



UNIVERSITÉ
LUMIÈRE
LYON 2

N°d'ordre NNT : 2016LYSE2093

THESE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

École Doctorale : ED 486

Sciences économiques et de gestion

Discipline : Sciences économiques

Soutenue publiquement le 30 septembre 2016, par :

Hind AISSAOUI

Choix de localisation résidentielle des ménages en milieu urbain

*Les apports récents des modèles de choix discrets en présence
d'un nombre élevé d'alternatives*

Devant le jury composé de :

Yves CROISSANT, Professeur des universités, Université de La Réunion, Président

Sophie MASSON, Professeure des universités, Université de Perpignan, Rapporteur

Nathalie GAUSSIÉ, Maître de conférences HDR, Université de Bordeaux, Rapporteur

Jean-Loup MADRE, Directeur de recherche émérite, IFSTTAR, Examineur

Louafi BOUZOUINA, Chargé de Recherche, École Nationale des Travaux Publics de l'État, Examineur

Patrick BONNEL, Docteur Ingénieur HDR, École Nationale des Travaux Publics de l'État, Directeur de thèse

Université Lumière Lyon 2

Faculté de Sciences Économiques et de Gestion

Choix de localisation résidentielle des ménages en milieu urbain

*Les apports récents des modèles de choix discrets en
présence d'un nombre élevé d'alternatives*

Thèse

pour l'obtention d'un

Doctorat de Sciences Économiques mention Économie des Transports

Présentée et soutenue publiquement par

Hind AISSAOUI

Le 30 septembre 2016

Membres du Jury :

M. BONNEL Patrick, Directeur de Recherches au Laboratoire d'Économie des Transports, ENTPE – Université Lumière Lyon 2, Directeur de Thèse

M. BOUZOUINA Louafi, Chargé de recherche à l'ENTPE, Co-directeur de Thèse

M. CROISSANT Yves, Professeur à l'Université de la Réunion

Mme. GAUSSIER Nathalie, Maître de Conférences en sciences économiques, Habilitée à Diriger des Recherches à l'université de Bordeaux, Rapportrice

M. MADRE Jean-Loup, Directeur de recherche émérite à l'IFSTTAR

Mme. MASSON Sophie, Professeure à l'université de Perpignan Via Domitia, Rapportrice

Remerciements

Je tiens à remercier en premier lieu Patrick Bonnel, pour la confiance qu'il m'a accordé en acceptant de diriger cette thèse. Au-delà de son appui scientifique, il a toujours été là pour m'écouter et m'encourager. J'aimerais également lui dire à quel point j'étais contente de travailler avec lui.

J'exprime aussi ma profonde gratitude à Louafi Bouzouina pour son suivi et son encadrement tout au long de la thèse.

Je remercie mesdames Nathalie Gaussier et Sophie Masson pour avoir accepté d'être rapportrices de cette thèse, et Messieurs Yves Croissant et Jean-Loup Madre pour avoir accepté de participer en tant que membres à ce jury.

Un grand Merci à Kiarash Motamedi, Jean Pierre Nicolas et Hakim Ouaras pour les relectures de certains chapitres.

Merci à Pascal Pochet, avec qui j'ai partagé le bureau, pour ses conseils, sa gentillesse et son soutien.

Mes remerciements s'adressent également à toute l'équipe du Laboratoire Aménagement Economie Transports qui a fait de ces années de thèse une véritable expérience humaine. Merci à Lourdes Diaz Olvera, Olivier Klein, Marie Lova , Florence Toilier et Didier Plat.

Merci à tous les anciens et actuels doctorants du LAET avec qui j'ai partagé des moments agréables, particulièrement Jorge, Jean, Quan Son, Berhanu et Hind.

Enfin, je vais terminer par remercier toutes les personnes qui m'ont soutenu pendant ces années de thèses. Mes chers parents, mes frères et sœurs, qui m'ont illuminée la vie de leurs précieux encouragements. Sans oublier de remercier mes amis au Maroc et à Lyon, qui m'ont aidé à surmonter les moments difficiles dans cette thèse.

Sommaire

Sommaire.....	5
Introduction	8
Partie I - Choix de localisation résidentielle : revue de littérature.....	14
Introduction	15
Chapitre 1 - Cadres théoriques du choix de localisation résidentielle	16
1.1. Le modèle standard de l'économie urbaine et ses extensions	18
1.2. L'apport de l'approche des interactions spatiales	22
1.3. La théorie de l'utilité aléatoire	24
L'approche d'utilité aléatoire comme cadre d'analyse de la localisation résidentielle	26
Chapitre 2 - Les déterminants des choix de localisation résidentielle : travaux empiriques.....	27
2.1. Arbitrage du marché : accessibilité vs prix de l'immobilier	27
2.2. Préférence pour les aménités spatiales et sociales	37
2.3. Caractéristiques des ménages : cycle de vie et accès à la propriété	44
Comparaison entre différents contextes : quels enseignements par rapport aux déterminants du choix de localisation ?	48
Conclusion	50
Partie II : Les modèles de choix discrets comme outil pour appréhender les choix de localisation résidentielle	52
Chapitre 3 - Les familles de modèles de choix discrets : structures et limites.....	54
3.1. Spécificités générales des modèles de choix discrets	55
3.2. La famille des modèles GEV	56
3.3. Modèle probit.....	64
3.4. Familles des modèles mixtes	65
3.5. D'autres structures.....	65

Synthèse : Modèles développés pour modéliser les choix de localisation	66
Chapitre 4 – Quels apports des développements récents pour les choix de localisation.....	69
4.1. La constitution de l'ensemble des choix	70
4.2. Échantillonnage des alternatives	71
4.3. Taille de l'échantillon	76
Quelle méthode pour réduire le nombre élevé d'alternatives	77
Conclusion	79
Partie III : Choix de localisation résidentielle dans l'aire urbaine de Lyon (1999, 2007)	80
Chapitre 5 – Aire urbaine de Lyon : Échelles et données.....	82
5.1. Données et constructions des variables.....	82
5.2. Échelles spatiales.....	94
Analyse empirique des choix de localisation : Données et échelles	101
Chapitre 6 - Analyse empirique des choix de localisation des ménages : un modèle logit Multinomial.....	103
6.1. Processus de l'estimation.....	104
6.2. Les critères de l'évaluation de la qualité du modèle.....	106
6.3. Estimation des modèles en 1999	108
6.4. Résultats et discussions.....	122
Un modèle logit multinomial avec échantillonnage aléatoire : Quelles variables et quelle taille d'échantillon ?.....	129
Chapitre 7 - Une structure hiérarchique pour modéliser les choix de localisation : Quel apport ?	131
7.1. Définition des nids et estimation d'un modèle emboîté	132
7.2. Comparaison avec d'autres structures de modèle	148
Estimation d'un logit emboîté : Quel apport ?.....	156
Chapitre 8 - Analyse diachronique des déterminants de localisation des ménages (1999, 2007)	157

8.1. Analyse descriptive.....	158
8.2. Résultats des estimations et comparaison.....	166
8.3. Analyse de la transférabilité temporelle.....	171
Le modèle estimé en 1999 est-il transférable ?.....	173
Conclusion	173
Conclusion générale.....	175
Bibliographie.....	179
Annexes	192
Liste des illustrations.....	212
Liste des tableaux.....	212
Liste des figures.....	213
Liste des graphiques.....	214
Liste des cartes.....	215
Table des matières	216

Introduction

« La maison est notre coin du monde. Elle est un refuge qui nous assure une première valeur de l'être : l'immobilité. ». G. Bachelard

Le foyer a toujours été l'endroit où à travers la sédentarité la personne cherche à la fois la sécurité, le confort, la proximité du lieu de production et de consommation et le besoin d'interactions sociales. Le choix de l'emplacement de ce foyer a été fortement orienté par des déterminants naturels dans l'ère agricole tels que la présence de cours d'eau ou de fertilité de la terre, principal facteur intrinsèque de différenciation du sol selon Ricardo. L'introduction de la rente foncière par Von Thünen et le renforcement de son poids dans l'ère industrielle a fortement modifié les structures urbaines et les facteurs de localisation dans un contexte de concurrence, de division fonctionnelle de l'espace et d'externalités négatives liées à la production. Avec le développement technologique et des moyens de communication de l'ère post-industrielle, les modes de vie ont connu de fortes mutations impliquant ainsi l'accroissement de la mobilité résidentielle et quotidienne. Au-delà de l'universalité de certains facteurs liés au marché, les choix de localisation résidentielle deviennent de plus en plus complexes en fonction des caractéristiques socioéconomiques des individus et des contextes urbains dans lesquels ils s'inscrivent.

Les conséquences des choix de localisations résidentielles, leurs interactions complexes avec les systèmes urbains ainsi que leurs mutations permanentes dans le temps et dans l'espace génèrent aujourd'hui un besoin de compréhension et d'anticipation qui se traduisent par un regain d'intérêt à la fois sur le plan politique et académique.

Le choix de localisation résidentielle : un regain d'intérêt...

La thématique des choix de localisation résidentielle n'est pas nouvelle dans la littérature économique, mais son développement s'est produit de manière distincte entre deux visées. La première est théorique et consiste à analyser les situations d'équilibre des structures urbaines (Alonso, 1964 ; Fujita, 1989 ; Anas et al., 1998). La deuxième est empirique et s'intéresse à la compréhension des facteurs de choix de localisation résidentielle. Cette dernière est beaucoup moins développée, notamment en France, et elle est souvent orientée vers les conséquences des choix plutôt que leurs déterminants. L'étalement urbain et la ségrégation résidentielle

occupent une grande partie de cette littérature (Glaeser et Kahn, 2004 ; Goffette-Nagot et Schaeffer, 2013). C'est dans le cadre des modèles d'interaction transport-usage du sol, communément appelés LUTI (Land Use and Transport Interaction) qu'une récente littérature sur les déterminants de localisation des ménages s'est développée à travers la mesure des effets des différents facteurs dans une perspective de modélisation et de prospective.

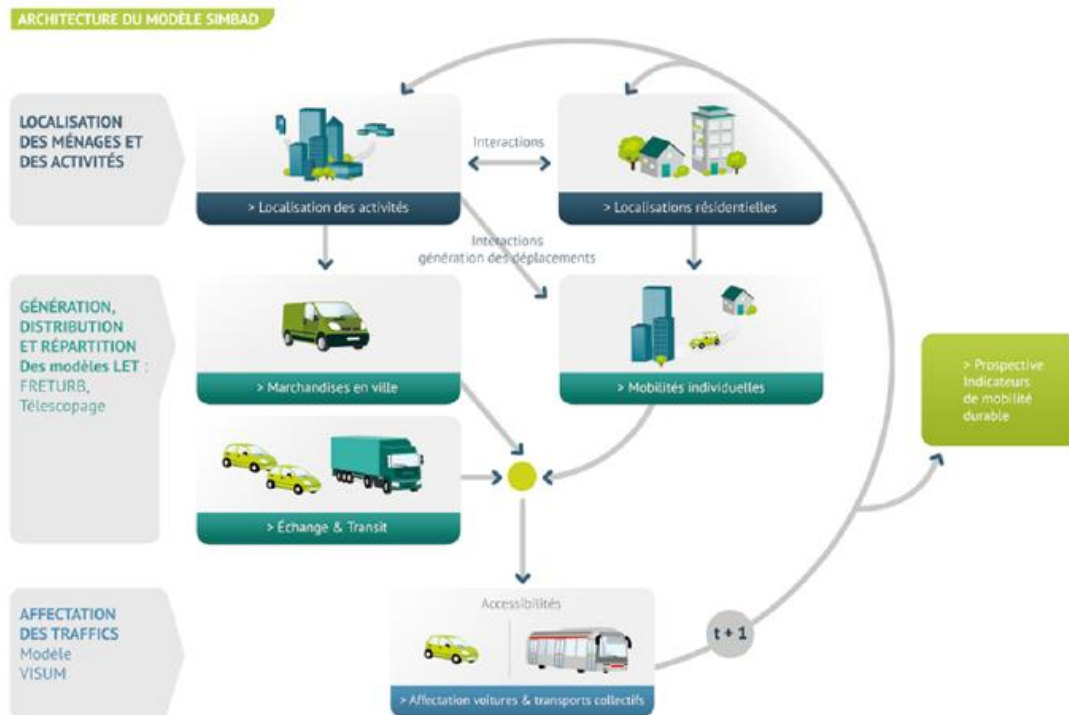
Appréhender les conséquences des choix de localisation résidentielle : étalement, ségrégation, croissance de mobilité

Depuis la fin des années 50, les choix de localisation des ménages de l'Europe et plus particulièrement de la France se font dans un contexte de permanence de la périurbanisation et d'un récent retour au centre. Ces choix de localisation sont aussi au cœur des problématiques liées à la périurbanisation, à la ségrégation urbaine et à la croissance de la mobilité, dans un contexte marqué par le renforcement des contraintes financière, énergétique et environnementale. Pour faire face à ces problèmes, la compréhension et l'anticipation des dynamiques urbaines deviennent indispensables.

Les modèles LUTI : outils d'anticipation et de prospective

L'organisation spatiale des villes est le résultat d'interactions complexes entre facteurs économiques, sociaux, culturels, politiques et technologiques. L'évolution du système de transport d'une part, des échanges économiques et des interactions sociales d'autre part se traduit par des évolutions d'accessibilités entre les zones urbaines, qui déterminent à long terme la configuration des localisations, que ce soit des résidences ou des autres activités. Toute prospective du système urbain nécessite donc une prise en compte des interactions entre sous-système des transports et déplacements et sous-système des localisations (Masson, 1998). Les travaux de recherche sur ces interactions ne sont pas nouveaux et ont conduit au développement des modèles d'interaction transport-urbanisme (LUTI) qui disposent aujourd'hui d'une abondante littérature (par exemple Wegener et Fürst, 1999 ; Simmonds et al., 1999 ; Waddell et al., 2001, Hunt et al., 2005). C'est dans cette perspective que s'est développé le modèle SIMBAD (Nicolas, 2010) dans lequel s'inscrit ce travail de thèse. La figure 1 illustre les interactions qui existent entre les différentes dimensions du système urbain et positionne celle en question dans ce travail de thèse, à savoir la localisation des ménages.

Figure 1 : Architecture du système modélisé par SIMBAD



Source : Nicolas, 2010

Un besoin de production de la connaissance

Le développement des modèles LUTI a conduit au regain d'intérêt pour analyser les choix de localisation, avec une analyse à un niveau désagrégé pour pouvoir mesurer directement les effets des déterminants de choix de localisation dans une perspective de modélisation à long terme.

Le développement des modèles LUTI, par leur visée opérationnelle, a également contribué à l'affranchissement de certaines contraintes méthodologiques liées à la gestion de données et à leurs analyses statistiques et économétriques.

...mais, des contraintes méthodologiques à surmonter

Les contraintes méthodologiques des modèles de choix de localisation résidentielle sont multiples. Elles sont liées au besoin de sources de données multiples pour caractériser à la fois les ménages et les lieux de localisation, à la méthode appropriée pour appréhender le nombre élevé d'alternatives et leur autocorrélation spatiale et à l'outil informatique pour mettre en œuvre le modèle de choix de localisation résidentielle.

Multitude des sources de données

Les données nécessaires à la modélisation des choix de localisation proviennent de plusieurs sources de données et leur rassemblement est confronté à plusieurs difficultés. La première tient à la disponibilité des données sur différentes échelles spatiales. La deuxième est liée aux manques de données relatives à certaines variables clés dans la modélisation des choix de localisation.

La méthode pour définir l'échelle spatiale du lieu de résidence...

La définition du zonage est centrale dans l'analyse des choix de localisation résidentielle, mais elle est confrontée au problème classique d'influence du découpage spatial (Modifiable Area Unit Problem) sur les résultats de modélisation (Openshaw, 1984). Quelle échelle spatiale faut-il utiliser pour analyser les choix de localisation des ménages ?

... et pour traiter le problème d'autocorrélation spatiale sous la contrainte du nombre élevé d'alternatives

La modélisation des choix discrets avec échantillon d'alternatives a été développée dès les premiers travaux de McFadden (1978) pour traiter le problème de la multitude des alternatives. Toutefois la consistance de ces méthodes d'échantillonnage était éprouvée seulement pour le modèle logit multinomial. Ce dernier possède la propriété d'Indépendance des Alternatives non Pertinentes, difficile à respecter dans les choix spatiaux comme les choix de localisation résidentielle. La question qui se pose est : comment développer un modèle qui permet à la fois de prendre en compte les autocorrélations spatiales et la multitude des alternatives de l'ensemble du choix de localisation ?

Récemment, Guevara et Ben-Akiva (2013a) proposent une nouvelle méthode qui permet de corriger le biais lié au tirage aléatoire des alternatives dans le cas des modèles de la valeur extrême généralisée (Generalised Extreme Value models-GEV). Néanmoins, l'application de cette méthode a pour le moment été limitée à des cas théoriques et à un jeu de données empiriques comprenant un nombre limité d'observations.

C'est dans cette perspective que nous avons choisi dans cette thèse de généraliser cette méthode sur des données complexes et de répondre aux différentes questions liées à l'application empirique de cette méthode.

L'analyse de la faisabilité de la mise en œuvre de cette méthode sur les données de choix de localisation va nous permettre également de tester d'autres types de modèles de la famille

GEV et d'évaluer à quel point l'application de nouvelles structures de modèles permet de traiter partiellement ou totalement les problèmes d'autocorrélation spatiale.

Objectifs et déroulement du travail

Cette thèse a pour objectif d'enrichir le corpus des travaux empiriques sur les déterminants de choix de localisation, en investiguant les développements récents de la modélisation des choix discrets en présence d'un nombre élevé d'alternatives.

Le premier apport est de comprendre les choix de localisation à travers une analyse synthétique de plusieurs travaux empiriques sur des contextes avec des spécificités bien définies afin de faire ressortir les principales dimensions à intégrer dans la modélisation des choix de localisation. Le travail sur deux dates (1999 et 2007) permet également de donner des éléments de réponses sur l'évolution des comportements des ménages en matière de choix de localisation.

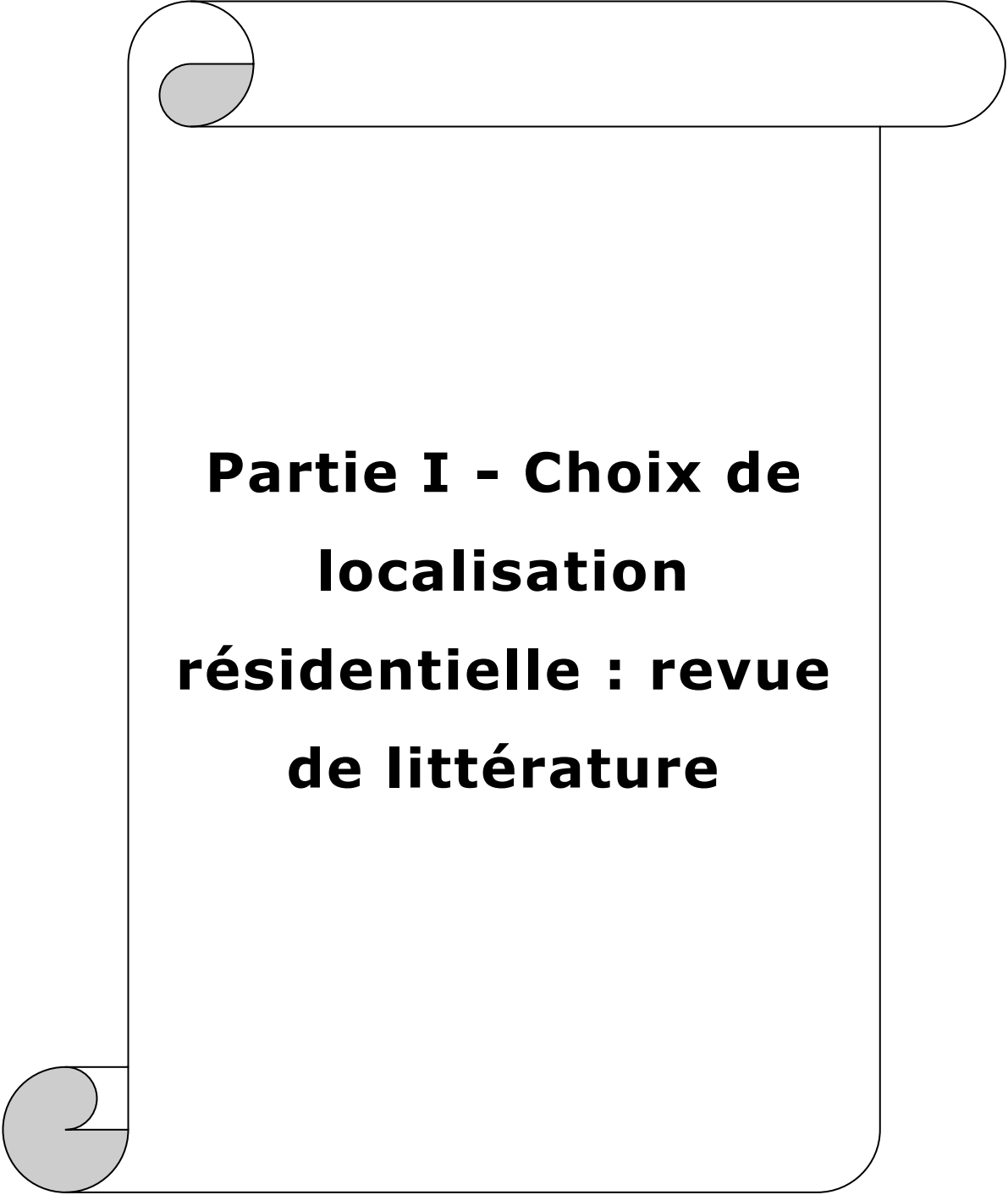
Le deuxième apport de cette thèse est d'ordre méthodologique et consiste à questionner plusieurs types de modèles de choix discrets afin de choisir celui qui permet à la fois de traiter le problème d'autocorrélation et la complexité des calculs et de l'appliquer ensuite pour modéliser les choix de localisation de l'aire urbaine de Lyon.

Nous avons procédé en deux temps. Dans un premier temps, nous nous sommes basés sur une revue synthétisée des déterminants des choix de localisation résidentielle pour estimer un modèle logit multinomial simple et identifier les facteurs qui expliquent les choix de localisation. Dans un second temps, nous avons opté pour un modèle logit emboîté avec un échantillon d'alternatives. L'idée est de prendre en compte la substitution des alternatives due aux autocorrélations spatiales. Cela permet également d'élargir les résultats des travaux récents (Guevara, 2010) sur des bases de données complexes et de répondre aux questions qui découlent de l'application de ces développements méthodologiques dans un contexte spatial, à savoir, la correction des biais liés à l'échantillonnage, la consistance du modèle, la constitution des nids et le nombre d'alternatives tirés à partir duquel les résultats d'estimation deviennent stables.

Cette thèse est organisée comme suit. La première partie de ce travail consiste à identifier les déterminants des choix de localisation résidentielle. Elle concerne non seulement la présentation des apports théoriques de l'économie urbaine et des interactions spatiales (chapitre 1), mais aussi la synthèse des travaux empiriques traitant les déterminants de la localisation résidentielle (chapitre 2).

La deuxième partie, est consacrée à la présentation des modèles de choix discrets traditionnellement utilisés (chapitre 3) et au questionnement des apports des développements récents et la possibilité de leur application pour améliorer le pouvoir prédictif des modèles de choix de localisation résidentielle (chapitre 4).

Après ce panorama interrogeant la faisabilité de l'application de différents types de modèles de choix discrets sur des contextes spatiaux, nous passons dans la troisième partie empirique à la confrontation des approches théoriques et des choix méthodologiques aux choix réels des ménages de l'aire urbaine de Lyon, prenant comme référence l'année 1999. L'ensemble des tests empiriques conduit, nous permet d'éclaircir les choix méthodologiques liés à l'application de ces modèles sur des aires urbaines avec des contraintes spécifiques. Le premier choix est lié à la constitution des variables caractérisant les ménages et les zones à partir de plusieurs sources de données et à la construction de l'échelle spatiale appropriée à l'estimation (chapitre 5). Ces variables sont ensuite testées afin de sélectionner celles qui représentent le mieux les choix observés des ménages (chapitre 6). Le deuxième choix consiste à adapter la structure hiérarchique retenue pour traiter les autocorrélations spatiales aux données réelles avec un nombre élevé d'alternatives et à la comparer avec les résultats du modèle traditionnellement utilisé (chapitre 7). Le dernier chapitre traite l'évolution des déterminants des choix de localisation résidentielle sur l'aire urbaine de Lyon, entre 1999 et 2007, en mettant l'accent sur la question de la transférabilité temporelle du modèle (chapitre 8).



**Partie I - Choix de
localisation
résidentielle : revue
de littérature**

Introduction

Avec l'émergence des recherches portant sur l'analyse des interactions transport-urbanisme, l'analyse et la compréhension des différentes composantes des décisions de localisation deviennent de plus en plus importantes.

Étant conscient de la complexité de réunir dans un seul macro-modèle tous les déterminants du sous-système de localisation (ménages et activités), nous avons choisi de nous concentrer sur la question des choix de localisation résidentielle afin d'apporter des éléments tangibles en nous basant sur des études théoriques ayant pour but de comprendre l'équilibre général des formes urbaines ou les causes de la croissance de la mobilité et de l'étalement urbain.

L'objectif de cette partie est de construire un cadre théorique des choix de localisation, à travers les différentes approches étudiées à savoir, le modèle standard de l'économie urbaine avec ses différentes extensions issues de la combinaison de la théorie des choix de localisation et celle du choix du consommateur, le modèle de l'interaction spatiale sous ses deux formes (gravitaire et maximisation de l'entropie) et enfin la théorie de l'utilité aléatoire. Une telle étude va nous permettre, dans un premier temps, de questionner l'opérationnalité de ces différentes approches pour modéliser les choix de localisation, et dans un second temps, de synthétiser les travaux empiriques qui ont été menés pour tester les facteurs de choix de localisation sur d'autres contextes urbains.

Nous présentons dans un premier chapitre les différentes approches théoriques utilisées dans la modélisation des choix de localisation, avant de nous pencher sur les différents facteurs conditionnant ces choix (chapitre 2).

Chapitre 1 - Cadres théoriques du choix de localisation résidentielle

” if urban modelers must wait for accurate empirical models of cities based on economic theory, it will take forever before we can analyze public policies or make useful predictions of urban change” (Cité par Anas (1982))

La thématique des choix de localisation résidentielle en milieu urbain repose sur un corpus théorique qui remonte à la théorie de la rente foncière de Von Thünen, qui explique la répartition des cultures en fonction de leur distance au marché situé au centre. En s'appuyant sur la rente d'enchère différentielle et des hypothèses simplistes concernant le ménage et l'espace, Alonso (1964) a développé le premier modèle permettant d'expliquer le choix de localisation résidentielle. Ce modèle standard de la microéconomie urbaine a donné suite à plusieurs extensions en relâchant successivement les hypothèses de départ. En parallèle, Lowry (1964) a développé des modèles opérationnels basés sur l'approche des interactions spatiales pour comprendre les articulations entre les décisions de localisation et la demande de transport. Le développement des modèles de choix discrets (McFadden, 1978) a permis de contourner la rigidité du modèle standard et de proposer un nouveau cadre théorique, sur la base de l'utilité aléatoire, pour des applications plus réalistes dans le domaine des choix de localisation des ménages.

L'objectif de ce chapitre est de présenter les différents cadres théoriques traitant de la localisation résidentielle, afin de positionner celui retenu pour appréhender la problématique de la thèse.

Dans sa classification des différentes approches de modélisation des choix de localisation résidentielle, Pagliara et al. (2010) distinguent les approches de l'interaction spatiale des approches économiques (Cf Figure 1.1). En suivant ce même agencement, nous commençons dans ce chapitre par la présentation du modèle standard de l'économie urbaine et ses différentes extensions (section 1.1). Cette approche économique, issue des modèles

mathématiques, repose sur la rente foncière. Nous présentons ensuite les modèles de l'interaction spatiale (section 1.2), avec les deux formes gravitaire et de maximisation de l'entropie. La dernière section (1.3) est consacrée à la présentation de la théorie économique de l'utilité aléatoire¹.

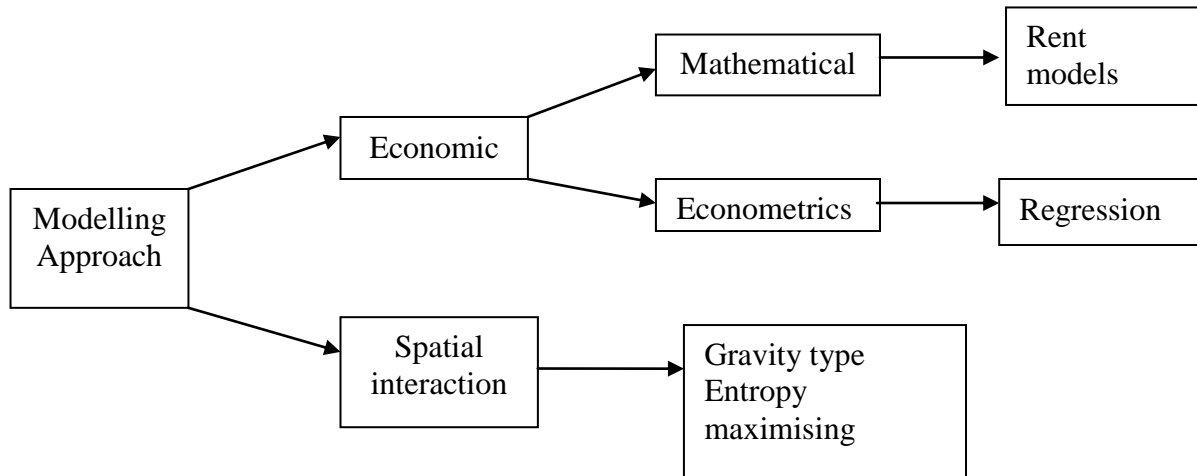


Figure 1.1 : Classification des approches de la modélisation des choix résidentiels

Source: Pagliara et al. (2010)

¹ Certains travaux utilisent l'approche de la maximisation de l'entropie pour développer leurs modèles. Toutefois, ils appliquent le logit multinomial pour calibrer leurs modèles sur des données réelles. Tout cela, peut faire appel à des modèles économétriques pour calculer des variables intermédiaires. C'est le cas de l'estimation du modèle hédonique à l'aide des régressions pour caler un modèle de choix de localisation. Pour cela, l'approche économique issue de l'économétrie ne sera pas présentée dans ce chapitre.

1.1. Le modèle standard de l'économie urbaine et ses extensions

Cette section est réservée au cadre théorique de l'économie urbaine. Toutefois, nous avons introduit aussi certains résultats empiriques qui ont montré les limites du modèle standard et qui ont été à la base des différentes extensions.

1.1.1. Le modèle standard d'Alonso

La question des choix de localisation a été abordée dans le modèle canonique d'Alonso (1964) comme un problème de maximisation de l'utilité sous contrainte budgétaire. En combinant la théorie de la localisation et la théorie du consommateur, il a repris le modèle de Von Thünen pour analyser les choix de localisation. Il s'agit de considérer un centre d'emploi appelé Central Business District (CBD) qui regroupe l'ensemble des emplois et de caractériser les lieux de résidence par une distance d qui les séparent du centre. Les ménages sont supposés maximiser leur utilité $U_{X,S}$ sous contrainte budgétaire. Le programme de maximisation s'écrit:

$$\text{Max } U_{X,S}(X;S) \text{ sous contrainte } Y=X+R(d)S+T(d) \quad -1.1-$$

Tel que :

Y : Revenu du ménage.

X : Quantité du bien composite consommé.

R(d) : Représente le prix d'une unité de superficie pour un logement situé à une distance d du centre.

S : Superficie du logement.

T(d) : Coût du trajet journalier domicile-travail, qui est supposé croître avec la distance au centre.

Ce modèle standard de localisation est basé sur deux hypothèses fondamentales. La première considère que l'espace est homogène autour d'un marché de l'emploi situé au centre. La seconde suppose que les ménages sont identiques.

La maximisation de ce système engendre un processus d'enchère et permet de dériver les conditions d'équilibre. Elle met en avant deux déterminants principaux : le prix de l'immobilier et la distance au centre qui représente l'accessibilité à l'emploi.

En considérant une localisation centrale des emplois, ce modèle standard adopte une logique d'équilibre partiel. Anas (1982) fait la distinction entre deux générations de modèles. La première génération est caractérisée par un équilibre partiel avec seulement le secteur résidentiel. Elle est centrée sur l'équilibre avec des préférences et des revenus non uniformes. Ce type de modèle permet de dériver des résultats plus complexes en analysant simultanément l'emploi, le logement et le secteur des transports d'une ville. L'objectif de ces modèles est de déterminer la taille et la capacité productive d'une ville monocentrique, mais ils constituent également un cadre théorique pour l'analyse des configurations urbaines.

La seconde génération de modèles est caractérisée par un équilibre général. Elle porte plus d'intérêt aux externalités et plus spécifiquement à la congestion et à l'interaction entre plusieurs secteurs, tels que le logement et l'offre de transport. Elle permet d'intégrer également le contrôle du secteur public, à travers les différentes politiques qui consistent à minimiser les coûts sociaux ou maximiser la fonction du bien-être collectif.

Très tôt, l'analyse des élasticités de Wheaton (1974) a montré que le rapport coût du transport/ quantité du sol est constant entre les différentes classes de revenu et qu'il ne peut pas être le seul déterminant de la localisation des ménages en fonction du revenu.

En confrontant les choix d'équilibre issus du modèle standard et les choix réels de localisation des ménages, certains travaux empiriques (Hamilton, 1982 ; Small et Song, 1992) montrent que la distance domicile-travail observée s'écarte de celle du modèle standard. Il s'agit de la littérature portant sur le « Wasteful commuting » qui a montré que les ménages ne cherchent pas forcément à minimiser la distance domicile-travail. Cela étant, la distance n'est qu'un proxy du temps et du coût du transport, qui dépend du mode utilisé pour aller au travail mais aussi pour réaliser les autres activités. Ce concept du « Wasteful commuting » a mis en avant l'effet de la structure urbaine et notamment l'émergence des formes polycentriques sur les comportements de déplacements et elle a donné lieu à de nombreuses extensions du modèle standard.

1.1.2. Extensions du modèle standard

Pour pallier aux principales critiques émises à l'encontre de ce modèle, à savoir l'exogénéité de la localisation de l'emploi et l'homogénéité de l'espace urbain et des ménages (Anas, 1990), plusieurs travaux ont relâché certaines hypothèses restrictives du modèle standard, et plus particulièrement la forme monocentrique, l'absence des aménités spatiales et sociales, la présence d'un seul mode de transport et la continuité de l'espace.

1 *Une forme non-monocentrique*

Henderson et Slade (1993) utilisent une approche de la théorie des jeux pour montrer que la configuration d'équilibre de l'usage du sol est le résultat d'un équilibre multiple. Il s'agit de considérer un jeu à un niveau pour deux types de villes (fermée et ouverte) et de tester, à l'équilibre de Nash, si les entreprises se localisent l'une à côté de l'autre (une ville monocentrique) ou si elles choisissent plutôt de se localiser séparément (une ville duocentrique).

Richardson (1977) a proposé une alternative au modèle standard. Il s'agit d'un modèle plus général avec une fonction d'utilité plus complexe et avec plus de contraintes. Les utilités marginales dérivées de l'équation d'équilibre du modèle théorique montrent que la rente d'enchère dépend non seulement de la distance au centre d'emplois (CBD), mais aussi de la distance aux différents centres secondaires. La minimisation du temps de déplacement ou la maximisation de la surface de localisation sont liées aux classes de revenu des ménages. Le processus du choix de localisation est donc le résultat d'un arbitrage complexe entre l'accessibilité et la qualité du voisinage, le temps de déplacements domicile travail et la surface du logement.

2 *Prise en compte des externalités*

En mettant en avant l'exemple de la répartition opposée des populations, riches et pauvres, entre Paris et Détroit, Brueckner et al. (1999) expliquent la différence de choix de localisation résidentielle par les préférences pour les aménités. Les auteurs distinguent entre les aménités exogènes, considérées comme caractéristiques spatiales de la zone (aménités naturelles ou historiques), et les aménités endogènes ou modernes qui dépendent en grande partie des conditions économiques du quartier et notamment du revenu de la population (restaurants, théâtres, équipements sportifs...). En s'appuyant sur le modèle standard d'Alonso, Muth et Mills (AMM) et en introduisant l'aménité exogène dans la différence des enchères des pauvres et des riches, les auteurs montrent que les populations riches habitent en périphérie si

l'aménité augmente moins vite que la quantité des sols quand le revenu augmente. Sinon ils optent pour une localisation centrale. C'est la différence de quantité d'aménités centrales et de leur valorisation par la population aisée dans les deux villes qui explique finalement cette répartition opposée entre Paris et Détroit. La sensibilité par rapport au type d'aménité peut être différente d'une ville à l'autre.

En parallèle, Hoang Huu et Wakely (2000) ont intégré les aménités sociales et les caractéristiques du logement, pour expliquer la distribution spatiale des ménages, en fonction de leurs statuts socioprofessionnels. Compte tenu du contexte urbain, ces aménités sociales peuvent être liées à la culture, à l'appartenance ethnique ou à la religion. A l'opposé du modèle standard de l'économie urbaine, ce modèle théorique (Hoang Huu et Wakely, 2000) a montré que les arbitrages du ménage sont de nature sociale plutôt qu'économique. Il met en avant un autre arbitrage, hors marché, entre les caractéristiques du logement et la qualité du voisinage dans sa composante sociale.

Dans une ville linéaire avec des aménités exogènes et centrales, deux pôles d'emploi, l'un au centre et l'autre en périphérie, Ng (2008) a considéré deux types de ménages. Le premier préfère les aménités et l'autre non afin d'expliquer la différence entre distance domicile-travail réelle et celle estimée par le modèle urbain standard. Il a trouvé que les ménages qui préfèrent les aménités acceptent de se localiser loin de leurs emplois pour être proche des aménités. Ces résultats expliquent le choix des ménages de parcourir des distances domicile-travail plus grandes que celles estimées par le modèle standard par l'importance de la proximité aux aménités.

3 Différentiation entre les modes de transports

En introduisant la différenciation entre les modes de transport pour tester son impact sur la répartition des ménages selon le revenu, LeRoy et Sonstelie (1983) montrent que l'introduction d'un nouveau mode rapide ne modifie pas la répartition initiale, les pauvres au centre et les riches en périphérie, si les pauvres ne peuvent pas y accéder financièrement. En revanche, si ce mode est coûteux pour les riches ou abordable pour les pauvres, alors la répartition des ménages est modifiée car les riches préfèrent plutôt une localisation centrale et les pauvres se trouvent contraints de se localiser en périphérie.

4 Une approche discrète de l'espace

La majorité des travaux cités ci-dessus considèrent un espace continu. Empiriquement, les zones de localisation sont discrètes, cela limite l'opérationnalité des modèles issus du modèle standard de l'économie urbaine. A l'opposé des modèles continus, les modèles discrets ont

plusieurs points d'équilibre. Cependant, avec un nombre important de population ces points d'équilibres convergent.

En considérant un espace discret, Herbert et Stevens (1960) ont développé un modèle dans la lignée du modèle d'Alonso. Il s'agit d'utiliser une représentation discrète de la zone et de maximiser la rente d'enchère selon un programme linéaire. Cette représentation discrète de l'espace permet de rendre les modèles standards de l'économie urbaine plus opérationnels. À l'optimum, il permet d'identifier la distribution optimale pour les résidences des ménages.

1.2. L'apport de l'approche des interactions spatiales

Avec l'émergence des recherches portant sur l'analyse de l'interaction transport-urbanisme, une approche modélisatrice basée sur le principe de l'interaction spatiale a été développée pour intégrer en un seul modèle urbain le sous-système des localisations et le sous-système de transport. Cette approche s'appuie sur des analogies entre la mobilité des ménages et la loi d'attraction de Newton en physique ou la maximisation de l'entropie en thermodynamique.

En se basant sur les principes du modèle gravitaire, Lowry (1964) a développé deux modèles de localisation : un modèle résidentiel et un modèle de localisation de service. Il s'agit de considérer que l'accessibilité à l'emploi (respectivement l'accessibilité à la clientèle potentielle) est le seul élément qui guide les choix de localisation résidentielle des ménages (respectivement choix de localisation des activités). En sortie, ce modèle permet d'estimer : i) la taille de la population urbaine et sa localisation dans les n zones de la ville, ii) le volume de l'emploi des services et sa localisation et iii) la demande de transport en ville (déplacement domicile-travail et domicile-commerce).

Wilson (1967) a montré que ces modèles gravitaires peuvent être dérivés statistiquement et le terme « modèle gravitaire » était remplacé par « modèle d'interaction spatiale ». Ces modèles s'inspirent du concept de la thermodynamique pour expliquer les choix résidentiels (Wilson, 1970). En considérant une matrice de déplacements entre un ensemble de zones d'origines et de zones de destinations d'une population P , cette approche consiste à maximiser l'entropie d'un système qui se représente sur trois échelles : i) macro-état, ii) méso-état et iii) micro-état. L'échelle du macro-état correspond à l'information de base et elle spécifie seulement le nombre total d'individus dans chacune des zones d'origines (O_i) et le nombre d'emplois dans

chaque zone de destination (D_j). L'échelle du méso-état correspond au nombre d'employés qui se déplacent entre une zone d'origine et une zone de destination (T_{ij}), c'est l'inconnue importante du système. La dernière échelle réfère à l'origine et à la destination de chaque individu au niveau du système. Elle est par nature inconnue mais somme toute peu importante.

En posant $P=T$, l'entropie du système (S) s'exprime de la manière suivante :

$$S(T_{ij}) = \frac{T!}{\prod_i \prod_j T_{ij}!} \quad -1.2-$$

La maximisation de cette fonction de l'entropie sous les trois contraintes définies par Wilson, à savoir la contrainte d'origine ($\sum_j T_{ij} = O_i$), la contrainte de destination ($\sum_i T_{ij} = D_j$) et la contrainte selon laquelle le coût global du transport (C) doit être fini ($\sum_i \sum_j c_{ij} T_{ij} = C$, tel que c_{ij} est le coût de transport de i à j), donne la solution analytique du système (T_{ij}).

L'avantage de ces modèles réside dans leurs capacités à considérer non seulement un espace urbain avec un seul centre d'emplois, mais aussi d'autres structures urbaines multipolaires et multifonctionnelles.

L'opérationnalité fait aussi la force de ce type de modèles, car ils utilisent une logique simplificatrice pour intégrer dans un seul modèle deux décisions différentes de localisation (des ménages et des activités).

Senior (1973) a identifié trois types d'améliorations pour développer les modèles issus de l'interaction spatiale. Le premier type s'intéresse à l'amélioration de la consistance de ces modèles, le deuxième type se concentre sur la désagrégation du modèle et le dernier type relâche l'aspect statique de l'équilibre de ces modèles.

Le modèle du bureau national de la recherche économique (Ingram et al., 1972 ; Kain et al., 1976 cité par Anas, 1982) reste le premier modèle de simulation à long terme fondé sur une théorie purement économique, conçue pour analyser les politiques du marché du logement. Ce modèle consiste à simuler l'offre et la demande de la mobilité résidentielle à travers un processus séquentiel, qui commence par l'estimation du nombre de ménages qui déménagent, et finit par les localiser. Selon Anas (1982), ce modèle reste meilleur que celui de Lowry (1964), même s'il souffre des problèmes de consistance comportementale et systémique.

En s'affranchissant des difficultés liées à l'application des modèles basés sur l'approche d'interaction spatiale, un travail récent de Putman (2010) présente le processus suivi pour

développer un modèle de localisation résidentielle des ménages de 30 régions métropolitaines aux Etats-Unis. En se basant sur le principe de la maximisation du surplus du consommateur (Cochrane, 1975 cité par Putman, 2010), ce modèle consiste à mesurer le bénéfice agrégé ressenti par les ménages pour une localisation donnée (au moins huit catégories de ménages). Deux méthodes se présentent pour mesurer ce surplus. La première utilise un modèle logit multinomial et calcule le surplus associé à une localisation comme étant l'utilité indirecte agrégée (McFadden, 1978) et la deuxième intègre directement la fonction de la demande du transport.

Les deux approches présentées jusqu'à présent et plus particulièrement les modèles à espace discret, comme ceux de programmation linéaire, ou ceux basés sur les interactions spatiales, permettent de proposer des modèles opérationnels. Toutefois, ces modèles s'appuient principalement sur l'accessibilité pour expliquer le choix de localisation.

Au-delà de l'importance de cet élément dans la détermination des choix de localisation, la décision du ménage en matière de choix d'une zone ne se limite pas à son niveau d'accessibilité, mais elle intègre également les spécificités sociales et spatiales de la zone et elle dépend aussi des caractéristiques du ménage. Devant la rigidité de ces modèles pour intégrer l'ensemble de ces dimensions dans l'analyse, une approche alternative et parallèle s'est développée. Il s'agit de la théorie de l'utilité aléatoire.

1.3. La théorie de l'utilité aléatoire

L'approche de l'utilité aléatoire est aujourd'hui la solution privilégiée par les modélisateurs pour combler l'écart entre la théorie de l'économie urbaine et les outils opérationnels de l'approche des interactions spatiales. Il s'agit d'une approche probabiliste fondée sur un mécanisme aléatoire. Selon le mécanisme aléatoire qui gouverne le choix, de Palma et Thisse (1987) distinguent des modèles avec des règles de décision stochastiques et des modèles avec utilité stochastique. Les premiers se caractérisent par des utilités déterministes mais un processus de choix probabiliste, ce qui conduit à ce que l'individu ne choisisse pas nécessairement l'alternative avec la probabilité la plus élevée. Il s'agit ainsi d'une rationalité limitée. Le deuxième type de modèle s'inscrit dans la tradition de la théorie néo-classique des choix. Deux interprétations sont proposées par ces modèles : la première est celle de la psychologie des choix (Thurstone, 1945 cité par de Palma et Thisse, 1987) et elle donne lieu au développement des modèles probit binaire (voir partie 2) pour mesurer le stimulus

économique (qui peut être interprété comme l'utilité). La seconde est celle développée à partir de l'axiome des choix par Luce (1959) et retenue ensuite par les économistes dans les développements des modèles de choix discrets (McFadden, 1978).

