiales.

Là où les résultats vont à l'encontre de nos hypothèses est en ce qui concerne la présence d'enfant(s). Nos attentes étaient à l'effet que les ménages dans lesquels il y a un enfant ou plus seraient davantage portés à maintenir une température ambiante plus élevée. La recherche d'un plus grand confort de la part des familles, surtout celles avec de jeunes enfants, explique largement nos attentes à ce sujet. Cependant, comme le laisse croire les valeurs obtenues, du moins pour le jour, la présence d'enfant(s) aurait pour effet de diminuer le niveau de la température intérieure choisi. Les autres variables ne sont pas significatives. Pour le soir et la nuit, plus aucune variable explicative semble déterminante dans le choix du niveau de la température intérieure fixée.

Nous remarquons aussi un phénomène particulier intervenant dans chacune des régressions, soit un très faible R^2 jumelé à des constantes très significatives. Nous croyons, tel que mentionné précédemment, que ce phénomène découle d'une trop faible variabilité dans les valeurs prises par la variable dépendante. Nous avons tenté d'y remédier en étirant l'échelle des valeurs de cette variable, mais il n'y eut aucune amélioration tangible. L'ajout des termes de corrections est justifiée par un test de Fisher au niveau de confiance de 95 % qui révèle que ces termes sont globalement significatifs.

5.1.3 Étape 3 : modèle de consommation moyenne annuelle

Cette dernière étape de notre modèle vise à intégrer les résultats précédents de l'analyse comportementale à une régression de la consommation moyenne annuelle totale inspirée de la théorie de l'analyse conditionnelle de la demande. Les résultats du modèle discret-continu sont inclus, sous la forme d'instruments, aux autres variables explicatives. Nous avons utilisé deux mesures pour représenter l'espace à chauffer, soit le nombre de pièces dans la résidence (donné directement dans le sondage) ainsi que la superficie de l'habitation que nous avons estimée à l'aide de variables déterminantes. Dans les deux cas, les résultats sont très semblables et sont présentés à l'annexe D.

Dans l'ensemble, les résultats vont dans le sens de nos hypothèses initiales, à l'exception des trois variables instrumentales. Le fait que ces variables reposent sur les valeurs estimées par notre modèle discret-continu, lui-même handicapé par certaines limitations, explique probablement en grande partie leur absence de significativité. Malheureusement, en raison de ces résultats peu probants statistiquement, il nous est impossible de conclure sur l'effet, indirect, de l'usage des divers thermostats sur la consommation annuelle totale. Même si l'objectif ultime ne peut être atteint, nous pouvons néanmoins tirer quelques conclusions sur les éléments affectant la consommation d'énergie de nos ménages types.

Parmi les éléments contribuant à une consommation d'électricité plus élevée, nous retrouvons le nombre d'individus dans le ménage, le revenu total, l'âge de la résidence (les habitations construites dans les décennies 70 et 80 semblent plus énergivores) et la valeur de la résidence. En plus de ces facteurs, l'utilisation de certains appareils électriques hausse significativement la consommation d'électricité. C'est le cas du dispositif utilisé pour chauffer la piscine, de la cuisinière auto-nettoyante, des réfrigérateurs (chaque ménage en possède au moins un) et de la pompe à eau électrique pour puits artésiens. À l'inverse, une résidence récente et le recours à des systèmes de chauffage en appoint au bois et à l'huile diminueraient la consommation d'électricité³.

Étant, à ce stade, en présence d'un échantillon très homogène, plusieurs variables possédaient un taux de pénétration trop élevé pour posséder un quelconque pouvoir explicatif sur la consommation d'électricité annuelle. C'est le cas, par exemple, des appareils ménagers fortement utilisés par la grande majorité des ménages types ainsi que des chauffe-eau électriques. Ceux-ci ne sont pas d'un intérêt particulier pour des ménages recourant uniquement à l'électricité comme source d'énergie.

³Il est à noter que les systèmes d'appoint au bois ont un effet tangible au Bas-Saint-Laurent, dans le centre du Québec et en Abitibi-Temiscamingue. Les systèmes d'appoint à l'huile agissent particulièrement sur la Côte-Nord, dans le centre du Québec et en banlieue de Montréal.