La fonction d'utilité s'exprime alors comme une variable aléatoire. Cela n'implique pas le manque de rationalité du ménage, mais plutôt le manque d'information concernant les caractéristiques de l'alternative (Manski, 1977).

La spécification de la fonction aléatoire de l'utilité est faite sous l'hypothèse selon laquelle le terme aléatoire est indépendamment et identiquement distribué. Cette hypothèse a permis de dériver des modèles avec des propriétés analytiques plus souples. Toutefois, Debreu (1960) a introduit le paradoxe du blue/red bus pour montrer la présence de situations où cette hypothèse n'est pas respectée.

Le travail de McFadden (1978) avait été précédé par un ensemble de travaux empiriques utilisant les modèles de choix discrets (Quigley, 1977 ; Lerman, 1976). Toutefois, il reste le pionnier qui a posé les jalons de la modélisation des choix de localisation résidentielle à partir d'une approche de l'utilité aléatoire.

Ces modèles de choix discrets consistent à considérer une population de ménages confrontés au même ensemble **A** d'actions et à chercher à déterminer le nombre de ménages qui choisissent une certaine action (pourcentage de population choisissant l'action **A**). Tout ménage **n** est supposé avoir une fonction d'utilité déterministe **U** définie sur **A**. Or, le modélisateur n'observe pas toutes les caractéristiques des actions influençant le comportement de choix du ménage. La fonction d'utilité se décompose en deux termes déterministe et aléatoire. Elle s'écrit alors :

$$U_{in} = V_{in} + \varepsilon_{in} \quad -1.3-$$

Tel que **V_{in}** est une fonction linéaire des caractéristiques de l'alternative **i** et du ménage **n**, et **ε_{in}** est le terme d'erreur.

Le ménage étant rationnel, la probabilité de choisir une alternative (zone de localisation) **i** parmi l'ensemble **C_n** des alternatives disponibles (**P(i/C_n)**), s'écrit de la manière suivante :

$$\begin{aligned} P(i / C_n) &= P(U_{in} = \max_{j \in C_n} U_{jn}) \\ &= P(U_{in} \geq U_{jn} \forall j \in C_n) \end{aligned} \quad -1.4-$$

En remplaçant **U_{in}** par la forme de l'équation (1.3) et quel que soit **j** appartenant à **C_n**, la probabilité (**P(i/C_n)**) s'écrit :

$$\begin{aligned}
P(i / C_n) &= P(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}) \\
&= P(V_{in} - V_{jn} \geq \varepsilon_{jn} - \varepsilon_{in}) \\
&= P(V_{in} - V_{jn} \geq \varepsilon)
\end{aligned}
\tag{-1.5-}$$

Pour obtenir un modèle opérationnel, il faut préciser la distribution du terme d'erreur. Deux hypothèses se présentent. La première est de supposer que les ε_{in} et ε_{jn} suivent une loi normale et la deuxième consiste à assimiler la distribution du terme d'erreur à une distribution de Gumbel. C'est le choix de cette distribution qui permet de dériver plusieurs familles de modèles (Partie II).

L'approche d'utilité aléatoire comme cadre d'analyse de la localisation résidentielle

Le cadre théorique de la microéconomie urbaine a réduit la décision du ménage à un simple résultat d'arbitrage entre le prix de l'immobilier et le coût du transport, tandis que les modèles basés sur l'approche des interactions spatiales se sont concentrés sur le principe de l'accessibilité généralisée pour simuler les choix de localisation et comprendre les interactions transport-usage du sol. Bien que les modèles de l'économie urbaine et de l'interaction spatiale apportent plusieurs éléments et justifications théoriques sur le rôle des facteurs de marché foncier et immobilier, de l'accessibilité et des aménités urbaines, c'est l'approche des choix discrets avec ses fondements théoriques basés sur l'utilité aléatoire qui semble la plus appropriée pour l'analyse empirique des choix de localisation résidentielle en milieu urbain.

Le travail pionnier de McFadden (1978), posant les jalons de la modélisation des choix résidentiels à partir d'une approche de l'utilité aléatoire a permis de s'affranchir de ces complexités et de multiplier les travaux faits dans ce sens. La force de ces modèles est leurs capacités à intégrer d'autres déterminants de choix de localisation résidentielle. Ils permettent également de quantifier l'effet de différentes caractéristiques de la zone de localisation, du logement et leurs interactions avec les caractéristiques des ménages.

Après cette revue de la littérature sur les fondements théoriques des choix de localisation, le chapitre suivant propose une synthèse des principaux déterminants de choix de localisation identifiés dans les travaux empiriques.

Chapitre 2 - Les déterminants des choix de localisation résidentielle : travaux empiriques

Les modèles de l'économie urbaine, malgré leurs hypothèses restrictives, mettent en avant le rôle d'un certain nombre de facteurs déterminants dans le choix de localisation des ménages, à savoir l'arbitrage entre coût de transport et coût de l'immobilier, ainsi que le rôle des aménités spatiales et sociales. Avec le développement des modèles basés sur l'utilité aléatoire, plusieurs travaux ont testé empiriquement et dans différents contextes l'effet de ces facteurs pour assurer leurs validations. Ces modèles empiriques ont permis également de tester d'autres facteurs de choix de localisation, en prenant en compte l'hétérogénéité des préférences des ménages via leurs caractéristiques socioéconomiques.

L'objectif de ce chapitre est de proposer une revue synthétique des travaux empiriques traitant les déterminants de choix de localisation. Nous distinguons trois groupes de déterminants : i) ceux liés à l'arbitrage du marché et mis en avant par les modèles d'économie urbaine, ii) ceux concernant les aménités spatiales et/ou sociales et enfin iii) les déterminants associés aux caractéristiques des ménages. Au-delà de leur signification, il s'agit également de comparer le poids de chaque déterminant dans le choix de localisation résidentielle entre les différents contextes de villes (géographique, structure urbaine, politiques de logement...).

2.1. Arbitrage du marché : accessibilité vs prix de l'immobilier

En utilisant différentes méthodes de calcul et différentes approches modélisatrices, les travaux empiriques sur les déterminants des choix de localisation (tableau 2.1) ont confirmé le rôle de l'accessibilité à l'emploi dans les arbitrages des ménages, en matière de choix d'une localisation (section 2.1.1). En parallèle, certains travaux ont montré le rôle déterminant du

prix de l'immobilier dans le processus de choix de localisation par rapport à l'accessibilité (section 2.1.2).

2.1.1. Accessibilité à l'emploi

En considérant un système multidimensionnel et en se limitant aux ménages motorisés avec un seul actif à temps plein et un seul groupe ethnique de la région de Dallas pour éviter la complexité des calculs, Waddell (1993) a testé trois modèles : i) le premier, en estimant conjointement le choix de l'emploi, de la résidence et du statut d'occupation à l'aide d'un logit multinomial, ii) le deuxième, en considérant que le choix de résidence se fait conditionnellement aux choix de l'emploi et du statut d'occupation iii) le dernier, en modélisant ces choix selon une structure hiérarchique séquentielle. La comparaison de la log-vraisemblance de ces trois structures montre que les choix de l'emploi et de la résidence sont déterminés conjointement, et met en évidence les limites d'un modèle de choix résidentiel avec un seul centre d'emploi. Les résultats ont montré également que les caractéristiques du logement (la taille et la surface) déterminent plus les préférences que les coûts du transport. Même s'ils nuancent les hypothèses du modèle standard, mettant en avant l'endogénéité de l'emploi, ces résultats ne rejettent pas pour autant le rôle de l'arbitrage entre l'accessibilité à l'emploi et le prix de l'immobilier dans les choix de localisation des ménages.

En testant plusieurs types d'accessibilité (à l'emploi, aux parcs et aux commerces) et leur effet sur la valeur des biens immobiliers, sur la même région (Dallas), Srour et al. (2002) montrent à l'aide d'un logit multinomial avec un échantillon d'alternatives que l'accessibilité approchée par la distance à l'emploi du chef du ménage est la variable la plus déterminante. Ils montrent également que l'accessibilité des opportunités cumulatives (cumulative opportunity) aux autres opportunités est négative dans les choix résidentiels. Ils expliquent cela par l'absence d'autres variables qui mesurent la qualité des écoles et le niveau de la criminalité mais aussi que l'effet de l'accessibilité aux autres opportunités est plutôt capitalisé dans la valeur du bien immobilier.

En combinant les arbitrages du marché (prix de l'immobilier/accessibilité à l'emploi) avec les aménités (spatiales ou sociales) de la zone de localisation, la comparaison faite par Waddell (1996) à Honolulu-Hawaii, entre la localisation des ménages avec un seul ou deux actifs et leurs interactions avec la localisation de l'emploi et le statut d'occupation, a montré que la distance à l'emploi (commuting distance) est la variable la plus déterminante dans les arbitrages des choix de localisation. L'effet est deux fois plus important pour les locataires

que pour les propriétaires et cela pour les deux types de ménages (avec un ou deux actifs). L'accessibilité à l'emploi comme variable mesurant à la fois l'accès à l'emploi, le degré d'urbanisation et la densité a un effet plus important chez les ménages locataires avec deux actifs que ceux avec un seul actif. Si l'accessibilité à l'emploi et aux aménités est toujours significative, l'accessibilité à l'emploi reste plus importante que l'accessibilité aux autres activités, ce qui rejoint le travail de Srour et al. (2002) sur la région de Dallas.

Toujours, en utilisant une approche d'utilité aléatoire, une étude effectuée aux Philippines, à Manille (Rivera et Tiglao, 2005) montre que l'accessibilité et plus particulièrement le temps de transport est beaucoup plus influent que les caractéristiques du logement et le coût moyen du foncier dans le quartier de résidence.

En utilisant les préférences révélées par les locataires ayant déménagé au cours des 5 dernières années dans le grand Zurich en Suisse, la considération de plusieurs sous modèles a permis d'améliorer le pouvoir prédictif du modèle. Bürgle (2006) a trouvé que la distance à l'emploi est significative pour tous les types de ménages. Le paramètre est plus élevé pour les ménages sans voiture. Le temps d'accès au centre est négatif sauf pour les familles de grande taille. L'accessibilité par les transports publics est significative pour les familles sans enfant et non motorisées. La présence des hôpitaux dans des zones commerciales ou dans des zones de mauvaise qualité résidentielle augmente le prix de l'immobilier.

Chen et al. (2008) montrent que la sensibilité à l'accessibilité à l'emploi diffère en fonction de la distance à l'emploi de l'ancienne localisation. Si les ménages évitent les localisations avec une grande distance de leur emploi, les résultats confirment que les ménages qui parcouraient des distances moins longues entre leur emploi et leur ancienne localisation, donnent moins d'importance à la distance à l'emploi. L'accessibilité aux espaces verts joue inversement par rapport à l'ancienne localisation. Si les ménages étaient à proximité des espaces verts, ils les préfèrent moins que les ménages qui étaient loin de ces espaces. Par contre l'accessibilité aux commerces de détails joue moins dans les arbitrages des ménages.

En modélisant la mobilité résidentielle sur l'ensemble des communes du territoire belge, Cornelis et al. (2012) montrent aussi que l'accessibilité à l'emploi joue plus que le prix de l'immobilier. Si le prix de l'immobilier était le principal déterminant des choix résidentiels, les travaux empiriques trouvent des signes contre-intuitifs de ce déterminant, ou moins significatif, en croisant le prix de l'immobilier avec le revenu.

Dans les travaux de Tillema et al. (2010) sur les Pays-Bas, au sein d'une région polynucléaire, le coût du transport apparaît comme un élément important dans les décisions de relocalisation des ménages.

Guo et Bhat (2007) ont montré que l'accessibilité est toujours présente dans les arbitrages des ménages de la baie de San Francisco indépendamment de l'échelle utilisée. Cela concerne plus particulièrement les arbitrages rendus par les femmes. Par contre, en croisant le temps de déplacement à l'emploi avec le revenu, les auteurs montrent que les riches préfèrent la qualité du logement à l'accessibilité à l'emploi.

En s'attaquant au problème d'endogénéité lors de l'analyse du choix résidentiel et du choix du mode du transport, DeSalvo et Huq (2005) ont conduit une approche en deux temps. Ils utilisent dans un premier temps une minimisation du coût de déplacement sous contrainte distance-temps et vitesse, avant d'utiliser la maximisation de l'utilité sous contrainte budget et temps. Les résultats du modèle de minimisation du coût montrent que les ménages à haut revenu utilisent les modes les plus rapides et dépensent moins de temps dans les déplacements et que les ménages habitant loin de leurs destinations utilisent des modes rapides et réduisent leurs coûts de transport. Les résultats du modèle de maximisation de l'utilité montrent que les ménages à haut revenu non salarial préfèrent se localiser loin du CBD. Ils exploitent aussi les effets d'un coût de déplacement exogène et trouvent que le coût marginal de déplacement diminue avec le taux de revenu, questionnant ainsi les résultats de LeRoy et Sonstelie (1983) sur la relation entre le choix du mode et la distance.

Si le signe négatif de la variable de l'accessibilité à l'emploi obtenu dans le cadre de la modélisation résidentielle des ménages de Londres (Zolfaghari et al., 2012) peut être expliqué par la nature du marché du logement à Londres, qui se caractérise par une demande des zones de logement loin des centres d'emploi. Ce signe négatif obtenu dans d'autres travaux (Guo, 2004), ou l'absence de l'accessibilité comme facteur déterminant des choix résidentiels, ont conduit à un réexamen de l'effet de l'accessibilité (Lee et al., 2010b). Il s'agit de proposer une nouvelle approche espace-temps (time-space prism approach) consistant à intégrer la localisation de l'emploi et du travail ainsi que la chaîne de déplacements et estimer cinq modèles pour déterminer l'approche qui donne les meilleurs résultats. En procédant ainsi, les auteurs ont trouvé que tous les types d'accessibilité sont significatifs. L'effet de l'accessibilité reste dominé par les caractéristiques du logement et plus particulièrement la taille, ce qui confirme les résultats de Waddell (1996) et de Weisbrod et al. (1980). Le temps domicile

travail comme proxy de l'accessibilité à l'emploi reste le plus important par rapport à l'accessibilité aux autres activités.

Avec le développement des nouvelles technologies et plus particulièrement le télétravail, la distance à l'emploi peut devenir moins significative dans les arbitrages des ménages. En étudiant les choix de localisation des ménages des Pays-Bas qui télétravaillent et en les comparant avec les autres, Ettema (2010) a trouvé qu'il existe deux types de télétravailleurs : ceux qui sont propriétaires de maison dans les zones rurales et ceux qui sont les moins jeunes. Les premiers sont plus sensibles à la distance au travail que les seconds.

2.1.2. Prix de l'immobilier

Dans leur étude sur les ménages de l'aire métropolitaine de Minneapolis-St Paul, ayant récemment déménagé, Weisbrod et al. (1980) montrent que le prix de l'immobilier est plus important dans le processus de la localisation résidentielle que le temps de transport pour accéder à l'emploi, que ce soit en voiture particulière ou en transport en commun. Abraham et Hunt (1997) trouvent aussi que le prix de l'immobilier joue plus que l'accessibilité dans les arbitrages des ménages. La modélisation des préférences déclarées des ménages de la ville d'Edmonton au Canada (Hunt, 2010) a montré que le type du logement est le facteur le plus déterminant dans les choix résidentiels, suivi par les externalités négatives (qualité de l'air et les nuisances sonores). De plus, les effets du prix de l'immobilier et de l'accessibilité diffèrent en fonction du revenu. Les ménages à haut revenu sont plus sensibles au temps d'accès en voiture. En comparant les travaux effectués sur six cas d'études aux Pays-Bas et en Belgique, Molin et Timmermans (2002) ont constaté que même si l'accessibilité reste importante, son effet reste minime par rapport à d'autres variables et plus spécifiquement les caractéristiques du logement et de la zone. Les Pays-Bas et la Belgique ont une géographie similaire. Par contre, les autorités belges sont caractérisées par le "laisser faire" concernant l'urbanisation, la mobilité et la politique de la ville. Par ailleurs, les Pays-Bas se caractérisent plutôt par des politiques combinées de transport, de logement et d'urbanisme. Dans un travail plus récent sur la région d'Ottawa-Caleton au Canada, Tayyaran et al. (2003) montrent également que le prix de l'immobilier est le plus déterminant dans le choix de la localisation résidentielle.

En plus des caractéristiques intrinsèques du logement, le prix de l'immobilier intègre également certaines caractéristiques du voisinage (Rosen, 1974). La plupart des travaux de choix de localisation des ménages mettent en avant une sensibilité plus importante aux

aménités spatiales et sociales, par rapport au prix de l'immobilier et à l'accessibilité à l'emploi (Gayda, 1998 ; Tayyaran et al., 2003 ; Kim et al., 2005).

Les modèles de l'économie urbaine et ceux de l'utilité aléatoire modélisent la demande de transport d'une manière endogène. Ben-Akiva et Bowman (1998) et Elisson (2010) ont développé des approches qui découlent d'une "activity based accessibility". Elles consistent à modéliser les choix discrets de programmes d'activités qui endogénéisent le choix des activités à réaliser et le choix de leurs horaires.

En utilisant des mesures descriptives de distance et de temps entre l'ancienne et la nouvelle localisation et le lieu de l'emploi des ménages de Seattle, Clark et al. (2003) montrent qu'à partir d'un certain seuil de distance, les ménages tendent à changer de résidence pour se rapprocher de leur emploi, plus particulièrement chez les femmes dont le couple est biactif. Même s'ils nuancent les hypothèses du modèle standard, mettant en avant l'hétérogénéité des ménages, de l'offre de logement et du transport, ces résultats ne rejettent pas pour autant le rôle de l'arbitrage entre l'accessibilité à l'emploi et le prix de l'immobilier dans les choix de localisation des ménages (tableau 2.1).

Auteurs	Aire d'étude	Arbitrage du marché							
		Temps de déplacement domicile-travail par VP	Temps de déplacement domicile-travail par TC	Coût de déplacement domicile-travail par VP	Coût de déplacement domicile-travail par TC	Accessibilité à l'emploi	temps au centre	Distance à l'emploi	Prix immobilier
Guo et Bhat (2007)	Baie de San Francisco	(-)				(-) Accessibilité gravitaire			(-) (La valeur médiane du bien immobilier) x (l'inverse du revenu du ménage)
Bürgle (2006)	Zurich						(-) Sauf pour les familles de grande taille	(-)	(-) Rapport entre prix de l'immobilier et du revenu
Clark et al. (2003)	Seattle	Ce n'est pas un modèle de choix discret							
Cornelis et al. (2012)	Belgique					(+) Nombre d'emplois à moins 30 min			(-) Moins significatif que l'accessibilité
Ettema (2010)	Pays-Bas							Le signe varie si l'actif	

								télétravail	
Gayda (1998)	Bruxelles	(+)							Plus important chez les propriétaires de couleur blanche
	Lyon								
Guo (2004)	Les régions nord-central de l'état de Texas : University Park, Highland Park et Dallas	(-)				(-)			
Hunt (2010)	Edmonton	(-)	(-)	(-)	(-)				(-)
Kim et al (2005)	Oxfordshire	(-)		(-)					(-)
Lee et al. (2010a)	Pugetopolis	(-)				(+) Accessibilité gravitaire			(+) La différence entre le revenu du ménage et le prix de l'immobilier
Molin et Timmermans (2002)	Pays bas et Belgique	Plusieurs cas d'études pour comparer le poids de l'accessibilité. Les gens donnent plus de poids aux prix de l'immobilier par rapport à l'accessibilité							
Rivera et Tiglaio (2005)	Manille (Philippines)	(-)		(-)		(-) Accessibilité			(-)

						gravitaire			
Srouf et al. (2002)	La région de Dallas					(+) Accessibilité des opportunités cumulatives		(-)	(-)
Tayyaran et al. (2003)	Ottawa-Caleton (Canada)	(-)							(-)
Waddell (1993)	La région métropolitaine de Dallas	(-) Pas de distinction entre modes				(-) Densité de l'emploi dans la zone de résidence	(+) La distance au carré est (-)		
Waddell (1996)	Honolulu-Hawai					(-) Accessibilité gravitaire		(-)	(+) La différence entre le revenu du ménage et le prix moyen du bien immobilier
Weisbrod et al. (1980)	Minneapolis-St Paul					(+) Accessibilité issue d'un modèle de choix discret			(-) Le rapport entre la valeur unitaire du prix de l'immobilier et le revenu
Zolfaghari et al.	Aire urbaine	(-)				(-)			(-)

(2012)	de Londres				Accessibilité gravitaire			
--------	------------	--	--	--	-----------------------------	--	--	--

Tableau 2.1 : Synthèse des travaux empiriques mettant en avant le rôle de l'arbitrage du marché

2.2. Préférence pour les aménités spatiales et sociales

Dans leur travail appliqué à plusieurs villes américaines, Glaeser et al. (2008) présentent une synthèse des critiques du modèle de localisation résidentielle résultant de l'arbitrage du marché de l'immobilier et du coût de transport. Ils mettent en avant trois points en particulier. Le premier concerne la limite d'application du modèle standard à un cas réel pour expliquer les localisations des ménages en fonction du revenu. Le deuxième point souligne le rôle des contraintes du marché du logement et plus spécifiquement l'offre du parc ancien dans l'explication de la localisation centrale des populations pauvres. Les politiques publiques orientent également les choix de cette catégorie de population à travers la localisation des logements sociaux. Le dernier point met l'accent sur l'impact indirect des transports publics et la desserte en métro en particulier sur les préférences des ménages aisés pour le centre. Les zones bien desservies par les transports en commun attirent les populations pauvres, ce qui fait fuir les ménages aisés vers des zones qui ne sont accessibles qu'en voiture particulière. Wang et Li (2004) ont montré que les préférences des ménages pour les aménités sociales ou spatiales ainsi que la proximité aux transports publics jouent plus que les contraintes du marché et plus spécifiquement le prix de l'immobilier dans les arbitrages des ménages en matière de choix d'un logement. Dans cette section, nous distinguons entre les travaux sur le rôle des aménités spatiales, à savoir la proximité aux services quotidiens, la qualité des écoles, et les espaces verts (2.2.1) et les travaux sur les aménités sociales (2.2.2).

2.2.1. Préférence pour les aménités spatiales

En analysant les travaux sur le rôle des préférences des ménages pour les aménités spatiales dans les choix de localisation (tableau 2.2), nous avons trouvé que ces travaux sont de deux types. Le premier type utilise des approches quantitatives et le deuxième des approches qualitatives basées sur des entretiens semi-directifs.

- **Approche quantitative**

En questionnant directement la localisation des ménages propriétaires dans le comté de Franklin dans l'Ohio, Bayoh et al. (2006) ont estimé un modèle de choix discret composé de deux types de variables. Le premier type de variables correspond au mécanisme de « fuite face à la rouille » (*flight from blight*) et le deuxième type mesure l'évolution de certains facteurs comme l'accessibilité, le revenu et le cycle de vie. En termes de résultats, ils ont trouvé que les variables liées aux *flight from blight* comme la qualité des écoles et les taxes ont un effet plus important sur les choix de localisation que les autres variables telles que le temps de déplacement domicile-travail ou l'âge du parc du logement qui ont aussi un effet significatif. Kim et al. (2005) trouvent aussi que la qualité des écoles est le facteur le plus déterminant de la localisation des ménages d'Oxfordshire, par rapport à l'accessibilité à l'emploi et à d'autres aménités telles que les supermarchés.

Les résultats de ces travaux, obtenus à partir de données en coupe instantanée, sont en grande partie confirmés par d'autres études prenant en compte la dynamique des choix résidentiels. L'analyse des localisations des ménages des Etats-Unis (Feridhanusetyawan et Kilkenny, 1996), à partir d'un panel de la dynamique de revenu sur plusieurs années (PSID), a montré que les aménités jouent plus que l'accessibilité à l'emploi dans les arbitrages des ménages. Ils ont montré aussi que la faible densité des zones rurales est l'élément le plus important qui motive les ménages à préférer les zones rurales. Même la taxe élevée des zones rurales ne permet pas de réduire la probabilité des ménages de se localiser dans ces zones.

Dans son travail de thèse, Kim (2010) généralise cet arbitrage à un arbitrage entre localisation et qualité de logement. Il s'agit d'estimer un modèle emboîté à travers lequel le ménage choisit d'abord l'appartenance à une communauté avant de choisir son logement au sein de la communauté. Le choix de la communauté se fait en fonction des aménités spatiales et sociales de la zone (qualité des écoles, taxes...). Les caractéristiques socioéconomiques des ménages ont été également intégrées dans le modèle en les croisant avec les attributs des alternatives pour tenir compte de l'hétérogénéité des préférences des ménages.

La sensibilité par rapport au type d'aménité peut être différente d'une ville à l'autre. Dans une comparaison des préférences déclarées des Lyonnais et des Bruxellois, Gayda (1998) a montré, par exemple, que la préférence pour la présence de commerces dans le quartier est plus importante à Bruxelles alors que la nature et l'environnement sont plus importants à Lyon.

- **L'approche qualitative**

D'autres travaux s'appuyant sur une approche qualitative ont confirmé le rôle des aménités spatiales dans les choix de localisation. A travers des entretiens avec des ménages de l'aire urbaine de Kitchener au Canada, Filion et al. (1999) ont analysé l'arbitrage entre l'accessibilité, la surface et les caractéristiques intrinsèques au logement et la proximité aux aménités spatiales dans une aire urbaine marquée par l'étalement et la dispersion de l'habitat. Parmi les résultats des scores donnés aux différentes caractéristiques évaluées lors d'un choix d'une zone de résidence, les critères de sécurité et de proximité aux centres commerciaux viennent en tête des préférences des ménages.

Les entretiens semi-directifs menés dans plusieurs communes du Val-de-Marne dans la région parisienne (Gueymard, 2006) ont permis de mettre en évidence le rôle des caractéristiques physiques de l'environnement comme la présence des espaces verts dans les choix résidentiels.

Bonvalet (2010) a questionné les logiques des choix résidentiels des habitants de l'Île de France et a montré que les arbitrages des choix résidentiels sont contraints par la logique institutionnelle. Certains ménages comme les bénéficiaires des logements HLM ne choisissent pas leurs localisations mais arbitrent entre quelques logements proposés, ce qui met l'accent sur le rôle des politiques publiques dans la détermination des choix de localisation.

2.2.2. Préférence pour les aménités sociales

Indépendamment des caractéristiques spatiales de la zone et des contraintes du marché (prix de l'immobilier et coût du transport), les préférences des ménages d'appartenir au même groupe social ou de se rapprocher de leur groupe ethnique expliquent la différence des choix de localisation (tableau 2.2). L'analyse factorielle des données de recensement de Montréal (Le Corre, 1983) a montré que le statut socio-économique est la variable qui explique le plus la différenciation spatiale observée dans la ville de Montréal.

Le travail de Waddell (1993) sur la région de Dallas a montré aussi une tendance de stratification en fonction de l'ethnicité. Chattopadhyay (2000) a confirmé ces résultats en modélisant les choix de localisation des ménages de la région métropolitaine de Chicago. En testant quatre structures hiérarchiques moyennant un modèle emboîté à trois niveaux, l'auteur montre que le consentement marginal à payer pour se localiser dans une zone où se concentre la population blanche est élevé. Quigley (1985) a modélisé les choix de localisation des ménages de la région métropolitaine de Pittsburgh en Pennsylvanie en trois étapes : le ménage

choisit d'abord la municipalité, puis le voisinage avant de choisir son logement. En procédant ainsi, l'auteur a montré que la composition ethnique de la population est clairement importante dans les choix des ménages pour une zone de localisation. Toute chose égale par ailleurs, la population noire a tendance à choisir des zones peuplées par des ménages de la même couleur, alors que la population blanche donne moins d'importance à ce critère. Galster (1988) s'oppose à ce type d'interprétation dans la lignée de Clark (1986) et considère que la concentration des populations noires dans les mêmes lieux n'est pas tant le résultat des préférences pour l'entre-soi, mais la conséquence de la discrimination raciale et de l'éviction sur le marché du logement (Gabriel et Rosenthal, 1989). Ce sont les populations blanches, plus aisées, qui choisissent leur localisation suivant des mécanismes de fuite, d'évitement et de recherche d'entre-soi. En France, Goffette-Nagot et Schaeffer (2013) montrent que les cadres et les indépendants préfèrent une localisation occupée par les populations riches. Cela vient confirmer le fait que la sensibilité aux caractéristiques de la zone d'habitat diffère d'une catégorie sociale à une autre (tableau 2.2) et dépend des caractéristiques des ménages.

Auteurs	Aire d'étude	Aménités spatiales et sociales							
		Proximité aux services quotidiens	Sécurité Et espace vert	Caractéristiques du logement	Services publics	Stratification en fonction			
						Revenu	Age	Ethnicité	Taille
Bayoh et al. (2006)	Columbus Ohio Area	(+) Qualité des écoles	(-) Les crimes						
Bonvalet (2010)	Île de France	L'analyse des entretiens semi-directifs							
Chattopadhyay (2000)	Chicago	(+) Dépense pour la scolarité		(√) Nombre de pièces, présence de garage...	(+) Les taxes dépensées	(√) Le revenu médian		(√) Pourcentage de la population de couleur blanche	(√) Présence d'un enfant
						Plusieurs croisements ont été testés			
Feridhanusetyawan et Kilkenny (1996)	Etats Unis		Les crimes + en métropole Et – ailleurs		*Taxe ((-)en métropole et (+) ailleurs) *Taux du chômage (-) *Taux d'emploi (-)				
Filion et al. (1999)	Kitchener	√	√						
	Les ménages donnent des scores élevés à ces deux variables lors d'un choix d'une zone de résidence								
Gabriel et Rosenthal (1989)	Washington							(√)	
Gayda(1998)	Lyon	*Les zones avec moins de facilités (-)	(-) Les zones rurales						

		*Les zones commerciales (-)						
	Bruxelles	(-) Les zones commerciales	*Les zones avec un trafic intense (-) *Les zones avec un trafic faible (+) *Les zones rurales (+)					
Goffette-Nagot et schaeffer (2013)	25 aires urbaines de plus de 300 000 habitants au recensement de 1999						(√)	
Gueymard (2006)	Val-de Marne		(√) Espace vert					
Kim et al. (2005)	Oxfordshire	(+) Qualité des écoles						
Quigley (1985)	La région métropolitaine de Pittsburgh			(√) Certaines caractéristiques ont un effet (+) comme la présence de la salle du bain et			(√)	

				d'autres ont un effet (-) comme l'âge de la maison				
Wang et Li (2004)	Beijing	(-)	(-)					
Waddell (1993)	La région de Dallas						(√)	

Tableau 2.2 : Synthèse des travaux empiriques mettant en avant le rôle des aménités spatiales et/ou sociales

2.3. Caractéristiques des ménages : cycle de vie et accès à la propriété

A l'opposé de l'approche de l'économie urbaine et de celle des interactions spatiales, seuls les modèles issus de l'utilité aléatoire permettent d'intégrer directement les caractéristiques socio-économiques des ménages dans la fonction d'utilité (tableau 2.3). Certains travaux intègrent leurs effets croisés avec les caractéristiques des zones. D'autres travaux utilisent des analyses typologiques pour saisir les éventuelles stratifications des choix de localisation en fonction des caractéristiques socio-économiques.

Très tôt, l'analyse descriptive conduite par Quigley (1977) a montré que la majorité des mobilités des ménages des Etats-Unis est intra-urbaine. La mobilité de la population noire est plus importante que celle de la population blanche. Le taux de locataires qui déménagent est quatre fois plus élevé que celui des propriétaires. La majorité des mobilités est due aux changements du cycle de vie (divorce, mariage...). Les travaux de Feridhanusetyawan et Kilkenny (1996) et de Wang et Li (2004) ont confirmé ensuite le rôle important du cycle de vie et des caractéristiques des ménages dans leurs décisions en matière de choix d'une localisation.

Waddell (1993) montre le rôle important des caractéristiques sociodémographiques dans la détermination du lieu de résidence. Les ménages à haut revenu et de grande taille préfèrent des maisons d'une grande taille. Il a montré également que la probabilité d'être propriétaire augmente avec le revenu, l'âge et la taille du ménage.

En testant empiriquement l'effet croisé des caractéristiques de la zone avec les caractéristiques des ménages pour déterminer les facteurs explicatifs des choix de localisation des ménages de la région métropolitaine de Dallas, Guo et Bhat (2001) montrent par exemple que les ménages ayant fait des études supérieures ont tendance à se localiser à proximité des écoles de bonne qualité.

En analysant les arbitrages des ménages en fonction de leur mode de vie, moyennant un modèle de choix à structure latente, Walker et Li (2007) ont distingué trois classes de ménages. La première, composée des ménages riches, est orientée vers les zones périphériques et donne plus de poids aux résidences avec une grande taille et au temps d'accès à l'emploi par voiture. La deuxième, regroupant des ménages jeunes et moins aisés, préfère

plutôt les zones urbaines et donne plus de poids au temps d'accès en transport en commun. Enfin, la troisième classe, composée d'adultes d'âge moyen et de retraités, est orientée plutôt vers les zones urbaines et privilégie plutôt le temps d'accès en voiture. En analysant l'effet du mode de vie et les caractéristiques démographiques sur les préférences des ménages résidant dans les logements collectifs de la grande ville de l'Etat de Caroline du Nord, à Charlotte, Lee et al. (2008) ont montré une corrélation entre la composition du ménage, la taille du ménage et les préférences de localisation. Les hommes et les familles d'une seule personne préfèrent le centre, alors que les familles nombreuses et les femmes déclarent plutôt une préférence pour les zones suburbaines. Par ailleurs, les ménages qui préfèrent les zones rurales donnent des scores plus importants à la taille du logement et aux facteurs environnementaux.

En modélisant les données de panel de Londres et du sud-est de l'Angleterre selon une structure hiérarchique où le ménage choisit de déménager avant de choisir de se localiser dans l'une des quatre zones (centre de Londres, reste du sud-est et reste de l'Angleterre), Andrew et Meen (2006) tentent d'expliquer la dynamique de localisation résidentielle de la population. Ils montrent que la probabilité de déménagement est plus élevée chez les populations appartenant au quartile supérieur de revenu et plus faible chez les personnes les plus âgées et les familles avec des enfants scolarisés. Au grand Toronto, Habib et Miller (2007) ont montré que ce sont l'âge et l'arrivée d'un enfant au sein de la famille qui incitent les ménages à changer de logement et guident ensuite leur choix de la nouvelle zone de localisation.

Clark et al. (2006) ont mis en avant une autre caractéristique du ménage dans la détermination du choix de localisation, à savoir le statut d'occupation. Ils ont trouvé que les préférences pour devenir propriétaire motivent les ménages néerlandais à changer de logement plutôt que la qualité du voisinage.

Si l'accessibilité est un déterminant important dans les choix résidentiels, l'ampleur de son effet varie en fonction des caractéristiques des personnes concernées. En effet, la valeur accordée au temps de transport chez une personne âgée ou retraitée n'est pas la même que chez un jeune actif. Ces valeurs varient également en fonction du sexe, du revenu, ou du statut social.

En faisant la distinction entre les choix résidentiels des femmes et des hommes de la baie de San Francisco, Sermons et Koppelman (2001) ont confirmé cette variation de l'effet de l'accessibilité en fonction des caractéristiques des ménages et plus particulièrement le sexe. Ils ont expliqué cette différence de sensibilité par la variation des responsabilités de l'homme et de la femme dans le ménage.

Aux Pays-Bas, en estimant un modèle hiérarchique à deux niveaux, où les ménages choisissent entre les régions du pays avant de passer au choix de la zone de résidence, Zondag et Pieters (2005) montrent que l'accessibilité détermine la décision de changer de logement. En effet, un ménage change de logement pour bénéficier de plus d'accessibilité, ce qui confirme les résultats précédents (2.1). Par contre, l'effet de l'accessibilité est non significatif chez les ménages actifs avec plus de trois personnes de moins de 65 ans à bas revenu, où il a un signe contre intuitif chez le même type de ménages à moyen ou à haut revenu. Pour l'expliquer, ils mettent en avant la spécificité de la structure spatiale des Pays-Bas. Avec la faible hétérogénéité entre les niveaux d'accessibilité, associée au niveau de maillage, ce facteur devient peu déterminant par rapport aux autres facteurs de structure urbaine.

L'ensemble des travaux empiriques résumés dans le tableau 2.3 propose une synthèse du rôle des différentes caractéristiques des ménages dans les décisions de choix de localisation.

Auteurs	Aires d'étude	Caractéristiques socio-économiques						Niveau d'étude
		Âge	Famille avec enfant	Revenu	Taille du ménage	Statut d'occupation	Sexe	
Andrew et Meen (2006)	Londres-Sud de l'Angleterre	(√)	(√)	(√)				
Clark et al. (2006)	Pays-Bas					(√)		
Guo et Bhat (2001)	Région métropolitaine de Dallas							(√)
Habib et Miller (2007)	Grand Toronto	(√)	(√)					
Lee et al. (2008)	Charlotte				(√)		(√)	
Sermons et Koppelman (2001)	Baie de San Francisco							(√)
Waddell (1993)	Région métropolitaine de Dallas	(√)		(√)	(√)			
Walker et Li (2007)*	Potland-Oregon	(√)		(√)				
Zondag et Pieters (2005)*	Pays Bas	(√)		(√)	(√)			

Tableau 2.3 : Synthèse des travaux empiriques mettant en avant le rôle des caractéristiques des ménages

* Ces variables ont été introduites pour construire des typologies et non pas comme des variables explicatives

Comparaison entre différents contextes : quels enseignements par rapport aux déterminants du choix de localisation ?

A travers l'analyse des travaux empiriques synthétisés dans les tableaux 2.1 à 2.3 traitant les déterminants des choix de localisation, nous avons mis en avant l'importance d'intégrer des variables rendant compte des arbitrages du marché, des aménités spatiales, de l'environnement social et des caractéristiques des ménages dans la modélisation des choix de localisation. Nous avons également retenu que la diversité des méthodes utilisées pour estimer les choix de localisation résidentielle, la diversité des contextes et des échelles utilisées compliquent la comparaison des préférences des ménages entre ces différents contextes. Cela étant, cette comparaison a permis de mettre en avant quelques enseignements par rapport aux effets des différents déterminants de choix de localisation résidentielle.

Les travaux empiriques sur des villes des pays en voie de développement montrent que le coût de transport, et plus particulièrement le temps d'accès à l'emploi, reste le facteur principal expliquant les choix de localisation résidentielle. Cependant, les ménages des villes des pays développés, que ce soit en Europe ou aux États-Unis, donnent plus d'importance aux aménités spatiales et sociales.

Par ailleurs, la comparaison des travaux portant sur des villes françaises et américaines apportent quelques nuances par rapport aux préférences des ménages pour les aménités sociales. Les travaux appliqués sur des villes françaises, mettent en avant une stratification des ménages en fonction du revenu et du statut social, alors que les travaux portant sur des villes américaines montrent plutôt une stratification en fonction de l'appartenance ethnique. Toutes choses étant égales par ailleurs, les ménages des villes françaises préfèrent se localiser dans des zones peuplées par des ménages appartenant à la même classe de revenu ou ayant le même statut socioéconomique. Si la recherche d'entre-soi et de l'évitement peut être testée chez des populations favorisées, moins contraintes par le marché et les logiques institutionnelles, elle est difficilement transposable aux ménages défavorisés dans le contexte

français. En plus, les travaux américains en faveur de cette hypothèse de préférence pour l'entre-soi reposent principalement sur le critère ethnico-racial. Ce dernier ne peut se substituer au critère économique car les deux se superposent (Gabriel et Rosenthal, 1989).

Enfin, les déterminants des choix de localisation diffèrent en fonction du contexte géographique et de la forme urbaine. Les résultats obtenus sur des villes polycentriques ne sont pas les mêmes que ceux obtenus dans un contexte de villes monocentriques. Toutefois, il est prématuré de conclure sur les différences liées à la forme urbaine puisque les méthodes utilisées, que ce soit pour la spécification du modèle ou la définition des échelles spatiales, sont relativement différentes. Les résultats des travaux réalisés sur des villes aux Pays-Bas (Tillema, Ettema, 2010 ; Zondag et Pieters, 2005 ; Clark et al., 2006) témoignent de cette difficulté et mettent l'accent sur l'importance des choix méthodologiques dans la modélisation des choix de localisation résidentielle en milieu urbain.

Conclusion

L'analyse des différents cadres théoriques traitant de la localisation résidentielle (chapitre 1) a conduit à retenir l'approche des choix discrets avec ses fondements théoriques basés sur l'utilité aléatoire pour analyser empiriquement les choix de localisation résidentielle en milieu urbain. Au-delà des justifications théoriques apportées par les modèles de l'économie urbaine et/ou de l'interaction spatiale sur le rôle des facteurs de marché foncier et immobilier, de l'accessibilité et des aménités urbaines, les modèles dérivés de l'utilité aléatoire disposent aussi des outils opérationnels nécessaires pour mesurer directement les effets de chaque déterminant de choix de localisation. Cela permettra ainsi de simuler ces choix de localisation en interaction avec les autres sous-systèmes dans le cadre des modèles LUTI, comme c'est le cas de SIMBAD dans lequel s'inscrit ce travail de thèse.

Pour ce qui est de l'analyse des travaux empiriques sur les déterminants de choix de localisation résidentielle (tableaux 2.1 à 2.3, chapitre 2), notre premier constat rejoint les résultats de l'analyse de la revue de littérature sur les choix résidentiels faite récemment par Schirmer et al. (2014) sur plusieurs points. Le premier point concerne le rôle des contraintes du marché, des aménités spatiales, de l'environnement social et des caractéristiques des ménages et l'importance d'intégrer des variables reflétant ces dimensions dans les modèles de localisation. Toutes choses égales par ailleurs, les ménages préfèrent se localiser dans des zones peuplées par des ménages du même profil (revenu, ethnie, couleur...). L'accessibilité et la proximité aux aménités sont aussi privilégiées. Ces préférences pour l'accessibilité et pour les aménités spatiales et sociales diffèrent en fonction des caractéristiques des ménages et elles sont contraintes par le prix de l'immobilier.

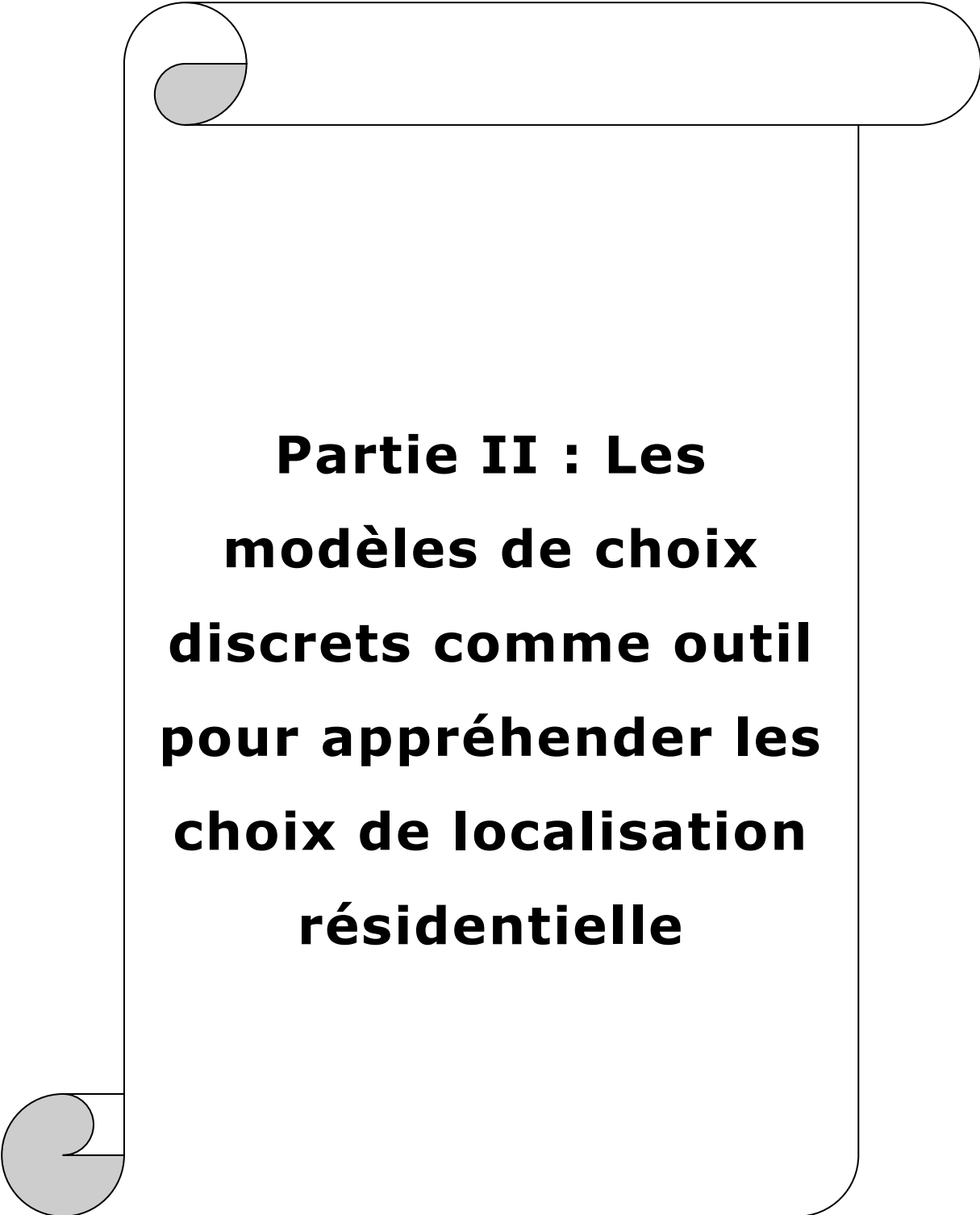
Le deuxième point concerne la variation des effets des variables explicatives des choix de localisation résidentielle en fonction de la forme urbaine et de la méthode utilisée. La majorité des travaux empiriques analysés donnent plus d'importance à l'analyse des choix résidentiels dans les villes multicentriques (les grandes aires métropolitaines nord-américaines de Chicago, de Los Angeles, Dallas ou Montréal, Seoul et Bogota). Pour cela, nous avons choisi dans ce travail de thèse de travailler sur une aire urbaine française parmi celles qui enregistrent de fortes dynamiques démographiques durant les quarante dernières années, à savoir l'aire urbaine de Lyon. Toutefois, la différence des effets des déterminants des choix de localisation n'est pas liée seulement à la forme urbaine ou au contexte de l'étude, mais aussi à

la spécification du modèle et aux échelles utilisées dans les estimations de choix de localisation.

Les questions qui se posent à ce niveau sont : quels sont les modèles de choix discrets qui permettent de mieux représenter spatialement les choix de localisation des ménages ? Quelle structure permet de traiter les problèmes d'autocorrélations spatiales liés à la dimension géographique des choix de localisation ? Comment dépasser la contrainte de la multitude d'alternatives de choix de localisation intra-urbaine ?

Dans la partie suivante, nous questionnons les structures et les limites des différentes familles de modèles de choix discrets, afin de répondre à ces questions pour choisir l'outil approprié pour appréhender les choix de localisation résidentielle.

La dernière partie est consacrée au choix de l'échelle spatiale adaptée pour analyser les choix de localisation des ménages de l'aire urbaine de Lyon et à la confrontation des approches théoriques et des choix méthodologiques aux choix observés des Lyonnais.



**Partie II : Les
modèles de choix
discrets comme outil
pour appréhender les
choix de localisation
résidentielle**

Cette deuxième partie est consacrée à l'analyse approfondie des différents modèles de choix discret permettant de prendre en compte la dimension géographique des choix de localisation dans la spécification des termes d'erreurs. Deux éléments sont généralement abordés lors de l'analyse de cette dimension géographique des choix de localisation résidentielle : i) l'autocorrélation spatiale et ii) l'hétérogénéité spatiale. Dans cette partie, nous nous focalisons exclusivement sur le premier point et plus particulièrement l'hypothèse d'indépendance entre les zones de choix de localisation. Cette dernière, lorsqu'elle n'est pas respectée, conduit à une incohérence avec les hypothèses restrictives des modèles de choix discrets les plus couramment utilisés.

L'application des modèles de choix discrets dans le domaine de la localisation des ménages est confrontée à une difficulté liée à la multitude d'alternatives de choix de localisation intra-urbaine. Pour dépasser ce problème et réduire le nombre d'alternatives considérées dans l'estimation, deux solutions se présentent. La première consiste à définir un sous ensemble d'alternatives à partir de celles disponibles pour chaque ménage et la deuxième propose des méthodes d'échantillonnage pour réduire le nombre d'alternatives de l'ensemble des choix.

Nous commençons par la présentation des différents types de modèles de choix discrets appropriés pour tester empiriquement les déterminants des choix de localisation résidentielle identifiés dans la première partie. Ces modèles sont analysés sous l'angle de leur aptitude à tenir compte des autocorrélations spatiales (chapitre 3). Le reste de cette partie aborde la présentation des développements des méthodes récentes pour réduire le grand nombre d'alternatives de choix de localisation (chapitre 4).

Chapitre 3 - Les familles de modèles de choix discrets : structures et limites

Le recours à l'approche de l'utilité aléatoire pour modéliser les choix de localisation des ménages nécessite un intérêt particulier au choix de la distribution des termes d'erreurs pour appréhender l'aspect spatial. Deux distributions sont traditionnellement utilisées. La première est la distribution de Gumbel et conduit à la famille des modèles de la valeur extrême généralisée (Generalised Extreme Value models-GEV). La deuxième est la distribution suivant une loi normale et conduit au modèle probit.

La considération d'une distribution normale donne plus de réalisme et de flexibilité à l'interdépendance entre les alternatives de choix de localisation. Toutefois, le calcul compliqué de ces modèles au-delà de quelques alternatives leur fait défaut. Bolduc et Ben-Akiva (1991) ont introduit des modèles hybrides ou mixtes. Il s'agit d'une combinaison entre les modèles logit et probit. Cette famille de modèles a l'inconvénient d'une forme mathématique non analytique. Pour dépasser cela, Bhat et Guo (2004) ont suivi une autre logique. Il s'agit de combiner la structure des modèles mixtes et des modèles GEV. Cependant, ces modèles ne peuvent pas être utilisés avec un nombre élevé d'alternatives.

Nous commençons par la présentation des aspects communs à toutes les familles de modèles de choix discrets (section 3.1). Nous passons ensuite à la présentation de la structure mathématique des modèles de la famille GEV qui peuvent être utilisés pour modéliser les choix de localisation (section 3.2). Les sections 3.3 et 3.4 sont consacrées à la présentation sommaire des autres familles de modèles, à savoir le modèle probit et la famille des modèles mixtes. Les modèles proposés récemment par Bhat et Guo (2004) sont discutés dans la cinquième section (3.5). Ces derniers modèles sont plus réalistes et plus flexibles pour modéliser les choix de localisation. Cependant, les difficultés méthodologiques liées à leur application sur un jeu de données avec plusieurs alternatives font d'eux des modèles moins opérationnels. La dernière section (3.6) est consacrée à une synthèse des travaux empiriques qui ont utilisé ces modèles pour appréhender l'aspect spatial des choix de localisation. Une

telle synthèse va nous permettre de sélectionner la ou les structures retenues pour analyser les choix de localisation des ménages de l'aire urbaine de Lyon.

3.1. Spécificités générales des modèles de choix discrets

Avant de passer à la présentation des principales familles de modèles de choix discrets, il est essentiel de commencer par la présentation de tous leurs aspects communs. Train (2009) résume ces aspects en deux points : i) Seule la différence des utilités est considérée lors de l'estimation des paramètres, ii) L'échelle de l'utilité est arbitraire.

Le premier aspect implique que la spécification du modèle de choix discrets passe par l'identification de la distribution de la différence des termes d'erreurs des utilités. De ce fait, seuls les paramètres mesurant la différence entre les alternatives peuvent être calculés.

Empiriquement, cet aspect se traduit par l'introduction d'une constante spécifique pour chaque alternative et la fixation du paramètre de l'une des alternatives (alternative de référence) à 0. L'interprétation se fait ensuite par rapport à l'alternative de référence. Pour mesurer l'effet moyen des variables non incluses dans la fonction d'utilité, les modélisateurs introduisent également un terme constant dans la fonction d'utilité.

Le deuxième aspect implique que la multiplication de l'utilité par un paramètre (μ) ne change pas la structure du modèle. La normalisation de la fonction d'utilité revient à la normalisation de la variance de son terme d'erreur ($\pi^2/6\mu$). Le paramètre μ est inversement proportionnel à la variance de la part stochastique de l'utilité. Conventionnellement, pour identifier les paramètres estimés d'un modèle logit, ce paramètre est normalisé à **1**.

La prise en compte de cette normalisation est nécessaire pour comparer des modèles de choix discrets avec une distribution des termes d'erreurs différente (probit et logit par exemple).

Pour illustrer l'importance de cette normalisation, nous avons repris l'exemple de Train (2009). Si le paramètre correspondant à l'accessibilité est égal à 0,1 pour un logit, et 0,2 pour un probit, cela ne veut pas dire que le logit implique moins de sensibilité que le modèle probit. Avant de faire la comparaison, il faut diviser la valeur du modèle logit par la racine carrée de la variance de la distribution du modèle logit ($\sqrt{\pi^2/6} = \sqrt{1,6}$). Cela est pris en compte

aussi pour comparer deux villes différentes qui n'ont pas la même variance des termes d'erreurs.

3.2. La famille des modèles GEV

Cette famille de modèles est connue aussi sous le nom de MEV (Multivariate Extreme Value). Initialement développée par McFadden (1978), la forme générale de ces modèles s'exprime de la manière suivante :

$$P(i / C_n) = \frac{e^{V_i} G_i(e^{V_1}, \dots, e^{V_J})}{\mu G(e^{V_1}, \dots, e^{V_J})} \quad -3.1-$$

En considérant un ensemble de choix C_n composé de J alternatives et tel que V_i est l'utilité de choisir une alternative i parmi l'ensemble des choix. La probabilité $P(i/C_n)$ de choisir une alternative i de l'ensemble des choix C_n est calculée à partir d'une fonction génératrice G et sa dérivée première G_i . Cette fonction doit être non négative et homogène. Selon le théorème d'Euler, cette probabilité peut être simplifiée et écrite sous cette forme :

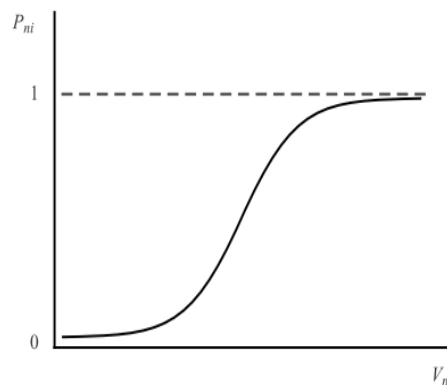
$$P(i / C_n) = \frac{e^{V_i + \ln G_i (\dots)}}{\sum_{j=1}^J e^{V_j + \ln G_j (\dots)}} \quad -3.2-$$

Le reste de cette section est organisé de la manière suivante. Nous commençons par la présentation de la structure mathématique du modèle le plus simple de la famille GEV, à savoir le modèle logit multinomial. La distribution des termes d'erreurs de ce modèle est basée sur des hypothèses restrictives pour simplifier leur structure mathématique et faciliter les calculs. Une telle restriction limite la prise en compte des autocorrélations spatiales entre les zones de localisation. Ces hypothèses ont été relâchées successivement pour développer des modèles plus réalistes. La deuxième sous section présente la structure du modèle logit emboîté qui permet le relâchement partiel de certaines hypothèses. Et la dernière présente d'autres types de modèles GEV susceptibles d'être utilisés pour appréhender l'autocorrélation spatiale.

3.2.1. Modèle logit Multinomial

Ce modèle reste le plus utilisé dans la modélisation des choix de localisation des ménages. Il se caractérise par une structure mathématique simple et facile à estimer et il a une probabilité sous forme de Sigmoide ou forme de S (graphique 3.1).

Graphique 3.1 : Représentation graphique de l'utilité en fonction de la probabilité du choix.



Source : Train (2009)

La probabilité de choisir une zone de localisation i parmi l'ensemble des zones de localisation C_n s'écrit de la manière suivante :

$$P(i / C_n) = \frac{e^{\mu V_i}}{\sum_{j=1}^J e^{\mu V_j}} \quad -3.3-$$

Tel que V_i est la composante déterministe de l'utilité de choisir une zone de localisation i et μ est un paramètre positif. Cette formule a initialement été dérivée par Luce (1959) à partir de l'axiome des choix. Ce modèle a été analysé par le travail de Marschak (1960), qui a montré sa compatibilité avec la maximisation de l'utilité aléatoire.

McFadden (1978) a montré ensuite que ce modèle peut être dérivé également à partir de la forme générale des modèles de la valeur extrême généralisée (GEV), en remplaçant la fonction génératrice (équation 3.2) par la forme suivante :

$$G(e^{V_1}, \dots, e^{V_J}) = \sum_{i=1}^J \left(e^{V_i} \right)^\mu \quad -3.4-$$

Les hypothèses de la forme mathématique du modèle logit multinomial sont les suivantes (Bhat, 2006) :

- Les termes d'erreurs sont distribués selon la loi de Gumbel : cette distribution est choisie car elle est proche de la distribution de la loi normale et se caractérise par une forme probabiliste fermée, ce qui facilite les calculs.
- Les termes d'erreurs sont indépendants et identiquement distribués (IID) entre les alternatives.
- Les termes d'erreurs sont indépendants et identiquement distribués entre les observations (les ménages dans notre cas).

L'hétérogénéité des observations ne contredit généralement pas l'hypothèse IID des termes d'erreurs entre les ménages. Ainsi on s'intéresse, dans ce travail de thèse, plus particulièrement à la deuxième hypothèse et à son incohérence avec la modélisation des choix de localisation.

En supposant que les termes d'erreurs sont IID entre les alternatives, le modélisateur accepte l'absence de facteurs inobservés communs entre les alternatives. Une telle hypothèse permet l'ajout et la suppression d'une alternative sans modifier les rapports de probabilité (Train, 2009). Cette propriété est connue sous le nom d'indépendance par rapport aux choix non retenus (Independence from Irrelevant Alternatives-IIA). Elle donne au modèle une certaine flexibilité et permet d'estimer un modèle sur un échantillon aléatoire d'alternatives sans biaiser les paramètres estimés. Cependant, les zones de localisation sont spatialement autocorrélées et il est difficile de respecter l'hypothèse dans les estimations empiriques des choix de localisation.

Train (2009) souligne trois points qui constituent à la fois la force et la limite du modèle logit multinomial :

- i) Le modèle logit peut représenter la variation des préférences systématiques (celles liées aux caractéristiques observées de l'individu). Toutefois, il ne permet pas d'intégrer la variation aléatoire des préférences (celles liées aux caractéristiques non observées de l'individu).
- ii) Le modèle logit implique des substitutions proportionnelles entre les alternatives. Cela est issu de la propriété d'indépendance par rapport aux choix non retenus. Mais, il n'est pas assez flexible pour saisir toutes les formes de substitutions.

- iii) Le modèle logit suppose que les facteurs sont indépendants dans le temps et par conséquent permet seulement de capturer la dynamique des choix répétés. Cependant, il ne permet pas de traiter les situations où les facteurs inobservés sont corrélés dans le temps.

3.2.2. Modèle emboîté

Dans le cas des choix de localisation, les alternatives de choix sont spatialement autocorrélées et le respect de la propriété principale du modèle logit multinomial (IIA) devient difficile. En relâchant partiellement cette propriété, le modèle emboîté présenté dans cette section peut être utilisé pour tenir compte de ces autocorrélations spatiales entre les zones de localisation.

Proposé initialement par Ben-Akiva (1973 ; 1974), ce modèle est une extension du modèle logit multinomial. Il est développé pour saisir les éventuelles corrélations qui peuvent exister entre les termes d'erreurs de certaines alternatives. Il consiste à diviser l'ensemble des choix en des nids (équation 3.5), tels que la propriété IIA tient pour les alternatives appartenant au même nid et ne tient pas pour les alternatives appartenant à des nids différents.

$$C_n = \bigcup_{m=1}^M C_{nm} \text{ et } C_{nm} \cap C_{nm'} = \emptyset \forall m \neq m' \quad -3.5-$$

En supposant une telle décomposition, la probabilité de choisir une zone **i** appartenant à une macro_zone **m** s'écrit alors comme étant le produit de deux logit multinomiaux simples : la probabilité marginale **P(i/C_{nm})** de choisir la zone **i** sachant que la macro_zone **m** est déjà choisie, et la probabilité **P(C_{nm})** de choisir la macro_zone **m** :

$$P(i) = P(i/C_{nm}) * P(C_{nm}) \quad -3.6-$$

Comme dans le cas du modèle logit multinomial, le modèle logit emboîté peut aussi être dérivé à partir des modèles GEV en prenant la fonction **G** suivante :

$$G(e^{V_1}, \dots, e^{V_J}) = \sum_{m=1}^M \left(\sum_{i \in C_{nm}} \left(e^{V_i} \right)^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \quad -3.7-$$

En pratique, l'estimation de ces modèles se fait d'une manière séquentielle ou simultanée. L'estimation simultanée estime les paramètres directement à partir de l'équation (3.2), en remplaçant **G** par la formule de l'équation (3.7). Par contre, l'estimation séquentielle suppose un choix en deux étapes. Le ménage **n** choisit d'abord entre plusieurs grandes zones

$\{C_{n1}, C_{n2}, \dots, C_{nM}\}$ avant de sélectionner sa zone de localisation i au sein de la grande zone sélectionnée C_{nm} . Et le processus de l'estimation se fait de la manière suivante :

- Calculer $P(i/C_{nm})$: $P(i/C_{nm}) = \frac{\exp(V_i \mu_m)}{\sum_{j \in C_{nm}} \exp(V_j \mu_m)}$ -3.8-

- Calculer les logsum I_m : $I_m = \ln \sum_{j \in C_{nm}} \exp(V_j \mu_m)$ -3.9-

- Calculer $P(C_{nm})$: $P(C_{nm}) = \frac{\exp(W_m + \frac{1}{\mu_m} I_m)}{\sum_{l=1}^M \exp(W_l + \frac{1}{\mu_l} I_l)}$ -3.10-

Un modèle où le paramètre μ est normalisé à 1 est un modèle normalisé du haut. Par contre, un modèle où le paramètre μ_m est normalisé à 1 est un modèle normalisé du bas.

3.2.3. D'autres familles de modèles GEV

La difficulté majeure du modèle emboîté concerne le regroupement des alternatives dans des clusters et la constitution des nids. D'autres fonctions génératrices ont été développées pour prendre en compte les interdépendances entre les alternatives du choix. Hunt et al. (2004) ont dérivé le modèle Logit emboîté généralisé (Generalized nested logit, GNL) à partir de la fonction génératrice suivante :

$$G(e^{V_1}, \dots, e^{V_J}) = \sum_{m=1}^M \left(\sum_{j \in C_{nm}} \left(\alpha_{jm} e^{V_j} \right)^{1/\mu_m} \right)^{\mu_m} \quad -3.11-$$

Tel que $\alpha_{im} \geq 0$ et $\sum_{m=1}^M \alpha_{im} = 1$

Il s'agit d'un modèle plus général de la famille GEV à partir duquel on peut dériver des modèles plus flexibles. La structure du modèle GNL se compose de m nids. Chaque nid est caractérisé par un paramètre de dissimilarité μ_m . Une valeur élevée de ce paramètre implique une faible corrélation entre les variations inobservées de deux alternatives. Les paramètres d'allocation α_{im} consistent à spécifier la proportion d'appartenance d'une alternative à un nid. Dans cette sous section, nous limitons la présentation de la structure des modèles à ceux qui ont été utilisés pour appréhender les autocorrélations spatiales. Il s'agit du modèle logit

combinatoire (Paired combinatorial logit-PCL), le modèle logit corrélé spatialement (Spatially Correlated logit-SCL), le modèle logit généralisé corrélé spatialement (Generalized Spatially Correlated Logit-GSCL) et le modèle emboîté croisé (Cross nested logit-CNL). La structure des deux premiers modèles permet de tester la présence d'autocorrélation spatiale entre toutes les paires d'alternatives. Tandis que les deux derniers modèles permettent à une zone de localisation d'appartenir à plusieurs regroupements.

- **Logit combinatoire par paire (Paired Combinatorial Logit-PCL)**

Contrairement au modèle emboîté qui considère que les alternatives appartenant au même nid ont la même covariance. Le modèle logit combinatoire par paire (PCL) alloue différentes covariances pour chaque paire d'alternatives. Sa fonction génératrice est de la forme suivante :

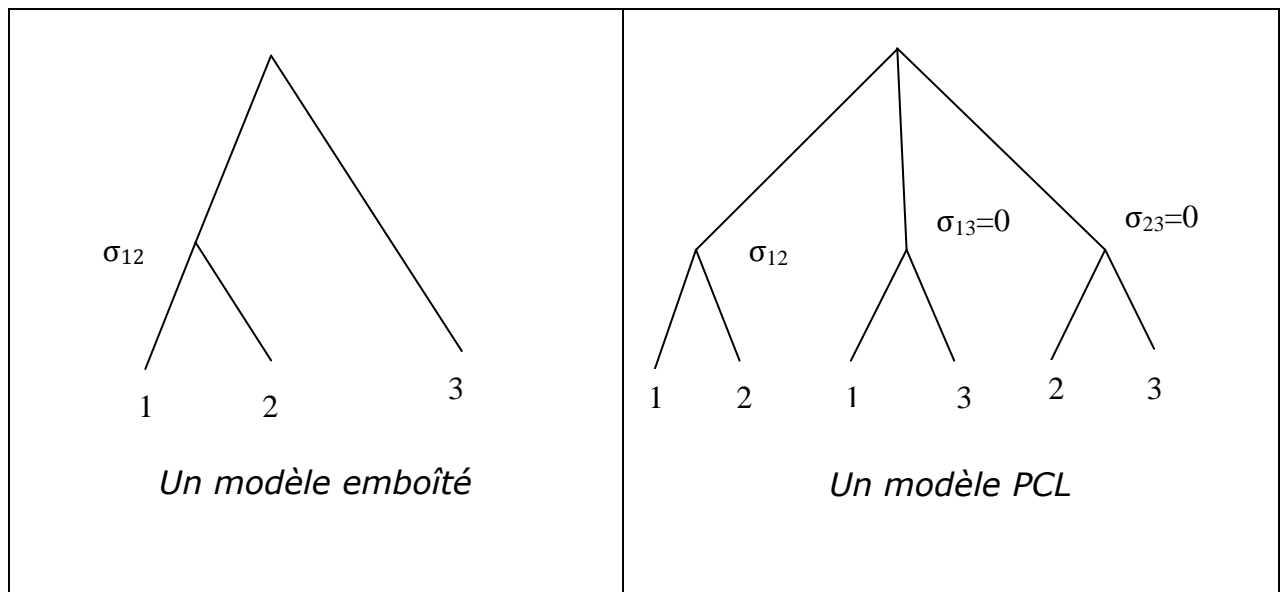
$$\sum_{i=1}^{J-1} \sum_{j=i+1}^J \left(\left(e^{V_i} \right)^{\frac{1}{1-\sigma_{ij}}} + \left(e^{V_j} \right)^{\frac{1}{1-\sigma_{ij}}} \right) \quad -3.12-$$

Tel que $0 \leq \sigma_{ij} < 1$

Ce modèle consiste à affecter proportionnellement chaque alternative à un nid et à estimer le paramètre de dissimilarité pour chaque nid. Chaque alternative est groupée avec une autre pour former un nid, et chaque alternative appartient à J-1 nids.

Le modèle emboîté n'est pas un cas particulier du modèle logit combinatoire par paire. Réciproquement, ce dernier n'est pas un cas particulier du modèle emboîté. La différence primaire entre ces deux modèles est la façon dont ils représentent les similarités entre les paires d'alternatives. Ils peuvent être représentés d'une manière approximativement mais pas exactement équivalente (Figure 3.1).

Figure 3.1 : Exemple illustrant l'équivalence entre un logit emboîté et un PCL



Source : Koppelman et Wen (2000)

Le nombre d'alternatives de choix de zones est grand. Le nombre de paires à considérer est donc encore plus grand, ce qui limite l'application de ce modèle dans la pratique. Les modèles développés concernent principalement le choix d'itinéraire (Koppelman et Wen, 2000 ; Pravinovongvuth et Chen, 2005 ; Gilbert, 2012 ; Li et al., 2013).

- **Logit corrélé spatialement (Spatially Correlated Logit-SCL)**

Le modèle SCL est dérivé à partir de la fonction génératrice suivante :

$$G(e^{V_{n,1}}, \dots, e^{V_{n,I}}) = \sum_{i=1}^{I-1} \sum_{j=i+1}^I \left[\left(\alpha_{i,j} e^{V_{n,i}} \right)^{1/\mu} + \left(\alpha_{j,i} e^{V_{n,j}} \right)^{1/\mu} \right]^\mu \quad -3.13-$$

Si le paramètre $\alpha_{i,j}$ est identique pour tous les nids, ce modèle se réduit au modèle PCL. La structure de ce modèle consiste à créer autant de nids qu'il y a de zones adjacentes.

$\alpha_{i,j}$ est défini comme le rapport des valeurs de la matrice de la contiguïté w_{ij} et de la somme de ces derniers.

- **Logit généralisé corrélé spatialement (Generalized Spatially Correlated Logit-GSCL)**

Ce modèle est une généralisation du modèle SCL. Il consiste à présenter le paramètre $\alpha_{i,j}$ sous une forme fonctionnelle (équation 3.14) qui permet de mesurer les différentes

interdépendances qui existent entre les zones de localisation. Le paramètre $\alpha_{i,ij}$ s'écrit alors de

la manière suivant :
$$\alpha_{i,ij} = \frac{\exp(\phi Z_{ij})}{\sum_k \phi Z_{ik}} \quad -$$

3.14-

Tel que ϕ est un paramètre à estimer et Z_{ij} est le vecteur des variables qui caractérisent les attributs en commun entre les zones de localisation (la distance, la contiguïté par exemple).

- **Logit emboîté croisé (Cross Nested Logit-CNL)**

Initialement présenté par McFadden (1978) comme cas du modèle GEV et utilisé ultérieurement par Small (1987). Le modèle logit emboîté croisé permet de traiter les cas complexes alors que le modèle emboîté ne permet pas de prendre en compte toutes les corrélations entre les termes d'erreurs des alternatives. En permettant à une alternative d'appartenir à plusieurs nids, le modèle logit emboîté croisé donne plus de flexibilité au regroupement des zones de localisation similaires.

Le modèle CNL peut être dérivé en choisissant plusieurs formes de la fonction génératrice. La forme la plus générale est celle développée par Ben-Akiva et Bierlaire (1999) :

$$G(e^{V_1}, \dots, e^{V_J}) = \sum_{m=1}^M \left(\sum_{j \in C_n} \alpha_{jm} \left(e^{V_j} \right)^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \quad -3.15-$$

La formulation initiale utilisée par Small (1987) pour dériver un modèle CNL est un cas particulier de la fonction génératrice de l'équation 3.15, en prenant $\mu=1$ et $\mu_m=1/\rho_m$. La formulation utilisée ensuite par Vovsha (1997) est aussi un cas particulier de celle de l'équation (3.15), en prenant $\mu_m=1$ pour \mathbf{m} .

En prenant $\mu=1$, le modèle CNL peut s'exprimer sous la forme de la fonction génératrice du modèle emboîté généralisé (Wen et Koppelman, 2001).

A partir de la fonction G du modèle logit emboîté généralisé, on peut également dériver d'autres types de modèles. Ces modèles ne sont pas détaillés dans cette sous-section.

Les modèles GEV permettent de relâcher partiellement l'hypothèse d'IID des termes d'erreurs. Toutefois, ces modèles maintiennent l'hypothèse de la distribution identique des termes d'erreurs au sein du sous-ensemble de l'univers de choix. Une telle structure leur permet une prise en compte partielle des interdépendances entre les alternatives, en maintenant la forme fermée des probabilités des choix.

3.3. Modèle probit

Nous avons vu que les modèles GEV sont de plusieurs types. En fonction du type du modèle GEV, l'hypothèse posée sur les termes d'erreurs et plus particulièrement une distribution indépendante limite la prise en compte partielle ou totale des corrélations entre les alternatives. Cependant, tous ces types de modèles supposent que les résidus sont distribués selon une loi de Gumbel.

Les modèles probit sont basés sur l'hypothèse que les termes d'erreurs suivent une distribution normale. Une telle distribution permet d'appréhender toutes les formes de corrélation et d'hétéroscédasticité entre les alternatives. Cela permet aussi de tenir compte des hétérogénéités des préférences des individus.

A l'opposé des modèles GEV, ces modèles se présentent sous forme d'intégrales multidimensionnelles de dimensions égales au nombre d'alternatives de l'ensemble du choix (équation 3.16).

$$\begin{aligned}
 P_{n,i} &= \text{Prob}(V_{n,i} + \varepsilon_{n,i} \succ V_{n,j} + \varepsilon_{n,j} \forall j \neq i) \\
 &= \int I(V_{n,i} + \varepsilon_{n,i} \succ V_{n,j} + \varepsilon_{n,j} \forall j \neq i) \Phi(\varepsilon_n) d\varepsilon_n,
 \end{aligned}
 \tag{3.16}$$

Tel que $I(\cdot)$ est un indicateur binaire qui vaut 1 si la formule entre parenthèses est vraie et 0 sinon. En posant $B_{n,i} = \{\varepsilon_n \text{ tel que } V_{n,i} + \varepsilon_{n,i} \succ V_{n,j} + \varepsilon_{n,j} \forall j \neq i\}$ l'équation (3.16) s'écrit comme suit : $P_{n,i} = \int_{\varepsilon_n \in B_{n,i}} \Phi(\varepsilon_n) d\varepsilon_n$

Il s'agit d'une forme mathématique non fermée et la simulation numérique reste le seul moyen pour calculer ces probabilités, ce qui complique les calculs. En n'imposant aucune hypothèse supplémentaire sur la distribution des termes d'erreurs, la deuxième difficulté est liée à l'estimation des paramètres de la matrice de variance-covariance. Malgré les tentatives faites en analyse numérique pour faciliter le calcul de cette matrice, ces développements n'ont pas apporté une réponse définitive sur la simulation qui peut être utilisée pour estimer un modèle probit dans le cas d'un grand nombre d'alternatives (Bolduc, 1993).

3.4. Familles des modèles mixtes

En supposant que les termes d'erreurs sont composés de deux parties, Bolduc et Ben-Akiva (1991) ont développé des modèles mixtes. La première partie dont la distribution suit une loi normale permet de prendre en compte toutes les corrélations et les hétéroscédasticités. La deuxième partie est distribuée selon la loi de Gumbel et repose sur l'hypothèse selon laquelle les termes d'erreurs sont identiquement et indépendamment distribués. Les modèles mixtes connus aussi sous le nom de modèles à composantes d'erreur (Brownstone et Train, 1999) ou Kernel logit (Chen et al., 2005) ont des structures plus flexibles. Ils permettent non seulement aux alternatives de tenir compte des corrélations entre elles, mais aussi de prendre en compte l'hétérogénéité des préférences des ménages.

La probabilité qu'un ménage \mathbf{n} choisisse une zone de localisation \mathbf{i} s'écrit de la manière

$$\text{suivante : } P_{n,i}(\theta) = \int_{-\infty}^{+\infty} L_{n,i}(\beta) f(\beta / \theta) d(\beta) \text{ avec } L_{n,i}(\beta) = \frac{e^{\beta V_{in}}}{\sum_j e^{\beta V_{jn}}} \quad -3.17-$$

V_{in} est la partie déterministe de la fonction d'utilité, β est une réalisation aléatoire de la fonction de distribution $f(\cdot)$ et θ est un vecteur des paramètres de la fonction $f(\cdot)$ du ménage \mathbf{n} de choisir une zone \mathbf{i} .

La difficulté des modèles mixtes est qu'ils n'estiment pas les paramètres des variables explicatives, mais plutôt les paramètres de la distribution d'une fonction $f(\cdot)$. Il s'agit encore une fois de simuler plutôt que de calculer la fonction de la log-vraisemblance.

3.5. D'autres structures

L'hétérogénéité des préférences des ménages peut être intégrée dans les modèles GEV en introduisant dans la fonction d'utilité des variables dichotomiques qui mesurent l'effet croisé d'une variable caractérisant la zone et les caractéristiques des ménages. Les modèles logit mixtes peuvent intégrer également les hétérogénéités non observées. En associant la structure des modèles mixtes (section 1.4) avec celle du modèle SCL, Bhat et Guo (2004) ont développé un modèle logit mixte spatialement corrélé (Mixed Spatially Correlated Logit-MSCL). Ce modèle consiste à utiliser une structure d'un modèle GEV pour traiter les autocorrélations entre les unités spatiales et superpose une distribution mixte pour tenir

compte de l'hétérogénéité des préférences des ménages. Il s'agit d'un cas du modèle GNL et plus spécifiquement une structure sous forme de paires, qui suppose que le paramètre de dissimilarité est égal entre les différents nids composés de paires. Chaque paire (nids) est composée d'une zone **i** et de son adjacente **j**.

Il est question de considérer une matrice de distance de l'ordre de 1 ($w_{ij}=1$ si **i** et **j** sont contiguës et 0 sinon). En appliquant cela à l'équation (3.13) le paramètre mesurant la part de l'alternative **i** dans le nid **m** (allocation of alternative) est exprimé comme suit :

$$\alpha_{ij} = \frac{w_{ij}}{\sum_{i=1}^I w_{ij}} \quad -3.18-$$

Ce paramètre est pondéré par le nombre de zones contiguës à l'alternative **i**. Plus la zone est contiguë à plusieurs zones plus le paramètre est faible.

Synthèse : Modèles développés pour modéliser les choix de localisation

A partir de la grille de lecture faite ci-dessus pour sélectionner les structures susceptibles d'être utilisées pour appréhender la présence des autocorrélations spatiales entre les alternatives de choix de localisation, on peut retenir deux enseignements. Premièrement, on constate qu'à l'exception du modèle Multinomial, les autres types de modèles GEV permettent aux termes d'erreurs d'être partiellement ou totalement corrélés, en maintenant l'hypothèse d'une distribution identique (des termes d'erreurs corrélés et identiquement distribués) au sein de sous-ensemble de l'univers de choix. En revanche, les modèles mixtes permettent aux termes d'erreurs des alternatives d'être corrélés et non identiquement distribués et aux paramètres liés aux variables explicatives de varier aléatoirement, en utilisant une distribution continue entre les individus (approche des coefficients aléatoires). Deuxièmement, il existe une relation inverse entre la complexité des calculs et la restriction des hypothèses posées sur la distribution des termes d'erreurs entre les alternatives de choix (figure 3.2).

En introduisant un terme aléatoire dans la fonction de l'utilité, les modèles mixtes permettent de prendre en compte l'hétérogénéité des préférences des ménages. Cependant, ces modèles ont des formes analytiques non fermées, ce qui demande des simulations numériques.

En mesurant l'effet croisé d'une caractéristique de l'alternative avec les caractéristiques des ménages, les modèles GEV permettent aussi de tenir compte de cette hétérogénéité.

Malgré leurs limites, les modèles de la famille GEV restent les plus utilisés dans les estimations des choix de localisation des ménages (Guevara, 2010 ; Sener et al, 2011 ; Ibeas et al, 2013).

Ozturk et Irwin (2001) sont parmi les rares auteurs à avoir utilisé le modèle probit pour appréhender l'aspect spatial des choix de localisation. Il s'agit d'estimer un modèle probit binaire spatial pour tester l'influence des services publics et les caractéristiques du voisinage sur la décision de relocalisation des propriétaires de la région d'Ohio. Ce modèle a permis d'intégrer la composante spatiale dans la décision de déménagement (la mobilité intra-urbaine). Cependant, l'application de ce modèle dans le cas d'un choix de localisation avec la présence de plusieurs choix de localisation reste limitée.

En parallèle, les seuls travaux qui ont utilisé les modèles logit mixtes pour modéliser les choix de localisations restent à notre connaissance ceux de Bhat et Guo (2004). Ces travaux n'ont pas utilisé des structures purement mixtes mais plutôt des structures combinées entre la forme des modèles GEV et mixte. Il s'agit d'associer la structure des modèles mixtes avec celle des modèles GEV. Ces structures combinées présentées dans la deuxième section permettent d'intégrer la dimension spatiale dans l'estimation des choix de localisation, tout en gardant une forme analytiquement fermée. Cependant, leur application à une échelle fine (un nombre élevé d'alternatives) comme c'est le cas dans notre travail de thèse, augmente encore le nombre de paramètres à estimer.

Dans le deuxième chapitre, nous mettons l'accent justement sur les méthodes développées pour réduire le nombre élevé d'alternatives dans le cas des modèles GEV.

Complexité des calculs

Figure 3.2 : Relation entre complexité des calculs et restriction des hypothèses des modèles de choix discrets avec un grand nombre d'alternatives

Famille des modèles	Type	Hypothèses principales
Famille GEV	MNL	Les termes d'erreurs entre les alternatives de choix sont IDD
	Emboîté	Permet aux alternatives appartenant au même nid d'être corrélées, en maintenant l'hypothèse d'une distribution identique
	PCL/SCL	Permet à toutes les paires d'alternatives d'être corrélées, en maintenant l'hypothèse d'une distribution identique
	GSCL/CNL	Permet à une alternative d'appartenir à plusieurs nids, en maintenant l'hypothèse d'une distribution identique
Famille des modèles mixtes	MMNL	Permet aux termes d'erreurs d'être corrélés et non identiquement distribués
Modèle probit	Probit	En supposant que les termes d'erreurs suivent une loi normale, ce modèle relâche l'hypothèse selon laquelle les alternatives sont identiquement et indépendamment distribuées.

Source : *Élaboration propre*

Restriction des hypothèses

Chapitre 4 – Quels apports des développements récents pour les choix de localisation

Le premier chapitre a permis de synthétiser l'ensemble des structures de modèles de choix discrets qui peuvent être utilisées pour analyser les choix de localisation des ménages. Afin que les zones de localisation soient aussi en concordance avec les structures retenues, ces zones doivent respecter certaines caractéristiques (Train, 2009). La première caractéristique est l'exclusivité mutuelle. Le choix d'une alternative implique que les autres alternatives ne sont pas choisies. La deuxième caractéristique est l'exhaustivité. Chaque individu doit nécessairement faire un choix. Et dernièrement, le nombre d'alternatives doit être fini. Si les deux premières caractéristiques ne sont pas respectées, le modélisateur peut revoir la constitution de l'ensemble du choix, en créant d'autres alternatives indiquant si le choix d'alternatives est fait seul ou avec une autre alternative pour respecter la première caractéristique ou ajouter une alternative « aucun choix » à l'ensemble de choix. Par contre, la violation de la dernière caractéristique nécessite d'utiliser un autre modèle que le modèle de choix discret.

Le nombre de zones de localisation est généralement fini. Il est donc en concordance avec la structure des modèles de choix discret. Toutefois, le problème majeur reste la définition de l'ensemble des choix disponibles aux ménages. Dans le cas des choix déclarés, le problème de la définition de cet ensemble est simple à résoudre. Dans le cas des choix révélés, le problème n'est pas simple et se complexifie avec l'augmentation du nombre d'alternatives (Bonnell, 2004).

Traditionnellement, les modélisateurs supposent que toutes les alternatives sont disponibles à tous les ménages. En analysant les impacts d'une spécification non correcte de l'ensemble de choix sur les résultats d'estimation, Swait et Ben-Akiva (1986) cités par Swait (2001) ont trouvé que l'inclusion d'alternatives non pertinentes dans l'ensemble des choix produit des paramètres biaisés. Ces biais sont dus principalement à la sous estimation de l'impact des variations dans les attributs des alternatives.

Pour corriger ces biais, deux courants de méthodologies ont été développés. Le premier consiste à définir l'ensemble des choix disponibles pour chaque ménage (section 4.1). Le deuxième généralise les méthodes d'échantillonnages proposés par McFadden (1978), en développant des méthodes pour corriger les biais d'échantillonnage dans le cas des modèles GEV (section 4.2).

4.1. La constitution de l'ensemble des choix

La définition de l'ensemble des choix dans le cas des choix de localisation nécessite une certaine formalisation. Thill (1992) donne une synthèse des approches utilisées pour définir l'ensemble des choix d'une manière endogène dans le cas des choix de destination. La première approche consiste à décomposer la probabilité de choisir une alternative en un produit de probabilité : la probabilité qu'une alternative appartienne à l'ensemble des choix C_n et la probabilité de choisir cette alternative conditionnellement à leur appartenance à l'ensemble C_n (Burnett et Hanson (1979 ; 1982) cités par Thill (1992)). La deuxième approche est basée sur le principe espace-temps (Lenntorp (1976), cité par Thill (1992)). Il s'agit d'algorithmes combinés, qui simulent l'ensemble des alternatives disponibles à un individu donné conditionnellement aux contraintes espace-temps définies par le modélisateur. La troisième approche est inspirée du modèle probabiliste de Manski (1977). Cette méthode est appliquée en deux étapes : les probabilités de toutes les combinaisons possibles entre les alternatives de l'ensemble des choix sont calculées d'abord avant de déduire la probabilité de choisir une alternative conditionnellement aux choix de l'ensemble défini. En supposant que les alternatives sont indépendamment et identiquement distribuées, la dernière approche (Forthringham, 1988) intègre la probabilité d'appartenance d'une alternative à l'ensemble des choix ($p_n(i \in C_n)$) de la manière suivante :

$$P_n(i) = \frac{e^{V_{in}} p_n(i \in C_n)}{\sum_k e^{V_{kn}} p_n(k \in C_n)} \quad -4.1-$$

Tel que V_{in} est la partie déterministe de choisir l'alternative i par le ménage n .

Le modèle GenL (choice set Generation Logit) de Swait (2001) est aussi utilisé comme une alternative du modèle emboîté. Il utilise une représentation réseau plutôt qu'une

représentation par arbre. Il s'agit de considérer plusieurs chemins entre les alternatives et le nid, chaque chemin correspond à un ensemble de choix latent possible. La complexité de ce modèle augmente avec l'augmentation du nombre d'alternatives. Déjà avec cinq alternatives, les estimations sont coûteuses. L'ensemble du choix est déterminé d'une manière endogène en fonction des préférences des individus pour un choix particulier.

Récemment, dans son travail de thèse, Zolfaghari (2013) a développé une méthode pour former l'ensemble des zones de localisation disponibles aux ménages. Il s'agit de considérer un seuil de distance domicile-travail et de définir l'ensemble des choix à partir des zones qui ne dépassent pas cette distance. En comparant cette méthode et celles susceptibles d'être utilisées pour limiter le nombre d'alternatives de l'ensemble des choix, les résultats de ce même travail de thèse ont montré que l'estimation d'un modèle multinomial avec les alternatives formées moyennant cette nouvelle méthode est équivalente à une estimation d'un MNL estimé avec un échantillon aléatoire d'alternatives.

4.2. Échantillonnage des alternatives

Les méthodes développées pour corriger les biais liés à l'échantillonnage doivent respecter les deux conditions suivantes (Wonnacott et Wonnacott, 1991) :

- Absence de Biais.

Cela peut se présenter sous deux formes : la première se traduit par une égalité entre l'espérance de l'estimateur obtenu après échantillonnage et celui calculé avec l'ensemble des choix, on parle dans ce cas d'un estimateur consistant. Et la deuxième est spécifique au cas des grandes tailles et l'absence de biais est seulement asymptotique (la normalité asymptotique).

- Efficacité

Elle se traduit par une variance minimale. La comparaison de deux estimateurs non biaisés se fait en choisissant celui ayant la variance minimale. Dans le cas de deux estimateurs un biaisé et un non, on fait appel plutôt à l'erreur quadratique moyenne.

Très tôt, l'efficacité et la consistance de l'estimation d'un modèle logit multinomial avec un échantillon d'alternatives ont été montrées (McFadden, 1978). Toutefois, la question restait non résolue pour les logit emboîtés et plus globalement les modèles GEV. Ce blocage

théorique explique le faible développement des modèles de choix discrets en matière de choix de localisation. Récemment, les travaux de Guevara (2010), de Hurtubia et al (2010) et de Guevara et Ben-Akiva (2013a) proposent une solution pour résoudre le problème. Leur application a pour le moment été limitée à des cas théoriques et à un jeu de données empiriques comprenant un nombre limité d'observations. Nous proposons dans cette thèse d'utiliser cette solution afin de pouvoir étendre les résultats de ces travaux.

Nous commençons par la présentation de la méthode d'échantillonnage utilisée pour estimer un modèle logit multinomial (4.2.1), et nous poursuivons par la méthode retenue pour corriger les biais liés à l'échantillonnage dans le cas d'un modèle emboîté (4.2.2).

4.2.1. MNL

La force du modèle logit multinomial peut être résumée par la propriété d'indépendance des alternatives non pertinentes (Ben-Akiva et Lerman, 1985). Selon cette propriété, le rapport des probabilités de deux alternatives est indépendant des autres alternatives disponibles, ce qui se traduit par la possibilité d'ajouter ou d'enlever une alternative de l'ensemble du choix sans devoir redéfinir le modèle, ce qui nous permet une estimation non biaisée sur un sous ensemble d'alternatives.

Soit \mathbf{D}_n le sous-ensemble d'alternatives de \mathbf{C}_n contenant \mathbf{i} choisit par le ménage \mathbf{n} et $\pi\left(\frac{D_n}{i}\right)$ la probabilité de construire ce sous-ensemble \mathbf{D}_n sachant que l'alternative \mathbf{i} est déjà choisie par le ménage \mathbf{n} . Alors, $\pi\left(\frac{D_n}{i}\right) = 0$ pour tout \mathbf{D}_n qui ne contient pas \mathbf{i} .

Selon le théorème de Bayes, la probabilité conditionnelle qu'un ménage \mathbf{n} choisisse l'alternative \mathbf{i} parmi un sous-ensemble \mathbf{D}_n est égale à

$$P_n\left(\frac{i}{D_n}\right) = \frac{P_{in} \pi_n\left(\frac{D_n}{i}\right)}{\sum_{j \in C_n} P_{jn} \pi_n\left(\frac{D_n}{j}\right)} = \frac{P_{in} \pi_n\left(\frac{D_n}{i}\right)}{\sum_{j \in D_n} P_{jn} \pi_n\left(\frac{D_n}{j}\right) + \sum_{j \notin D_n} P_{jn} \pi_n\left(\frac{D_n}{j}\right)} \quad -4.2-$$

Or $\pi\left(\frac{D_n}{i}\right) = 0$ pour tout sous-ensemble \mathbf{D}_n qui ne contient pas \mathbf{i} , la somme

$\sum_{j \notin D_n} P_{jn} \pi_n\left(\frac{D_n}{j}\right)$ est égale à 0 et l'équation (4.2) s'écrit alors :

$$P_n \left(\frac{i}{D_n} \right) = \frac{P_{in} \pi_n \left(\frac{D_n}{i} \right)}{\sum_{j \in D_n} P_{jn} \pi_n \left(\frac{D_n}{j} \right)} \quad -4.3-$$

En remplaçant P_{in} par la Formule (3.3) on obtient :

$$P_n \left(\frac{i}{D_n} \right) = \frac{e^{\mu W_{in}} \pi_n \left(\frac{D_n}{i} \right)}{\sum_{j \in D_n} e^{\mu W_{jn}} \pi_n \left(\frac{D_n}{j} \right)} = \frac{e^{\mu W_{in} + \ln \left(\pi_n \left(\frac{D_n}{i} \right) \right)}}{\sum_{j \in D_n} e^{\mu W_{jn} + \ln \left(\pi_n \left(\frac{D_n}{j} \right) \right)}} \quad -4.4-$$

En maximisant l'équation (4.4), on obtient un estimateur convergent et non biaisé (McFadden, 1978).