Chapitre 6

Conclusion

À ce jour, très peu d'écrits avaient été consacrés à l'étude de l'usage des thermostats électroniques non-programmables ou programmables pour contrôler le chauffage des locaux par des plinthes électriques dans le secteur résidentiel. Nous avons donc entrepris une étude empirique de l'efficacité des thermostats programmables pour le secteur résidentiel québécois. La présente analyse apporte une contribution particulière aux recherches à ce sujet puisqu'elle propose une méthodologie nouvelle et fondamentalement différente des recherches antérieures. De plus, elle se concentre sur les ménages dont la principale source d'énergie est l'électricité et dont les plinthes électriques sont le principal système de chauffage. Cette catégorie d'utilisateurs n'a jamais réellement fait l'objet d'une étude poussée quant à l'efficacité de l'usage de thermostats électroniques.

L'objectif de notre étude était double, d'une part, aborder le problème de l'efficacité des thermostats électroniques dans une toute autre perspective, davantage axée sur la théorie économique, et, d'autre part, de répondre à la question suivante : l'expérience actuelle des clients résidentiels d'Hydro-Québec Distribution permet-elle de faire ressortir des gains pour les consommateurs liés à l'utilisation des thermostats électroniques ? Alors que la spécification de notre modèle nous permet d'affirmer que le premier objectif a été atteint, la réalisation du second objectif ne laisse pas place à autant d'enthousiasme. La déception et la frustration associées à des résultats statistiquement non probants, menant à l'impossibilité de conclure sur la problématique étudiée, étaient peut être l'envers de la médaille à une tentative d'explorer de nouvelles avenues.

En effet, peu importe la qualité des résultats, notre modélisation était une tentative d'analyser l'efficacité des thermostats électroniques en utilisant de nouvelles techniques de la théorie économique permettant de prendre en considération le comportement des utilisateurs au lieu des simulations de modèles de circulation d'énergie traditionnellement

employées. Le comportement des agents était ainsi intégré dans notre modèle et non considéré comme un facteur externe. Cette spécification semble pertinente étant donné que l'essence même de l'intérêt de l'efficacité du thermostat électronique (programmable) découle de l'usage que le ménage en fait. C'est ce même ménage qui décide du type de thermostat utilisé et de la température qu'il y maintient.

Les résultats peu convaincants sont le reflet que cette tentative ne s'est pas faite sans problème. Plusieurs éléments nous sont apparus comme des limitations, dont une analyse ultérieure du même ordre devrait s'affranchir. Ces limitations ont été traitées plus haut, il suffit de rappeler que l'ajout de variables pour le prix et pour les caractéristiques des alternatives est un enjeu important. Nous avons aussi constaté que le modèle discret-continu ne représentait peut-être pas pleinement la réalité et qu'un modèle entièrement à choix discrets, tel que le logit emboîté, pourrait être une voie favorisée.

Nous terminerons notre discussion par la question suivante : quels résultats faut-il espérer d'une analyse de l'efficacité des thermostats électroniques si la grande majorité des ménages utilisent ces thermostats d'une façon inefficace ? Autrement dit, nous cherchons peut être à faire ressortir des gains liés à l'utilisation des thermostats électroniques alors que l'utilisation de ces thermostats se fait de telle manière à ne permettre aucun gain. Une telle situation exigerait non seulement de promouvoir les thermostats électroniques plus efficaces, mais aussi d'informer les usagers, présents et futurs, de la façon efficace de les utiliser.

Bibliographie

- [1] Ben-Akiva, M. (1973), « The Structure of Travel Demand Models », thèse de doctorat, MIT.
- [2] Bernard, Jean-Thomas, Denis Bolduc et Donald Bélanger (1996), « Québec Residential Electricity Demand : A Microeconomic Approach », *The Canadian Journal of Economics* 29(1) : 92-113.
- [3] Bierlaire, Michel (2001), « A theoretical analysis of the cross-nested logit model », *Report RO-011218*.
- [4] Bouchelle, Matthew P., Danny S. Parker et Michael T. Anello (2000), « Factors Influencing Space Heat and Heat Pump Efficiency from a Large-Scale Residential Monitoring Study », *Proceedings of the 2000 Summer Study on Energy Efficiency in Buildings*, Washington : American Council for an Energy-Efficient Economy (ACEEE).
- [5] Dubin, Jeffrey A. et Daniel L. McFadden (1984), « An Econometric Analysis of Residential Electric Appliance Holdings and Consumption », *Econometrica* 52(2) : 345-362.
- [6] Greene, William H. et David A. Hensher (2001), « The Mixed Logit Model : The State of Practice and Warnings for the Unwary », *Working Paper*, School of Business, University of Sidney.
- [7] Heckman, J. (1978), « Dummy Endogenous Variables in a Simultaneous Equation System », *Econometrica* 46 : 931-959.
- [8] Heckman, J. (1979), « Sample Selection Bias as a Specification Error », *Econometrica* 47 : 153-162.
- [9] Hydro-Québec (2001), *Plan stratégique 2002-2006*, Vice-présidence - Recherche et planification stratégique.
- [10] Hydro-Québec (2003), *Plan global en efficacité énergétique (PGÉE) 2003-2006*.
- [11] Hydro-Québec (2003), *Rapport annuel 2002*, Affaires corporatives et secrétariat général.