Une procédure aléatoire d'échantillonnage se traduit mathématiquement par

$$\pi_n \left(\frac{D_n}{i} \right) = \pi_n \left(\frac{D_n}{j} \right) \quad -4.5-$$

Et par conséquent, on peut simplifier par $\ln \left(\pi_n \left(\frac{D_n}{i} \right) \right)$ dans l'équation (4.4), elle devient alors :

$$P_n \left(\frac{i}{D_n} \right) = \frac{e^{\mu W_{in}}}{\sum_{j \in D_n} e^{\mu W_{jn}}} \quad -4.6-$$

D'après les équations (3.3) et (4.6), on obtient : $P_n \left(\frac{i}{D_n} \right) = P_n \left(\frac{i}{C_n} \right)$

Cela signifie que la probabilité de choisir une alternative parmi l'ensemble des alternatives (C_n) est équivalente à la probabilité de choisir une alternative parmi le sous-ensemble (D_n), à condition que ce dernier ne soit pas le même pour tous les ménages et soit tiré aléatoirement.

Empiriquement, plusieurs travaux ont montré la pertinence des paramètres estimés à partir d'un modèle logit Multinomiale avec un échantillon d'alternatives. La comparaison des approches permettant de traiter le problème d'un grand nombre d'alternatives (Chen et al., 2005) a montré que l'estimation d'un logit multinomial avec un échantillon aléatoire donne de meilleurs résultats. Plus récemment, le travail de Zolfaghari et al. (2012) a montré une équivalence entre les résultats d'estimation d'un modèle logit multinomial, en considérant

l'ensemble des choix, un échantillon aléatoire de l'ensemble des choix, ou une sélection de l'ensemble des choix avec la méthode importance sampling avec correction.

4.2.2. GEV

Rappelons que la probabilité qu'un ménage \mathbf{n} choisisse une alternative \mathbf{i} selon une forme **GEV** s'écrit comme suit :

$$P(i / C_n) = \frac{e^{V_i} G_i(e^{V_1}, \dots, e^{V_J})}{\mu G(e^{V_1}, \dots, e^{V_J})} \quad -4.7-$$

Dans ce travail de thèse, le modèle emboîté retenu pour estimer les choix de localisation se dérive de la fonction suivante :

$$G(e^{V_1}, \dots, e^{V_J}) = \sum_{m=1}^M \left(\sum_{i \in C_{nm}} \left(e^{V_i} \right)^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \quad -4.8-$$

μ_m est le paramètre de dissimilarité du nid \mathbf{m}

\mathbf{J} est le nombre d'alternatives de l'ensemble du choix \mathbf{C}_n

\mathbf{G}_i est la dérivée première de la fonction \mathbf{G}

La probabilité devient alors :

$$P(i / C_n) = \frac{e^{\frac{\mu_m}{\mu} V_i} \left(\sum_{j \in C_{nm}} e^{\frac{\mu_m}{\mu} V_j} \right)^{\frac{\mu}{\mu_m} - 1}}{\sum_{m=1}^M \left(\sum_{j \in C_{nm}} e^{\frac{\mu_m}{\mu} V_j} \right)^{\frac{\mu}{\mu_m}}} \quad -4.9-$$

Or \mathbf{G} est homogène, selon le théorème d'Euler, cette formule peut être écrite sous cette forme :

$$P(i / C) = \frac{e^{V_i + \ln G_i(e^{V_1}, \dots, e^{V_J})}}{\sum_{j \in C_m} e^{V_j + \ln G_j(e^{V_1}, \dots, e^{V_J})}} \quad -4.10-$$

En remplaçant $\ln \mathbf{G}_i$ par sa formule, l'équation (4.10) devient :

$$P(i / C_n) = \frac{e^{V_{in} + \left(\frac{\mu}{\mu_m} - 1\right) \left(\ln \sum_{j \in C_{nm}} e^{\mu_m V_{jn}} \right) + \ln \mu + (\mu_m - 1) V_{in}}}{\sum_{j \in C_{nm}} \left(e^{V_{jn} + \left(\frac{\mu}{\mu_m} - 1\right) \left(\ln \sum_{j \in C_{nm}} e^{\mu_m V_{jn}} \right) + \ln \mu + (\mu_m - 1) V_{jn}} \right)} \quad -4.11-$$

$\ln \sum_{j \in C_m} e^{\mu_m V_{jn}}$ est le seul terme qui dépend de l'ensemble C_{nm} (l'ensemble des alternatives du nid m). En tirant un échantillon aléatoire d'alternatives D_{nm} de l'ensemble C_{nm} , Guevara (2010) a proposé de multiplier cette somme des exponentielles des utilités systématiques par un facteur w_i . Quatre méthodes peuvent être utilisées pour calculer ce facteur et obtenir un estimateur consistant et non biaisé respectant la condition suivante :

$$E \left(\sum_{j \in C_{nm}} e^{\mu_m V_{jn}} \right) - E \left(\sum_{j \in D_{nm}} w_j e^{\mu_m V_{jn}} \right) = 0.$$

Dans cette thèse, nous avons utilisé la méthode qui a donné les meilleurs résultats dans le travail de Guevara (2010), à savoir la méthode des probabilités itératives élargies (Expanded Iterative Probability). Elle consiste à un calcul itératif à quatre étapes du facteur w_i .

Pour un nid m donné, on ne considère que les alternatives appartenant au même nid qui ont le même facteur w_i . La valeur initiale de ce facteur se calcule comme étant le rapport entre le nombre de ménages qui ont choisi des alternatives appartenant au nid m et le total des ménages. Le processus itératif se déroule comme suit :

Première étape : $k=0$

Calculer la part du marché de l'alternative i (w_i)

$$w_i^k = \frac{1}{w_j + \frac{\tilde{J}_m - 1}{J_m - 1} \sum_{\substack{l \in C_{nm} \\ l \neq j}} w_l + \left(1 - \sum_{\substack{l \in C_{nm} \\ l \neq j}} w_l\right) \frac{\tilde{J}_m}{J_m}} \quad -4.12-$$

Tel que J_m est le nombre d'alternatives appartenant au nid m et \tilde{J}_m est le nombre d'alternatives tirées aléatoirement du nid m .

Deuxième étape :

Pour chaque ménage \mathbf{n} : calculer la probabilité des alternatives de l'ensemble de choix des ménages (les alternatives tirées aléatoirement pour un ménage donné plus l'alternative choisie).

En remplaçant le facteur w_j par sa formule et en pondérant par $\ln \pi(\mathbf{D}_m / i)$ pour corriger le biais de l'échantillonnage stratifié, la probabilité s'écrit alors :

$$P_n(i / D_{nm}) = \frac{e^{V_{in} + \left(\frac{\mu}{\mu_m} - 1\right) \left(\ln \sum_{j \in D_{nm}} w_j e^{\mu_m V_{jn}} \right) + \ln \mu + (\mu_m - 1) V_{in} + \ln \pi(D_{nm} / i)}}{\sum_{\substack{l \in D_{nm} \\ l \neq j}} e^{w_l^k V_{jn} + \left(\frac{\mu}{\mu_m} - 1\right) \left(\ln \sum_{j \in D_{nm}} w_j e^{\mu_m V_{jn}} \right) + \ln \mu + (\mu_m - 1) V_{jn} + \ln \pi(D_{nm} / j)}} \quad -4.13-$$

Troisième étape :

Calculer le facteur w_j à la $k+1^{\text{ème}}$ itération :

$$w_i^{k+1} = \frac{1}{\hat{P}^k(j) + \frac{\tilde{J}_m - 1}{J_m - 1} \sum_{\substack{l \in D_m \\ l \neq j}} w_l^k \hat{P}^k(l) + \frac{\tilde{J}_m}{J_m} \left(1 - \sum_{\substack{l \in D_m \\ l \neq j}} w_l^k \hat{P}^k(l) \right)} \quad -4.14-$$

Quatrième étape :

$k \leftarrow k+1$ et on répète les trois étapes jusqu'à l'obtention de la convergence : $\max | \hat{P}^k(j) - \hat{P}^{k+1}(j) | \leq 1 / (10J)$. Sinon, le processus s'arrête à la 50^{ème} itération.

4.3. Taille de l'échantillon

L'estimation du modèle multinomial avec échantillon d'alternatives consiste à tirer aléatoirement un sous ensemble d'alternatives plus l'alternative choisie par chaque ménage et cela parmi les alternatives de l'ensemble du choix (\mathbf{C}_n). L'échantillon d'alternatives sélectionné (\mathbf{D}_n) est différent pour chaque ménage. Cela ne permet pas de prendre une alternative de référence et d'interpréter les résultats par rapport à cette dernière, comme dans le cas des modèles multinomiaux avec des alternatives identifiées et dont le nombre est réduit.

Il s'agit ensuite de calculer pour chaque ménage les probabilités de choisir chacune des alternatives de l'échantillon \mathbf{D}_n et de déduire la log vraisemblance :

$$L_n = \sum_{i=1}^J \ln(P_{in} * Ch_{in}) \quad -4.15-$$

Tel que \mathbf{P}_{in} est la probabilité qu'un ménage \mathbf{n} choisisse une alternative \mathbf{i} , \mathbf{Ch}_{in} est une variable dichotomique qui vaut 1 si l'alternative \mathbf{i} est choisie par le ménage \mathbf{n} et 0 sinon, et \mathbf{J} est la taille de l'échantillon \mathbf{D}_n .

Il s'agit enfin de maximiser la somme des log vraisemblances : $L = \sum_{n=1}^N L_n$

Concernant l'effet de la taille de l'échantillon d'alternatives sur la consistance de l'estimateur, la théorie n'a pas apporté de réponse pour l'instant.

Empiriquement, il y a quelques tentatives pour déterminer la taille de l'échantillon à utiliser pour estimer un logit multinomial avec un échantillon d'alternatives et obtenir des résultats fiables. A titre d'exemple le travail de Parsons et Kealy (1992) a montré qu'une taille d'échantillon de six alternatives donne de bonnes estimations et le travail de Nerella et Bhat (2004) suggère d'utiliser entre 1/8 et 1/4 de l'ensemble des alternatives pour obtenir des valeurs plus précises des paramètres estimés. Toutefois, ces travaux restent basés sur des données synthétiques qui ne peuvent pas considérer tous les contextes avec les différents nombres d'alternatives et de variables. Ainsi, les résultats obtenus restent indicatifs face à l'absence d'une justification théorique. Par ailleurs, le nombre d'alternatives dans les choix de localisation est large (des centaines d'alternatives) et l'application par exemple de la règle de Nerella et Bhat (2004) implique de sélectionner un échantillon avec une taille qui reste grande (des cinquantaines). Or, avec le nombre élevé de variables explicatives et d'observations, l'estimation devient fastidieuse. Pour pallier cette difficulté, nous proposons dans la suite du travail de tester plusieurs tailles d'échantillons afin de retenir une plage de tailles conduisant à des résultats stables.

Quelle méthode pour réduire le nombre élevé d'alternatives

Le choix des modèles appropriés pour modéliser les choix de localisation en milieu urbain est fait sous la contrainte du nombre élevé d'alternatives de choix. Pour réduire ce nombre et

dépasser la complexité des calculs, deux méthodes se présentent. La première consiste à définir l'ensemble des choix disponibles pour chaque ménage et la deuxième est basée sur l'échantillonnage d'alternatives. Les travaux empiriques réalisés pour comparer entre ces deux méthodes ont montré que la première méthode est équivalente à la méthode d'échantillonnage aléatoire des alternatives. Si l'efficacité et la consistance de l'estimation d'un modèle logit multinomial avec un échantillon d'alternatives ont été démontrées depuis le travail de McFadden (1978), la question restait non résolue pour les logit emboîtés et plus globalement les modèles GEV. Cela explique en partie l'utilisation par la majorité des travaux empiriques traitant les déterminants de choix de localisation d'un modèle logit multinomial, malgré les limites pour prendre en compte la présence d'autocorrélation spatiale entre les zones de localisation. Les développements récents faits dans ce sens permettent de résoudre ce problème. Toutefois, leur application a pour le moment été limitée à des cas théoriques et à un jeu de données synthétiques comprenant un nombre limité d'observations.

A travers ce chapitre, nous avons constaté également une absence de réponse théorique concernant la taille de l'échantillon optimale à utiliser pour estimer les modèles logit multinomiaux et les modèles de la famille GEV en général.

En l'absence de réponses théoriques pour estimer nos modèles, nous posons la question suivante : l'adaptation des méthodes développées récemment permet-elle de traiter les autocorrélations spatiales et d'améliorer la portée modélisatrice du modèle multinomial ?

Conclusion

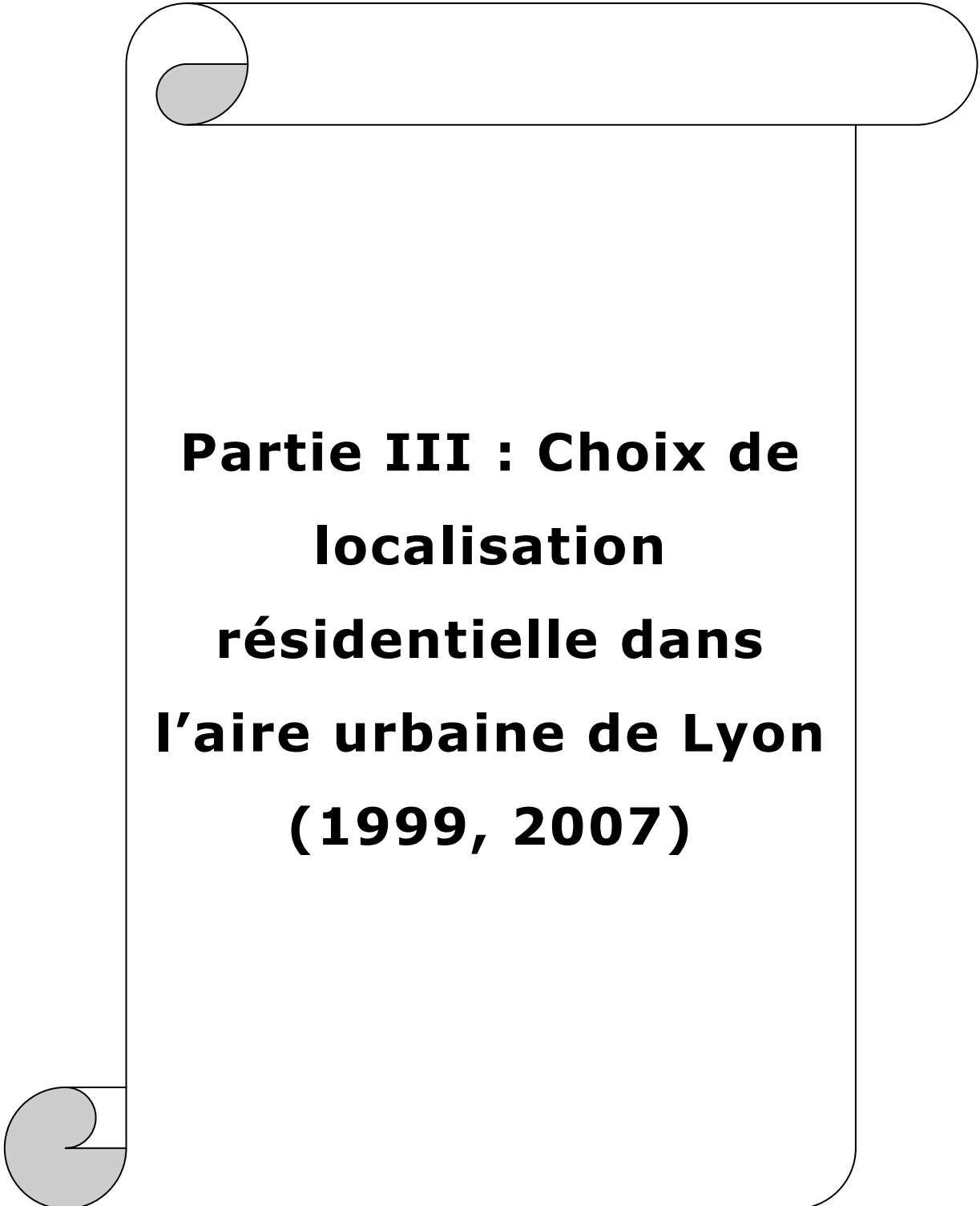
En analysant les différents travaux développés pour appréhender l'aspect spatial des choix de localisation, nous avons montré que théoriquement la famille des modèles mixtes est la plus appropriée pour traiter les problèmes d'autocorrélation spatiale. Toutefois, la forme non analytique de cette famille de modèle complexifie les calculs au plan empirique.

Étant contraint dans ce travail de thèse par un nombre élevé d'alternatives, nous avons privilégié la famille des modèles GEV et plus particulièrement une structure hiérarchique selon un logit emboîté. La flexibilité de cette structure nous permet à la fois de relâcher partiellement les propriétés restrictives du modèle logit multinomial, à savoir l'indépendance par rapport aux choix non retenus, et de réduire le nombre élevé d'alternatives de l'ensemble des choix. D'un point de vue statistique, les méthodes d'échantillonnage restent les plus appropriées pour réduire le nombre d'alternatives, surtout avec les développements récents qui ont généralisé les méthodes de corrections des biais liés à l'échantillonnage sur tous les modèles de la famille GEV. Cependant, l'absence d'une réponse théorique concernant la taille de l'échantillon soulève la question suivante : quelle est la taille optimale pour estimer nos modèles ?

A cette question s'ajoute celle liée à la définition du zonage et le problème classique d'influence du découpage spatial (Modifiable Area Unit Problem) sur les résultats de modélisation.

L'objectif de cette thèse étant de tester l'apport d'une structure hiérarchique pour traiter les problèmes d'autocorrélation spatiale, nous sommes confrontés à un problème lié à l'absence de méthode bien établie permettant de regrouper les zones de localisation et de constituer les nids.

Dans la partie suivante, nous traitons empiriquement l'ensemble de ces problèmes afin d'apporter des éléments de réponses à ces questionnements.



**Partie III : Choix de
localisation
résidentielle dans
l'aire urbaine de Lyon
(1999, 2007)**

En conciliant les points forts des différentes structures de modèles de choix discrets qui peuvent être utilisées pour expliquer les choix de localisation et la faisabilité de leur application sur une base de données volumineuse, nous avons retenu deux modèles de la famille GEV avec échantillon d'alternatives pour modéliser les choix de localisation des ménages de l'aire urbaine de Lyon. Il s'agit d'un modèle logit multinomial et d'un modèle logit emboîté.

Cette thèse s'inscrit dans le cadre du projet Simbad. Ce dernier vise à modéliser et simuler les interactions transport-usage du sol sur l'aire urbaine de Lyon, prenant comme référence l'année 1999. Le choix de cette année de référence, relativement ancienne, est en grande partie lié aux objectifs de ce projet et aux contraintes de disponibilité et de production de données.

La suite de cette partie est structurée de la manière suivante. Le cinquième chapitre présente nos choix méthodologiques. Le sixième chapitre, présente le processus d'estimation empirique d'un modèle multinomial avec échantillon aléatoire des alternatives pour déterminer les facteurs expliquant les choix de localisation des Lyonnais en 1999. Il s'agit ensuite d'utiliser plusieurs structures hiérarchiques pour tester l'apport de l'application d'un modèle logit emboîté pour modéliser les choix de localisation (chapitre 7). Le dernier chapitre consiste à comparer les comportements de choix de localisation en 1999 à ceux en 2007.

Chapitre 5 – Aire urbaine de Lyon : Échelles et données

Toute tentative de modélisation des choix de localisation résidentielle se trouve confrontée à plusieurs difficultés méthodologiques liées à la multi-dimensionnalité des facteurs explicatifs, à la disponibilité des données et des outils de modélisation nécessaires pour les appréhender, au grand nombre d’alternatives de choix et à l’échelle appropriée pour les analyser. Ces multiples contraintes expliquent en grande partie le nombre réduit de modèles développés en France pour estimer les choix de localisation à un niveau désagrégé (de Palma et al., 2007).

L’objectif de notre travail est de dépasser ces contraintes méthodologiques et d’analyser les déterminants de choix de localisation des ménages de l’aire urbaine de Lyon. A travers ce chapitre, nous cherchons à contourner progressivement les difficultés rencontrées lors de l’estimation. La première section (5.1) discute les différentes sources de données susceptibles d’être exploitées pour construire les déterminants de choix de localisation identifiés dans la première partie. Toutefois, ces données sont disponibles à différentes échelles spatiales, ce qui nous conduit à poser d’autres questions liées à la définition du zonage. Plusieurs échelles spatiales ont donc été testées (commune, quartier...), afin de choisir celle la plus adaptée à l’aire urbaine de Lyon (section 5.2)

5.1. Données et constructions des variables

Cette section consiste à présenter d’une manière détaillée les données utilisées pour analyser empiriquement les choix de localisation résidentielle. Après une brève description de l’aire urbaine de Lyon, nous nous penchons sur la présentation des principales bases de données utilisées pour construire les déterminants des choix de localisation identifiés dans les travaux empiriques. En suivant la classification adoptée dans le deuxième chapitre (sous-sections 5.1.1-5.1.4) nous questionnons les sources de données disponibles et les variables issues de leurs traitements. Quatre groupes de variables sont modélisés : arbitrage du marché, aménités spatiales, environnement social et caractéristiques des ménages.

L'aire urbaine de Lyon est la plus importante en termes de population après l'aire urbaine parisienne. Selon les résultats du recensement de l'INSEE correspondant à l'année de référence, à savoir 1999, 1 648 216 personnes, constituant 662 249 ménages, habitent l'aire urbaine de Lyon. Ces populations sont réparties de manière relativement concentrique, où l'on peut distinguer le centre (Lyon et Villeurbanne), la couronne urbaine regroupant les autres communes du grand Lyon (périmètre des transports urbains) et la couronne périurbaine à dominante résidentielle (figure 5.1). Le choix de la définition de l'aire urbaine de 1999 et le découpage qui en découle sont en grande partie contraints par la disponibilité des données à cette échelle d'observation pour les deux années d'analyse (1999 et 2007) et par certains choix méthodologiques propres à la plateforme SIMBAD que cette thèse alimente à travers la modélisation des choix de localisation résidentielle.

L'aire urbaine lyonnaise a enregistré de fortes dynamiques démographiques durant les quarante dernières années (figure 5.1). Elles illustrent que l'analyse des choix de localisation des ménages lyonnais, en 1999, se fait dans un contexte de permanence de la périurbanisation et d'un récent retour au centre (figure 5.1).

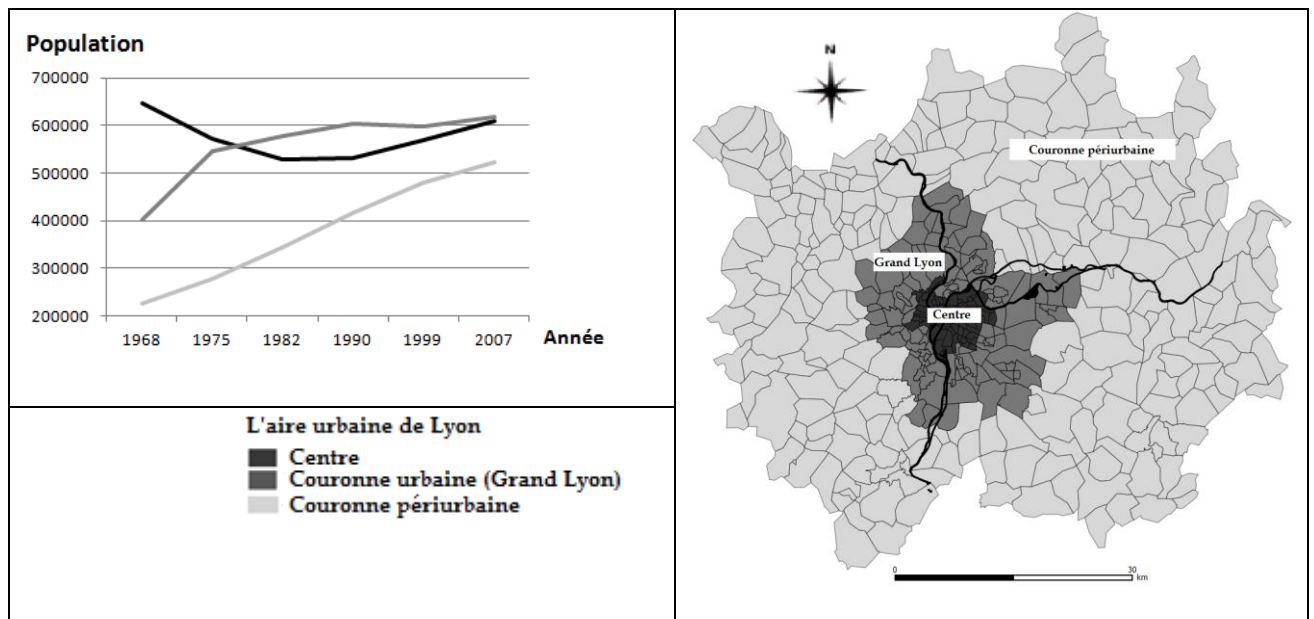


Figure 5.1 : Évolution de la population de l'aire urbaine de Lyon dans le centre et les deux couronnes urbaine et périurbaine (1968-2007)

Source : Recensements INSEE 1968, 1975, 1982, 1990, 1999, 2007

5.1.1. Arbitrage du marché : accessibilité et prix de l'immobilier

Les deux variables relatives aux contraintes du marché mobilisées dans l'estimation du modèle de choix de localisation résidentielle sont l'accessibilité à l'emploi et le prix de l'immobilier. Chacune des deux variables a été mesurée en combinant plusieurs sources de données disponibles à différentes échelles.

- Accessibilité à l'emploi

La distance au centre représente la forme la plus simplifiée de l'accessibilité à l'emploi introduite dans plusieurs modèles de localisation résidentielle (de Palma et al., 2007 ; Goffette-Nagot et Schaeffer, 2013). Or, ce proxy considère que l'ensemble des emplois sont concentrés au centre et ne permet pas de prendre en compte la distribution multipolaire des emplois. Il ne prend pas non plus en compte l'hétérogénéité de la desserte en transport pour accéder aux emplois. L'accessibilité peut être représentée à travers le nombre d'emplois atteignables à partir de n'importe quelle zone de localisation à l'intérieur d'un seuil de distance ou de temps. Il s'agit alors de l'indicateur d'accessibilité isochrone. Les indicateurs d'accessibilités gravitaires permettent de prendre en compte les deux dimensions relatives à la

distribution des emplois et des temps de déplacement permettant de les atteindre à partir des différents modes de transport. Plusieurs travaux utilisent cet indicateur pour expliquer la localisation résidentielle des ménages (Waddell, 1996 ; Guo et Bhat, 2001).

– Échelle de disponibilité

Les temps de transport nécessaires au calcul des accessibilités en voiture et en transport en commun ont été modélisés à l'échelle de l'IRIS dans le cadre du projet SIMBAD (Godinot et al., 2008 ; Nicolas, 2010 ; Cabrera Delgado, 2013). Openshaw (1977) propose une méthode qui permet d'utiliser ces temps pour calculer des matrices temps à un niveau plus agrégé. Il s'agit de déterminer à la fois le centroïde de l'IRIS et celui de la zone de localisation en question (commune, quartier), et de considérer l'IRIS dont le centroïde est le plus proche du centroïde de la zone de localisation comme étant l'IRIS de référence.

– Mesure

Pour cette catégorie de variables, nous nous sommes limités au calcul des accessibilités gravitaires. Elles sont faciles à interpréter et les données nécessaires pour les calculer sont disponibles. D'autres mesures d'accessibilité existent, mais elles demandent des données plus difficiles à collecter ou elles sont difficiles à calculer et à interpréter.

La formule utilisée pour calculer les accessibilités gravitaires est celle de Hansen (1959). Elle

est de la forme suivante :
$$A_i = \sum_j D_j \exp(-\beta t_{ij}) \quad i, j \in Z \quad -5.1-$$

A_i est l'accessibilité de la zone de localisation i aux emplois D_j de la zone j , à l'intérieur du périmètre de l'aire urbaine Z ; t_{ij} est le temps généralisé de déplacement de la zone i vers la zone j et β est le paramètre de pondération reflétant la résistance au temps de déplacement.

La valeur du paramètre β (cf. Équation 5.1) est de 0,13 en 1995 et 0,12 en 2006. Elles sont obtenues à partir du calibrage d'un modèle de distribution des déplacements, pour le motif travail, à l'aide des données des enquêtes ménages déplacements lyonnaises de 1995 et 2006 (Cabrera Delgado, 2013). Ce paramètre étant relativement stable et afin de permettre une analyse diachronique de l'impact de l'accessibilité sur le choix de la localisation résidentielle, nous avons opté pour une valeur stable de (0,13) pendant la période.

A partir de cette formule, on peut calculer plusieurs mesures d'accessibilité gravitaire, en fonction des types d'emplois considérés (industriel, enseignement, service, etc). On parle dans ce cas des accessibilités gravitaires aux établissements de service (acc_grav_G1), aux établissements industriels (acc_grav_G2), aux établissements d'enseignement (acc_eng), etc.

Le mode de transport (VP ou TC) considéré fait aussi la distinction entre une accessibilité gravitaire à l'emploi en voiture particulière (ac_g_VP) ou en transport en commun (ac_g_TC).

– Données

Les données d'emplois utilisées pour mesurer l'indicateur d'accessibilité gravitaire pour les deux années 1999 et 2007 proviennent de la base de données des établissements SIRENE (1999, 2006). Les établissements identifiés par leur numéro SIRET et leurs adresses ont été regroupés par IRIS car c'est à cette échelle fine que la plupart des données de l'INSEE sont disponibles. Le nombre d'emplois a été estimé pour chaque établissement en fonction de son appartenance à une tranche de taille et d'autres variables présentes dans le fichier SIRENE telles que le secteur d'activité et le type du local de l'établissement selon la méthodologie développée par Gérardin et al. (2000). Cela permet de calculer directement le nombre d'emplois par IRIS ou par regroupement de plusieurs IRIS.

- Prix de l'immobilier

- Échelles de disponibilité

Les données sur le marché foncier peuvent être extraites à partir de la « cote annuelle des valeurs vénales immobilières et foncières ». Il s'agit d'un ouvrage qui propose des valeurs vénales approchées pour toutes les communes de la France. Cet ouvrage est riche en données et permet de faire la distinction entre le prix de location et de ventes. Toutefois, toutes les données fournies sont établies à l'échelle de la commune. A la même échelle, on trouve également les données en ligne de l'agence immobilière Efficity (www.efficity.com).

D'autres données sont disponibles et permettent de mesurer le prix de l'immobilier à une échelle plus fine. Il s'agit du fichier « Marché Immobilier des Notaires » (MIN)-PERVAL et de la base « Observatoire des Transactions Immobilières et Foncières » (OTIF)-COURLY et c'est à partir de ces deux sources que le prix de l'immobilier est estimé au niveau de l'IRIS.

- Mesure

Pour le prix de l'immobilier en 1999, nous avons utilisé les valeurs estimées dans le cadre de la thèse d'Homocianu (2009). Il s'agit d'une régression linéaire en fonction du temps d'accès par voiture à l'IRIS central (la préfecture), de la proportion des ménages appartenant au premier quintile de revenu et de la proportion des ménages appartenant au cinquième quintile de revenu.

Pour le prix de l'immobilier 2007, nous nous sommes basés sur les données d'Efficity rassemblées par le LET au niveau de la commune pour estimer un modèle linéaire du prix de l'immobilier de 2007 en fonction du prix de l'immobilier de 1999.

– Données

Le prix de l'immobilier utilisé dans les modèles de choix de localisation résidentielle pour les deux années 1999 et 2007 provient principalement du fichier (MIN)-PERVAL et de la base (OTIF)-COURLY, complétés par des données en ligne de l'agence immobilière Efficity.

5.1.2. Aménité spatiale : services quotidiens et équipements publics

Les aménités spatiales et plus particulièrement la présence des services quotidiens et des équipements publics à proximité de la zone de localisation sont aussi déterminantes dans les choix de localisation.

– Échelle de disponibilité

La construction des variables caractérisant les zones de localisation peut être faite de deux façons. A partir de la Base Permanente des Equipements (BPE) de l'INSEE, cette base fournit le niveau d'équipements et de services des données rapportées à une zone géographique, l'IRIS. Ces équipements sont réparties en trois classes : les équipements de proximité (école maternelle, pharmacie, boulangerie, bureau de poste), intermédiaires (collège, orthophoniste, supermarché, police, gendarmerie), et les équipements de la gamme supérieure (lycée, maternité, hypermarché, agence pôle emploi).

La première BPE disponible sur le site de l'INSEE porte sur les données de l'année 2007. Elle se substitue ainsi à l'inventaire communal, dont le dernier a été réalisé en 1998. Comparables sur le volet présence/absence des équipements, les données de la BPE sont actualisées chaque année, au lieu de tous les 7-8 ans pour l'inventaire communal. De plus, la BPE répertorie une plus grande variété d'équipements (179 au lieu de 36) ; leur localisation est également plus affinée (IRIS) et les coordonnées géographiques précises sont fournies pour certains d'entre eux.

Comme son nom l'indique l'inventaire communal recense la présence ou l'absence d'un équipement au niveau de la commune, ce qui va nous limiter et ne permet pas de tester des zones plus fines. L'absence à la fois d'un inventaire communal et d'un BPE pour l'année 1999 constitue une limite par rapport à l'objectif de la thèse.

– Mesure

Nous nous sommes basés sur la base SIRENE et nous avons compté le nombre d'établissements ayant comme activité principale le service en question.

L'ensemble des services qui ont été constitués à partir de la base SIRENE (1999 ou 2005) est le suivant :

- ❖ Les supermarchés,
- ❖ Les boulangeries et les pâtisseries,
- ❖ Les magasins d'alimentation spécialisé (viande, poissons, tabacs, produits laitiers...),
- ❖ Les services quotidiens : ils regroupent les trois services de bases citées ci-dessus,
- ❖ Les établissements d'enseignement primaire,
- ❖ Les établissements d'enseignement secondaire,
- ❖ Les services de santé,
- ❖ Les banques et postes,
- ❖ Les équipements culturels : cinéma, bibliothèque et maison de la culture (musée, association...),
- ❖ Les équipements sportifs : salle de sports, autres activités sportives,
- ❖ Les équipements récréatifs : manèges forains et parcs d'attractions, bals et discothèques, restauration de type traditionnel, restauration de type rapide, café tabac, autres activités récréatives, jeux de hasard et d'argent, les aires des jeux.

A partir de ces services, deux types de variables ont été constitués et testés. Le premier type mesure le nombre de services et d'équipements au sein de la zone. Le deuxième type indique la proximité à ces services ou ces équipements, en regardant leurs présences ou leurs absences dans la zone de localisation. Il s'agit d'une variable dichotomique qui prend la valeur 1 si la zone de localisation contient au moins un établissement fournissant le service en question et 0 sinon.

Nous avons construit également des variables pour mesurer le niveau de desserte de la zone de localisation par les transports en commun. Il s'agit des accessibilités isochroniques indiquant le nombre de stations de métro ou/et le nombre d'arrêts de bus desservis par des lignes de forte fréquence à l'intérieur d'un seuil de temps. Le temps de marche à pied au sein des différentes zones de localisation utilisées dans nos estimations ne dépasse pas 10 minutes. Pour cela, nous avons construit une variable binaire qui prend la valeur 1 s'il y a une station

de métro (funiculaire inclus) ou un arrêt de bus avec une fréquence de 5 minutes maximum et 0 sinon.

L'effet de l'accessibilité diffère en fonction des préférences des ménages. Pour cela, nous avons construit des variables croisées qui mesurent l'effet combiné de l'accessibilité avec quelques caractéristiques des ménages.

- Données

A l'exception des accessibilités isochroniques obtenues à partir de la codification des lignes de transports collectifs (Cabrera Delgado, 2013), l'ensemble des variables de cette catégorie est construit à partir de la base de données des établissements SIRENE (1999, 2005).

5.1.3. Environnement social : niveau de vie et logements sociaux

L'environnement social a un rôle important dans les choix de localisation (Partie I), mais il est difficile de le caractériser car les critères peuvent être multiples (éthique, social, économique). En optant pour un critère socioéconomique, on distingue deux types de données caractérisant l'environnement social. Le premier concerne la préférence pour l'entre soi et le deuxième la fuite face aux externalités négatives.

- L'entre soi (attiré par externalités sociales positives)

- Échelle de disponibilité

Les travaux appliqués sur des villes américaines ont utilisés des données relatives à l'origine raciale ou ethnique pour mesurer l'entre soi. Toutefois, le recueil de ces données est interdit par la loi française et il n'est pas aisé de distinguer la recherche d'entre soi de la discrimination positive. Pour tester l'effet de cette externalité sociale positive sur les choix de localisation des Lyonnais, nous nous sommes basés sur des variables relatives au revenu et d'autres caractéristiques des ménages, à savoir la catégorie socioprofessionnelle, la taille et l'âge.

- Mesure

Généralement, le revenu du ménage est croisé avec le taux de ménages de la même classe. Une valeur positive du paramètre indique la préférence pour l'entre soi. Ce déterminant peut aussi être mesuré à partir de la différence entre le revenu des ménages et le revenu médian de la zone de localisation (Guo et Bhat, 2001).

– Données

L'ensemble de ces données est obtenue en fusionnant les données de la base logement ou de la base du recensement et les revenus fiscaux DGI-INSEE de l'INSEE.

• Fuites face aux externalités négatives

La préférence des ménages aisés à se localiser dans des zones peuplées par des ménages du même profil et à être entre-soi dans les grandes villes françaises se combine avec un autre mécanisme de fuite et d'évitement des quartiers les moins favorisés. A partir des études menées au niveau de la France, Maurin (2004) a montré que les ménages cherchent à fuir les classes sociales qui leur sont inférieures. Ces dernières se trouvent alors exclues. Il a constaté que généralement le niveau socioprofessionnel des familles qui s'installent dans un quartier est le même que celui des familles qui habitent déjà le quartier. La concentration des logements sociaux (HLM) dans certains quartiers peut également conduire à une concentration des populations à faible revenu et une dévalorisation du sol, ce qui peut faire fuir les ménages riches. En effet, les bénéficiaires des logements HLM ne choisissent pas leurs localisations mais arbitrent entre quelques logements proposés par les gestionnaires de logements sociaux (Bonvalet, 2010).

– Échelle de disponibilité

Ces constats, nous ont conduit à faire appel à la base logement pour calculer le taux des logements sociaux (HLM). Il s'agit d'une base agrégée au niveau infra-communal (IRIS) et qui indique entre autres le nombre de logements total et le nombre de logements HLM. Les données agrégées du recensement général de la population sont aussi utilisées pour extraire le nombre de ménages étrangers au niveau de l'IRIS.

– Mesure

L'intégration directe du taux de ménages à faible revenu et le taux de logements HLM permettent de mesurer les externalités perçues comme négatives. Le nombre d'étrangers dans la zone de localisation est aussi testé pour mesurer l'effet de la nationalité du ménage dans les choix de localisation.

– Données

La fusion des données de la base logement et de la base du recensement général de la population a permis d'obtenir des variables caractérisant le quartier.

5.1.4. Caractéristiques des ménages : catégorie socioprofessionnelle, revenu...

En parallèle des variables caractérisant les zones de localisation, les choix de localisation des ménages peuvent être expliqués par des variables relatives au cycle de vie des ménages : revenu, statut d'occupation du logement, âge, taille du ménage....

- Échelle de disponibilité

Les variables caractérisant le ménage peuvent être extraites de deux bases de données. La base logement ou celle des ménages. Les deux bases sont constituées par l'INSEE à partir du Recensement Général de la Population (RGP).

- Base logement

Il s'agit d'un fichier dont chaque ligne renvoie à un logement décrit par les caractéristiques intrinsèques au logement, sa localisation au niveau de l'IRIS, et les caractéristiques du ménage qui y réside. Cette base est plus riche d'information concernant les caractéristiques des ménages et du logement que le recensement. Toutefois, le niveau de la localisation du logement est la tranche d'unité urbaine. En 2007, cette base était enrichie avec un niveau géographique plus fin pour les communes découpées en IRIS.

Pour obtenir les caractéristiques des ménages, nous avons utilisé le poids du logement (IPONDL) arrondi pour dupliquer les lignes de cette base. Il s'agit de créer autant de lignes (avec les mêmes caractéristiques des ménages et du logement) que le nombre obtenu à partir de la valeur arrondie de l'IPONDL. En faisant l'hypothèse qu'un logement est occupé par un seul ménage, nous sommes passés d'une table de logement à une table de ménages. La somme des IPONDL donne le nombre de ménages en 2007 obtenu à partir des données agrégées de l'INSEE.

À partir de cette base, nous avons extrait les ménages de l'aire urbaine de Lyon ayant déménagé entre 2006 et 2009. Le tableau 5.1 résume la répartition des ménages entre ces années.

Tableau 5.1 : pourcentage des ménages ayant emménagé entre 2006 et 2009

Année	Pourcentage des ménages
2006	45%
2007	33%
2008	20%
2009	2%

Source : *Élaboration propre à partir des données de la base logement 2007*

- Base ménage

En 1999, l'absence d'une variable indiquant l'IRIS du logement dans la base logement nous a conduit à utiliser la base ménage commandée à l'INSEE. Il s'agit d'une base exhaustive décrivant les ménages à partir des variables suivantes :

- ❖ Statut d'occupation : propriétaire, locataire ou sous locataire d'un logement HLM, locataire ou sous locataire d'un logement non HLM, locataire logé gratuitement, et locataire ou sous locataire logé en meublé ou en chambre d'hôtel. Dans nos estimations, nous avons gardé que les propriétaires et les locataires et sous locataires (HLM et Non HLM) et nous avons supprimé les ménages ayant d'autres statuts d'occupation (6% du nombre total des ménages) ;
 - ❖ Année de l'emménagement : pour isoler l'effet des variables explicatives des déménagements, de celui des variables explicatives de localisation, nous avons utilisé cette variable pour sélectionner les ménages ayant récemment emménagé (entre 1998 et 1999). Nous avons obtenu au total 112 112 ménages ;
 - ❖ Age du chef du ménage en 5 classes ;
 - ❖ Nombre de personnes du ménage ;
 - ❖ Nombre de personnes actives du ménage ;
 - ❖ Taux de motorisation (0 VP, 1, 2 ou +) ;
 - ❖ Type d'activité du chef du ménage : actif ayant un emploi, ancien actif, autre inactif, chômeur, militaire, élève ou étudiant ou stagiaire non rémunéré ;
 - ❖ Temps d'activité : sans objet, à temps partiel moins d'un mi-temps, à temps partiel plus d'un mi-temps, à temps complet.
- Mesure

Le revenu peut influencer le choix de localisation à travers la discrimination entre les riches et les pauvres sur le marché de l'immobilier ou à travers la différence de préférence pour les accessibilités, pour les aménités, ou pour l'entre soi.

Les données extraites de la base logement ou celle des ménages ont été complétées par l'affectation d'un quintile de revenu à chaque ménage. Il s'agit d'un processus d'imputation réalisé à partir des données fiscales de l'INSEE-DGI agrégées au niveau de l'IRIS.

La base des revenus fiscaux diffusée par l'INSEE-DGI (2007) n'est pas complète. Pour cela nous avons complété cette base pour estimer les revenus manquants. Ce travail a permis d'estimer les déciles des 275 IRIS manquants à l'aide d'une régression linéaire simple (stepwise). D'autres variables ont été calculées à partir des données initiales comme le taux de chômage et le taux de retraités et ont été utilisés dans l'estimation du premier décile.

Après ce travail d'estimation des revenus manquants dans la base des données fiscales de l'INSEE-DGI (2007), nous sommes passés à l'affectation des revenus aux ménages.

Pour assurer la comparabilité des données entre 1999 et 2007, nous avons appliqué en 2007 la même méthode que celle utilisée pour affecter les revenus des ménages en 1999. Il s'agit d'affecter les revenus selon une procédure d'ajustement proportionnel itératif (IPF : Iterative Proportional Fitting).

L'IPF est une procédure mathématique, initialement développée pour combiner les informations à partir de deux ou plusieurs bases de données. Dans notre cas, nous nous sommes basés sur les revenus agrégés au niveau de l'IRIS (INSEE-DGI, 2007), les données agrégées de l'INSEE (famille) et certaines caractéristiques que nous avons aussi agrégées au niveau de l'IRIS à partir de la base logement (nombre de personnes et nombre de pièces), pour affecter une classe de revenu (décile) à chaque ménage.

- Données

Les données de cette catégorie de variables caractérisant le ménage proviennent principalement de la base du recensement général de la population (RGP) pour 1999 et l'enquête logement pour 2007.

5.2. Échelles spatiales

La définition du zonage est centrale dans l'analyse des choix de localisation résidentielle, mais elle est confrontée au problème classique d'influence du découpage spatial (Modifiable Area Unit Problem) sur les résultats de modélisation (Openshaw, 1984).

En testant plusieurs méthodes de délimitation des quartiers de la baie de San Francisco, Guo et Bhat (2007) montrent que les résultats peuvent changer significativement d'un zonage à un autre. Même s'ils n'apportent pas de réponse sur la meilleure échelle, ils soulignent les limites des unités administratives et mettent en avant l'intérêt d'effectuer des regroupements de données désagrégées, prenant en compte les ruptures spatiales. La manière de définir le quartier dépend du contexte urbain, mais elle reste largement contrainte par la disponibilité des données.

Dans cette section, nous avons testé plusieurs échelles spatiales. La première sous-section (5.2.1) discute les résultats obtenus au niveau d'un zonage très fin en carroyage d'une centaine de mètres (Cellule). La deuxième sous-section (5.2.2) présente les estimations établies à l'échelle de l'IRIS. Les découpages administratifs en commune (sous-section 5.2.3) et en Grand quartiers (sous-section 5.2.4) sont aussi testés.

5.2.1. Le zonage le plus fin : carroyage

Les plates-formes de simulation urbaine développées dans le contexte des villes américaines, telles qu'URBANSIM (Waddell, 2000), privilégient généralement un zonage très fin en carroyage d'une centaine de mètres (cellule).

En France, la production de ce type de données par l'INSEE est encore très embryonnaire et limitée à quelques caractéristiques de la population. La non-disponibilité d'informations sur la plupart des facteurs de localisation impose une désagrégation de données à partir d'une échelle plus large (IRIS). Étant donné que les estimations des choix de localisation du projet SIMBAD sont menées au niveau de la cellule, nous avons commencé par tester ce type de zonage. En suivant les choix faits dans le cadre de ce projet, nous avons sélectionné certaines variables parmi celles retenues par Homocianu (2009) dans son modèle final pour estimer un modèle logit multinomial avec un échantillon aléatoire composé de 7 alternatives (six alternatives plus la cellule déjà choisie par le ménage) à l'aide d'un programme que nous avons développé sous R.