- [12] Hydro-Québec (2003), *Plan stratégique 2004-2008*, Vice-présidence - Recherche et planification stratégique.
- [13] Hydro-Québec Distribution (2002), *Utilisation de l'électricité dans le marché résidentiel, Édition 2002*, Planification et Recherche Commerciale : Direction Planification et Contrôle.
- [14] Jacques, Christiane et Martin Perron (2000), *Travaux sur la modélisation de l'ACD en 1999*, INRS : Énergie et Matériaux.
- [15] Jehle, Geoffrey A. et Philip J. Reny (2001), *Advanced Microeconomic Theory*, 2e édition, Addison Wesley.
- [16] Lacroix, Geneviève (2004), *Analyse conditionnelle de la demande appliquée au secteur résidentiel québécois en 1989, 1994 et 1999 : nouvelle analyse statistique*, Mémoire de maîtrise, Université Laval.
- [17] Lee, B. (1999), « Calling Patterns and Usage of Residential Toll Service under Self-Selecting Tariffs », *Journal of Regulatory Economics* 16 : 45-82.
- [18] Nelson, Lorne W. et J. Ward MacArthur (1978), « Energy Savings through Thermostat Setbacks », *ASHRAE Transactions* 83(1) : 319-333.
- [19] Nevius, Monica J. et Scott Pigg (2000), « Programmable Thermostats that Go Berserk? Taking a Social Perspective on Space Heating in Wisconsin », Energy Center of Wisconsin et University of Wisconsin.
- [20] Parti, M. et C. Parti (1980), « The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector », *Bell Journal of Economics* 11 : 309-324.
- [21] Plourde, André (2003), « Programmable Thermostats as Means of Generating Energy Savings : Some Pros and Cons », CBEEDAC 2003-RP-01, Edmonton, Alberta.
- [22] Train, Kenneth E. (1986), *Qualitative Choice Analysis : Theory, Econometrics, and an Application to Automobile Demand*, The Massachusetts Institute of Technology Press, Cambridge, MA.
- [23] Train, Kenneth E. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA.
- [24] Varian, Hal R. (1997), *Introduction à la microéconomie*, 4e édition, De Boeck Université, Paris.
- [25] Vekeman, Francis (2004), *Choix de véhicules et demande de kilométrage : une approche microéconomique*, Mémoire de maîtrise, Université Laval.
- [26] Woolridge, Jeffrey M. (2000), *Introductory Econometrics : A Modern Approach*, South-Western College Publishing.

Annexe A

Maximisation de l'utilité

Dans cette annexe, nous revenons plus en détail sur les notions de maximisation de l'utilité, d'utilité directe et indirecte et sur la relation qui les unie. Rappelons tout d'abord qu'au sein de la théorie classique de la demande, les axiomes sur les préférences définissent la rationalité des consommateurs. Les préférences sont ordinales pour un agent, ce qui signifie que le niveau d'utilité (plus élevé ou plus bas) n'est pas chiffré. Les préférences sont supposées complètes, réflexives, transitives, continues, monotones (non-satiété) et convexes. Voilà pour les prémisses de base. Tournons-nous maintenant vers la maximisation de cette utilité.