Les variables utilisées sont les suivantes :

- ❖ L'accessibilité gravitaire aux établissements de service (acc_grav_G1),
- ❖ L'accessibilité gravitaire aux établissements industriels (acc_grav_G2),
- ❖ L'accessibilité gravitaire aux établissements d'enseignement (acc_eng),
- ❖ Une variable croisée entre les ménages de moins 30 ans et le taux de ménages appartenant à la même classe d'âge dans la cellule (age1*taux1),
- ❖ Une variable croisée entre les ménages de plus de 75 ans et le taux de ménages appartenant à la même classe d'âge dans la cellule (age5*taux5),
- ❖ Une variable croisée entre l'appartenance du ménage aux deux premiers quintiles du revenu et le taux de ménages des mêmes quintiles de revenu dans la cellule (taux1*rev1),
- ❖ La variable croisée entre l'appartenance du ménage aux deux derniers quintiles du revenu et le taux de ménages des mêmes quintiles de revenu dans la cellule (taux3*rev3),
- ❖ Le prix de l'immobilier (prix_imm). Le tableau 2 résume les résultats obtenus.

Tableau 5.2 : Résultats de l'estimation du modèle de localisation des ménages à l'échelle de la cellule

Variabes	Coef	Se	t_test
acc_grav_G1	1,14E-04	4,94E-07	2,30E+02
acc_grav_G2	4,40E-05	1,40E-06	3,13E+01
acc_eng	1,18E-07	5,83E-04	2,02E-04
prix_imm	7,17E-07	1,53E-05	4,70E-02
age1*taux1	5,78E-10	3,95E-02	1,46E-08
age5*taux5	8,63E-11	1,11E-01	7,77E-10
taux1*rev1	7,72E-10	2,58E-02	3,00E-08
taux3*rev3	3,17E-10	2,56E-02	1,24E-08

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-1 183 551</i>
<i>Log likelihood zero</i>	<i>-1 288 677</i>
<i>R² ajusté</i>	<i>0,081</i>

Source : Traitement Aissaoui sous R

La comparaison entre les résultats obtenus à l'aide du script R et les résultats implémentés sous Urbansim nous a permis de revoir les fonctions d'optimisation utilisées initialement, jusqu'à l'obtention de valeurs semblables.

Les effets de la majorité des paramètres estimés sont trop faibles (des t_test trop faibles). Cela peut être expliqué par la désagrégation des données. Mais, au-delà des résultats obtenus, il n'est pas certain que la décision de choisir une localisation s'effectue à une échelle aussi fine pour les ménages. Pour cela, nous avons testé les mêmes variables, mais cette fois ci avec une échelle spatiale plus large (la commune).

5.2.2. Le zonage administratif : Commune

Dans le travail réalisé sur la région parisienne, de Palma et al. (2007) mettent bien en avant l'intérêt de cibler la commune comme une échelle d'estimation pour analyser les choix de localisation. Toutefois, dans le cadre de l'aire urbaine de Lyon, le nombre limité de communes (304) et la grande taille des communes de la partie centrale limite la pertinence de ce découpage administratif. L'hétérogénéité des communes centrales est trop grande pour saisir les éventuelles homogénéités de certains quartiers en termes de structure sociodémographique.

Les données caractérisant la zone de localisation sont établies à l'échelle de l'IRIS. En agrégeant ces données, nous avons estimé un modèle à une échelle plus large, la commune. Les résultats obtenus sont les suivants :

Tableau 5.3 : Résultats de l'estimation du modèle de localisation des ménages à l'échelle de la commune

Variables	Coef	Se	t_test
acc_grav_G1	2,15E-04	6,17E-07	3,49E+02
acc_grav_G2	8,94E-05	6,45E-07	1,38E+02
acc_eng	1,94E-07	3,45E-04	5,62E-04
prix_imm	3,97E-06	1,50E-05	2,65E-01
age1*taux1	4,13E-10	1,37E-01	3,02E-09
age5*taux5	5,74E-11	2,36E-01	2,44E-10
taux1*rev1	4,24E-10	6,04E-02	7,02E-09
taux3*rev3	2,75E-10	4,86E-02	5,65E-09

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-885 287</i>

<i>Log likelihood zero</i>	-1 288 677
<i>R² ajusté</i>	0,313

Source : Traitement Aissaoui sous R

La valeur du R^2 obtenue est relativement élevée. Cependant, les effets de la majorité des paramètres estimés sont trop faibles et non statistiquement significatives (des t_{test} trop faible). Cela peut être liée aux biais d'agrégation et plus particulièrement l'utilisation de l'accessibilité gravitaire moyenne au lieu de générer une matrice des temps au niveau de la commune et de recalculer les accessibilités gravitaires. Avant de faire ces calculs, nous voulons nous assurer que ces biais sont liés seulement à la façon d'agréger les variables d'accessibilité. Pour cela, avons enlevé les variables de l'accessibilité gravitaires du modèle (tableau 5.4)

Tableau 5.4 : Résultats de l'estimation du modèle de localisation des ménages à l'échelle de la commune, sans variable d'accessibilité

Variabes	Coef	Se	t_test
prix_imm	4,10E-03	8,62E-06	4,75E+02
age1*taux1	2,35E+01	1,09E-01	2,15E+02
age5*taux5	2,11E+01	1,72E-01	1,23E+02
taux1*rev1	1,87E+01	5,38E-02	3,48E+02
taux3*rev3	-1,94E+00	3,73E-02	-5,20E+01

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-994 672</i>
<i>Log likelihood zero</i>	<i>-1 288 677</i>
<i>R² ajusté</i>	<i>0,228</i>

Source : Traitement Aissaoui sous R

Suite à la suppression des variables de l'accessibilité gravitaire, nous avons obtenus des paramètres statistiquement significatifs. Toutefois, la valeur du R^2 reste relativement élevée vu le nombre des variables intégrées dans le modèle.

En testant seulement la variable du prix de l'immobilier, calculé comme étant la moyenne des prix de l'immobilier au niveau des IRIS, nous avons une valeur de R^2 égale à 0,1.

Tableau 5.5 : Résultats de l'estimation d'un modèle de localisation des ménages à l'échelle de la commune avec seulement le prix de l'immobilier

Variables	Coef	Se	t_test
prix_imm	3,58E-03	7,13E-06	5,02E+02

<i>Nombre d'observations</i>	112 112
<i>Log likelihood</i>	-1 152 834
<i>Log likelihood zero</i>	-1 288 677
<i>R² ajusté</i>	0,105

Source : Traitement Aissaoui sous R

Toutefois, on obtient presque la même valeur, en introduisant quatre variables caractérisant le ménage (tableau 5.6). Les valeurs élevées des R^2 obtenus sont alors dues aux biais d'agrégation.

Tableau 5.6 : Résultats de l'estimation d'un modèle de localisation des ménages à l'échelle de la commune avec seulement les caractéristiques des ménages

Variables	Coef	Se	t_test
age1*taux1	2,81E+01	1,05E-01	2,69E+02
age5*taux5	2,18E+01	1,57E-01	1,39E+02
taux1*rev1	1,43E+01	5,24E-02	2,72E+02
taux3*rev3	6,04E+00	3,43E-02	1,76E+02

<i>Nombre d'observations</i>	112 112
<i>Log likelihood</i>	-1 118 890
<i>Log likelihood zero</i>	-1 288 677
<i>R² ajusté</i>	0,132

Source : Traitement Aissaoui sous R

L'utilisation d'un zonage administratif à un niveau communal donne des valeurs élevées du R^2 et parfois des paramètres statistiquement non significatifs. Ces biais d'estimation sont liés

principalement à la structure de l'aire urbaine de Lyon. En effet, l'hétérogénéité de leurs communes centrales est trop grande pour saisir les éventuelles homogénéités de certains quartiers en termes de structure sociodémographique et limite par conséquent la pertinence de ce découpage.

5.2.3. Le zonage statistique : IRIS

Les IRIS (Ilots Regroupés pour l'Information Statistique) garantissent une meilleure homogénéité intra-zone, prenant en compte le type et la continuité du bâti et les caractéristiques sociodémographiques des populations résidentes. C'est à ce niveau infracommunal que toutes les données sont disponibles. Ainsi, l'aire urbaine de Lyon contient 777 IRIS dont chacun abrite environ 2 000 habitants.

Tableau 5.7 : Résultats de l'estimation du modèle de localisation des ménages à l'échelle de l'IRIS

Variables	Coef	Se	t_test
acc_grav_G1	-2,97E-16	4,03E-07	-7,37E-10
acc_grav_G2	7,58E-16	9,47E-07	8,00E-10
acc_eng	1,73E-18	3,51E-04	4,93E-15
prix_imm	3,86E-17	8,24E-06	4,68E-12
age1*taux1	1,44E-20	3,38E-02	4,27E-19
age5*taux5	2,13E-21	7,68E-02	2,77E-20
taux1*rev1	1,66E-20	2,47E-02	6,71E-19
taux3*rev3	1,64E-20	2,51E-02	6,52E-19

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-1 288 677</i>
<i>Log likelihood zero</i>	<i>-1 288 677</i>
<i>R² ajusté</i>	<i>6,21E-06</i>

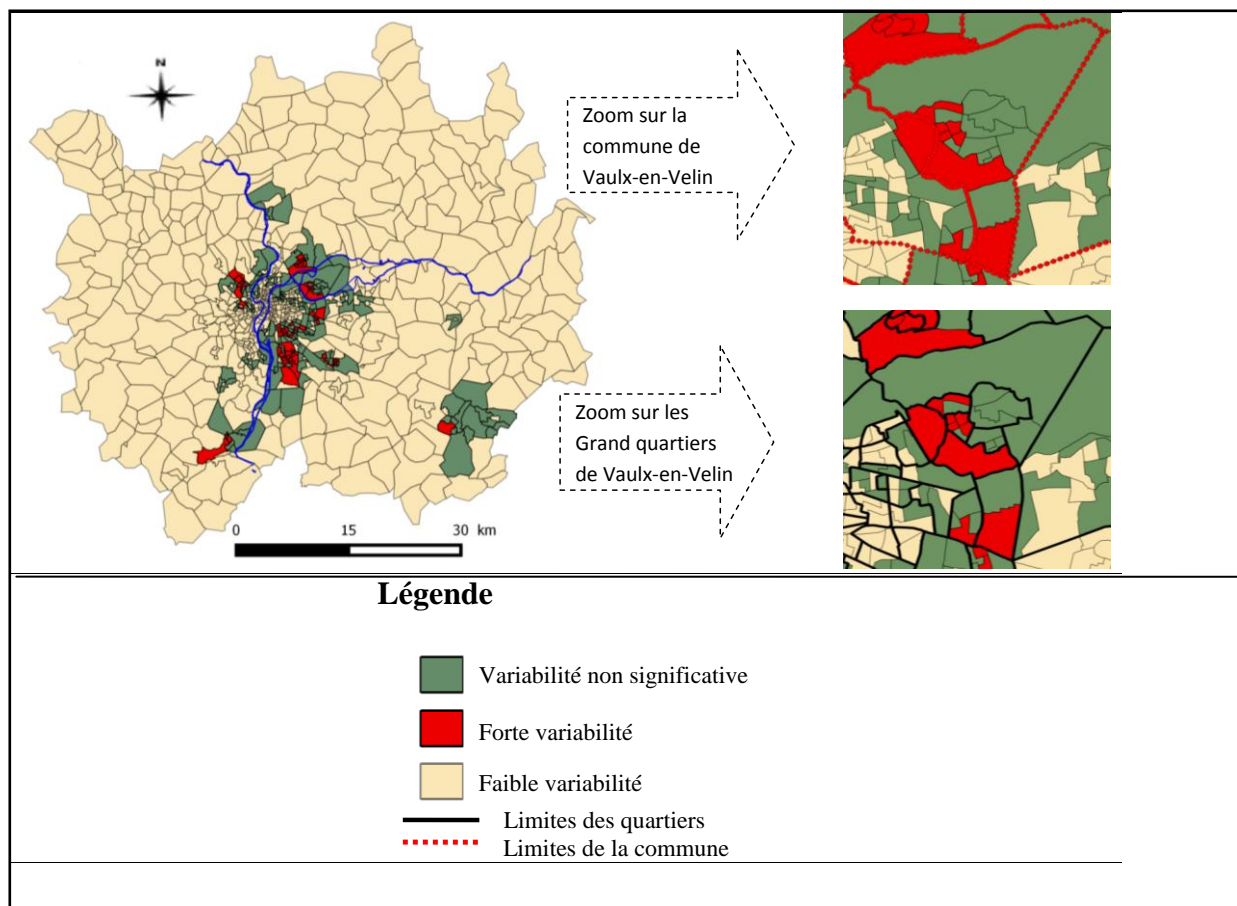
Source : Traitement Aissaoui sous R

De nouveau, ce découpage pose problème lors de l'estimation et donne des paramètres de valeurs trop faibles. Nous avons expliqué cela par la répartition relativement uniforme du nombre de ménages entre ces différents IRIS. Une telle répartition peut donner des probabilités de choix proches et par suite des valeurs très faibles du R².

5.2.4. Le Grand quartier comme alternative intermédiaire

Pour contourner ce problème, nous avons testé un niveau intermédiaire, plus large que l'IRIS et moins grand que la commune. Il s'agit d'utiliser les « Grands quartiers » de l'INSEE, qui sont des IRIS agrégés par contiguïté et selon leur ressemblance. Ils représentent 432 Grands quartiers dont la taille peut aller jusqu'à plusieurs milliers d'habitants dans les zones les plus denses. En se basant sur des critères liés à l'environnement social, et plus particulièrement le taux de logements sociaux et le taux de ménages à faible revenu, nous avons utilisé une analyse spatiale pour tester l'homogénéité de ces Grands quartiers en termes de structure socioprofessionnelle. Cette analyse, nous a montré qu'un tel regroupement permet de constituer des entités géographiques relativement distinctes entre elles et homogènes pour chacune d'elle et respecter ainsi la propriété fondamentale pour l'estimation d'un modèle logit multinomial, à savoir l'indépendance par rapport aux choix non retenus (IIA Independence from Irrelevant Alternatives). La carte (5.1) illustre un exemple des homogénéités saisies par le Grand quartier et non pas par la commune. Il s'agit du résultat obtenu à partir du pourcentage de logements sociaux. D'autres analyses ont été faites, en utilisant la distribution des quintiles de revenu à l'intérieur de chaque IRIS. Bien que ce zonage n'ait pas résolu tous les problèmes d'autocorrélation spatiale puisqu'on observe que certains Grands quartiers regroupent des IRIS non homogènes, il permet de traiter partiellement le problème.

Carte 5.1 : Un exemple de l'analyse spatiale illustrant la différence entre les différents zonages testés



Source : Traitement à partir des caractéristiques des IRIS sous Geoda et superposition avec les différents zonages (commune, Grand quartier) sous QGIS.

Analyse empirique des choix de localisation : Données et échelles

Grâce à notre travail d'exploitation de plusieurs sources de données en fonction de leur échelle de disponibilité, nous avons pu constituer des variables qui peuvent mesurer les effets relatifs aux quatre dimensions de choix de localisation mis en avant par les travaux empiriques traitant les choix de localisation (chapitre 2). L'accessibilité et le prix de l'immobilier ont été calculés pour définir l'effet de l'arbitrage du marché. La constitution des variables indiquant la présence des services (banque, poste, supermarché, etc), des établissements de scolarité, et des équipements (culturels, sportifs, etc) dans la zone permet de mesurer les effets liés aux aménités spatiales. La concentration des logements sociaux et des

ménages de la même classe socioprofessionnelle ou du même revenu dans la zone est construite pour caractériser l'environnement social. Plusieurs variables caractérisant le ménage sont aussi rassemblées afin de mesurer l'effet du cycle de vie dans les choix de localisation. L'annexe 15 résume l'ensemble des variables construites et leur description.

En se basant sur des méthodes itératives, ce travail de rassemblement de données en fonction de leur échelle de disponibilité était complété par l'estimation d'une variable clé dans les choix de localisation, le revenu. Nous avons affecté à chaque ménage un quintile de revenu afin de pouvoir mesurer directement l'effet de cette variable dans les arbitrages des ménages.

La deuxième question traitée dans ce chapitre concerne la définition du zonage à utiliser pour estimer les choix de localisation. Les travaux faits jusqu'à maintenant pour définir le zonage n'ont pas apporté de solutions valables pour tous les contextes urbains. A travers ce chapitre, nous avons testé plusieurs zonages, du plus fin (carroyage) au plus grand (la commune) et nous avons mis en avant l'importance d'utiliser le Grand quartier de l'INSEE pour modéliser les choix de localisation de l'aire urbaine de Lyon. Il s'agit d'une échelle intermédiaire entre l'IRIS et la commune qui permet de constituer des entités géographiques relativement distinctes entre elles et homogènes pour chacune d'elle et ainsi de respecter au moins en partie la propriété d'indépendance par rapport aux choix non retenus fondamentale pour l'estimation d'un modèle logit multinomial. Ce choix est justifié à travers les résultats des tests empiriques conduits et les analyses spatiales faites pour comparer les différentes échelles.

Chapitre 6 - Analyse empirique des choix de localisation des ménages : un modèle logit Multinomial

En s'appuyant sur la revue synthétisée des déterminants de choix de localisation résidentielle présentée au chapitre 2 (cf. Tableaux 2.1 à 2.3), l'objectif de ce chapitre est d'utiliser un modèle logit multinomial pour tester empiriquement les déterminants du choix de localisation des ménages de l'aire urbaine de Lyon en 1999.

En retenant le Grand quartier comme échelle pour analyser ces choix, nous avons exploité plusieurs sources de données pour produire deux bases importantes. La première rassemble les variables caractérisant les Grands quartiers et la deuxième regroupe celles décrivant les ménages. Ces deux bases, nous ont permis de tester l'effet de l'ensemble des facteurs qui peuvent influencer la décision des ménages en matière de choix d'un quartier de résidence et d'enrichir le corpus des travaux sur les déterminants de choix de localisation résidentielle.

L'utilisation de l'ensemble des Grands quartiers de l'aire urbaine de Lyon, en prenant le centre (la préfecture) comme référence pour modéliser les choix de localisation, multiplie les paramètres à estimer et complexifie l'interprétation des résultats et l'identification des déterminants de choix de localisation. Pour cela, nous avons opté pour l'utilisation d'un échantillon d'alternatives.

Dans la première section (6.1), nous présentons en détail le processus de l'estimation d'un modèle logit multinomial avec un échantillon d'alternatives. La deuxième section présente les critères de l'évaluation de la qualité du modèle. La troisième section (6.3) aborde les différentes hypothèses testées pour sélectionner les variables explicatives à intégrer dans la fonction d'utilité. La dernière section (6.4) présente les résultats et leur interprétation.

6.1. Processus de l'estimation

Les modèles de choix discrets (logit multinomial et logit emboîté) avec un échantillon d'alternatives ont été retenus pour mesurer l'effet de chaque déterminant de choix de localisation et tester nos différentes hypothèses.

Le processus d'estimation commence par tirer aléatoirement et pour chaque ménage un échantillon des alternatives (X_{j-1}) parmi celles non choisies. L'ajout de l'alternative choisie donne un ensemble de choix (D_n) composé de X_j alternatives au total. Une fois l'ensemble de choix défini pour chaque ménage, on passe au calcul de l'utilité de chacune des alternatives de cet ensemble et de leur probabilité d'être choisie. On définit ensuite une matrice binaire des choix égale 1 si l'alternative est choisie et 0 dans le cas contraire (x_{Ch}) et on finit par maximiser la fonction de log vraisemblance. Une fois atteint l'optimum, on passe au calcul des paramètres liés aux variables explicatives ($\beta_1, \beta_2, \dots, \beta_m$) et des indicateurs d'évaluation du modèle et on trace le graphique qui permet d'illustrer l'effet de chaque alternative. La figure 6.1 résume le processus d'estimation qui permet la programmation d'un modèle logit multinomial avec échantillonnage d'alternatives.

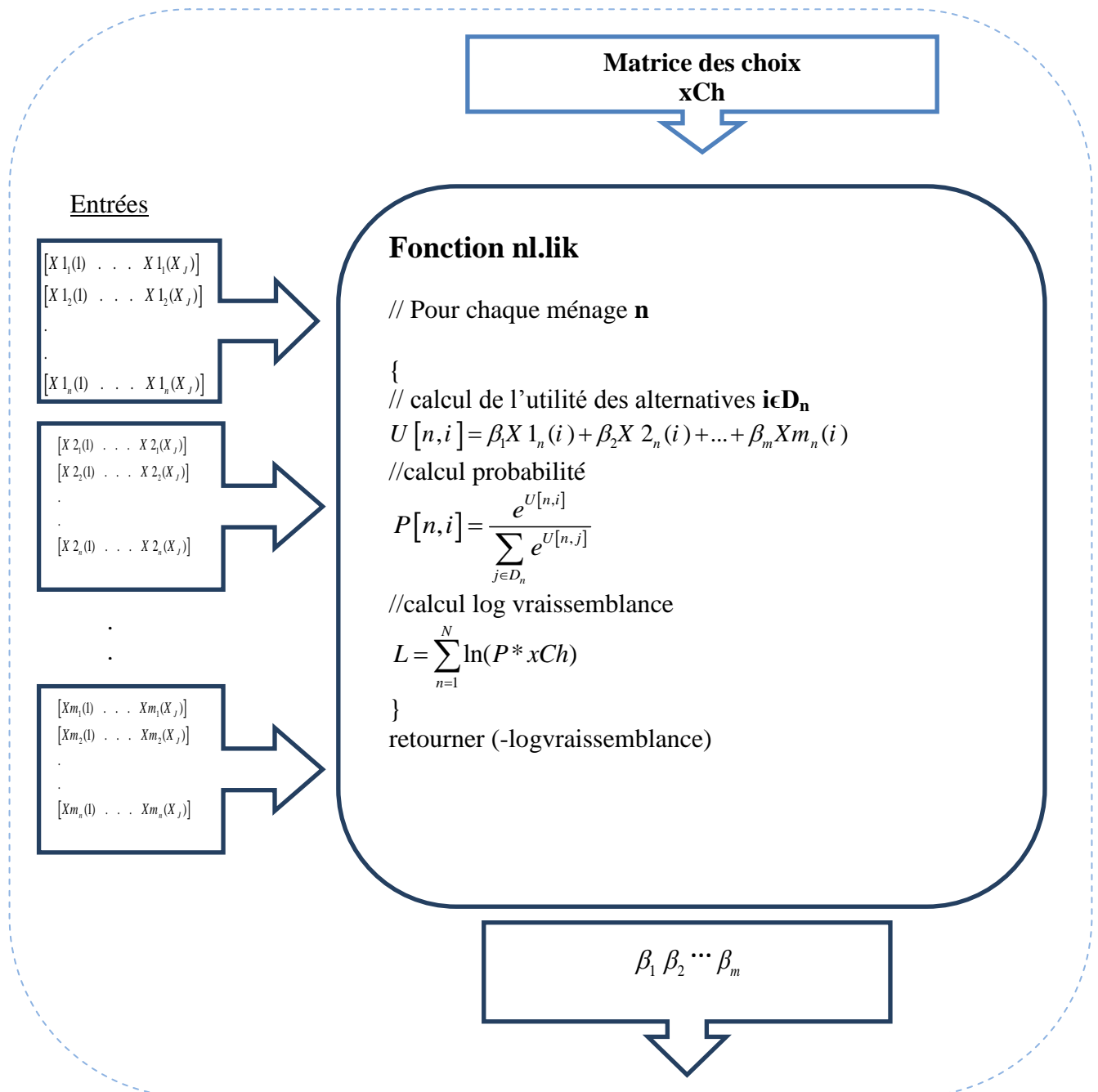
Le module de localisation intégré dans SIMBAD permet de faire ces estimations. Toutefois, l'interaction de ce module avec d'autres modules a été souvent critiquée comme étant semblable à une boîte noire, et dont l'utilisation opérationnelle n'est pas encore claire. Pour cela, nous avons choisi dès le départ de concevoir nos modèles indépendamment de la plateforme Urbansim.

Les logiciels commerciaux (Stata, Sas, Xlstat, Gauss, alogit, etc) se dotent de commandes préprogrammées pour analyser des choix avec peu d'alternatives. Ces commandes n'estiment pas le modèle logit (multinomial ou emboîté) avec un échantillon d'alternatives et ne traitent pas les bases de données volumineuses. Toutefois, ils permettent à l'utilisateur de programmer son propre modèle à l'aide des macros et d'augmenter la mémoire utilisée. Le programme développé par Bierlaire (Biogème) permet aussi certaines flexibilités. Mais, l'échantillonnage d'alternatives se fait en amont de l'estimation.

Dans cette thèse, nous avons choisi de travailler sous R. Il s'agit d'un logiciel libre avec plusieurs bibliothèques qui nous permet de faire à la fois la gestion des données, les traitements statistiques, les analyses spatiales et la production des cartes. Il a aussi l'avantage d'utiliser l'espace mémoire de l'ordinateur, et de traiter de grosses bases de données. Des travaux récents montrent la faisabilité de l'utilisation parallèle de la mémoire de plusieurs

ordinateurs, ce qui va nous permettre dans nos travaux futurs de réduire les temps de calculs. Tous ces avantages n'impliquent pas forcément les limites de mémoire ou de capacité des autres langages de programmation (python par exemple). Ces derniers peuvent également être utilisés pour programmer le processus de modélisation.

Figure 6.1 : Processus d'estimation d'un MNL avec un échantillon d'alternatives



Source : Élaboration propre

6.2. Les critères de l'évaluation de la qualité du modèle

Les tests de significativité des coefficients servent à indiquer l'apport d'une ou de plusieurs variables explicatives dans l'explication de la variable dépendante. En passant par des hypothèses de nullité d'une ou de plusieurs variables indépendantes, ces tests aident le modélisateur à évaluer l'apport de certaines variables dans la fonction d'utilité. Ils peuvent être présentés sous deux formes. Les deux sous-sections (6.2.1 et 6.2.2) présentent la formulation de ces deux formes. Nous poursuivons par l'interprétation des signes des paramètres estimés et le calcul des élasticités (6.2.3). Et nous finissons par la présentation du test de la propriété d'indépendance par rapport aux choix non retenus (sous-section 6.2.4)

6.2.1. Test de Wald

En se basant sur la normalité asymptotique des estimateurs, ce test permet de tester la nullité des paramètres estimés. Il s'agit de calculer la matrice des dérivées partielles, dite matrice Hessienne. L'inverse de cette matrice correspond à la matrice des variance-covariance. A partir de cette matrice, on peut calculer l'erreur standard et le t-test. Ce test est le plus utilisé par les logiciels.

6.2.2. Ratio de vraisemblance

Ce test est moins utilisé par les logiciels parce qu'il demande des temps élevés de calcul. Il se base sur la comparaison des valeurs prises par le logarithme de la vraisemblance selon que le modèle est contraint ou non contraint. Il est calculé comme suit :

$$-2*(LL_0 - LL) \quad -6.1-$$

Il suit asymptotiquement la loi du χ^2 (avec \mathbf{K} degré de libertés, \mathbf{K} étant le nombre de variables explicatives ajouté dans le modèle (LL) par rapport au modèle de base (LL₀).

Pour tester la significativité d'une variable indépendante. Il suffit d'estimer un modèle avec et sans la variable en question et de calculer ensuite la différence entre les deux valeurs de la fonction du log vraisemblance obtenues selon l'équation 6.1.

En estimant un modèle avec des coefficients nuls à l'exception des constantes et un autre avec les paramètres estimés, ce test permet également de tester la nullité globale des coefficients.

6.2.3. L'élasticité

Il s'agit de l'effet marginal d'une variable explicative (X) sur la variable à expliquer

(quartier). Elle est égale à : $E_x^{quartier} = \beta(1 - P_n(quartier))X$ -6.2-

L'interprétation est différente selon que la variable est simple ou croisée. Dans le cas des variables simples, qui représentent seulement les caractéristiques de la zone de localisation, un signe négatif signifie un effet négatif de la variable explicative sur l'utilité de choisir une zone de localisation. Les variables croisées articulent une modalité d'une variable caractéristique du ménage avec une caractéristique du quartier. Le croisement signifie que la variable caractéristique du quartier s'applique qu'aux ménages appartenant au segment de la modalité considérée. Un signe négatif indique un effet négatif d'une telle caractéristique du quartier sur l'utilité du ménage d'un tel groupe à choisir ce quartier.

Si les signes de notre modèle GEV (multinomial ou emboîté) avec un échantillon d'alternatives sont directement interprétables à partir des élasticités (cf. Équation 6.2), les paramètres estimés nécessitent un calcul moins simple des élasticités.

Lee et al. (2010a) proposent un calcul de la différence des utilités pour quantifier l'effet de chaque variable explicative. Il s'agit de calculer deux utilités pour chaque variable. La première correspond au cinquième percentile et la deuxième au 95ème percentile de la valeur de la variable en question. La différence de ces deux utilités nous permet de mesurer et de comparer le poids de chaque facteur par rapport aux autres variables (figure 6.2, section 6.3.1). Dans ce travail, nous avons choisi de passer par ce calcul de différence des utilités pour mesurer l'effet de chaque variable sur le choix de localisation résidentielle des ménages.

6.2.4. Test de la propriété d'IIA

Ce test consiste à retirer une alternative de l'ensemble des choix et de comparer les paramètres obtenus avec ceux du modèle complet. Hausman et McFadden (1984) fournissent

une statistique pour valider ou rejeter l'hypothèse selon laquelle les paramètres ne changent pas significativement (IIA est valide). Il s'agit d'un test khi 2 avec K degrés de liberté où K est le nombre des paramètres du modèle restreint.

6.3. Estimation des modèles en 1999

En se basant sur les données disponibles sur l'aire urbaine de Lyon, le premier chapitre de cette partie (chapitre 5) nous a permis de définir la majorité des variables identifiées par les travaux théoriques ou empiriques comme déterminants du choix de localisation. A travers cette section, nous cherchons à tester ces variables à l'aide d'un modèle logit multinomial avec échantillon d'alternatives. Ce modèle de choix discrets ne permet pas de prédire la zone choisie par le ménage. Toutefois, il donne la probabilité de choix des différentes zones et ainsi fournit la répartition des ménages par zone de localisation. Cela permet d'anticiper la composition spatiale des quartiers afin d'orienter les politiques publiques en fonction des objectifs recherchés.

La première question qui se pose à ce niveau est : quelle est la méthode à utiliser pour sélectionner les variables explicatives à tester ? Cela peut être fait d'une manière automatisée ou manuelle. La première solution est purement statistique. Elle consiste à intégrer l'ensemble des variables explicatives dans le modèle et à enlever progressivement les variables non significatives, la méthode stepwise. Il est également possible d'ajouter les variables une à une et de déterminer un critère d'arrêt, la méthode forward. L'inconvénient de ces deux méthodes automatisées est qu'elles sont variantes d'un échantillon à l'autre et d'un critère à l'autre. Parfois, les optimums obtenus peuvent être juste les résultats des fluctuations aléatoires dans les données (Rakotomalala, 2014).

La perspective modélisatrice de ce travail est de comprendre et de prévoir les choix de localisation. Pour cela, nous avons opté pour la deuxième solution et nous avons sélectionné les variables à tester de manière manuelle. A partir d'un ensemble d'hypothèses, notre idée est d'établir une liste de l'ensemble des variables liées aux différents arbitrages faits par les ménages en matière de choix d'une zone de localisation et de mesurer ensuite leurs poids.

- Arbitrage du marché

Partant du principe du modèle classique d'Alonso, le coût de déplacement domicile-travail et le prix de l'immobilier sont les premiers indicateurs mis en avant par ce modèle. Le premier arbitrage à tester est lié à la contrainte du marché et plus particulièrement la présence d'une proportionnalité entre le prix de l'immobilier et l'accès à l'emploi. Devant l'absence d'une base de données qui indique la zone de localisation du lieu du travail de la personne de référence du ménage. Nous avons choisi de passer par des variables qui nous permettent de mesurer l'effet de la proximité aux emplois sur les choix de localisation des ménages.

En se basant sur des mesures gravitaires, nous avons testé l'accessibilité à l'emploi. Cette variable est calculée de plusieurs façons, en utilisant le temps TC et le temps VP.

Tableau 6.1 : Résultats de l'estimation du modèle de localisation des ménages, avec seulement les variables relatives à l'arbitrage du marché

Variables	Coef	Se	t_test
acc_vp_g1	0,00003	0,000002	19,0
acc_vp_g2	0,00005	0,000001	37,2
acc_tc_g1	0,00002	0,000004	4,4
acc_tc_g2	-0,00001	0,000004	-1,9
prix_imm	0,00001	0,000017	0,4

<i>Nombre d'observations</i>	112 112
<i>Log likelihood</i>	-179 376
<i>Log likelihood zero</i>	-218 160
<i>R² ajusté</i>	0,178

Source : Traitement Aissaoui sous R

En faisant la distinction entre deux groupes d'emploi (voir annexe 1), nous avons trouvé que le paramètre associé à l'accessibilité gravitaire à l'emploi du groupe 2 en transport en commun (acc_tc_g2) a un signe négatif. Cela signifie que l'utilité des ménages de se localiser dans une zone diminue quand l'accessibilité gravitaire aux emplois du groupe 2 augmente. Or, les emplois de ce groupe se localisent en périphérie, cela est dû plutôt à la désutilité de se localiser en périphérie pour les ménages utilisant les transports en commun. Le prix de l'immobilier a aussi un signe non intuitif. Toutefois, ces deux paramètres (acc_tc_g2 et prix_imm) sont statistiquement non significatifs. Nous avons testé dans un deuxième temps un modèle avec le prix de l'immobilier et l'accessibilité gravitaire à tous les types d'emplois en TC et en VP.

Tableau 6.2 : Résultats de l'estimation du modèle de localisation des ménages, avec les accessibilités gravitaires TC et VP et le prix de l'immobilier

Variables	Coef	Se	t_test
ac_g_VP	0,0207	0,00015	138,2
ac_g_TC	-0,0037	0,00053	-7,0
prix_imm	-0,0003	0,00002	-16,8

<i>Nombre d'observations</i>	112 112
<i>Log likelihood</i>	-179 603
<i>Log likelihood zero</i>	-218160
<i>R² ajusté</i>	0,177

Source : Traitement Aissaoui sous R

Les résultats obtenus montrent que les accessibilités gravitaires à l'emploi en VP ont un effet positif sur les choix de localisation. En enlevant la variable non significative, nous avons obtenus les résultats suivants :

Tableau 6.3 : Résultats de l'estimation du modèle de localisation des ménages, avec l'accessibilité gravitaire VP et le prix de l'immobilier

Variables	Coef	Se	t_test
ac_g_VP	0,0199	0,00009	214,0
prix_imm	-0,0003	0,00002	-17,0

<i>Nombre d'observations</i>	112 112
<i>Log likelihood</i>	-179 626
<i>Log likelihood zero</i>	-218160
<i>R² ajusté</i>	0,177

Source : Traitement Aissaoui sous R

La non significativité de l'accessibilité gravitaire à l'emploi en TC peut être expliquée par la présence de corrélations entre les deux accessibilités. Nous les avons testés séparément.

Tableau 6.4 : Résultats de l'estimation du modèle de localisation des ménages, avec l'accessibilité gravitaire TC et le prix de l'immobilier

Variables	Coef	Se	t_test
ac_g_TC	0,0563	0,00033	172,2
prix_imm	0,0004	0,00002	23,4

<i>Nombre d'observations</i>	112 112
<i>Log likelihood</i>	-189 087
<i>Log likelihood zero</i>	-218160
<i>R² ajusté</i>	0,133

Source : *Traitement Aissaoui sous R*

Les résultats obtenus en intégrant l'accessibilité gravitaire à l'emploi en TC donne des résultats avec des R^2 moins élevés et le prix de l'immobilier a changé de signe. Cela est expliqué par le fait que le prix de l'immobilier capte aussi l'effet de l'accessibilité aux transports en commun.

En intégrant seulement les variables liées aux arbitrages du marché, nous avons obtenu un R^2 relativement élevé (tableau 6.3). Ce résultat confirme l'importance de l'arbitrage entre le prix de l'immobilier et l'accessibilité à l'emploi dans les choix de localisation, même si l'ajout d'autres variables permet d'améliorer légèrement le modèle.

- Aménité spatiale

En suivant le même ordre chronologique de développement du modèle standard, nous avons introduit des variables liées aux aménités spatiales pour tester l'effet de cette composante sur les choix de localisation. Ces aménités peuvent être naturelles ou non (historiques, modernes, services et équipements de base...). Dans ce travail, nous avons donné plus d'importance à la mesure des aménités non naturelles et plus particulièrement celles qui varient à l'échelle intra urbaine. La proximité aux transports publics était le premier équipement de base que nous avons testé. Deux variables ont été testées. La première indique le nombre d'arrêts bien desservis dans la zone. Nous avons testé également le nombre de stations de métro à proximité. Ces variables ont été transformées ensuite en des variables binaires pour tester seulement l'effet de la présence d'une station du métro et/ou un arrêt de bus bien desservi dans le Grand quartier.

En se basant sur l'activité principale de l'établissement, la base SIRENE nous a permis de définir avec précision le nombre d'établissements primaires (crèche et école), secondaires (collège et lycée) au sein du Grand quartier. Toutefois, nous n'avons pas pu identifier d'indicateurs qui indiquent la qualité de ces écoles. Ce calcul n'est pas aisé car la qualité réelle ou perçue d'un établissement ne dépend pas uniquement du niveau de dépenses, proxy souvent mis en avant dans les travaux anglo-saxons, mais surtout de la composition sociale du périmètre de la carte scolaire. La proximité aux espaces verts avec jeux pour les enfants est aussi testée pour mesurer la préférence des ménages avec enfant à ce type d'aménité.

Dans un premier modèle, nous avons intégré toutes les variables qui peuvent être considérées comme aménité positive ou négative. Le pourcentage de logements sociaux est intégré dans ce modèle comme une aménité négative. Le croisement de cette variable avec les classes de revenu des ménages s'utilise plutôt pour mesurer l'entre soi.

Tableau 6.5 : Résultats de l'estimation du modèle de localisation des ménages, avec les variables relatives aux aménités spatiales

Variable	Coef	Se	t_test
prox_metro	0,5402	0,01265	42,7
prox_bus	0,1730	0,00810	21,4
café_tabac	0,0551	0,00763	7,2
Boulang	-0,0091	0,00217	-4,2
supermarché	-0,0281	0,00602	-4,7
Magasin	0,0122	0,00601	2,0
serv_part	0,0227	0,00033	69,6
serv_santé	0,0004	0,00027	1,4
banque_poste	-0,0192	0,00126	-15,3
etb_prim	0,0331	0,00115	28,8
etb_sec	0,0579	0,00392	14,8
Ciné	-0,2797	0,00807	-34,7
resto_trad	-0,0346	0,00078	-44,3
rest_rapide	-0,0356	0,00155	-22,9
Disco	0,0259	0,00420	6,2
Biblio	0,1593	0,01453	11,0
autre_cult	-0,1114	0,00946	-11,8
salle_sport	0,0575	0,00602	9,6
spot_autre	0,0034	0,00127	2,6
acc_jeux	-0,0455	0,01001	-4,5
Hlm	1,4586	0,02029	71,9

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	-169743
<i>Log likelihood zero</i>	-218160
<i>R² ajusté</i>	0,222

Source : Traitement Aissaoui sous R

Le nombre de variables relatives aux aménités spatiales est relativement élevé, nous avons construit deux nouvelles variables (serv_base1 et serv_base2). La première variable vaut 1 s'il y a un bureau de tabac (café_tabac) ou une boulangerie (boulang) dans la zone et 0 dans le cas contraire. La deuxième variable prend la valeur 1 s'il y a un supermarché ou un magasin dans la zone et 0 sinon. Ce deuxième modèle n'est pas contraint par rapport au premier et par suite le test de Log-vraisemblance ne peut pas être fait pour rejeter ou garder le modèle. La comparaison des deux modèles s'est limitée à la comparaison des R².

Tableau 6.6 : Résultats de l'estimation du modèle de localisation des ménages avec les variables relatives aux aménités spatiales, en remplaçant les services de bases par des indicateurs agrégés

Variables	Coef	Se	t_test
prox_metro	0,5345	0,01185	45,1
prox_bus	0,1048	0,00794	13,2
serv_base1	0,6042	0,01538	39,3
serv_base2	0,4459	0,01334	33,4
serv_part	0,0194	0,00029	67,2
serv_santé	-0,0007	0,00026	-2,8
banque_poste	-0,0393	0,00115	-34,2
etb_prim	0,0261	0,00111	23,6
etb_sec	0,0388	0,00366	10,6
equ_recr	-0,0290	0,00055	-53,1
equ_cult	0,0195	0,00691	2,8
Sport	0,0045	0,00119	3,8
acc_jeux	-0,1073	0,00971	-11,0
Hlm	1,2725	0,02062	61,7

<i>Nombre d'observations</i>	<i>112 112</i>
Log likelihood	-168 410
Log likelihood zero	-218 160
R ² ajusté	0,228

Source : Traitement Aissaoui sous R

La valeur du R^2 est légèrement plus grande. Nous avons gardé ce deuxième modèle. Nous avons enlevé ensuite les variables moins significatives. Les résultats obtenus sont résumés dans le tableau suivant :

Tableau 6.7 : Résultats de l'estimation du modèle de localisation des ménages avec les variables relatives aux aménités spatiales, en remplaçant les services de bases par des indicateurs agrégés et en enlevant les variables moins significatives

Variabiles	Coef	Se	t_test
prox_metro	0,5302	0,01158	45,8
prox_bus	0,1084	0,00787	13,8
serv_base1	0,6086	0,01530	39,8
serv_base2	0,4495	0,01331	33,8
serv_part	0,0190	0,00022	87,5
banque_poste	-0,0398	0,00106	-37,5
etb_prim	0,0282	0,00104	27,1
etb_sec	0,0386	0,00348	11,1
equ_recr	-0,0282	0,00043	-66,1
acc_jeux	-0,1019	0,00957	-10,6
Hlm	1,2421	0,02010	61,8

<i>Nombre d'observations</i>	<i>112 112</i>
Log likelihood	-168430
Log likelihood zero	-218160
R^2 ajusté	0,228

Source : Traitement Aissaoui sous R

Certaines variables n'ont pas les signes attendus. En effet, l'omission des variables caractérisant les ménages dans ce modèle fait que les variables (Hlm et acc_jeux) qui dépendent de certaines caractéristiques des ménages comme le revenu et la présence des enfants sont corrélées aux termes d'erreurs et crée des problèmes d'endogénéité. En enlevant ces variables, nous avons obtenu les résultats suivants :

Tableau 6.8 : Résultats de l'estimation du modèle de localisation des ménages avec les variables relatives aux aménités spatiales retenues

Variable	Coef	Se	t_test
prox_metro	0,7055	0,01116	63,2
prox_bus	0,1640	0,00778	21,1
serv_base1	0,6960	0,01503	46,3
serv_base2	0,5821	0,01290	45,1
serv_part	0,0045	0,00008	56,0
etb_prim	0,0741	0,00091	81,8
etb_sec	0,0549	0,00354	15,5

<i>Nombre d'observations</i>	<i>112 112</i>
Log likelihood	-172 293
Log likelihood zero	-218 160
R ² ajusté	0,210

Source : Traitement Aissaoui sous R

Ce dernier modèle affiche des paramètres significatifs et avec le bon signe.

L'intégration des variables liées aux aménités spatiales donne un R² plus élevé que celui d'un modèle avec des variables relatives aux contraintes du marché.

- Environnement social

Le croisement du revenu du ménage avec le taux de ménages appartenant à la même classe de revenu dans le Grand quartier permet de mesurer la présence d'une stratification en fonction du revenu. Cependant, cette stratification peut être attribuée à une discrimination entre les riches et les pauvres sur le marché de l'immobilier ou à la différence selon les classes de revenu des préférences pour les aménités spatiales. Pour tester la première hypothèse, nous avons intégré des variables croisées pour mesurer l'effet combiné du revenu et du taux de logements sociaux dans la zone de localisation. Ces indicateurs permettent de tester à la fois l'effet de la concentration de l'habitat social sur les ménages à haut et faible revenu et le rôle des politiques de logement.

Nous avons intégré également d'autres variables qui mesurent la stratification des ménages en fonction de la taille du ménage et de l'âge de la personne de référence du ménage.

Tableau 6.9 : Résultats de l'estimation du modèle de localisation des ménages, avec les variables relatives aux aménités sociales

Variables	Coef	Se	t_test
%hlm*rev1	2,1753	0,06016	36,2
%hlm*rev2	2,3684	0,04515	52,5
%hlm*rev3	1,7511	0,03564	49,1
%hlm*rev4	2,3583	0,05237	45,0
%hlm*rev5	2,1086	0,06075	34,7
taux1*rev1	1,5786	0,11915	13,2
taux2*rev2	-3,6406	0,23728	-15,3
taux3*rev3	8,1748	0,19154	42,7
taux4*rev4	10,7889	0,25640	42,1
taux5*rev5	6,8642	0,08825	77,8
age1*taux1	7,9189	0,11063	71,6
age2*taux2	9,7977	0,34199	28,6
age3*taux3	36,4312	1,15591	31,5
age4*taux4	63,2347	3,08567	20,5
age5*taux5	227,1505	8,20300	27,7
taux_person1*person1	12,4755	0,11094	112,4
taux_person2*person2	32,0061	0,40946	78,2
taux_person3*person3	12,5711	0,94071	13,4
taux_person4*person4	4,4919	1,10148	4,1
taux_person5*person5	35,8499	2,22994	16,1

<i>Nombre d'observations</i>	112 112
Log likelihood	-166 079
Log likelihood zero	-218 160
R ² ajusté	0,239

Source : Traitement Aissaoui sous R

Les paramètres obtenus sont tous significatifs. Toutefois, la tendance des ménages à se localiser dans des zones peuplées par des ménages de la même taille ou du même âge que la personne de référence peut être le résultat de la préférence des ménages ayant des caractéristiques semblables pour les mêmes aménités.