Supposons un agent, dont le revenu est fixé à y , et qui fait face à un vecteur de prix donné \mathbf{p} correspondant à un vecteur de biens \mathbf{x} . Si la consommation de x_i unités du bien i au prix p_i pour chaque unité exige de dépenser $p_i x_i$ dollars, la contrainte (budgétaire) que la dépense ne dépasse pas le revenu est notée par $\mathbf{p} \cdot \mathbf{x} = y$. Le problème de maximisation de l'utilité de l'agent s'écrit donc

$$\max_{\mathbf{x}} u(\mathbf{x}) \quad \text{s.c.} \quad \mathbf{p} \cdot \mathbf{x} = y. \quad (\text{A.1})$$

Parce que la solution de (A.1) est unique pour des valeurs données de \mathbf{p} et y , nous pouvons la voir comme une fonction de l'ensemble des prix et du revenu vers l'ensemble des quantités X . Nous pouvons alors écrire cette solution (la quantité choisie) par $x_i^* = x_i(\mathbf{p}, y)$, $i = 1, \dots, n$. Sous cette forme, les solutions au problème de maximisation sont connues en tant que fonctions de demande ordinaires ou marshalliennes.

La fonction d'utilité directe, $u(\mathbf{x})$, représente les préférences de l'agent directement. Selon les prix \mathbf{p} et le revenu y , l'agent choisit un ensemble de consommation maximisant son utilité ($\mathbf{x}(\mathbf{p}, y)$). La relation entre les prix, le revenu et la valeur maximisée de l'utilité se traduit par une fonction définie comme suit :

$$v(\mathbf{p}, y) = \max_{\mathbf{x}} u(\mathbf{x}) \quad \text{s.c.} \quad \mathbf{p} \cdot \mathbf{x} = y. \quad (\text{A.2})$$

La fonction $v(\mathbf{p}, y)$ est appelée fonction d'utilité indirecte et elle correspond au niveau d'utilité le plus élevé qui peut être atteint lorsque $\mathbf{x}(\mathbf{p}, y)$ est choisi.

$$v(\mathbf{p}, y) = u(\mathbf{x}(\mathbf{p}, y)). \quad (\text{A.3})$$

On peut voir $v(\mathbf{p}, y)$ comme une mesure du niveau d'utilité de la courbe d'indifférence la plus élevée que l'agent peut atteindre, étant donné les prix et le revenu (Jehle et Reny, 2001).

Nous sommes donc en présence de deux fonctions d'utilité : (1) $u(\mathbf{x})$, qui donne l'utilité que l'agent obtient pour des quantités données de chaque bien et qui est appelée fonction d'utilité directe ; (2) $v(\mathbf{p}, y)$, qui donne l'utilité que l'agent obtient à des prix et un revenu donnés une fois qu'il a choisi les quantités maximisant son utilité (directe) sous contrainte budgétaire pour des prix et un revenu donnés (Train, 1986). Ainsi, nous pouvons spécifier une fonction d'utilité indirecte pour représenter les préférences d'un agent en sachant qu'une fonction d'utilité directe lui est implicite.

En soi, il est peu intéressant, bien souvent, de connaître la forme de la fonction d'utilité. De plus, la procédure pour y arriver exige de résoudre le problème de maximisation de l'utilité donné par (A.1) qui implique de spécifier le Lagrangien, calculer les dérivées et les fixer à zéro, pour ensuite résoudre pour les quantités associées à chaque bien. Ce processus est complexe et difficile à manipuler. Cependant, comme nous l'avons vu plus haut, il est beaucoup plus facile de travailler avec l'utilité indirecte pour dériver des fonctions de demande. Nous utilisons alors l'identité de Roy, qui établit que la demande marshallienne pour le bien i est le ratio des dérivées partielles de l'utilité indirecte par rapport au prix du bien i et au revenu, suite à un changement de signe.

$$-\frac{\partial v(\mathbf{p}, y)/\partial p_i}{\partial v(\mathbf{p}, y)/\partial y} = x_i(\mathbf{p}, y)$$

où $x_i(\mathbf{p}, y)$ est la fonction de demande marshallienne ou ordinaire et $v(\mathbf{p}, y)$ est la fonction d'utilité indirecte.

Annexe B

Définitions des variables

Modèle de choix discrets

abit_temis : Abitibi-Temiscamingue ;
bsl_gas : Bas-Saint-Laurent-Gaspésie ;
collegial : niveau de scolarité de la personne répondante ;
cotenord : Côte-Nord ;
enfant : présence d'enfant(s) dans le ménage ;
nordqc : Nord-du-Québec ;
one : constante spécifique ;
primaire : niveau de scolarité de la personne répondante ;
qcmaurichau : Québec-Mauricie-Chaudière-Appalaches ;
ql1 : nombre de personnes dans le ménage ;
regionmtl1 : Canton-de-l'Est-Outaouais-Lanaudière-Laurentides ;
regionmtl2 : Montréal-Laval-Montérégie-Haut-Saint-Laurent-Centre-du-Québec ;
revenu : revenu total du ménage ;
sag_lac : Saguenay-Lac-Saint-Jean ;
secondaire : niveau de scolarité de la personne répondante ;
universite : niveau de scolarité de la personne répondante ;
vha_100k : valeur de l'habitation inférieure à 100 000\$;
vha100_200k : valeur de l'habitation entre 100 000\$ et 200 000\$;
vha200k_ : valeur de l'habitation supérieure à 200 000\$.

Modèle d'usage

collegial : niveau de scolarité de la personne répondante ;
corr2,3,4,5,6,7 : corrections pour le biais d'auto-sélection ;
enfant : présence d'enfant(s) ;
millirevenu : revenu divisé par mille ;
primaire : niveau de scolarité de la personne répondante ;
ql1 : nombre de personnes dans le ménage ;
totalJreel : nombre de degrés-heures réels extérieurs pour le jour ;
totalNreel : nombre de degrés-heures réels extérieurs pour la nuit ;
totalSreel : nombre de degrés-heures réels extérieurs pour le soir ;
universite : niveau de scolarité de la personne répondante.

Modèle de consommation

chaufpisci : chauffe piscine électrique ;
cuisiautonet : cuisinière électrique avec four autonettoyant ;
hab_1970 : habitation construite avant 1970 ;
hab1971_80 : habitation construite entre 1971 et 1980 ;
hab1981_90 : habitation construite entre 1981 et 1990 ;
hab1991_02 : habitation construite entre 1991 et 2002 ;
millirevenu : revenu total du ménage divisé par mille ;
nbrfrigoma : nombre de réfrigérateurs à dégivrage manuel ;
nbrpers : nombre de personnes dans le ménage ;
piec_totalJ : nombre de pièces multiplié par degrés-heures de jour calculés avec les réponses ;
piec_totalN : nombre de pièces multiplié par degrés-heures de nuit calculés avec les réponses ;
piec_totals : nombre de pièces multiplié par degrés-heures de soir calculés avec les réponses ;
piec_totalJcal : nombre de pièces multiplié par degrés-heures de jour calculés avec les réponses estimées ;
piec_totalNcal : nombre de pièces multiplié par degrés-heures de nuit calculés avec les réponses estimées ;
piec_totalScal : nombre de pièces multiplié par degrés-heures de soir calculés avec les réponses estimées ;
pompeauelec : possède une pompe à eau électrique (puits artésien ou sub-pompe) ;

sup_totalJ : superficie imputée multipliée par degrés-heures de jour calculés avec réponses;

sup_totalN : superficie imputée multipliée par degrés-heures de nuit calculés avec réponses;

sup_totalS : superficie imputée multipliée par degrés-heures de soir calculés avec réponses;

sup_totalJcal : superficie imputée multipliée par degrés-heures de jour calculés avec réponses estimées;

sup_totalNcal : superficie imputée multipliée par degrés-heures de nuit calculés avec réponses estimées;

sup_totalScal : superficie imputée multipliée par degrés-heures de soir calculés avec réponses estimées;

sysappbois1 : possède un système d'appoint au bois (Bas-Saint-Laurent-Gaspésie);

sysappbois2 : possède un système d'appoint au bois (Côte-Nord);

sysappbois3 : possède un système d'appoint au bois (Saguenay-Lac-Saint-Jean);

sysappbois4 : possède un système d'appoint au bois (Nord-du-Québec);

sysappbois5 : possède un système d'appoint au bois (Québec-Mauricie-Chaudière-Appalaches);

sysappbois6 : possède un système d'appoint au bois (Canton-de-l'Est-Outaouais-Lanaudière-Laurentides);

sysappbois7 : possède un système d'appoint au bois (Montréal-Laval-Montérégie-Haut-Saint-Laurent-Centre-du-Québec);

sysappbois8 : possède un système d'appoint au bois (Abitibi-Temiscamingue);

sysapphuil1 : possède un système d'appoint à l'huile (Bas-Saint-Laurent-Gaspésie);

sysapphuil2 : possède un système d'appoint à l'huile (Côte-Nord);

sysapphuil3 : possède un système d'appoint à l'huile (Saguenay-Lac-Saint-Jean);