Partant de l'hypothèse que les ménages d'une taille supérieure à 3 auront les mêmes préférences en matière de choix d'une localisation. Nous avons testé un deuxième modèle en regroupant les ménages d'une taille de 3, 4 et 5 personnes dans une seule variable (person_sup3) que nous avons croisé avec le taux de ménages d'une taille supérieure à 3 dans la zone (taux_sup3).

Tableau 6.10 : Résultats de l'estimation du modèle de localisation des ménages avec les variables relatives aux aménités sociales, en regroupant les ménages d'une taille supérieure ou égale à 3 dans une seule variable

Variables	Coef	Se	t_test
%hlm*rev1	2,1307	0,06035	35,3
%hlm*rev2	2,3512	0,04506	52,2
%hlm*rev3	1,7494	0,03561	49,1
%hlm*rev4	2,3387	0,05239	44,6
%hlm*rev5	2,0862	0,06071	34,4
taux1*rev1	1,7060	0,11927	14,3
taux2*rev2	-3,6561	0,23703	-15,4
taux3*rev3	8,1435	0,19161	42,5
taux4*rev4	10,6256	0,25621	41,5
taux5*rev5	6,8142	0,08816	77,3
age1*taux1	7,9551	0,11070	71,9
age2*taux2	9,4734	0,34359	27,6
age3*taux3	37,1790	1,15479	32,2
age4*taux4	63,5227	3,08515	20,6
age5*taux5	227,3815	8,20101	27,7
taux_person1*person1	12,4443	0,11089	112,2
taux_person2*person2	31,9496	0,40922	78,1
taux_sup3*person_sup3	6,2528	0,32005	19,5

<i>Nombre d'observations</i>	112 112
Log likelihood	-166 101
Log likelihood zero	-218 160
R ² ajusté	0,239

Source : Traitement Aissaoui sous R

Les résultats obtenus donnent les mêmes valeurs de R². En prenant uniquement les classes extrêmes de l'âge de la personne de référence du ménage (age1, age5), nous avons testé un troisième modèle pour cette catégorie de variables.

Tableau 6.11 : Résultats de l'estimation du modèle de localisation des ménages avec les variables relatives aux aménités sociales retenues

Variables	Coef	Se	t_test
%hlm*rev1	2,1216	0,06044	35,1
%hlm*rev2	2,4784	0,04477	55,4

%hlm*rev3	1,8456	0,03542	52,1
%hlm*rev4	2,4145	0,05215	46,3
%hlm*rev5	2,2662	0,06030	37,6
taux1*rev1	1,8975	0,11941	15,9
taux2*rev2	-4,2327	0,23787	-17,8
taux3*rev3	7,9187	0,19190	41,3
taux4*rev4	10,4272	0,25624	40,7
taux5*rev5	7,0080	0,08780	79,8
age1*taux1	7,5500	0,11092	68,1
age5*taux5	227,6522	8,27626	27,5
taux_person1*person1	12,9226	0,11069	116,7
taux_person2*person2	34,5459	0,40379	85,6
taux_sup3*person_sup3	10,0330	0,30529	32,9

<i>Nombre d'observations</i>	112 112
Log likelihood	-167 119
Log likelihood zero	-218 160
R ² ajusté	0,235

Source : Traitement Aissaoui sous R

Les résultats montrent une légère baisse du R². Toutefois, les R² obtenus à partir des variables mesurant l'entre soi reste les plus élevés par rapport à ceux obtenus précédemment, en utilisant des variables du marché ou des aménités. A partir de ces résultats, on peut prévoir que les variables de l'environnement social auront un poids élevé dans les arbitrages des ménages en matière de choix d'une zone de localisation.

- Caractéristiques des ménages

Le croisement des caractéristiques des ménages (type d'activité, taille) avec la zone de localisation (centrale ou périphérique), nous a permis de tester l'effet du cycle de vie sur le choix d'une localisation centrale ou périurbaine. Pendant les études, les ménages préfèrent une localisation centrale pour être proche des universités et bénéficier des aménités du centre.

Tableau 6.12 : Résultats de l'estimation du modèle de localisation des ménages, avec certaines caractéristiques de la zone et la catégorie socioprofessionnelle du ménage

Variabes	Coef	Se	t_test
prox_metro	0,1964	0,01173	16,8
serv_base1	0,1313	0,00158	83,1
serv_base2	0,0205	0,00133	15,5
centre*Etud	3,2277	0,03468	93,1

<i>Nombre d'observations</i>	112 112
Log likelihood	-180 932
Log likelihood zero	-218 160
R ² ajusté	0,170

Source : Traitement Aissaoui sous R

Avec l'arrivée des enfants et l'élargissement du ménage, ils optent plutôt pour un logement en périphérie pour avoir plus de surface.

Tableau 6.13 : Résultats de l'estimation du modèle de localisation des ménages, avec certaines caractéristiques de la zone et la taille du ménage

Variabes	Coef	Se	t_test
prox_metro	0,0238	0,01175	2,0
serv_base1	0,1174	0,00158	74,4
serv_base2	0,0188	0,00131	14,3
retseurb*person5	-0,6740	0,03005	-22,4
centre*person1	1,7229	0,01235	139,5

<i>Nombre d'observations</i>	112 112
Log likelihood	-178 888
Log likelihood zero	-218 160
R ² ajusté	0,180

Source : Traitement Aissaoui sous R

L'information n'étant pas disponible pour 1999, nous n'avons pas pu introduire les variables du type de logement (maison/appartement) et de ses caractéristiques intrinsèques (surface, confort). Ce sont globalement ces facteurs, combinés à une offre réduite dans le centre (Cornuel, 2010), qui incitent à la périurbanisation dans une optique d'accès à la propriété. En partant de l'hypothèse que les maisons se concentrent dans les périphéries et les appartements au centre. Nous nous sommes limités à la mesure de l'effet combiné de la zone de localisation et le statut d'occupation (propriétaire/locataire).

Tableau 6.14 : Résultats de l'estimation du modèle de localisation des ménages, avec certaines caractéristiques de la zone et le statut d'occupation du logement

Variabes	Coef	Se	t_test
prox_metro	0,0037	0,01174	0,3
serv_base1	0,1139	0,00157	72,5

serv_base2	0,0190	0,00130	14,7
resturb *propriétaire	0,3980	0,01589	25,1
centre*loc	1,2578	0,00964	130,5

<i>Nombre d'observations</i>	112 112
Log likelihood	-180 887
Log likelihood zero	-218 160
R ² ajusté	0,170

Source : Traitement Aissaoui sous R

En intégrant des variables qui mesurent l'effet croisé du statut d'occupation avec une localisation centrale (centre) ou périphérique (resturb), nous avons remarqué une baisse de la significativité de la variable mesurant la proximité au métro.

En intégrant dans le même modèle l'ensemble des variables testées séparément pour analyser la préférence d'une localisation au centre ou en périphérie en fonction de la taille du ménage, du statut socioprofessionnel et du statut d'occupation du ménage, nous avons obtenu les résultats suivant :

Tableau 6.15 : Résultats de l'estimation du modèle de localisation des ménages, avec certaines caractéristiques de la zone et du ménage

Variables	Coef	Se	t_test
prox_metro	-0,1150	0,01197	-9,6
serv_base1	0,1050	0,00159	65,9
serv_base2	0,0159	0,00131	12,2
centre*Etud	2,3292	0,03580	65,1
resturb*person5	0,7608	0,03002	25,3
centre*person1	0,8858	0,01508	58,7
resturb*propriétaire	0,3930	0,01651	23,8
centre*loc	0,7830	0,01151	68,0

<i>Nombre d'observations</i>	112 112
Log likelihood	-172 557
Log likelihood zero	-218 160
R ² ajusté	0,209

Source : Traitement Aissaoui sous R

La variable mesurant la proximité au métro a affiché un signe négatif. Cela peut être lié à la présence des corrélations avec des dimensions non observées. En remplaçant cette variable

par la proximité aux arrêts de bus bien desservi (prox_bus). Nous avons trouvé que l'utilité d'un ménage de se localiser dans une zone augmente avec le nombre d'arrêts bien desservis à proximité.

Tableau 6.16 : Résultats de l'estimation du modèle de localisation des ménages avec certaines caractéristiques de la zone et du ménage, en remplaçant la proximité aux métros par celle aux arrêts bien desservis

Variabes	Coef	Se	t_test
prox_bus	0,9860	0,01012	97,4
serv_base1	0,0832	0,00155	53,7
serv_base2	0,0093	0,00127	7,3
centre*Etud	2,2192	0,03562	62,3
restseurb*person5	0,3833	0,03020	12,7
centre*person1	0,7150	0,01512	47,3
restseurb*propriétaire	-0,0579	0,01708	-3,4
centre*loc	0,4650	0,01163	40,0

<i>Nombre d'observations</i>	112 112
Log likelihood	-167 638
Log likelihood zero	-218 160
R ² ajusté	0,232

Source : Traitement Aissaoui sous R

Cette variable est conditionnée aussi par le taux de motorisation du chef du ménage. Pour cela, nous avons testé dans un premier temps l'effet croisé de la présence ou non d'une station de métro avec le taux de motorisation du ménage. Cette variable, a ensuite été améliorée en créant une variable binaire mesurant la présence d'arrêt de métro ou d'un arrêt bien desservi. L'effet croisé de cette variable avec les ménages à bas revenu est aussi testé.

En testant séparément les variables de chaque catégorie, nous avons eu une idée sur leurs significativités et leurs poids dans les arbitrages des ménages.

Nous avons estimé ensuite un modèle qui intègre l'ensemble des variables et nous avons supprimé ensuite les variables qui ont des signes non attendus ou qui ne sont pas significatives (Annexe 3).

6.4. Résultats et discussions

6.4.1. Présentation et interprétation du modèle retenu

Les résultats de l'estimation du modèle confirment l'effet des principaux facteurs relevés dans la revue de la littérature, avec les signes attendus (tableau 6.17). Toutes choses égales par ailleurs, l'utilité diminue en fonction du prix de l'immobilier (prix_imm) et augmente avec l'accessibilité à l'emploi (ac_g_VP) et aux aménités. Toutefois, l'effet de l'accessibilité aux transports en commun reste conditionné par le taux de motorisation, puisque les ménages non motorisés sont les plus sensibles à ce mode de transport. Par ailleurs, les ménages motorisés (2 voitures ou plus) ont tendance à choisir une localisation en couronne périurbaine, alors que les ménages non motorisés sont attachés au centre.

L'effet du cycle de vie est aussi confirmé et semble le plus déterminant entre une localisation centrale ou en couronne périphérique. Les ménages sont généralement réticents à une localisation périurbaine, qu'ils soient propriétaires ou locataires, montrant leur attachement aux zones les plus centrales. C'est ce que traduit le signe négatif des variables « peri*propr » et « peri*loc », indiquant la désutilité des deux catégories à se localiser en couronne périurbaine.

L'effet croisé du revenu des ménages avec le taux de ménages de la même classe de revenu dans la zone a un signe positif, ce qui montre une tendance de stratification des ménages en fonction du revenu. Cette stratification peut être le résultat des mécanismes de recherche d'un environnement social favorable et de l'entre-soi des différents groupes. Mais elle peut être également un simple effet grégaire lié à la préférence pour les mêmes aménités spatiales, au déterminisme du marché immobilier ou à l'orientation de l'offre de logements, notamment pour les populations à faible revenu. En effet, le pourcentage des populations à faible revenu dans le quartier est fortement corrélé au prix de l'immobilier et au taux de logements sociaux (hlm). Avec l'introduction de ces deux facteurs, l'effet croisé se limite aux seuls revenus moyens et aisés (taux3*rev3 , taux4*rev4 et taux5*rev5). C'est la concentration de l'habitat social qui conduit dans certains quartiers à une concentration des populations à faible revenu et à une dévalorisation du sol, ce qui met en avant les logiques institutionnelles et le rôle des politiques de logement (Bonvalet, 2010). Le croisement du revenu avec le taux de logements

hlm montre un signe négatif concernant la dernière classe (%hlm*rev5), ce qui confirme la désutilité des ménages à haut revenu à habiter dans des quartiers avec un pourcentage élevé de logements sociaux. La préférence des ménages aisés à se localiser dans des zones peuplées par des ménages du même profil et à être entre-soi dans les grandes villes françaises (Goffette-Nagot et Schaeffer, 2013 ; Pinçon et Pinçon-Charlot, 2004) se combine avec un autre mécanisme de fuite et d'évitement des quartiers les moins favorisés (Maurin, 2004). L'environnement social n'est donc pas seulement le résultat mais aussi un déterminant important des choix de localisation des ménages (Schelling, 1969). Si les ménages pauvres de l'aire urbaine sont plus sensibles à l'accessibilité (Kryvobokov et Bouzouina, 2014), les ménages aisés semblent attirés davantage par la qualité de l'environnement social et de ses aménités endogènes (Brueckner et al., 1999).

Tableau 6.17 : Résultats de l'estimation du modèle final de localisation des ménages en 1999

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0110***	0,00016	68,8
	prix_imm	-0,0002***	0,00003	-7,0
Aménité spatiale	prox_serv	0,8026***	0,01906	42,1
	etb_prim	0,0994***	0,00101	98,7
	etb_sec	0,0654***	0,00336	19,5
	acc_jeux	0,0990***	0,00979	10,1
Environnement social	taux3*rev3	7,3756***	0,19848	37,2
	taux4*rev4	8,2172***	0,27584	29,8
	taux5*rev5	4,1076***	0,09966	41,2
	%hlm*rev1	1,6929***	0,04006	42,3
	%hlm*rev2	0,7672***	0,05101	15,0
	%hlm*rev3	0,1436**	0,04420	3,2
	%hlm*rev4	0,4679***	0,06033	7,8
	%hlm*rev5	-0,5958***	0,07135	-8,4
Caractéristiques des ménages	peri*2VP+	0,7031***	0,02181	32,2
	centre*VP0	0,8031***	0,02080	38,6
	TC*VP0	0,3947***	0,01800	21,9
	TC*VP1	0,1518***	0,01256	12,1
	peri*loc	-0,6294***	0,01800	-35,0
	peri*propr	-0,4947***	0,02512	-19,7
	centre*loc	0,1193***	0,01459	8,2
	centre*propr	-0,4266***	0,02156	-19,8
	peri*tail>=3	0,2864***	0,01978	14,5

	centre*Etud	2,0870***	0,03666	56,9
--	-------------	-----------	---------	------

<i>Nombre d'observations</i>	112 112
Log likelihood	-153 371
Log likelihood zero	-218 160
R ² ajusté	0,297

Source : Traitement Aissaoui sous R

Pour le prix de l'immobilier, nous avons obtenu le signe attendu. Toutefois, l'effet est un peu faible. Pour cela, nous avons testé un autre modèle sans le prix de l'immobilier (tableau 6.18).

Tableau 6.18 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, sans le prix de l'immobilier

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0109	0,00016	70,2
Aménité spatiale	prox_serv	0,8284	0,01903	43,5
	etb_prim	0,1038	0,00102	102,0
	etb_sec	0,0576	0,00334	17,3
	acc_jeux	0,1000	0,00975	10,3
Environnement social	taux3*rev3	6,4980	0,13725	47,3
	taux4*rev4	3,6719	0,16313	22,5
	taux5*rev5	3,5606	0,08637	41,2
	%hlm*rev1	1,8645	0,03469	53,7
	%hlm*rev2	0,9027	0,04681	19,3
	%hlm*rev3	0,3785	0,03966	9,5
	%hlm*rev4	0,1191	0,05176	2,3
	%hlm*rev5	-0,7461	0,06964	-10,7
Caractéristiques des ménages	peri*2VP+	0,7694	0,02057	37,4
	centre*VP0	0,8235	0,02098	39,2
	TC*VP0	0,3637	0,01912	19,0
	TC*VP1	0,2070	0,01491	13,9
	peri*loc	-0,7287	0,01604	-45,4
	peri*prop	-0,5831	0,02381	-24,5
	centre*loc	0,1078	0,01455	7,4
	centre*propr	-0,4228	0,02152	-19,6
	peri*tail >=3	0,3090	0,01977	15,6
	centre*Etud	2,1072	0,03669	57,4

<i>Nombre d'observations</i>	112 112
------------------------------	---------

Log likelihood	-153 488
Log likelihood zero	-218 160
R ² ajusté	0,297

Source : Traitement Aissaoui sous R

Pour tester la significativité du prix de l'immobilier, nous avons calculé le ratio de vraisemblance entre les deux modèles (tableau 6.17 et 6.18) et nous avons conduit un test de Khi2 à 1 degré de liberté (tableau 6.19).

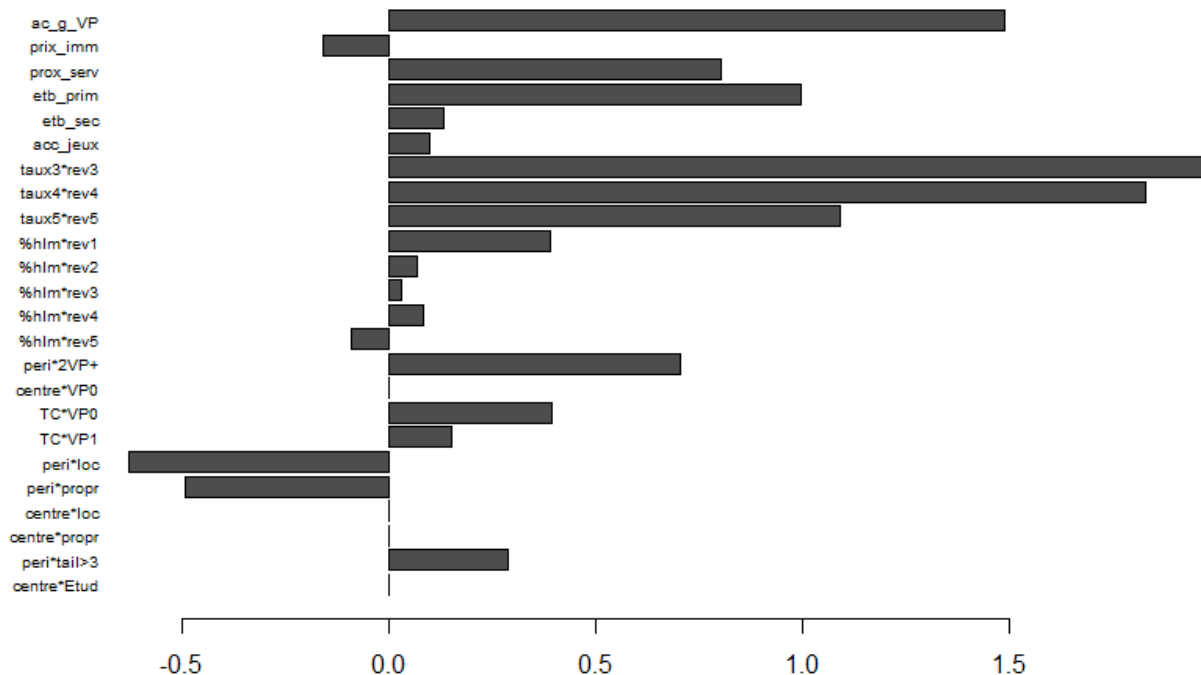
Tableau 6.19 : Résultats du test du ratio de vraisemblance du modèle avec et sans le prix de l'immobilier

Khi-deux	Degré de liberté	Signif à 1%
234	1	6,6

Source : Traitement Aissaoui sous R

Les résultats du test de Khi2 rejettent au risque de 1% ($234 \gg 6,6$) l'hypothèse de la nullité du coefficient relatif au prix de l'immobilier. Nous avons retenu alors le modèle avec le prix de l'immobilier (tableau 6.17). La figure 6.2 présente les effets marginaux de ce modèle.

Figure 6.2 : Différence d'utilités entre le 5ème et le 95ème percentile des valeurs des variables du modèle retenu



Source : Traitement Aissaoui sous R

Le poids de l'accessibilité dans le choix de localisation des ménages est nettement plus important que celui du prix de l'immobilier (cf. Figure 6.2). Les aménités spatiales et plus particulièrement la proximité aux crèches / écoles primaires et aux services quotidiens ont également un effet très important dans les choix résidentiels, mais il est moins prononcé que celui de l'accessibilité gravitaire à l'emploi. Cela étant, les effets de l'entre-soi et la qualité de l'environnement social sont les plus importants et plus particulièrement pour les classes favorisées.

Au-delà des limites liées à la différence des contextes urbains et à la caractérisation des ménages et des échelles spatiales, la comparaison de nos résultats avec ceux obtenus sur d'autres villes européennes et américaines met en avant de fortes similitudes, mais aussi des spécificités des comportements de choix de localisation des ménages. Tout d'abord, l'effet de l'accessibilité à l'emploi est positif et significatif, ce qui est concordant avec les résultats des différentes études de cas mentionnées dans la revue de littérature. C'est pour cela d'ailleurs que ce facteur est central dans les modèles intégrés transport-usage du sol (Anas, 1994 ; Waddell, 2000 ; Nicolas, 2010). En revanche, son poids dans l'arbitrage du marché est nettement plus important à Lyon et dans d'autres villes européennes (de Palma et al., 2007 ; Kim et al., 2005) comparativement aux villes américaines (Weisbrod et al., 1980 ; Guo et

Bhat, 2001). Les ménages lyonnais, et pas seulement les plus pauvres, donnent plus d'importance à l'accessibilité dans leur choix de localisation par rapport à la contrainte du prix de l'immobilier. Par ailleurs, l'attrait des ménages pour les aménités spatiales est également confirmé mais il apparaît moins important par rapport à celui mis en avant par les travaux américains voire même européens. Parmi les aménités spatiales introduites dans notre modèle, la proximité des crèches et des établissements primaires est le facteur le plus significatif, ce qui rejoint les résultats des travaux antérieurs (Oates, 1969 ; Guo et Bhat, 2001 ; Kim et al., 2005). Cela étant, ce facteur est probablement insuffisant et mérite d'être affiné dans les futurs travaux, en distinguant les écoles selon leur qualité et leur réputation. Néanmoins, cet exercice n'est pas aisé car la qualité réelle ou perçue d'un établissement ne dépend pas uniquement du niveau de dépenses, proxy souvent mis en avant dans les travaux anglo-saxons, mais surtout de la composition sociale du périmètre de la carte scolaire. Enfin, la recherche d'environnement social favorable, qui constitue un des facteurs les plus importants de la ségrégation résidentielle aux Etats-Unis (Gabriel et Rosenthal, 1989 ; Chattopadhyay, 2000), apparaît également comme étant le plus déterminant dans les choix de localisation des ménages en France (de Palma et al., 2007 ; Dantan et Picard, 2009), même s'il est parfois difficilement dissociable de la préférence pour les aménités spatiales.

Les résultats de ce premier travail réalisé sur l'aire urbaine de Lyon montrent que les arbitrages des ménages entre accès à l'emploi et prix de l'immobilier sont étroitement liés aux aménités sociales et spatiales dans le quartier. Les enjeux liés à la répartition des logements sociaux et des populations pauvres, mais aussi des équipements et des services, sont extrêmement importants pour l'orientation des choix de localisation résidentielle et la lutte contre les mécanismes de ségrégation sociale par les politiques publiques.

6.4.2. Effet de la taille de l'échantillon

La théorie n'a pas apporté de réponse formelle concernant le nombre d'alternatives à tirer aléatoirement. Pour cela, nous avons testé plusieurs tailles d'échantillon (X_j) et nous nous sommes arrêtés quand les résultats deviennent stables. Nous présentons ci-dessous un tableau des tailles d'échantillons testées :

Tableau 6.20 : Variation de la valeur de log-vraisemblance en fonction de la taille de l'échantillon d'alternatives

Nombre d'alternatives	54	35	22	14	9	7	4
Pourcentage de l'ensemble de l'échantillon	12.5%	8%	5%	3%	2%	1.5%	1%
Log-vraisemblance	-368 710	-321 177	-271 037	-223 134	-178 059	-153 371	-101 544
Log-vraisemblance normalisé	-368 710	-368 716	-368 738	-368 801	-368 930	-369 061	-369 644

Source : Traitement Aissaoui sous R

La capacité de l'ordinateur utilisé ne nous a pas permis d'aller au-delà de 54 alternatives, vu le nombre de variables explicatives introduites dans le modèle (une trentaine environ) et le nombre d'observations (112 112 ménages). En Annexe (4 et 5), nous présentons les résultats des tailles extrêmes (4 et 54).

On remarque que la valeur de log-vraisemblance augmente quand la taille de l'échantillon diminue. La comparaison des valeurs de log-vraisemblance des modèles avec des tailles d'échantillon d'alternatives différentes ne se fait pas directement. Pour cela, la log-vraisemblance normalisé est calculée. Il s'agit d'utiliser les paramètres obtenus avec un ensemble composé de X_j alternatives tirées aléatoirement pour calculer la log-vraisemblance pour l'ensemble des choix. Dans notre travail, nous sommes arrêtés à 54 alternatives et par suite nous avons fait la normalisation sur la base de 54 alternatives. Nous avons sélectionné pour chaque ménage 54 alternatives et nous avons calculé la valeur de log-vraisemblance (log-vraisemblance normalisé) à partir des paramètres estimés en tirant successivement 4, 7, 9, 14, 22, 35 et 54.

Les estimations faites sur un modèle sans prix de l'immobilier ont montré une stabilisation des coefficients, à partir d'une taille d'échantillon de sept alternatives (6 alternatives tirées aléatoirement plus le choix déjà fait par le ménage). En intégrant le prix de l'immobilier, la variation de la taille de l'échantillon déstabilise les valeurs des coefficients mesurant les aménités sociales et le prix de l'immobilier. Cela peut être expliqué par les fortes corrélations qui existent entre ce dernier et les variables liées à l'environnement social.

D'autres transformations des variables explicatives ont aussi été testées. La variable du prix de l'immobilier peut être croisée avec d'autres caractéristiques du ménage pour mesurer l'effet combiné entre le prix de l'immobilier et la classe de revenu (Annexe 2).

Avec plus de variables et en tirant aléatoirement 35 alternatives (Annexe 6), nous avons obtenu une valeur de R^2 ajusté moins élevée et certaines variables n'ont pas eu les signes attendus (%hlm*rev1). Le signe négatif de la variable taux__men_rev2*rev2 montre la désutilité des ménages du deuxième quintile de se localiser dans une zone peuplée par des ménages de la même classe de revenu. Toutefois, le signe positif de la variable taux1*rev1 peut être interprété de deux façons : i) La préférence des ménages pauvres pour l'entre soi, ii) La stratification des ménages à hauts revenus dans certaines zones peut conduire aussi à éviter les ménages appartenant au premier quintile qui se trouvent obligés de se localiser dans la même zone.

Un autre modèle est aussi estimé mais en prenant seulement les classes extrêmes. L'annexe 7 résume les résultats. Les résultats obtenus sont à nouveau avec des signes non attendus (taux1*rev1, taux5*rev5).

Toujours dans un souci d'améliorer la portée modélisatrice du modèle retenu (tableau 6.17) et d'éviter les problèmes de corrélations entre les variables du prix de l'immobilier et le taux de revenu et le taux du logement sociaux, nous avons testé plusieurs combinaisons (Annexes 8 à 14). Les résultats obtenus nous conduisent toutefois à conserver le modèle du tableau 6.17 avec un échantillon aléatoire de 7 alternatives.

Un modèle logit multinomial avec échantillonnage aléatoire : Quelles variables et quelle taille d'échantillon ?

La synthèse de la littérature théorique et empirique (partie I) met bien en avant la complexité des choix de localisation résidentielle des ménages. Ces derniers ne sont pas la conséquence d'un simple arbitrage entre coût du transport et prix de l'immobilier, mais ils sont étroitement dépendants des caractéristiques des ménages et de leurs préférences pour les aménités spatiales et sociales. L'analyse faite dans ce chapitre pour modéliser les choix de localisation

des ménages de l'aire urbaine de Lyon a permis non seulement de confirmer ce constat, mais aussi de mesurer le poids de chaque déterminant dans le contexte d'une ville marquée à la fois par l'étalement urbain et par le retour au centre des populations.

La modélisation des choix de localisation des ménages montre que la qualité de l'environnement social est le facteur le plus déterminant, mais ne remet pas en cause le poids toujours important de l'accessibilité à l'emploi. Tout en cherchant à être entre-soi et en évitant des quartiers où se concentrent les populations modestes, les ménages de classes moyennes et aisés choisissent également des quartiers accessibles à l'emploi et proches des services et des équipements centraux. Ainsi, nous sommes loin du contexte américain où les préférences pour les aménités sociales conduisent les ménages favorisés à donner moins d'importance à l'éloignement des emplois. L'importance accordée simultanément aux facteurs d'accessibilité et à l'environnement social par les ménages peut être une opportunité pour les acteurs publics permettant d'articuler politiques de transport et politiques de logement et renouveler l'attractivité des quartiers les moins favorisés. La ségrégation socio-spatiale est à la fois conséquence et cause des choix de localisation résidentielle.

Le test de plusieurs tailles d'échantillon a montré une stabilisation des coefficients, à partir d'une taille d'échantillon de sept alternatives (6 alternatives tirés aléatoirement plus le choix déjà fait par le ménage).

Tous les modèles estimés dans ce chapitre sont basés sur le modèle logit multinomial. Étant conscients des limites de l'application d'un tel modèle dans un contexte spatial, nous proposons à travers le chapitre suivant une modélisation hiérarchique, afin de tester si cette structure permet de dépasser les limites du modèle logit multinomial et de prendre en compte la dépendance entre les quartiers choisis et non choisis par les ménages.

Chapitre 7 - Une structure hiérarchique pour modéliser les choix de localisation : Quel apport ?

La ressemblance entre les zones contigües, considérées comme des alternatives différentes dans les choix de localisation résidentielle, peut créer des problèmes d'autocorrélation spatiale entre ces zones et conduit par suite à la violation de la propriété d'indépendance par rapport aux choix non retenus. D'où l'intérêt d'utiliser des modèles plus flexibles qui permettent de relâcher partiellement cette propriété. Toutefois, l'absence de méthodes, jusqu'à une date récente, permettant d'estimer ces modèles avec échantillon d'alternatives explique en partie le choix de l'utilisation des modèles logit multinomiaux pour estimer les choix de localisation. Le développement récent des méthodes permettant d'estimer des modèles GEV non biaisés avec échantillon d'alternatives sur données théoriques et de petites tailles (Guevara, 2010) a ouvert des perspectives d'application et d'adaptation à différents contextes urbains.

En utilisant les mêmes variables du modèle retenu (section 6.3.1), l'objectif de cette section est d'adapter la méthode proposée aux données réelles de l'aire urbaine de Lyon et de tester l'apport de l'utilisation d'un modèle logit emboîté avec échantillon d'alternatives pour modéliser les choix de localisation résidentielle. Cela va nous permettre également de répondre aux difficultés liées à la consistance des estimations dans un contexte où les alternatives sont des zones spatiales parfois difficiles à distinguer les unes des autres. Toutefois, la principale difficulté d'une telle modélisation réside justement dans la définition de la structure hiérarchique permettant de prendre en compte ces autocorrélations spatiales. Plusieurs structures ont été testées. Dans cette section, nous commençons par expliquer les hypothèses utilisées pour regrouper les alternatives qui se ressemblent et construire des nids (section 7.1) et nous finissons par la présentation des résultats obtenus et leur interprétation (section 7.2).

7.1. Définition des nids et estimation d'un modèle emboîté

La définition des nids constitue l'étape primordiale dans l'estimation des modèles logit emboîté. Si la distinction apparaît relativement évidente dans le domaine des transports, en distinguant les modes motorisés des modes non-motorisés (Koppelman et Wen, 1998 ; Yao et al, 2002), la construction des nids est beaucoup moins évidente quand il s'agit de classer des zones de localisation résidentielle. La principale raison est liée à l'autocorrélation spatiale entre les zones et la multitude des critères susceptibles d'intervenir dans la construction des nids. Ces critères pouvant être corrélés entre eux, leur mobilisation dans la construction des nids et dans l'estimation du choix des zones peut également poser problème.

La proximité spatiale induit des problèmes d'autocorrélation spatiale selon la loi de géographie : « Everything is related to everything, but nearest things are more related to one another (Tobler, 1970) ». Si ces problèmes ont entraîné l'abandon de l'hypothèse fondamentale d'observations indépendantes, l'introduction de l'espace dans les modèles économétriques n'est ni neutre, ni immédiate (Le Gallo, 2000). En faisant appel à la matrice des poids, Bhat et Guo (2004) ont essayé d'intégrer dans la fonction de l'utilité un paramètre spatial qui mesure le degré de l'autocorrélation spatiale entre les zones de localisation. Toutefois, ces modèles développés ont une structure qui combine à la fois la structure de la famille des modèles mixtes et celle de la famille GEV et la théorie n'a pas encore apporté des réponses concernant la correction des biais liés à l'échantillonnage des alternatives dans le cas de ce type de modèle.

La littérature traitant du regroupement des alternatives qui se ressemblent pour construire des nids dans une perspective d'estimation d'un modèle logit emboîté est peu abondante. Les travaux empiriques utilisent plusieurs méthodes et différents critères pour définir les nids et faire ainsi la distinction entre les zones de localisation, mais ces choix méthodologiques ne sont pas toujours explicites. En s'appuyant sur le travail de Van Der Haegen et al. (1996), Cornelis et al. (2012) distinguent quatre types de quartiers pour définir les nids dans l'ensemble de la Belgique : agglomération urbaine, banlieue, zone migratoire et zone rurale. Cette typologie est basée sur trois critères : faible densité, discontinuité spatiale et monofonctionnalité. Au Portugal, Guevara (2010) mobilise le travail de Rayle (2008) qui

utilise une analyse factorielle pour faire la distinction entre les municipalités de la ville d'Odivelas et de celle d'Amadora et les autres municipalités de la région de Lisbonne.

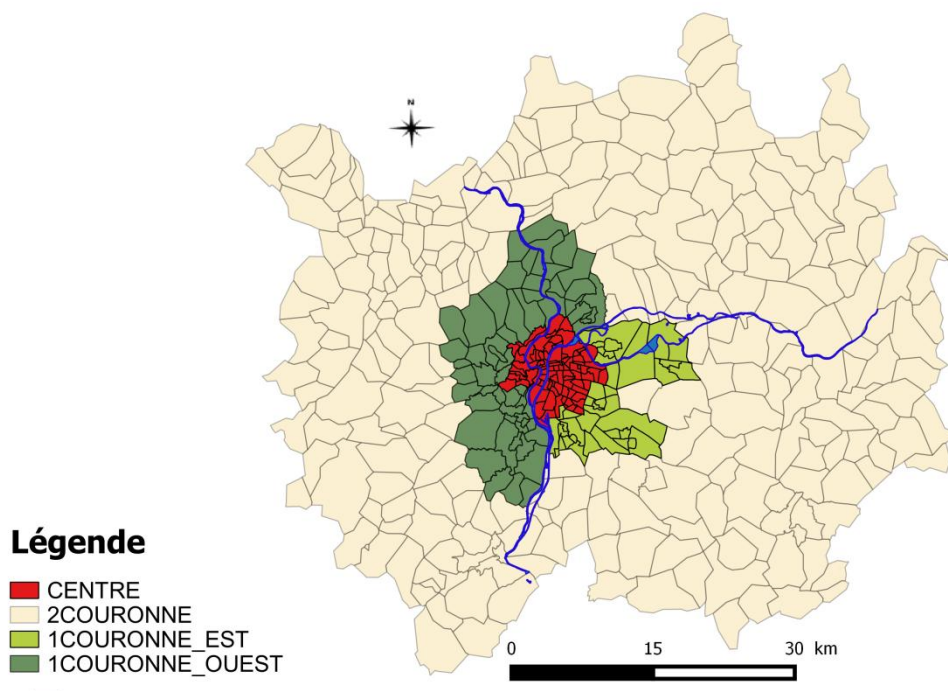
La difficulté dans la construction des nids est semblable à celle rencontrée par les travaux empiriques visant à distinguer les zones de résidence dans la perspective de mesurer la ségrégation résidentielle (Gaschet et Le Gallo, 2005). Le regroupement des zones pour constituer l'échelle de mesure pertinente de la ségrégation peut se faire *a priori* sur une connaissance sociologique approfondie du terrain ou *a posteriori* en faisant appel à une matrice de contiguïté pour détecter la ressemblance entre les zones de manière automatique.

Devant l'absence d'une méthode précise pour constituer les nids, nous nous sommes appuyés sur cette distinction en mobilisant plusieurs hypothèses pour regrouper les alternatives qui se ressemblent. Nous avons retenu quatre structures de nids : la première repose sur la connaissance de la répartition des revenus ; la deuxième est automatique et s'appuie sur le critère de densité de population et de distance au centre ; la troisième et la quatrième sont également issues de regroupements automatiques, mais elles sont multicritères et s'appuient sur des matrices de contiguïté pour identifier les ressemblances.

7.1.1. Structure centrée sur la répartition des revenus

La première identification des nids est issue d'une analyse de la ségrégation résidentielle réalisée à partir de la répartition des classes de revenu sur les quartiers IRIS de l'aire urbaine de Lyon (Bouzouina, 2008). En s'appuyant sur les résultats et la connaissance fine de ce travail, nous avons ainsi testé une structure en quatre grandes zones : le centre, la première couronne-est, la première couronne-ouest et la deuxième couronne (Carte 1).

Carte 7.1 : Structure centrée sur la répartition des revenus (4 zones)



Source : Traitement Aissaoui

Les résultats obtenus (tableau 7.1) montrent que les paramètres estimés sont significativement différents de zéro. Toutefois, certaines variables n'ont pas le signe attendu et les logsums ($1/\mu_1$, $1/\mu_2$, $1/\mu_3$ et $1/\mu_4$) sont supérieurs à 1.

Tableau 7.1 : Résultats du modèle emboîté selon la structure centrée sur la répartition des revenus (4 zones), avec les variables du modèle multinomial

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0147	0,00026	56,9
	prix_imm	-0,0018	0,00004	-45,1
Aménité spatiale	prox_serv	-0,0143	0,01255	-1,1
	etb_prim	0,1235	0,00167	73,8
	etb_sec	0,0651	0,00371	17,6
	acc_jeux	0,1940	0,01050	18,5
Environnement social	taux3*rev3	4,7167	0,15345	30,7
	taux4*rev4	4,8276	0,19210	25,1
	taux5*rev5	4,9936	0,10157	49,2
	%hlm*rev1	0,9140	0,03595	25,4

	%hlm*rev2	-0,0748	0,04851	-1,5
	%hlm*rev3	-1,0300	0,04787	-21,5
	%hlm*rev4	-1,2368	0,05633	-22,0
	%hlm*rev5	-2,0377	0,07678	-26,5
Caractéristiques des ménages	peri*2VP+	0,9264	0,01906	48,6
	centre*VP0	0,7067	0,01706	41,4
	TC*VP0	-0,1403	0,01717	-8,2
	TC*VP1	-0,0188	0,01243	-1,5
	peri*loc	-0,4843	0,02399	-20,2
	peri*propr	-0,3248	0,02886	-11,3
	centre*loc	-0,7120	0,02162	-32,9
	centre*propr	-1,2549	0,02705	-46,4
	peri*tail>3	0,4210	0,01860	22,6
	centre*Etud	1,5034	0,02785	54,0
Logsum	μ_1	0,8293	0,01492	55,6
	μ_2	0,7560	0,01246	60,7
	μ_3	0,9462	0,01461	64,8
	μ_4	0,7041	0,00853	82,5

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-390 573</i>
<i>Log likelihood zero</i>	<i>-419 476</i>
<i>R² ajusté</i>	<i>0,193</i>

Source : Traitement Aissaoui sous R

A l'exception des variables non significatives (prox_serv, %hlm*rev2 et TC*VP1), la majorité des paramètres a le signes attendu. Les signes de certaines variables liées à l'environnement social (%hlm*rev3, %hlm*rev4) ou aux caractéristiques des ménages (centre*loc) sont différents des signes obtenus dans le cas du modèle multinomial. Ce regroupement a donné des logsums avec des valeurs supérieures à 1 (généralement considéré comme un modèle inconsistant). En effet, ce regroupement est basé sur le revenu du ménage et on trouve que les variables liées à l'environnement social intégrées dans le modèle capturent les mêmes effets. Avant de tester d'autres structures hiérarchiques, nous avons enlevé les variables de l'environnement social pour voir si la suppression de ces variables va améliorer la consistance du modèle. Le tableau 7.2 résume les résultats obtenus.

Tableau 7.2 : Résultats du modèle emboîté selon la structure centrée sur la répartition des revenus (4 zones) avec les variables du modèle multinomial, mais sans l'environnement social

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0044	0,00007	65,9
	prix_imm	-0,0003	0,00001	-33,0
Aménité spatiale	prox_serv	-0,0250	0,00407	-6,1
	etb_prim	0,0389	0,00076	51,3
	etb_sec	0,0229	0,00134	17,0
	acc_jeux	0,0700	0,00270	25,9
Caractéristiques des ménages	TC*VP0	-0,0390	0,00582	-6,7
	TC*VP1	-0,0159	0,00402	-4,0
Logsum	μ_1	3,4214	0,07766	44,1
	μ_2	2,9209	0,07653	38,2
	μ_3	3,0909	0,06796	45,5
	μ_4	1,7319	0,02035	85,1

Nombre d'observations	112 112
Log likelihood	-331 394
Log likelihood zero	-390 573
R ² ajusté	0,152

Source : Traitement Aissaoui sous R

Suite à la suppression des variables relatives à l'environnement social, nous avons obtenu un modèle consistant (des valeurs de logsum inférieur à 1). Toutefois, les variables comme prox_serv, TC*VP0 et TC*VP1 ont des signes contre intuitifs.

En enlevant ces deux variables, la valeur du R² a légèrement baissé (tableau 7.3). Cependant, le test de khi2 a donné une valeur de 48>>>7,8.

Tableau 7.3 : Résultats du modèle emboîté selon la structure centrée sur la répartition des revenus (4 zones) avec les variables du modèle multinomial, sans les variables liées à l'environnement social et aux caractéristiques des ménages

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0044	0,00006	72,0
	prix_imm	-0,0003	0,00001	-37,1
Aménité spatiale	prox_serv	-0,0255	0,00407	-6,3
	etb_prim	0,0388	0,00070	55,0
	etb_sec	0,0227	0,00135	16,9
	acc_jeux	0,0686	0,00278	24,7
Logsum	μ_1	3,4201	0,07305	46,8
	μ_2	2,9229	0,07258	40,3
	μ_3	3,0967	0,06481	47,8

	μ_4	1,7305	0,01912	90,5
--	---------	--------	---------	------

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-331 418</i>
<i>Log likelihood zero</i>	<i>-390 573</i>
<i>R² ajusté</i>	<i>0,151</i>

Source : Traitement Aissaoui sous R

En enlevant aussi la variable mesurant la proximité aux services quotidiens (prox_serv), nous avons obtenu la même valeur du R² (tableau 7.4).

Tableau 7.4 : Résultats du modèle emboîté selon la structure centrée sur la répartition des revenus (4 zones) avec les variables du modèle multinomial, sans les variables liées à l'environnement social, aux caractéristiques des ménages et la proximité des services de base

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0041	0,00013	30,8
	prix_imm	-0,0002	0,00001	-14,8
Aménité spatiale	etb_prim	0,0387	0,00153	25,3
	etb_sec	0,0223	0,00130	17,2
	acc_jeux	0,0658	0,00100	65,6
Logsum	μ_1	3,4136	0,15286	22,3
	μ_2	2,9133	0,13945	20,9
	μ_3	3,0952	0,12873	24,0
	μ_4	1,7198	0,03827	44,9

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-331 515</i>
<i>Log likelihood zero</i>	<i>-390 573</i>
<i>R² ajusté</i>	<i>0,151</i>

Source : Traitement Aissaoui sous R

Pour vérifier que les valeurs du logsum supérieur à 1 sont liées aussi à la manière de regrouper les alternatives, nous avons testé d'autres structures hiérarchiques.

7.1.2. Structure centrée sur la densité de population et la distance au centre

Dans la deuxième hypothèse, nous avons fait la distinction entre le centre et la périphérie à partir de la distribution des densités de population. Une telle structure n'est pas basée sur des variables liées à l'environnement social et va nous permettre de les garder dans le modèle (vu leur poids dans l'explication des choix de localisation).

La limite entre le centre et la périphérie correspond au périmètre à partir duquel la densité commence à fléchir. En utilisant le modèle de Bussière amendé (Tabourin et al., 1995) (cf. Encadré 1), nous avons déterminé le point d'inflexion **b** et nous avons défini le centre comme étant l'ensemble des Grands quartiers dont la distance est inférieure à $1/b$. Pour ce faire, nous avons procédé de la manière suivante :

- Identification des coordonnées géographiques (les x et y) des centroïdes des Grands quartiers à l'aide d'un SIG ;
- Calcul de la distance à vol d'oiseau de chaque Grand quartier par rapport au Grand quartier central (préfecture) ;
- Calcul du volume de population au niveau de chaque Grand quartier ;
- Calcul du volume de population cumulée en fonction de la distance au centre ;
- Calage du modèle de Bussière et mesure des coefficients A, b, et K.