sysapphuil4 : possède un système d'appoint à l'huile (Nord-du-Québec);

sysapphuil5 : possède un système d'appoint à l'huile (Québec-Mauricie-Chaudière-Appalaches);

sysapphuil6 : possède un système d'appoint à l'huile (Canton-de-l'Est-Outaouais-Lanaudière-Laurentides);

sysapphuil7 : possède un système d'appoint à l'huile (Montréal-Laval-Montérégie-Haut-Saint-Laurent-Centre-du-Québec);

sysapphuil8 : possède un système d'appoint à l'huile (Abitibi-Temiscamingue);

sysappgaz : possède un système d'appoint au gaz (pour toutes les régions);

vha_100k : valeur de l'habitation inférieure à 100 000\$;

vha100_200k : valeur de l'habitation entre 100 000\$ et 200 000\$;

vha200k_ : valeur de l'habitation supérieure à 200 000\$.

Annexe C

Résultats du modèle de choix discrets

TABLEAU C.1: Résultats du modèle à choix discrets

Variables par alternative	Coefficients estimés	Statistiques t	Statistiques t (robustes)
Constantes			
Réguliers	-	-	-
Non-programmables	-2,930	-3,969*	-4,233*
Programmables	-3,064	-5,595*	-5,227*
Non-prog. et prog.	-3,401	-5,952*	-6,033*
Réguliers et non-prog.	-4,740	-6,532*	-7,008*
Réguliers et prog.	-2,941	-6,863*	-7,070*
Réguliers,prog. et non-prog.	-4,222	-5,412*	-5,463*
ql1			
Réguliers	-	-	-
Non-programmables	-0,576	-2,624*	-3,739*
Programmables	-0,127	-0,811	-0,737
Non-prog. et prog.	-0,006	-0,032	-0,036
Réguliers et non-prog.	0,016	0,105	0,110
Réguliers et prog.	0,082	0,871	0,846
Réguliers,prog. et non-prog.	0,107	0,820	0,943
enfant			
Réguliers	-	-	-
Non-programmables	0,813	1,752	1,920
Programmables	-0,145	-0,399	-0,367

suite à la page suivante

TABLEAU C.1 – suite

Variables par alternative	Coefficients estimés	Statistiques t	Statistiques t (robustes)
Non-prog. et prog.	0,075	0,172	0,198
Réguliers et non-prog.	0,010	0,026	0,029
Réguliers et prog.	0,256	1,120	1,091
Réguliers,prog. et non-prog.	-0,348	-1,031	-1,096
secondaire			
Réguliers	-	-	-
Non-programmables	-0,436	-0,649	-0,650
Programmables	-0,592	-1,208	-1,194
Non-prog. et prog.	-1,836	-3,414*	-3,524*
Réguliers et non-prog.	0,158	0,246	0,241
Réguliers et prog.	0,294	0,747	0,762
Réguliers,prog. et non-prog.	1,030	1,387	1,389
collegial			
Réguliers	-	-	-
Non-programmables	0,442	0,661	0,658
Programmables	0,058	0,116	0,116
Non-prog. et prog.	-1,207	-2,320*	-2,453*
Réguliers et non-prog.	0,170	0,253	0,244
Réguliers et prog.	0,492	1,214	1,244
Réguliers,prog. et non-prog.	1,233	1,622	1,634
universite			
Réguliers	-	-	-
Non-programmables	-0,587	-0,798	-0,807
Programmables	-0,441	-0,834	-0,826
Non-prog. et prog.	-1,176	-2,187*	-2,299*
Réguliers et non-prog.	8,682e-05	1,252e-04	1,170e-04
Réguliers et prog.	0,285	0,681	0,706
Réguliers,prog. et non-prog.	1,185	1,535	1,531
bsl_gas			
Réguliers	-	-	-
Non-programmables	-0,194	-0,296	-0,303
Programmables	0,049	0,103	0,100