Les résultats obtenus sont les suivant :

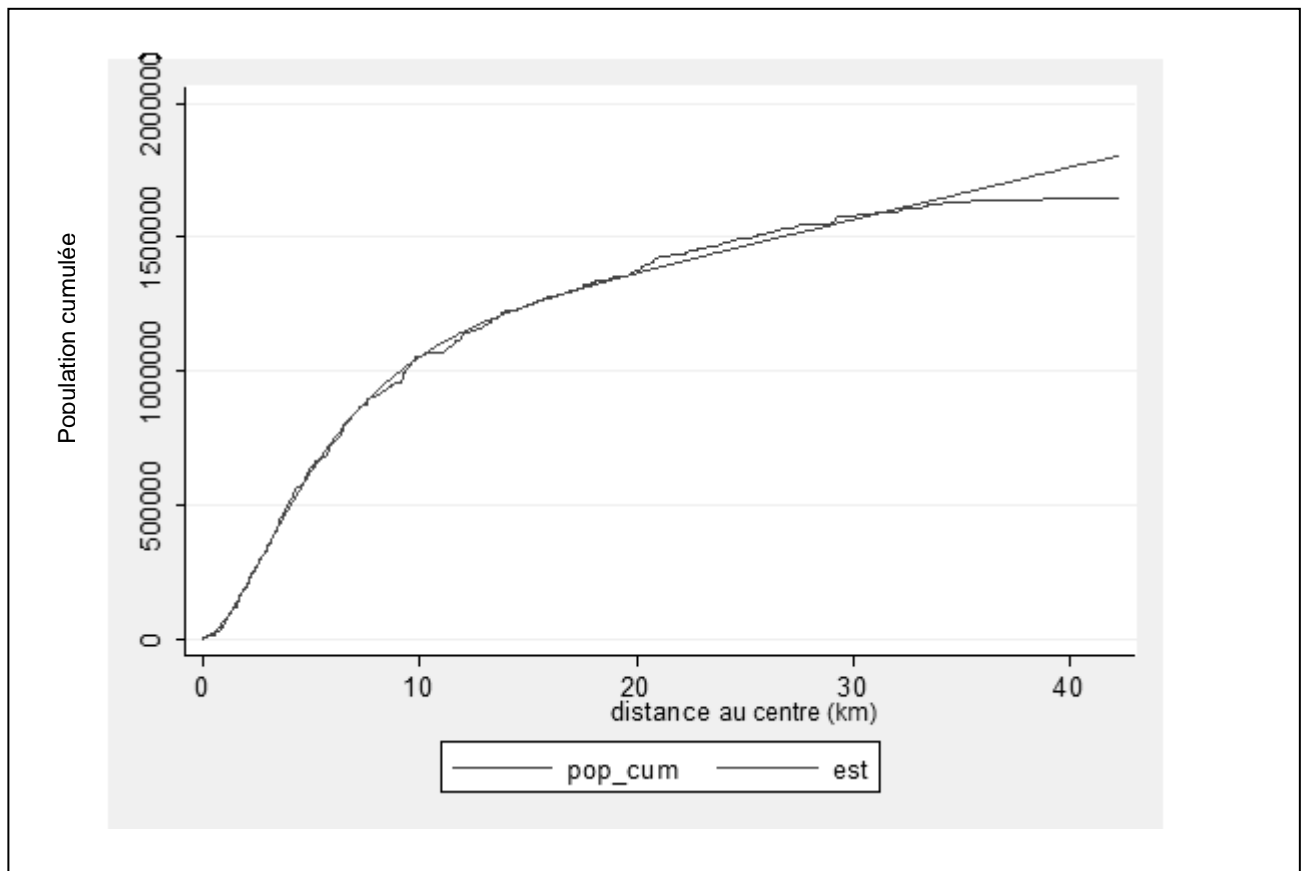
Tableau 7.5 : Résultats de l'estimation du modèle de Bussière

Paramètres	Coef	Se	T
A	19243,35	185,926	103,5
B	0,338756	0,002	146,7
K	16738,05	244,744	68,4

<i>Nombre d'observations</i>	432
R^2	0,9996
R^2 ajusté	0,9995

Source : Traitement Aissaoui sous Stata à partir des données RGP 1999

Figure 7.1 : Résultats du modèle Bussière appliqué à l'aire urbaine de Lyon



Source : *Élaboration propre sous stata à partir des données RGP 1999*

En s'inspirant du modèle monocentrique d'Alonso, le modèle de René Bussière est développé initialement pour définir les limites de la suburbanisation des ménages et des entreprises. Il s'agit de la maximisation du système de l'entropie suivant :

$$P(r) = \frac{2\pi A}{b^2} * [1 - (1 + br)e^{-br}]$$

Tel que :

r : la distance au centre.

A : la densité extrapolée au centre de la ville

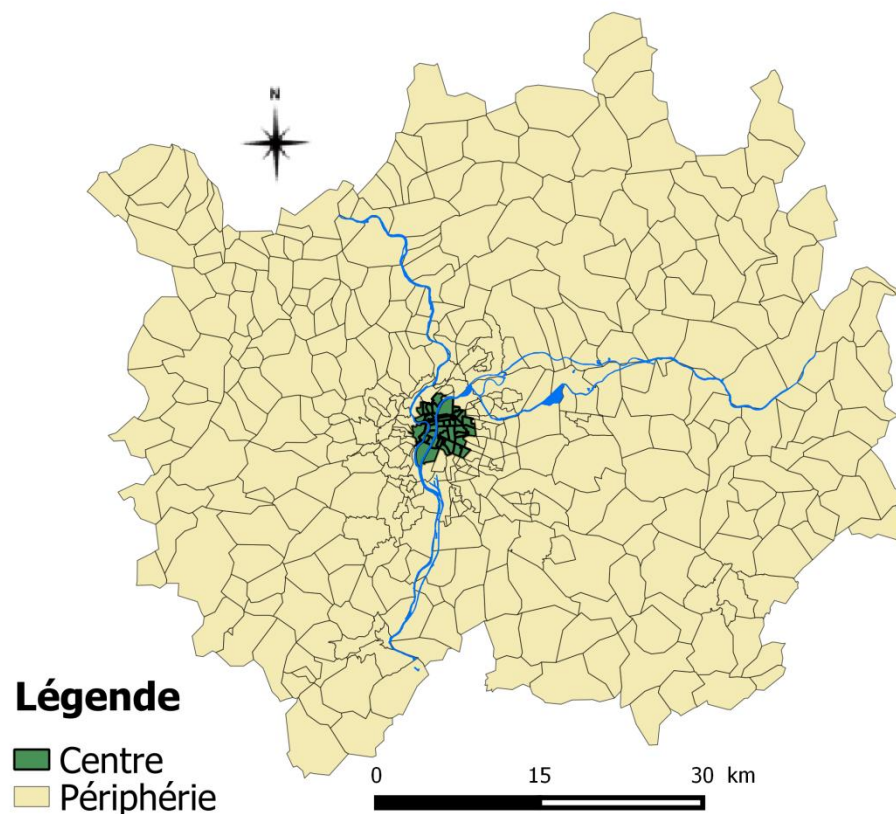
b : le taux de décroissance exponentielle de la densité par rapport à la distance au centre. Son inverse représente le point d'inflexion de la courbe de la population cumulée (P(r)).

La minimisation de la somme des écarts au carré entre les données réelles et les données estimées à partir de la distance au centre (P(r)) donne A et b.

Le travail de Tabourin et al. (1995) a montré que l'amélioration des moyens de transport donne une asymptote non horizontale. Pour traiter ce problème, ils proposent un amendement. Il s'agit d'ajouter une fonction linéaire (Kr) qui capture l'effet de déconcentration périphérique lié à l'effet des infrastructures de transport. Le modèle Bussière amendé s'écrit comme suit :

$$P(r) = \frac{2\pi A}{b^2} * [1 - (1 + br)e^{-br}] + Kr$$

Carte 7.2 : Distinction centre-périphérie à partir de la densité de population et la distance au centre (2 zones)



Source : *Traitement Aissaoui*

Tableau 7.6 : Résultats du modèle emboîté selon une structure centre-périphérie définie à partir de la densité de population et de la distance, avec les variables du modèle multinomial

Catégorie	Variables	Coef	Se	t_test
-----------	-----------	------	----	--------

Arbitrage du marché	ac_g_VP	0,0212	0,00028	74,6
	prix_imm	-0,0009	0,00004	-23,8
Aménité spatiale	prox_serv	0,5636	0,01752	32,2
	etb_prim	0,1406	0,00166	84,9
	etb_sec	0,0706	0,00390	18,1
	acc_jeux	0,1114	0,01275	8,7
Environnement social	taux3*rev3	10,0117	0,23231	43,1
	taux4*rev4	12,6826	0,35995	35,2
	taux5*rev5	5,0783	0,11195	45,4
	%hlm*rev1	1,0335	0,04731	21,8
	%hlm*rev2	-0,1292	0,06040	-2,1
	%hlm*rev3	-0,9436	0,05412	-17,4
	%hlm*rev4	-0,1806	0,07558	-2,4
Caractéristiques des ménages	%hlm*rev5	-1,8159	0,08780	-20,7
	peri*2VP+	1,1334	0,02616	43,3
	centre*VP0	1,0110	0,01989	50,8
	TC*VP0	0,1445	0,02040	7,1
	TC*VP1	0,0698	0,01593	4,4
	peri*loc	-1,1603	0,02373	-48,9
	peri*propr	-1,0125	0,03317	-30,5
	centre*loc	-0,1735	0,01703	-10,2
	centre*propr	-0,7548	0,02512	-30,1
	peri*tail>3	0,5914	0,02483	23,8
centre*Etud	2,3606	0,03516	67,1	
Logsum	μ_1	0,6872	0,00675	101,8
	μ_2	0,6489	0,00720	90,2

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-251 071</i>
<i>Log likelihood zero</i>	<i>-325 227</i>
<i>R² ajusté</i>	<i>0,228</i>

Source : Traitement Aissaoui sous R

L'utilisation d'une structure basée sur la distribution de la densité de la population donne des résultats significatifs. Toutefois, les signes de certaines variables explicatives sont différents de ceux du modèle multinomial et les valeurs des logsum sont toujours supérieures à 1. Cela peut être lié à l'introduction des variables faisant la distinction entre le centre et la périphérie dans le modèle. En enlevant ces dernières, nous avons obtenus le modèle suivant :

Tableau 7.7 : Résultats du modèle emboîté selon une structure centre-périphérie définie à partir de la densité de population et de la distance, en enlevant les variables de localisation (centre, périphérie)

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0196	0,00029	67,4
	prix_imm	-0,0005	0,00003	-15,9
Aménité spatiale	prox_serv	0,4368	0,01484	29,4
	etb_prim	0,1119	0,00169	66,2
	etb_sec	0,0650	0,00320	20,3
	acc_jeux	0,0660	0,01024	6,4
Environnement social	taux3*rev3	8,2602	0,19531	42,3
	taux4*rev4	11,1015	0,31295	35,5
	taux5*rev5	3,4551	0,09364	36,9
	%hlm*rev1	1,0188	0,03742	27,2
	%hlm*rev2	0,1260	0,04816	2,6
	%hlm*rev3	-0,3797	0,04234	-9,0
	%hlm*rev4	0,2437	0,06026	4,0
%hlm*rev5	-1,5117	0,07364	-20,5	
Caractéristiques des ménages	TC*VP0	0,1448	0,01601	9,1
	TC*VP1	0,2775	0,01304	21,3
Logsum	μ_1	0,8633	0,01212	71,2
	μ_2	0,8204	0,01274	64,4

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-263 388</i>
<i>Log likelihood zero</i>	<i>-300 067</i>
<i>R² ajusté</i>	<i>0,122</i>

Source : Traitement Aissaoui sous R

A l'opposé de la première structure, la suppression des variables liées au centre ou en périphérie n'a pas amélioré la valeur des logsums. Et la valeur du R² a fortement baissée.

La structure centrée sur la densité de population et la distance au centre n'a pas permis d'améliorer la consistance du modèle. Les valeurs des logsum sont toujours supérieures à 1. Pour cela, nous avons passé au test d'une autre structure multicritère basée sur la matrice de contiguïté qui permet de tenir compte d'autres variables pour regrouper les alternatives qui se ressemblent.

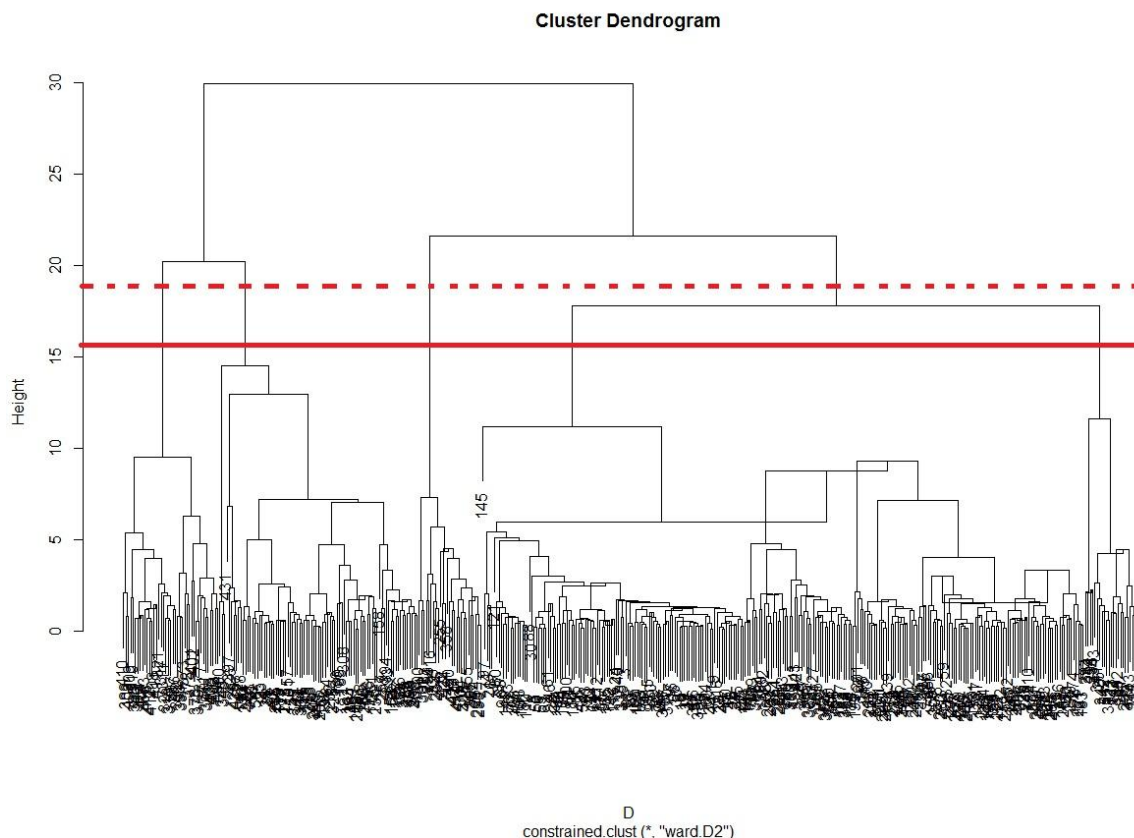
7.1.3. Structure multicritère sur la base d'une matrice de contiguïté

La troisième hypothèse est inspirée de la méthode de Bhat et Guo (2004). Il s'agit de faire une classification hiérarchique spatiale pour regrouper sous le même nid les alternatives qui se ressemblent. Cette classification est faite de la manière suivante :

- Détermination de la liste des Grands quartiers voisins selon une contiguïté du mouvement du fou dans un jeu d'échec ;
- Calcul du vecteur des distances entre deux quartiers et de la matrice de proximité. Cette matrice est calculée en utilisant le pourcentage de logements sociaux, la densité, le prix de l'immobilier, le taux de ménages appartenant au premier et au cinquième quintile de revenu ;
- Regroupement des quartiers dont la distance entre eux est la plus petite ;
- Recalcul du vecteur des distances.

A l'issue de cette classification, nous avons obtenu le dendrogramme suivant avec deux structures de nids (à 4 classes et à 5 classes) :

Figure 7.2 : Dendrogramme de la classification hiérarchique



Source : Traitement Aissaoui sous R

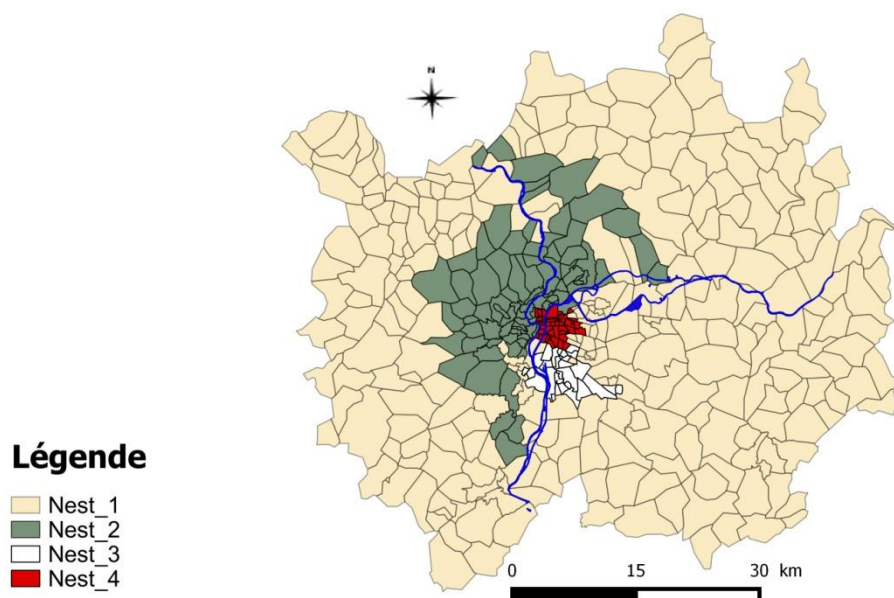
Le développement d'un programme sous R, nous a permis de visualiser sur des cartes plusieurs découpages du dendrogramme. Seuls les découpages en quatre et en cinq classes nous permettent de garder la forme concentrique de l'aire urbaine de Lyon.

- En quatre classes

Le découpage du dendrogramme, au niveau de la ligne en pointillé, nous donne un regroupement en quatre classes :

Ce découpage permet de faire sortir la dimension centre-périphérie et la polarisation Est-Ouest. En revanche, il ne permet pas de distinguer entre le Nord et l'Ouest de la première couronne comme c'est le cas du découpage centré sur la répartition des revenus.

Carte 7.3 : Structure multicritère sur la base d'une matrice de contiguïté (4 classes)



Source : Traitement Aissaoui

Les résultats obtenus sont les suivants :

Tableau 7.8 : Résultats du modèle emboîté selon la structure multicritère sur la base d'une matrice de contiguïté (4 classes), avec les variables du modèle multinomial

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0124	0,00025	48,9
	prix_imm	-0,0023	0,00004	-60,7
Aménité spatiale	prox_serv	0,3189	0,01353	23,6
	etb_prim	0,1220	0,00134	91,3
	etb_sec	0,1141	0,00388	29,4
	acc_jeux	0,1529	0,01199	12,8
Environnement social	taux3*rev3	4,8641	0,17143	28,4
	taux4*rev4	5,9016	0,24323	24,3
	taux5*rev5	6,0770	0,10060	60,4
	%hlm*rev1	0,9891	0,03886	25,5
	%hlm*rev2	0,0507	0,05081	1,0

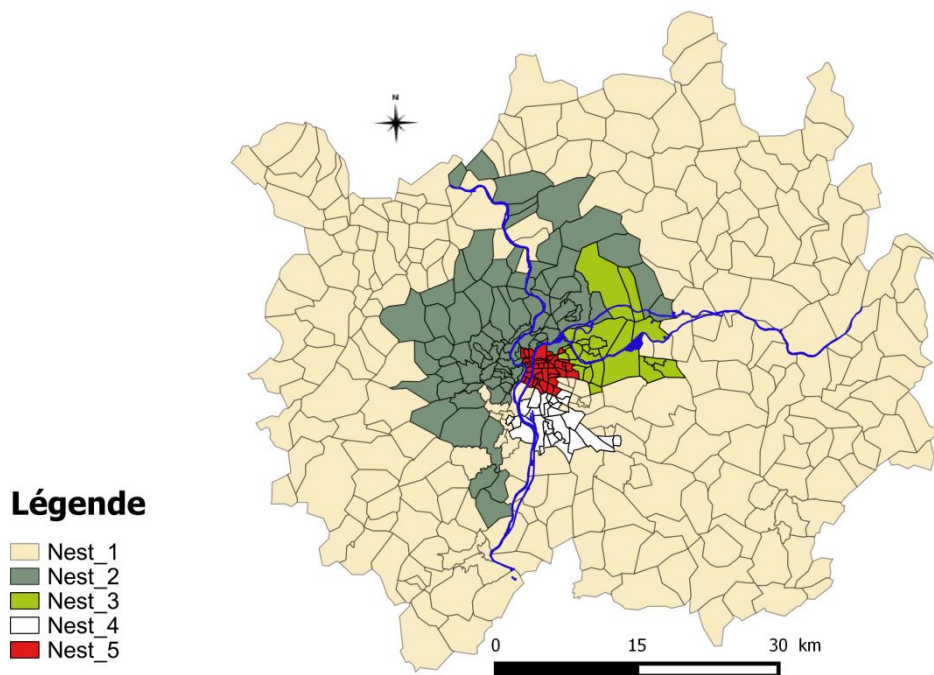
	%hlm*rev3	-0,7348	0,04718	-15,6	
	%hlm*rev4	-0,7858	0,05978	-13,1	
	%hlm*rev5	-1,0906	0,07861	-13,9	
Caractéristiques des ménages	peri*2VP+	0,9491	0,02171	43,7	
	centre*VP0	1,0007	0,01916	52,2	
	TC*VP0	0,1384	0,02375	5,8	
	TC*VP1	0,2136	0,01813	11,8	
	peri*loc	-1,3368	0,02011	-66,5	
	peri*propr	-1,1215	0,02769	-40,5	
	centre*loc	-0,2329	0,01597	-14,6	
	centre*propr	-0,7910	0,02397	-33,0	
	peri*tail>3	0,4350	0,02095	20,8	
	centre*Etud	1,9970	0,03094	64,6	
	Logsum	μ_1	0,7534	0,00713	105,7
		μ_2	0,6632	0,00686	96,7
μ_3		0,6609	0,00812	81,4	
μ_4		0,5838	0,00564	103,5	

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-323 349</i>
<i>Log likelihood zero</i>	<i>-397 846</i>
<i>R² ajusté</i>	<i>0,187</i>

Source : Traitement Aissaoui sous R

En coupant le dendrogramme au niveau de la ligne rouge, nous avons testé un regroupement en cinq classes illustré dans la carte 7.4. A l'opposé du premier découpage (en quatre classes), ce découpage a la particularité de faire la distinction du Nord et du Sud en première couronne, tout en gardant la structure concentrique de l'aire urbaine.

Carte 7.4 : Structure multicritère sur la base d'une matrice de contiguïté (5 classes)



Source : Traitement Aissaoui

Tableau 7.9 : Résultats du modèle emboîté selon la structure multicritère sur la base d'une matrice de contiguïté (5 classes), avec les variables du modèle multinomial

Catégorie	Variables	Coef	se	t_test
Arbitrage du marché	ac_g_VP	0,0136	0,00024	54,7
	prix_imm	-0,0018	0,00004	-42,6
Aménité spatiale	prox_serv	0,7494	0,02187	34,3
	etb_prim	0,1194	0,00127	93,9
	etb_sec	0,0902	0,00351	25,7
	acc_jeux	0,1066	0,01143	9,3
Environnement social	taux3*rev3	5,8417	0,20450	28,6
	taux4*rev4	8,3798	0,30124	27,8
	taux5*rev5	5,7670	0,09821	58,7
	%hlm*rev1	0,9256	0,03965	23,3
	%hlm*rev2	-0,0689	0,05151	-1,3
	%hlm*rev3	-0,7609	0,04669	-16,3
	%hlm*rev4	-0,4578	0,06396	-7,2
%hlm*rev5	-1,0307	0,07740	-13,3	
Caractéristiques des ménages	peri*2VP+	0,8238	0,02270	36,3
	centre*VP0	0,9125	0,02035	44,8
	TC*VP0	0,4221	0,01923	21,9
	TC*VP1	0,1423	0,01439	9,9

	peri*loc	-1,0106	0,02071	-48,8
	peri*propr	-0,8794	0,02753	-31,9
	centre*loc	-0,0290	0,01562	-1,9
	centre*propr	-0,6019	0,02229	-27,0
	peri*tail>3	0,3653	0,02004	18,2
	centre*Etud	2,3647	0,03537	66,9
Logsum	μ_1	0,7665	0,00646	118,6
	μ_2	0,7066	0,00628	112,5
	μ_3	0,6933	0,00804	86,2
	μ_4	0,6750	0,00703	96,0
	μ_5	0,6711	0,00561	119,7

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-340 957</i>
<i>Log likelihood zero</i>	<i>-419 476</i>
<i>R² ajusté</i>	<i>0,187</i>

Source : Traitement Aissaoui sous R

A l'exception des paramètres non significatifs (%hlm*rev2, centre*loc), tous les paramètres ont les signes attendus. En revanche, les valeurs des logsums sont toujours supérieures à 1.

L'ensemble des structures hiérarchiques testées donne des valeurs de logsums supérieures à 1. A l'opposé des deux premières hypothèses, la troisième et plus spécifiquement la quatrième hypothèse est réalisée en intégrant à la fois les variables liées à l'environnement social (revenu, pourcentage des logements sociaux) et la densité de population. Cette structure a permis de garder la forme concentrique de l'aire urbaine de Lyon et de faire la distinction entre l'est et l'ouest lyonnais. Ce résultat nous a conduit à la retenir pour comparer les résultats d'une structure hiérarchique selon un modèle logit emboîté avec ceux d'un modèle multinomial. Cette structure de modèle emboîté servira de référence dans la comparaison avec les autres structures de modèles choix discret.

7.2. Comparaison avec d'autres structures de modèle

Dans cette partie, nous présentons les résultats obtenus en retenant 5 nids (4^{ème} structure). Nous avons tirés aléatoirement 7 alternatives (6 alternatives + alternative choisie pour le nid contenant la zone choisie) de chaque nid.

La modélisation d'une structure hiérarchique à l'aide d'un modèle emboîté avec échantillonnage d'alternatives revient à un échantillonnage stratifié. Pour cela, la comparaison va être faite avec un modèle multinomial avec échantillon stratifié. Ce dernier sera aussi comparé avec un modèle multinomial avec échantillon aléatoire composé de 35 alternatives. Un tel nombre est retenu pour pouvoir comparer une structure stratifiée avec 5 classes (7 alternatives au niveau de chaque nid).

7.2.1. Un MNL avec un échantillonnage aléatoire ou un MNL avec un échantillonnage stratifié

Tableau 7.10 : Résultats du modèle MNL avec un échantillonnage aléatoire de 35 alternatives

Catégorie	Variable	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0119	0,00013	89,8
	prix_imm	-0,0004	0,00003	-16,0
Aménité spatiale	prox_serv	0,7443	0,01679	44,3
	etb_prim	0,0959	0,00078	123,1
	etb_sec	0,0464	0,00260	17,8
	acc_jeux	0,0819	0,00861	9,5
Environnement social	taux3*rev3	6,7370	0,16892	39,9
	taux4*rev4	8,2254	0,24094	34,1
	taux5*rev5	3,8523	0,07957	48,4
	%hlm*rev1	1,3284	0,03199	41,5
	%hlm*rev2	0,4783	0,04224	11,3
	%hlm*rev3	-0,1267	0,03811	-3,3
	%hlm*rev4	0,2214	0,05213	4,2
Caractéristiques des ménages	%hlm*rev5	-0,8601	0,06238	-13,8
	peri*2VP+	0,7190	0,01974	36,4
	centre*VP0	0,7649	0,01734	44,1
	TC*VP0	0,4095	0,01462	28,0
	TC*VP1	0,1416	0,01076	13,2
	peri*loc	-0,6459	0,01624	-39,8
	peri*propr	-0,5435	0,02277	-23,9
	centre*loc	0,0762	0,01237	6,2
	centre*propr	-0,4355	0,01869	-23,3
	peri*tail>3	0,3186	0,01797	17,7
centre*Etud	2,0535	0,03057	67,2	

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-321 177</i>
<i>Log likelihood zero</i>	<i>-398 597</i>
<i>R² ajusté</i>	<i>0,194</i>

Source : Traitement Aissaoui sous R

Le modèle logit multinomial avec un échantillon stratifié est estimé de deux manières. La première consiste à appliquer la méthode développée par McFadden (1978) et la deuxième utilise celle proposée par Guevara (2010) en supposant que les logsums (μ_m) sont égaux à 1. En effet, un modèle logit emboîté revient à un modèle logit stratifié avec des logsums égaux à 1. Or μ est déjà normalisée à 1, la valeur du terme $\ln(f(\mathbf{B}_i(\mathbf{D}_n)))$ est égale à 0.

Théoriquement les deux probabilités sont les mêmes (tableau 7.11). Toutefois, le calcul numérique de ces probabilités ne se fait pas de la même manière. Dans la méthode de Guevara (2010), il se fait selon un processus itératif (p.72 et 73). Dans celle de McFadden (1978), le calcul se fait directement en utilisant le correctif $\ln\pi(\mathbf{D}_n/i)$.

Tableau 7.11 : La correction des biais liés à l'échantillonnage des alternatives

Logit	Probabilité des choix sur l'ensemble des alternatives (C_n)	Probabilité des choix sur un échantillon d'alternative (D_n)	La log-vraisemblance
Multinomial	$P_n(i) = \frac{e^{U_{in}}}{\sum_{j \in C_n} e^{U_{jn}}}$	$P_n(i) = \frac{e^{U_{in}}}{\sum_{j \in D_n} e^{U_{jn}}}$	$L = \sum_{n=1}^N \ln \frac{e^{u_{in}}}{\sum_{j \in D_n} e^{u_{jn}}}$
Multinomial stratifié	$P_n(i) = \frac{e^{U_{in}}}{\sum_{j \in C_n} e^{U_{jn}}}$	$P_n(i) = \frac{e^{U_{in} + \ln \pi(D_n/i)}}{\sum_{j \in D_n} e^{U_{jn} + \ln \pi(D_n/j)}} = \frac{e^{U_{in} + \ln(1/q_i)}}{\sum_{j \in D_n} e^{U_{jn} + \ln(1/q_j)}}$ <p>Si l'alternative i appartient au nid m $q_i = (\text{le nombre d'alternatives tirés du nid } m) / (\text{le nombre total d'alternatives du nid } m)$</p>	$L = \sum_{n=1}^N \ln \frac{e^{u_{in} + \ln \pi(D_n/i)}}{\sum_{j \in D_n} e^{u_{jn} + \ln \pi(D_n/j)}}$
Emboîté	$P_n(i) = \frac{e^{U_{in} + \ln(f(B_i(C_n)))}}{\sum_{j \in C_n} e^{U_{jn} + \ln(f(B_j(C_n)))}}$	$P_n(i) = \frac{e^{U_{in} + \ln(f(B_i(D_n))) + \ln(1/q_i)}}{\sum_{j \in D_n} e^{U_{jn} + \ln(f(B_j(D_n))) + \ln(1/q_j)}}$	$L = \sum_{n=1}^N \ln \frac{e^{u_{in} + \ln f(B_i(D_n)) + \ln \pi(D_n/i)}}{\sum_{j \in D_n} e^{u_{jn} + \ln f(B_j(D_n)) + \ln \pi(D_n/j)}}$
		$\ln(f(B_i(D_n))) = \left(\frac{\mu}{\mu_m} - 1 \right) \left(\ln \sum_{j \in C_{mn}} e^{\mu_m U_{jn}} \right) + \ln \mu + (\mu_m - 1) U_{in}$	

Source : *Élaboration propre*

Dans notre analyse, les deux méthodes ont donné des signes différents pour les variables suivantes : %hlm*rev2, %hlm*rev4. Ces différences peuvent être expliquées par le processus itératif utilisé dans la deuxième méthode de Guevara. Le tableau (7.12) présente les résultats obtenus en utilisant la première méthode, et tableau (7.13) pour la deuxième méthode.

Tableau 7.12 : Résultats du modèle MNL avec un échantillonnage stratifié et correctif selon la méthode proposée par McFadden (1978)

Catégorie	Variable	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0121	0,00013	92,7
	prix_imm	-0,0005	0,00003	-17,0
Aménité spatiale	prox_serv	0,7848	0,01770	44,3
	etb_prim	0,0965	0,00081	119,8
	etb_sec	0,0436	0,00256	17,0
	acc_jeux	0,0731	0,00863	8,5
Environnement social	taux3*rev3	6,8211	0,16796	40,6
	taux4*rev4	8,4602	0,24179	35,0
	taux5*rev5	3,7850	0,07723	49,0
	%hlm*rev1	1,2372	0,03149	39,3
	%hlm*rev2	0,4543	0,04198	10,8
	%hlm*rev3	-0,1808	0,03782	-4,8
	%hlm*rev4	0,2021	0,05179	3,9
Caractéristiques des ménages	%hlm*rev5	-0,8965	0,06199	-14,5
	peri*2VP+	0,7370	0,01985	37,1
	centre*VP0	0,7665	0,01705	45,0
	TC*VP0	0,4203	0,01435	29,3
	TC*VP1	0,1377	0,01069	12,9
	peri*loc	-0,6692	0,01632	-41,0
	peri*propr	-0,5358	0,02294	-23,4
	centre*loc	0,0643	0,01227	5,2
	centre*propr	-0,4440	0,01850	-24,0
	peri*tail>3	0,3252	0,01807	18,0
centre*Etud	2,0638	0,03027	68,2	

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-340 639</i>
<i>Log likelihood zero</i>	<i>-419 475</i>
<i>R² ajusté</i>	<i>0,188</i>

Source : Traitement Aissaoui sous R

Récemment, Daly et al (2014) ont simulé empiriquement la méthode développée par Guevara (2010) et ils ont montré que le modèle logit multinomial avec un échantillon stratifié est non efficace (variance élevée).

Dans notre cas, la comparaison des deux modèles montre que le modèle multinomial avec un échantillon aléatoire donne des résultats meilleurs que celui avec un échantillon stratifié. Il y a des différences légères des paramètres estimés. Toutefois, les signes des variables explicatives sont les mêmes pour les deux modèles.

7.2.2. Un NL ou un MNL avec un échantillonnage stratifié

La comparaison du modèle logit emboîté (tableau 7.9) et du logit multinomial avec échantillonnage stratifié estimé à partir de la méthode de McFadden (tableau 7.12) montre un changement des signes des mêmes variables. Ces variables sont les mêmes qui ont changé de signes en estimant un logit multinomial avec échantillon stratifié de deux méthodes différentes. Cela n'est pas liée à la structure du modèle, mais plutôt à la méthode itérative utilisée pour corriger le biais d'échantillonnage dans le cas du modèle emboîté. La valeur du R^2 du modèle logit emboîté est légèrement moins élevée que celle du modèle logit multinomial avec un échantillon stratifié.

En revanche, la comparaison du modèle logit emboîté (tableau 7.9) et du logit multinomial échantillonnage stratifié estimé à partir de la méthode de Guevara (tableau 7.13) montre la stabilité des signes des paramètres estimés. La valeur du R^2 du modèle logit emboîté est plus élevée que celle du modèle logit multinomial stratifié estimé à partir de la méthode de Guevara. Toutefois, cette valeur reste moins élevée par rapport à celle obtenue en estimant un logit multinomial avec un échantillonnage aléatoire.

Tableau 7.13 : Résultats du modèle MNL avec un échantillonnage stratifié (emboîté avec logsum=1, estimation selon la méthode itérative de Guevara (2010))

Catégorie	Variable	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0126	0,00013	95,5
	prix_imm	-0,0012	0,00003	-45,9
Aménité spatiale	prox_serv	0,5315	0,01566	34,0

	etb_prim	0,0958	0,00081	118,8
	etb_sec	0,0509	0,00259	19,7
	acc_jeux	0,0805	0,00866	9,3
Environnement social	taux3*rev3	3,5985	0,15816	22,8
	taux4*rev4	5,9380	0,22858	26,0
	taux5*rev5	4,6412	0,07561	61,4
	%hlm*rev1	0,6778	0,03072	22,1
	%hlm*rev2	-0,0974	0,04157	-2,3
	%hlm*rev3	-0,6951	0,03675	-18,9
	%hlm*rev4	-0,6378	0,05019	-12,7
	%hlm*rev5	-1,0223	0,06098	-16,8
Caractéristiques des ménages	peri*2VP+	0,7666	0,01975	38,8
	centre*VP0	0,7946	0,01715	46,3
	TC*VP0	0,3375	0,01424	23,7
	TC*VP1	0,0808	0,01061	7,6
	peri*loc	-0,9228	0,01588	-58,1
	peri*propr	-0,7770	0,02261	-34,4
	centre*loc	0,0486	0,01231	3,9
	centre*propr	-0,4486	0,01854	-24,2
	peri*tail>3	0,2927	0,01798	16,3
	centre*Etud	2,0995	0,03029	69,3

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-342 868</i>
<i>Log likelihood zero</i>	<i>-419 476</i>
<i>R² ajusté</i>	<i>0,183</i>

Source : Traitement Aissaoui sous R

Pour que le modèle respecte la consistance avec la maximisation de l'utilité, Daly et Zachary (1979) identifient un ensemble de critères. Empiriquement, le critère le plus difficile à maintenir est celui qui garantit une distribution de densité non négative.

En normalisant μ à 1, $\frac{1}{\mu_m}$ doit être inférieur à 1. Dans le cas contraire, Kling et Herriges

(1995) ont montré que le modèle peut être consistant même avec des valeurs $\frac{1}{\mu_m}$

supérieures à 1 s'il respecte les deux conditions suivantes :

i) $1/\mu_m \leq V_{1m}$

ii) $1/\mu_m \leq V_{2m}$

V_{1m} et V_{2m} sont respectivement la première et la deuxième dérivée de l'utilité de choisir le nid m . Empiriquement, le calcul de ces deux dérivées peut être mené de deux manières. La première, consiste à calculer ces dérivées pour chaque observation et à compter le nombre des observations qui respectent les deux conditions ci-dessus. La deuxième, se limite au calcul des deux dérivées de l'utilité pour la moyenne des variables explicatives et à tester ensuite les deux conditions.

Les résultats de l'analyse de la consistance en utilisant la première méthode montre que 0% des observations respectent la condition i). En calculant la première dérivée de l'utilité de la moyenne, nous avons trouvé une valeur égale à 0.2. Cette valeur est inférieure de la valeur des

$1/\mu_m$ et par suite la condition n'est pas respectée.

Train et al. (1989) ont retenu une structure hiérarchique en deux niveaux avec des $1/\mu_m > 1$ puisqu'elle donne des résultats meilleurs qu'un modèle multinomial simple. Ils expliquent cela par une substitution des alternatives entre nids supérieure aux alternatives au sein d'un nid donné. En effet, l'augmentation du prix d'une alternative au sein d'un nid augmente la probabilité que les gens choisissent un autre nid plutôt que de choisir une alternative au sein du même nid.

Un travail de Holland et Sutinen (2000) a aussi retenu des structures hiérarchiques avec des $1/\mu_m$ supérieures à 1 et les résultats ont été interprétés statistiquement dans la lignée des travaux de Train et al. (1989).

Dans notre travail, nous avons choisi de retenir un modèle multinomial. Le modèle emboîté donne des paramètres semblables à ceux obtenus en utilisant un logit multinomial avec échantillon d'alternatives. Ce dernier est plus consistant et demande moins de temps de calcul. Généralement, les travaux empiriques utilisent les modèles logit multinomiaux pour leur simplicité de calculs. A travers ce travail, nous avons montré que l'utilisation d'une échelle spatiale homogène (Grand quartier) pour modéliser les choix de localisation permet de résoudre les principaux problèmes d'autocorrélation sans faire appel à une structure hiérarchique complexe. Ces résultats rejoignent ceux du travail de Ibeas et al. (2013) qui ont

montré que le logit multinomial est plus performant que le modèle emboîté pour modéliser les choix de localisation des ménages de la baie de Santander.

Estimation d'un logit emboîté : Quel apport ?

Grâce au développement des méthodes récentes, nous avons été en mesure de tester l'apport de l'utilisation d'un modèle logit emboîté pour analyser les choix de localisation de l'aire urbaine de Lyon.

Théoriquement, ce modèle apparaît très séduisant. Il permet à la fois de dépasser les problèmes d'autocorrélation spatiale et ceux liés à la multitude des alternatives, en sélectionnant aléatoirement un échantillon d'alternatives. Toutefois, au plan empirique ce modèle pose plusieurs problèmes.

Au-delà des temps de calculs (jusqu'à quatre jours) qui limitent le nombre de modèles et d'hypothèses à tester, la difficulté de l'application d'un tel modèle est le regroupement des zones homogènes et la construction des nids.

Malgré les différents groupements définis pour réduire les problèmes d'autocorrélation, nous n'avons pas pu dépasser la difficulté de séparer la corrélation en espace et nids. En effet, la prise en compte de l'espace à travers la construction des nids dans le modèle, donne des paramètres associés aux logsums qui ne sont pas à l'intérieur de l'intervalle (0,1) et par suite pose des problèmes de consistance du modèle. En réduisant la portée modélisatrice du modèle à des variables non liées à l'environnement social, nous avons obtenus des structures hiérarchiques consistantes. Toutefois, ces modèles ne respectent pas l'une des exigences d'opérationnalité, à savoir la pertinence (Bonnaïfous, 1989).

Dans un souci de trouver un compromis entre la triple exigence d'opérationnalité à savoir, la pertinence, la cohérence et la mesurabilité, nous avons choisi de passer par un modèle logit multinomial et d'utiliser une échelle spatialement homogène (Grand quartier).

En retenant la structure du modèle logit multinomial et en utilisant les mêmes variables du modèle retenu dans le chapitre 6 (tableau 6.17), le chapitre suivant analyse l'évolution, entre 1999 et 2007, des comportements des ménages en matière de choix de localisation.

Chapitre 8 - Analyse diachronique des déterminants de localisation des ménages (1999, 2007)

L'objectif de ce chapitre est d'analyser l'évolution des comportements de ménages en matière de choisir une zone de localisation. L'idée de départ était d'introduire une dimension temporelle à travers un panel de cinq ans. Mais devant l'indisponibilité des données, nous nous sommes limités à des analyses descriptives entre deux dates pour donner une idée de l'impact des variations des accessibilités et des aménités entre 1999 et 2007 sur le changement de comportements de localisation des ménages.

Les estimations faites sur les données de 2007 vont nous permettre également de valider la pertinence de notre modèle pour prévoir les choix de localisation. Il s'agit de simuler les choix de localisation résidentielle de l'année 2007 à partir des données de 1999 et d'analyser la transférabilité temporelle de notre modèle et inversement. L'objectif de ce chapitre est double d'une part en termes de compréhension de l'évolution des comportements de choix de localisation résidentielle entre 1999 et 2007 et d'autre part méthodologique en termes d'analyse de la stabilité et de la transférabilité des résultats de la modélisation.

Nous débutons par une analyse descriptive des données en 1999 et 2007 pour mettre en avant les principaux changements tels que le retour au centre, la croissance du prix de l'immobilier et l'amélioration du réseau du transport en commun (section 8.1). La comparaison des estimations porte ensuite sur l'analyse des variations des poids des facteurs explicatifs des choix de localisation résidentielle entre 1999 et 2007 (section 8.2). La dernière section, est consacrée à l'analyse de la transférabilité temporelle du modèle 1999.

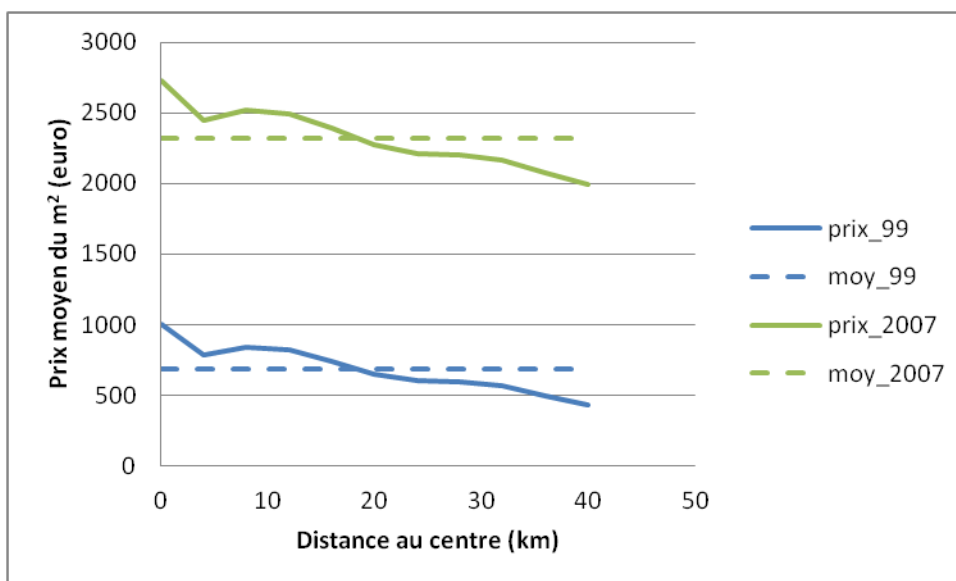
8.1. Analyse descriptive

En s'appuyant toujours sur la typologie adoptée pour classer les déterminants de choix de localisation dans le deuxième chapitre, nous effectuons dans cette section une analyse descriptive des variables explicatives du modèle retenu.

8.1.1. Arbitrage du marché

L'analyse des choix de localisation des ménages en 1999 a montré l'importance de deux facteurs de cette catégorie : le prix de l'immobilier et l'accessibilité gravitaire à l'emploi en voiture particulière. Le prix de l'immobilier a triplé entre 1999 et 2007 (graphique 1). A l'exception d'une légère augmentation observée à 5 km du centre, le prix de l'immobilier diminue avec la distance au centre en 1999 et 2007.

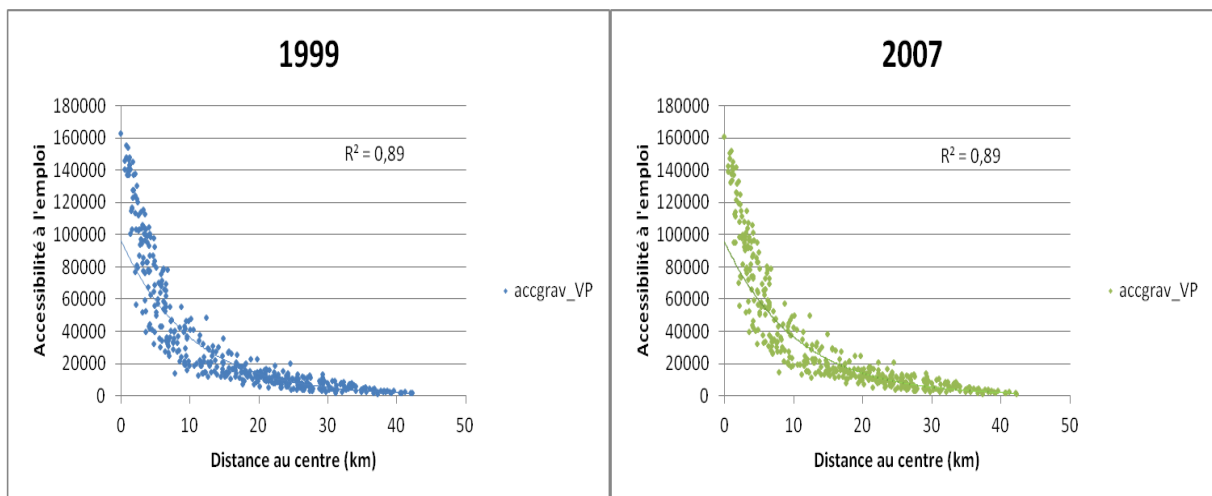
Graphique 8.1 : Variation du prix de l'immobilier (en fonction de la distance au centre (1999 et 2007))



Source : Traitement Aissaoui à partir des données du prix de l'immobilier estimés en 1999 et 2007

En prenant la valeur moyenne des caractéristiques des zones sur un intervalle de 4 km, nous avons établi l'ensemble des graphiques. Toutefois, l'accessibilité gravitaire moyenne n'a pas de sens. Pour cela, nous avons choisi de passer par des nuages de points (graphique 8.2).