suite à la page suivante

TABLEAU C.1 – suite

Variables par alternative	Coefficients estimés	Statistiques t	Statistiques t (robustes)
Non-prog. et prog.	-0,539	-0,702	-0,693
Réguliers et non-prog.	1,120	2,284*	2,180*
Réguliers et prog.	-0,045	-0,134	-0,135
Réguliers,prog. et non-prog.	-0,263	-0,522	-0,526
cotenord			
Réguliers	-	-	-
Non-programmables	-5,204	-0,641	-18,345*
Programmables	-0,047	-0,084	-0,086
Non-prog. et prog.	-0,204	-0,269	-0,280
Réguliers et non-prog.	0,712	1,224	1,219
Réguliers et prog.	0,165	0,483	0,481
Réguliers,prog. et non-prog.	0,063	0,125	0,124
sag_lac			
Réguliers	-	-	-
Non-programmables	-0,100	-0,130	-0,133
Programmables	-0,636	-0,849	-0,863
Non-prog. et prog.	-4,724	-0,678	-16,710*
Réguliers et non-prog.	0,464	0,714	0,707
Réguliers et prog.	-0,332	-0,789	-0,784
Réguliers,prog. et non-prog.	-1,545	-1,506	-1,508
qcmaurichau			
Réguliers	-	-	-
Non-programmables	-0,241	-0,571	-0,590
Programmables	0,322	1,103	1,087
Non-prog. et prog.	0,524	1,624	1,636
Réguliers et non-prog.	0,603	1,616	1,588
Réguliers et prog.	0,246	1,246	1,234
Réguliers,prog. et non-prog.	-0,138	-0,454	-0,458
regionmtl1			
Réguliers	-	-	-
Non-programmables	-0,507	-1,227	-1,256
Programmables	-0,831	-2,232*	-2,202*

suite à la page suivante

TABLEAU C.1 – suite

Variables par alternative	Coefficients estimés	Statistiques t	Statistiques t (robustes)
Non-prog. et prog.	-1,018	-2,208	-2,261
Réguliers et non-prog.	0,065	0,168	0,165
Réguliers et prog.	-0,153	-0,761	-0,749
Réguliers,prog. et non-prog.	-0,335	-1,153	-1,144
nordqc			
Réguliers	-	-	-
Non-programmables	-2,449	-2,912*	-3,736*
Programmables	-2,981	-0,380	-4,988*
Non-prog. et prog.	-0,456	-0,155	-0,655
Réguliers et non-prog.	-1,792	-0,272	-2,893*
Réguliers et prog.	0,389	0,349	0,336
Réguliers,prog. et non-prog.	-3,313	-0,376	-5,868*
abit_temis			
Réguliers	-	-	-
Non-programmables	-0,197	-0,257	-0,256
Programmables	-1,207	-1,175	-1,179
Non-prog. et prog.	-4,720	-0,688	-17,810*
Réguliers et non-prog.	0,577	0,889	0,877
Réguliers et prog.	-1,469	-2,012*	-2,008*
Réguliers,prog. et non-prog.	-0,223	-6,498e-04	-41,615*
vha100_200k			
Réguliers	-	-	-
Non-programmables	0,543	1,589	1,673
Programmables	0,180	0,676	0,710
Non-prog. et prog.	0,618	1,832	1,816
Réguliers et non-prog.	0,558	1,866	1,780
Réguliers et prog.	0,229	1,385	1,378
Réguliers,prog. et non-prog.	0,456	1,801	1,834
vha200k_			
Réguliers	-	-	-
Non-programmables	-2,922	-0,899	-7,723*
Programmables	0,019	0,032	0,033

suite à la page suivante

TABLEAU C.1 – suite

Variabes par alternative	Coefficients estimés	Statistiques t	Statistiques t (robustes)
Non-prog. et prog.	0,670	1,122	1,117
Réguliers et non-prog.	0,439	0,650	0,635
Réguliers et prog.	-1,209	-1,952	-1,946
Réguliers,prog. et non-prog.	0,663	1,319	1,415
revenu			
Réguliers	-	-	-
Non-programmables	1,062e-05	1,651	1,849
Programmables	1,342e-05	2,795*	2,856*
Non-prog. et prog.	1,626e-05	2,901*	2,929*
Réguliers et non-prog.	1,108e-05	2,054*	2,075*
Réguliers et prog.	9,910e-06	3,174*	3,531*
Réguliers,prog. et non-prog.	7,650e-07	0,159	0,172
(*) Significatif au niveau de confiance de 95%.			

Annexe D

Résultats du modèle de consommation annuelle

TABLEAU D.1: Résultats du modèle de consommation avec le nombre de pièces

Variable	Coefficient estimé	Statistique t
piec_totalJ	-0,035	-1,10
piec_totalS	0,068	1,23
piec_totalN	0,015	0,40
nbrpers	1147,627	7,91
millirevenu	38,518	5,23
(hab_1970)	–	–
hab1971_80	1221,674	2,71
hab1981_90	387,751	0,81
hab1991_02	-1327,267	-2,34
(vha_100k)	–	–
vha100_200k	367,884	0,99
vha200k_	2784,291	3,02
chaufpisci	7225,809	4,36
cuisiautonet	1054,021	2,96
nbrfrigoma	1322,718	3,15
pompeauelec	2262,577	5,14
sysappbois1	-3699,251	-3,34
sysappbois2	37,591	0,03

suite à la page suivante

TABLEAU D.1 – suite

Variable	Coefficient estimé	Statistique t
sysappbois3	-1535,663	-1,16
(sysappbois4)	–	–
sysappbois5	-1436,648	-2,50
sysappbois6	-333,177	-0,54
sysappbois7	208,243	0,42
sysappbois8	-3805,237	-1,93
sysapphuil1	3121,615	0,61
sysapphuil2	-15448,130	-2,83
sysapphuil3	-7747,889	-1,52
(sysapphuil4)	–	–
sysapphuil5	-5691,864	-2,71
sysapphuil6	-873,827	-0,31
sysapphuil7	-3895,892	-1,92
sysapphuil8	-4539,334	-0,89
sysappgaz	1573,049	0,38
constante	11914,890	17,18
$R^2 = 0,184$		
$n = 1931$		
Instrumentés :	piec_totalJ, piec_totalS, piec_totalN	
Instruments :	piec_totalJcal, piec_totalScal, piec_totalNcal, nbrpers, millirevenu, hab1971_80, hab1981_90, hab1991_02, vha100_200k, vha200k_, chaufpisci, cuisiautonnet, nbrfrigoma, pompeauelec, sysappbois1, sysappbois2, sysappbois3, sysappbois4, sysappbois5, sysappbois6, sysappbois7, sysappbois8, sysapphuil1, sysapphuil2, sysapphuil3, sysapphuil4, sysapphuil5, sysapphuil6, sysapphuil7, sysapphuil8, sysappgaz	

TABLEAU D.2: Résultats du modèle de consommation avec la superficie

Variable	Coefficient estimé	Statistique t
sup_totalJ	-0,942e-04	-0,84
sup_totalS	0,262e-03	1,33
sup_totalN	-0,157e-04	-0,13
nbrpers	1213,980	8,35
millirevenu	45,811	6,36
(hab_1970)	–	–
hab1971_80	1279,674	2,84
hab1981_90	152,135	0,32
hab1991_02	-1462,255	-2,57
(vha_100k)	–	–
vha100_200k	397,865	1,05
vha200k_	2030,846	2,13
chaufpisci	7143,539	4,25
cuisiautonet	1131,869	3,15
nbrfrigoma	1258,292	2,98
pompeauelec	2206,642	4,99
sysappbois1	-3492,855	-3,30
sysappbois2	109,313	0,08
sysappbois3	-1564,939	-1,18
(sysappbois4)	–	–
sysappbois5	-1289,336	-2,21
sysappbois6	-363,042	-0,58
sysappbois7	-53,221	-0,11
sysappbois8	-3378,937	-1,69
sysapphuil1	4052,458	0,78
sysapphuil2	-14112,610	-2,60
sysapphuil3	-7426,046	-1,44
(sysapphuil4)	–	–
sysapphuil5	-5426,361	-2,57
sysapphuil6	-874,799	-0,31
sysapphuil7	-3139,667	-1,53
sysapphuil8	-2906,109	-0,57
sysappgaz	-635,260	-0,15
constante	12517,740	18,18

suite à la page suivante

TABLEAU D.2 – suite

Variable	Coefficient estimé	Statistique t
$R^2 = 0,172$		
$n = 1931$		
Instrumentés :	sup_totalJ, sup_totalS, sup_totalN	
Instruments :	sup_totalJcal, sup_totalScal, sup_totalNcal, nbrpers, millirevenu, hab1971_80, hab1981_90, hab1991_02, vha100_200k, vha200k_, chauffisci, cuisiautonet, nbrfrigoma, pompeauelec, sysappbois1, sysappbois2, sysappbois3, sysappbois4, sysappbois5, sysappbois6, sysappbois7, sysappbois8, sysapphuil1, sysapphuil2, sysapphuil3, sysapphuil4, sysapphuil5, sysapphuil6, sysapphuil7, sysapphuil8, sysappgaz	