Graphique 8.2 : Variation de l'accessibilité gravitaire à l'emploi en VP (en fonction de la distance au centre (1999 et 2007))



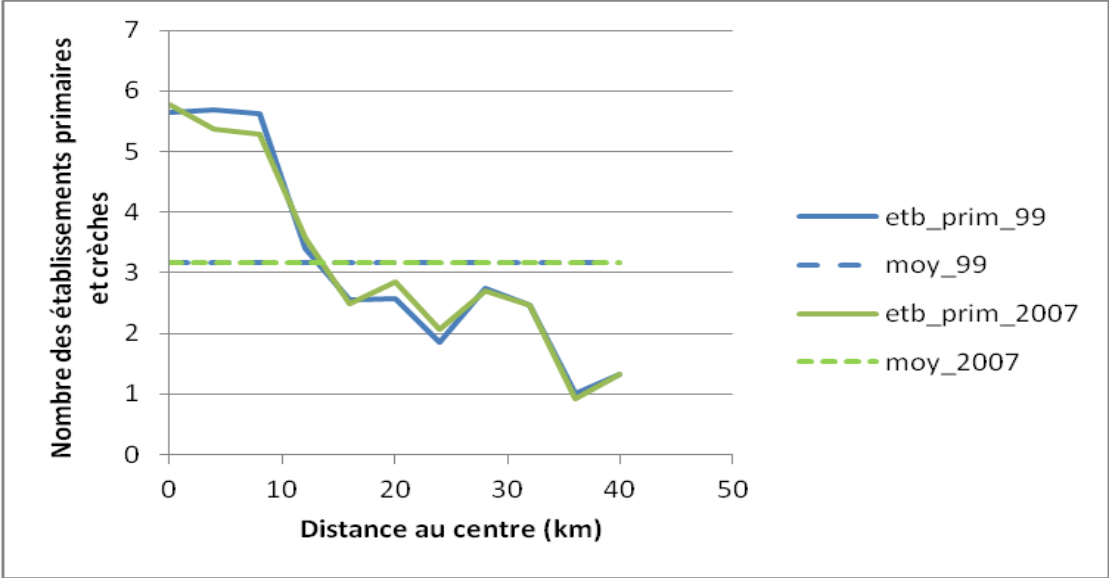
Source : *Élaboration propre ; données accessibilité gravitaire (en nombre d'emplois) : SIRENE 1999, 2005; réseau VP et EMD 1999 et 2006.*

L'accessibilité en voiture est exponentiellement négative. Le temps de déplacements est globalement stable en voiture entre ces deux dates. L'emploi moyen n'a pas trop changé, il est passé de 784 595 emplois en 1999 à 829 636 en 2007. Ceci explique la stabilité des données entre les deux dates.

8.1.2. Aménité spatiale

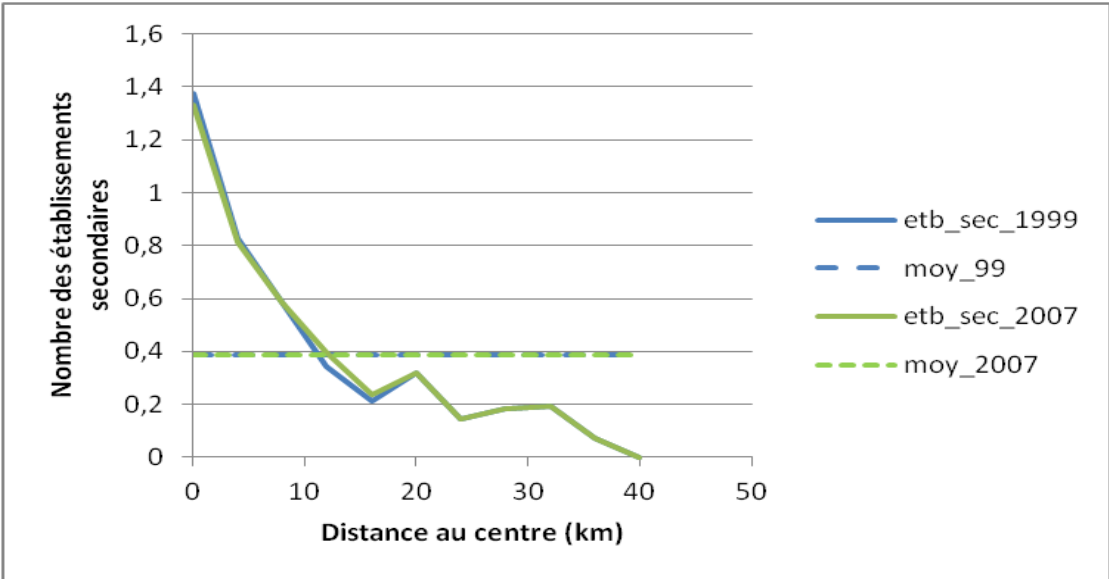
A l'exception du nombre de parcs de jeux dans la zone qui a légèrement baissé en 2007, les autres aménités sociales n'ont pas changé entre 1999 et 2007.

Graphique 8.3 : Nombre d'établissements primaires et crèches (en fonction de la distance au centre (1999 et 2007))



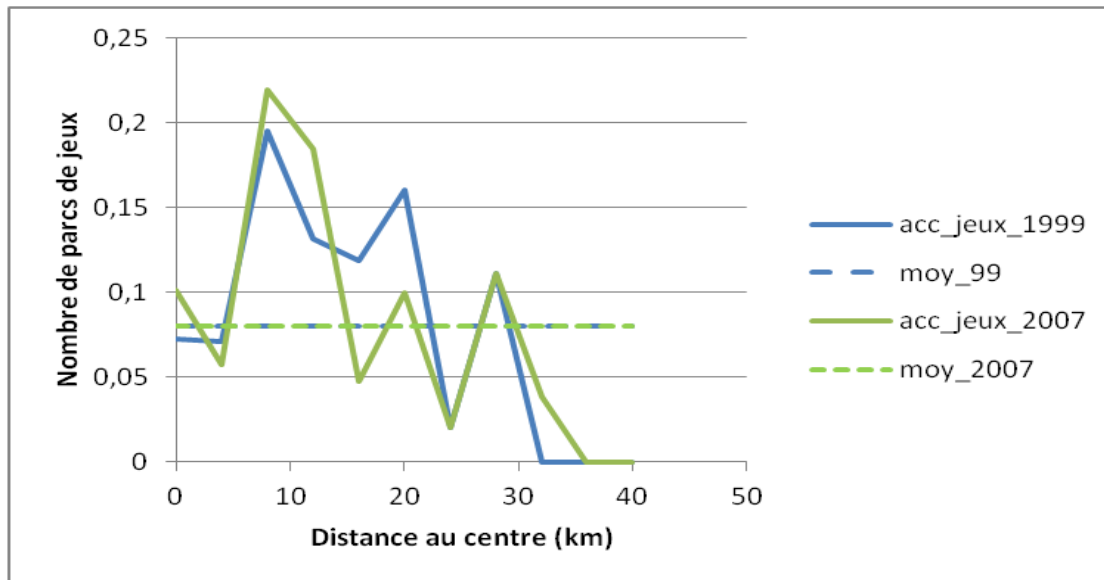
Source : Traitement Aissaoui à partir de la base SIRENE 1999, 2005

Graphique 8.4 : Nombre d'établissements secondaires (en fonction de la distance au centre (1999 et 2007))



Source : Traitement Aissaoui à partir de la base SIRENE 1999, 2005

Graphique 8.5 : Nombre de parcs de jeux (en fonction de la distance au centre (1999 et 2007))

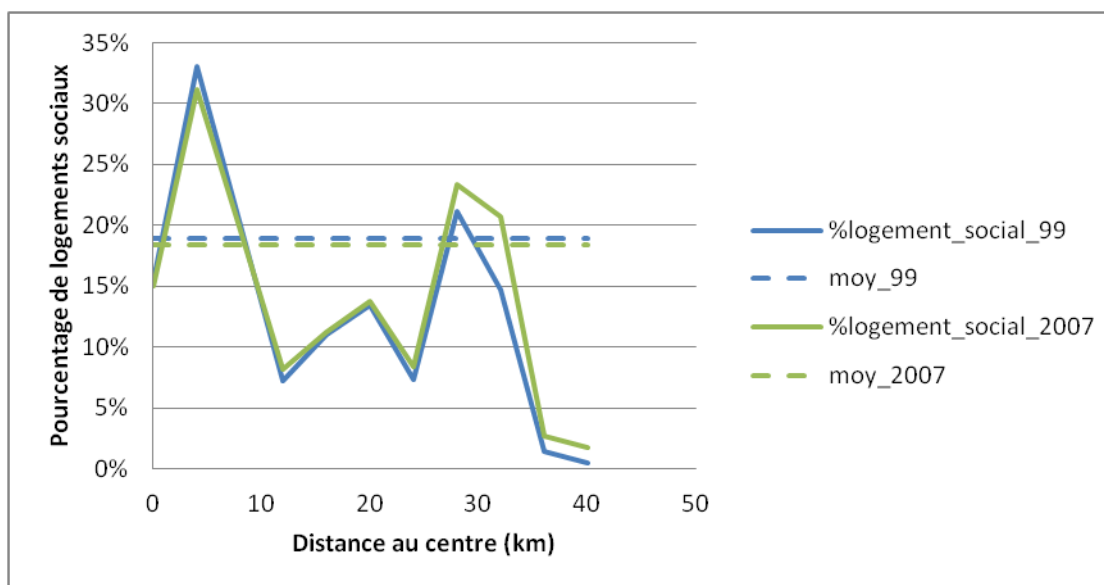


Source : Traitement Aissaoui à partir de la base SIRENE 1999, 2005

8.1.3. Aménité sociale

Le parc de logement social a connu une légère augmentation en 2007 dans les zones qui se situent entre 30 et 40 km du centre.

Graphique 8.6 : Concentration des logements sociaux (en fonction de la distance au centre (1999 et 2007))



Source : Traitement Aissaoui à partir de données sur les logements sociaux fournies par l'INSEE en 1999 et 2007

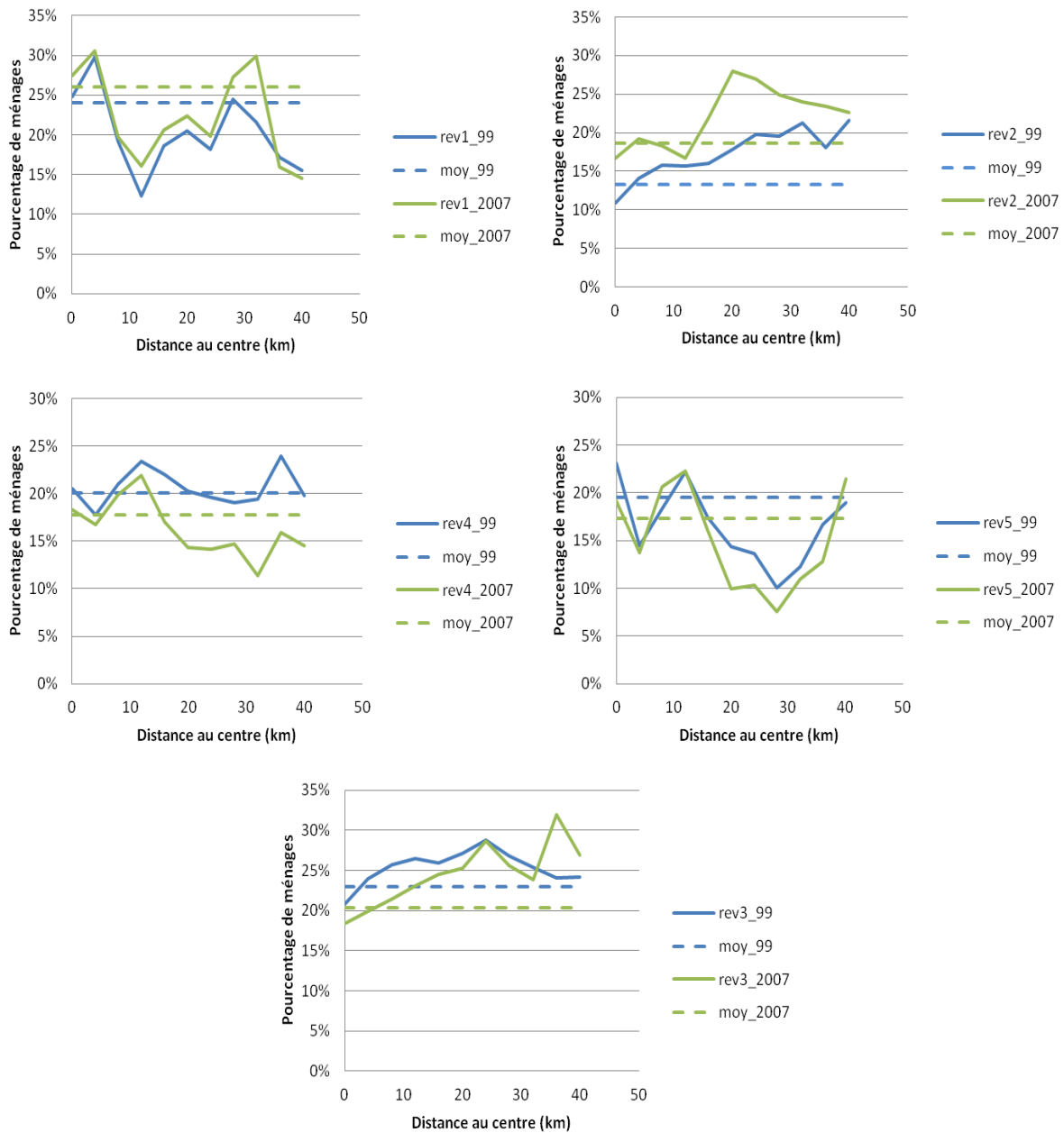
Malgré les destructions que certains quartiers ont connu entre 1999 et 2007, la concentration des logements sociaux par rapport au centre n'a pas trop changé entre les deux dates.

La loi n° 2000-1208 du 13 décembre 2000 relative à la solidarité et au renouvellement urbains couramment appelée loi SRU a mis en place un quota de logements sociaux. Il s'agit des seuils que chaque commune de plus de 1 500 habitants en île de France et de plus de 3 500 habitants pour les autres régions doit respecter (20% de logement sociaux). Toutefois, la majorité des villes préfèrent payer les pénalités plutôt que de respecter ce quota pour éviter le mécontentement de leurs habitants.

A l'opposé des taux de ménages appartenant au 4^{ème} et au 5^{ème} quintile qui ont diminué en 2007, ceux des ménages appartenant au 1^{er} et 2^{ème} quintile ont augmenté en 2007.

Pour les ménages à revenu moyen, on constate que leur taux devient de plus en plus important dans les zones qui se situent au-delà de 35 km.

Graphique 8.7 : Variation des taux des ménages de chaque quintile de revenu (en fonction de la distance au centre (1999 et 2007))



Source : Élaboration propre ; les revenus fiscaux DGI-INSEE

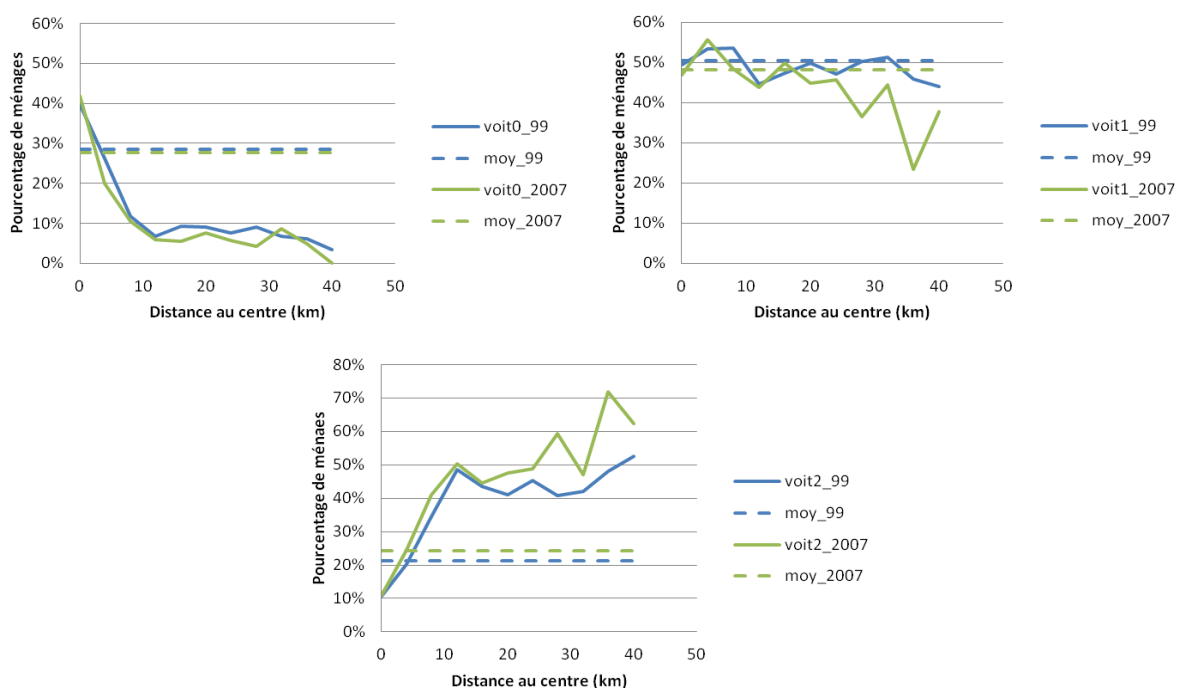
8.1.4. Caractéristiques des ménages

Concernant les variables liées à l'accessibilité, l'année 2007 est marquée par le prolongement du métro B vers Gérard, la mise en service des lignes de tramway T1 et T2 (en 2001), la

création de nouvelles lignes de bus. Le modèle retenu en 1999 intègre seulement des variables qui mesurent l'effet croisé de la proximité aux transports en communs avec le taux de motorisation des ménages. Ce choix n'explique pas le fait que les ménages non motorisés sont concentrés dans un rayon de 10 km autour du centre. Au-delà d'une distance de 10 km autour du centre, le taux de ménages non motorisés baisse.

En 2007 et au-delà de 10 km, le taux des ménages possédant une voiture a baissé et celui des ménages ayant plus de 2 voitures a augmenté (graphique 8.8).

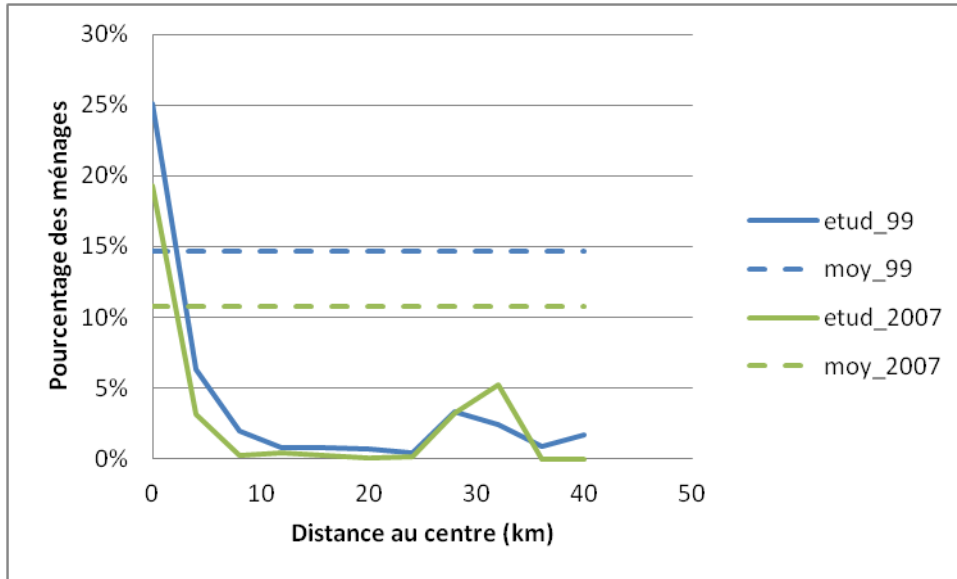
Graphique 8.8 : Variation des taux de motorisation (en fonction de la distance au centre (1999 et 2007))



Source : Élaboration propre ; à partir de la base ménages en 1999 et la base logement en 2007-INSEE

Pour ce qui de catégorie socioprofessionnelle, on constate qu'au-delà de 5 km la concentration des étudiants converge vers 0. Les étudiants se concentrent également entre 25 km et 35 km.

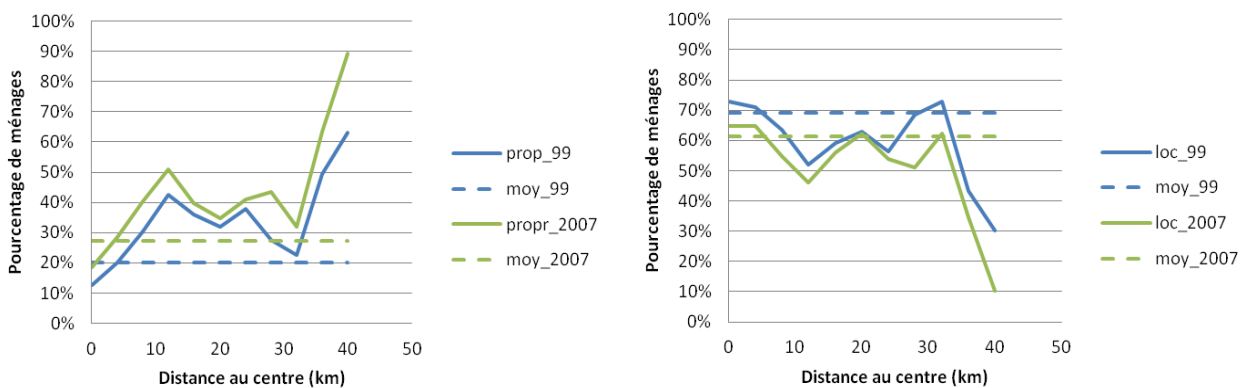
Graphique 8.9 : Concentration des étudiants (en fonction de la distance au centre (1999 et 2007))



Source : Élaboration propre ; à partir de la base ménages en 1999 et la base logement en 2007-INSEE

Le pourcentage de propriétaires est plus important au-delà de 30 km du centre, avec une légère augmentation entre 1999 et 2007.

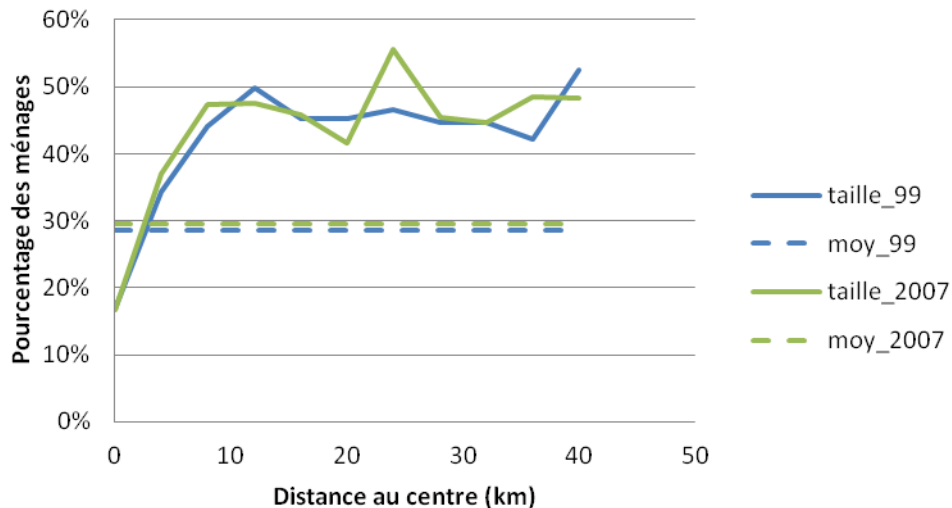
Graphique 8.10 : Variation des taux de locataires et de propriétaires (en fonction de la distance au centre (1999 et 2007))



Source : Élaboration propre ; à partir de la base ménages en 1999 et la base logement en 2007-INSEE

En 1999 et en 2007, on constate la concentration des familles nombreuses d'une taille supérieure ou égale à trois personnes dans les zones situées à plus que 10km

Graphique 8.11 : Variation des taux de ménages composés de 3 personnes ou plus (en fonction de la distance au centre (1999 et 2007))



Source : Élaboration propre ; à partir de la base ménages en 1999 et la base logement en 2007-INSEE

L'analyse des différentes variables entre 1999 et 2007 montre que le prix de l'immobilier et le réseau de transport en commun sont les variables qui ont connu un grand changement. L'analyse des résultats des estimations dans la section suivante va nous permettre de mesurer l'effet de ces changements sur les paramètres estimés.

8.2. Résultats des estimations et comparaison

Le travail de modélisation des choix de localisation des ménages en 1999 (Aissaoui et al., 2015) a montré que la qualité de l'environnement social est le facteur le plus déterminant, sans remettre en cause le poids toujours important de l'accessibilité à l'emploi. L'analyse des données de 2007 a montré que le prix de l'immobilier a triplé entre les deux dates. L'objectif de cette section est de tester l'effet de ces changements sur les comportements de localisation des ménages de l'aire urbaine. En estimant un modèle logit multinomial avec échantillonnage

des alternatives (6 alternatives+alternative choisie) et avec les mêmes variables que celles du modèle 1999 (tableau 6.17), nous obtenons les résultats suivants :

Tableau 8.1 : Résultats de l'estimation du modèle MNL en 2007

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0129	0,00017	76,3
	prix_imm	-0,0008	0,00002	-34,5
Aménité spatiale	prox_serv	0,9111	0,02090	43,6
	etb_prim	0,0987	0,00111	89,2
	etb_sec	0,0886	0,00386	22,9
	acc_jeux	0,0026	0,01018	0,3
Environnement social	taux3*rev3	2,6385	0,14315	18,4
	taux4*rev4	8,3862	0,19991	42,0
	taux5*rev5	3,9183	0,07619	51,4
	%hlm*rev1	0,8145	0,04275	19,1
	%hlm*rev2	-0,1788	0,05017	-3,6
	%hlm*rev3	-0,7822	0,05419	-14,4
	%hlm*rev4	-0,5245	0,07271	-7,2
	%hlm*rev5	-1,4602	0,07644	-19,1
Caractéristiques des ménages	peri*2VP+	0,6971	0,02185	31,9
	centre*VP0	1,3089	0,02330	56,2
	TC*VP0	0,3217	0,01973	16,3
	TC*VP1	0,2152	0,01315	16,4
	peri*loc	-0,6782	0,01876	-36,1
	peri*propr	-0,7995	0,02410	-33,2
	centre*loc	-0,0467	0,01582	-3,0
	centre*propr	-0,4634	0,01965	-23,6
	peri*tail>3	0,2700	0,01984	13,6
	centre*Etud	2,3978	0,05346	44,8

<i>Nombre d'observations</i>	<i>103 256</i>
<i>Log likelihood</i>	<i>-142 731</i>
<i>Log likelihood zero</i>	<i>-200 927</i>
<i>R² ajusté</i>	<i>0,290</i>

Source : Traitement Aissaoui sous R

Les résultats montrent que les coefficients ont gardé les mêmes signes et les mêmes valeurs.

L'analyse du poids de chaque déterminant (cf. Figure 8.1) montre que l'effet du prix de l'immobilier est devenu plus important. Cela peut être expliqué par l'augmentation du prix de l'immobilier entre 1999 et 2007.

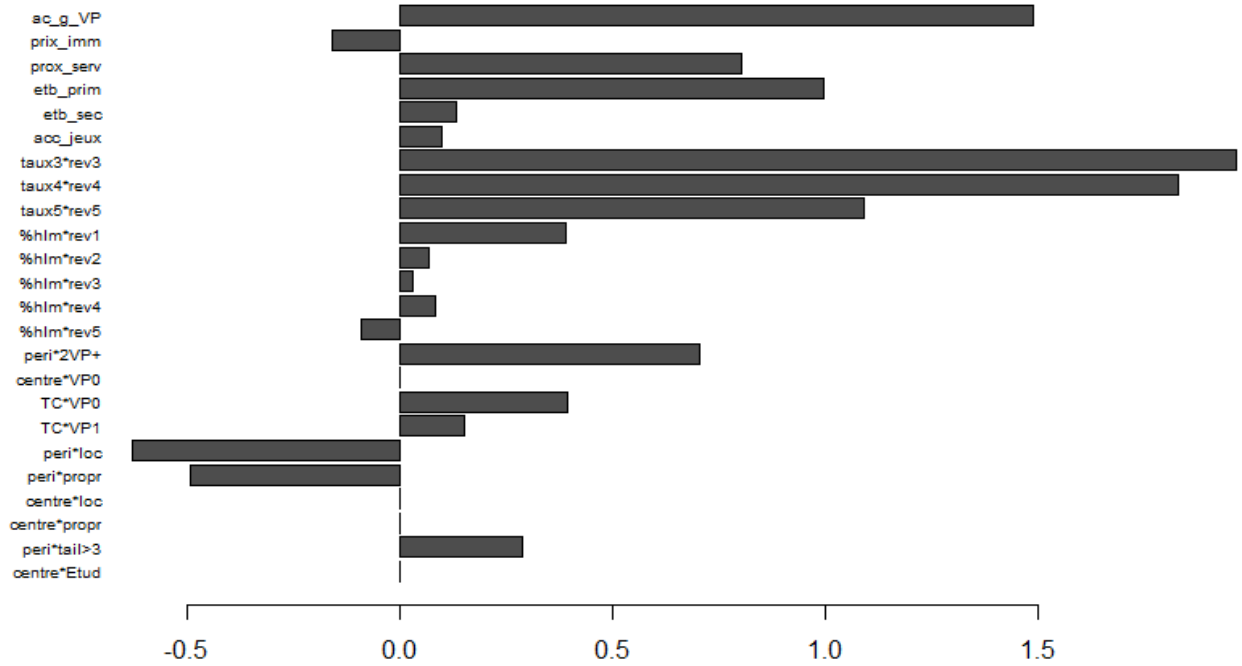
L'effet de l'environnement social a aussi changé. C'est probablement lié aux variations des taux de ménages appartenant aux différents quintiles entre les deux dates.

Le poids des autres déterminants est resté stable.

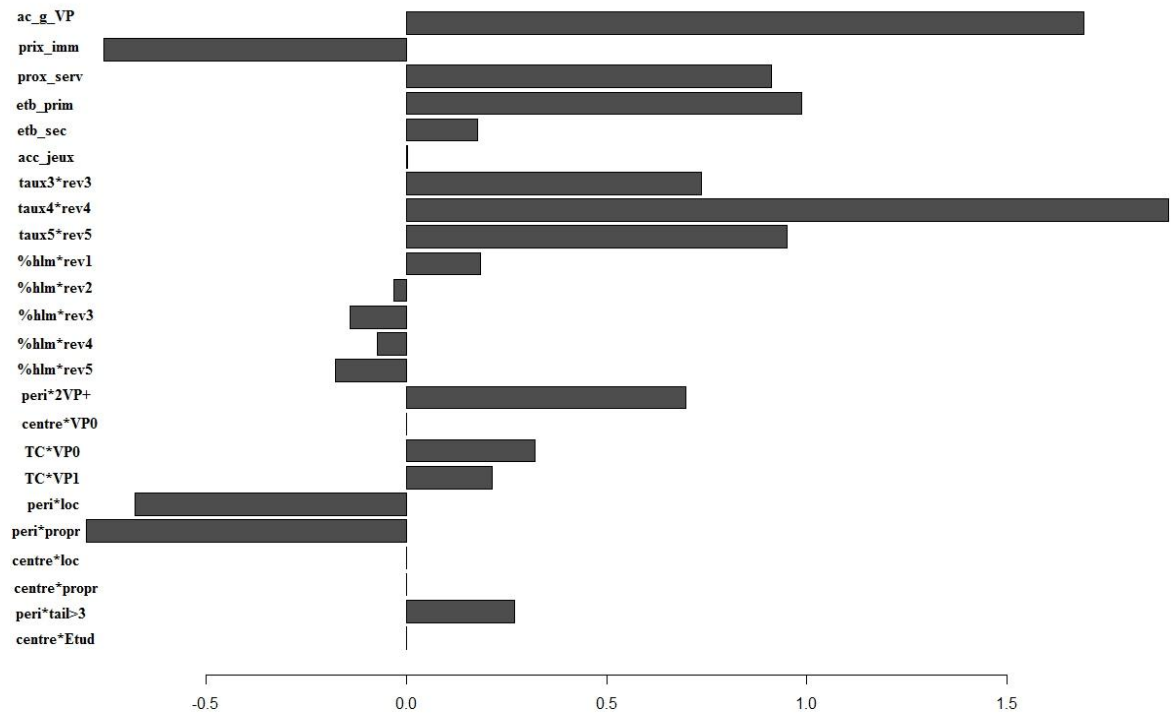
L'accessibilité isochronique est introduite à travers une variable binaire mesurant l'effet croisé du taux de motorisation du ménage avec le niveau de desserte en transport collectif de la zone de localisation. Les résultats obtenus confirment la préférence des ménages non motorisés à se localiser dans les zones bien desservies en transport en commun. L'effet de cette variable a légèrement baissé par rapport à 1999. Cela est dû à l'amélioration du réseau entre 1999 et 2007. Quant à l'effet relatif de l'accessibilité gravitaire à l'emploi en VP, il est resté stable. En effet, le nombre d'emplois n'a pas trop changé et le temps en voiture est resté relativement stable entre les deux dates.

Figure 8.1 : Différence d'utilités entre le 5ème et le 95ème percentile des valeurs des variables du modèle 1999 et 2007

1999



2007



Source : Traitement Aissaoui sous R

Nous avons également testé un modèle emboîté (cf. Tableau 8.2).

Tableau 8.2 : Résultats de l'estimation du modèle emboîté en 2007

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0148	0,00025	58,7
	prix_imm	-0,0015	0,00003	-46,8
Aménité spatiale	prox_serv	1,1984	0,02684	44,7
	etb_prim	0,1117	0,00132	84,8
	etb_sec	0,1091	0,00398	27,4
	acc_jeux	0,0475	0,01081	4,4
Environnement social	taux3*rev3	3,6166	0,14772	24,5
	taux4*rev4	10,0152	0,22617	44,3
	taux5*rev5	4,0693	0,06885	59,1
	%hlm*rev1	0,5194	0,04233	12,3
	%hlm*rev2	-0,4923	0,05097	-9,7
	%hlm*rev3	-0,9790	0,05670	-17,3
	%hlm*rev4	-0,5727	0,07772	-7,4
Caractéristiques des ménages	%hlm*rev5	-2,2394	0,08387	-26,7
	peri*2VP+	0,8356	0,02271	36,8
	centre*VP0	1,4063	0,02280	61,7
	TC*VP0	0,3521	0,02013	17,5
	TC*VP1	0,2573	0,01500	17,2
	peri*loc	-0,7473	0,02131	-35,1
	peri*propr	-0,9633	0,02647	-36,4
	centre*loc	-0,1477	0,01647	-9,0
	centre*propr	-0,6676	0,02013	-33,2
	peri*tail>3	0,3676	0,01997	18,4
centre*Etud	2,3685	0,04874	48,6	
Logsum	μ_1	0,8071	0,00703	114,7
	μ_2	0,7320	0,00675	108,5
	μ_3	0,7319	0,00902	81,1
	μ_4	0,7195	0,00787	91,4
	μ_5	0,7027	0,00607	115,8

<i>Nombre d'observations</i>	<i>103 256</i>
<i>Log likelihood</i>	<i>-313 357</i>
<i>Log likelihood zero</i>	<i>-385 544</i>
<i>R² ajusté</i>	<i>0,187</i>

Source : Traitement Aissaoui sous R

A l'opposé du modèle estimé en 1999, tous les paramètres sont significatifs. Toutefois, les valeurs des logsums sont toujours supérieures à 1.

La comparaison des effets des variables explicatives des deux dates montre un impact fort et stable dans le temps des variables liées à l'environnement social et à l'accessibilité. En revanche, d'autres facteurs tels que le prix de l'immobilier et la variable croisée mesurant l'effet combiné du statut d'occupation (propriétaire/locataire) et de la localisation (central/périphérique) et plus particulièrement la catégorie des propriétaires de la périphérie prennent de l'importance.

8.3. Analyse de la transférabilité temporelle

La comparaison des modèles de choix de localisation résidentielle des ménages de l'aire urbaine de Lyon en 1999 et 2007 montrent que la variation des variables explicatives entre ces deux dates implique une variation des effets relatifs à ces variables. La question qui se pose à ce niveau concerne la transférabilité de ces modèles. La littérature distingue deux types de transférabilité. La première concerne la transférabilité contextuelle et la deuxième la transférabilité temporelle. Dans cette section, on s'intéresse au deuxième type de transférabilité. Nous commençons par la présentation des tests qui peuvent être utilisés pour conduire une telle analyse (sous-section 8.3.1) et nous finissons par l'analyse des résultats de ces tests (sous-section 8.3.2).

8.3.1. Les tests de transférabilité

Deux tests peuvent être utilisés pour tester l'hypothèse de l'égalité des paramètres du modèle transféré et du modèle estimé :

- La statistique de test de transférabilité (TTS)

$$TTS = -2(LL_j(\theta_i) - LL_j(\theta_j)) \quad -8.1-$$

Cette statistique permet de tester l'hypothèse selon laquelle la valeur des paramètres calculés à partir d'une année i est égale à la valeur des paramètres estimée en une année j . Elle est égale au double de la valeur absolue de la différence de la log-vraisemblance calculée à partir des paramètres du modèle de l'année i et les données de l'année de prédiction j ($LL_j(\theta_i)$) et la log-vraisemblance estimée à partir des données de l'année j ($LL_j(\theta_j)$). Toutes choses égales par ailleurs, la log-vraisemblance du modèle transféré et la log-vraisemblance du modèle de l'année j sont égaux. Ce test suit une loi de khi 2 à k degrés de liberté telle que k est le nombre des paramètres estimés.

- L'indice de transfert (TI)

Cette statistique permet d'examiner dans quelle mesure la log-vraisemblance $LL_j(\theta_i)$ du modèle transféré est meilleure que la log-vraisemblance du modèle de référence $LL_j(\theta_c)$. Il s'agit de mesurer les informations apportées par rapport au modèle de référence à l'année j .

Une valeur négative implique que le modèle transféré est moins bon que le modèle appliqué avec les données.

$$TI = \frac{(LL_j(\theta_i) - LL_j(\theta_c))}{(LL_j(\theta_j) - LL_j(\theta_c))} \quad -8.2-$$

8.3.2. Résultats des tests

L'objectif étant de tester la capacité du modèle calibré avec les données de 1999 à prédire les choix de localisation en 2007, nous avons estimé un modèle avec les paramètres de l'année 1999 et les données de 2007. Nous avons obtenu une valeur de la log-vraisemblance égale à -144 407.

Le calcul de TTS a donné une valeur de 3351, ce qui implique le rejet de l'hypothèse de l'égalité des paramètres. L'indice TI est égale à 0,97, ce qui implique que l'information apportée par les coefficients du modèle estimé en 1999 est élevée.

L'étude faite par Sermons (2002) cité par Karasmaa (2003) a montré que les améliorations de la spécification du modèle pour mieux représenter la variation des préférences des ménages n'améliore pas nécessairement la transférabilité du modèle. En effet, les modèles complexes sont difficiles à transférer. L'inclusion des caractéristiques des ménages améliore la portée modélisatrice du modèle. Toutefois les ménages qui ont déménagé entre 1998 et 1999 non pas forcément les mêmes caractéristiques que ceux qui ont déménagé entre 2006 et 2007. Malgré

que les paramètres ne soient pas temporellement stables, l'indice TI ainsi que la comparaison des estimations des deux dates nous montrent que le modèle transféré peut donner des informations utiles pour 2007.

Les travaux sur la transférabilité temporelle montrent un arbitrage entre la transférabilité du modèle et leur portée modélisatrice. En utilisant plusieurs spécifications des modèles de choix, l'analyse faite par Badoe et Miller (1995) a montré que le changement de la spécification du modèle peut améliorer leur transférabilité. L'objectif de cette thèse étant plutôt de comprendre les choix de localisation, nous avons privilégié le modèle qui explique mieux les choix de localisation.

Le modèle estimé en 1999 est-il transférable ?

La comparaison des résultats de l'analyse empirique des choix de localisation en 1999 et 2007 confirme l'impact fort et stable dans le temps de l'accessibilité et de l'environnement social. Elle a montré également que l'effet du prix de l'immobilier et du statut d'occupation croisé à la localisation (central/périphérique) et plus particulièrement pour la catégorie des propriétaires de la périphérie a pris de l'importance.

Les résultats des tests menés pour analyser la transférabilité du modèle montre que le modèle est non transférable tout en apportant des informations utiles sur les tendances pour 2007.

Si ces résultats mettent en avant la stabilité de l'accessibilité sur laquelle repose les modèles LUTI et donnent d'ores et déjà une idée sur l'évolution des comportements des ménages en matière de choix d'une zone de localisation, la question de la transférabilité temporelle reste à creuser afin d'apporter des réponses sur le pouvoir prédictif de notre modèle.

Conclusion

A travers cette partie, nous avons essayé de répondre empiriquement aux différentes questions évoquées dans les deux premières parties, à savoir la question liée à la compréhension des arbitrages faits par les ménages en matière de choix d'une localisation et celle d'ordre

méthodologique et plus particulièrement la structure du modèle qui permet de mieux représenter spatialement et temporellement les choix de localisation. En retenant l'aire urbaine de Lyon comme cas d'étude pour confronter les approches théoriques aux choix observés des ménages, les chapitres de cette partie ont permis d'analyser empiriquement les choix de localisation des ménages en 1999 et 2007. A travers le chapitre 5 nous avons montré que l'utilisation d'une échelle intermédiaire (Grand quartier) entre un zonage administratif (commune) et un zonage statistique (IRIS) permet de résoudre partiellement le problème d'autocorrélations spatiales et de construire des zones spatialement homogènes. Plusieurs sources de données sont ensuite exploitées afin de rassembler dans une seule base de données les différents facteurs susceptibles d'expliquer les choix des ménages en matière de localisation à une échelle qui correspond au Grand quartier. La mesure du poids de chaque déterminant à l'aide d'un modèle logit multinomial simple (chapitre 6) a montré que le poids des paramètres liés à la qualité de l'environnement social est le plus élevé ce qui confirme que le choix de localisation n'est pas le résultat d'un simple arbitrage entre le coût du transport et le prix de l'immobilier mais est lié aussi aux caractéristiques des ménages et à leurs préférences pour les aménités sociales. En effet, les ménages à haut revenu accordent une importance simultanée à l'accessibilité et aux aménités et non pas une préférence aux aménités au détriment de l'accessibilité comme c'est le cas dans le contexte américain.

L'analyse des différents types de modèles de choix discrets afin de trouver un compromis entre leurs hypothèses restrictives et leur complexité de calculs (partie II) a montré que théoriquement le modèle logit emboîté reste le plus approprié pour traiter à la fois les problèmes d'autocorrélations spatiales et de multitude d'alternatives de choix de localisation. Toutefois, l'application empirique d'un tel modèle pour modéliser les choix de localisation résidentielle des ménages de l'aire urbaine de Lyon en 1999 (chapitre 7) présente plusieurs difficultés et ne débouche pas sur une amélioration de la portée modélisatrice du modèle. Ces résultats sont certes à prendre avec une certaine prudence car le développement méthodologique n'est pas encore stabilisé.

Bien que les résultats des tests de la transférabilité temporelle ont montré que notre modèle est non transférable dans le temps, la modélisation des choix de localisation des ménages de l'aire urbaine de Lyon en 2007 a montré l'impact toujours fort et stable de l'accessibilité et de l'environnement social, ce qui confirme que les choix des ménages ne sont pas seulement guidés par les contraintes du marché mais ils sont aussi le résultat des préférences des ménages pour l'environnement social.

Conclusion générale

Les aires urbaines connaissent de plus en plus une accélération du processus de métropolisation. Dans le contexte d'étalement urbain, les distances parcourues et les déplacements deviennent plus importants, et par conséquent la mobilité quotidienne évolue. Cette évolution n'est pas sans coûts liés principalement à la dépendance de l'automobile : augmentation de l'émission de CO₂, du coût du carburant, problème d'équité sociale, etc. Pour faire face à ces impacts environnementaux, économiques et sociaux, une planification à long terme devient indispensable.

C'est dans cette perspective que se situe ce projet de thèse. Il propose dans un premier temps de modéliser les choix de localisation résidentielle, en s'interrogeant sur l'intérêt d'une différenciation des ménages en fonction de leurs caractéristiques socioéconomiques dans le cadre des réflexions stratégiques de politiques d'aménagement urbain : qui réside où, et quels enjeux cela représente-t-il en matière de cohésion sociale, de ségrégation ? Dans un deuxième temps, nous questionnons l'apport de l'utilisation d'une structure hiérarchique pour appréhender l'aspect spatial des choix de localisation. Enfin, nous abordons la question de la transférabilité temporelle des modèles et de l'évolution éventuelle des comportements au fil du temps.

La modélisation de ces choix est confrontée à plusieurs difficultés liées à la multidimensionnalité des facteurs explicatifs du choix de localisation, au choix du modèle adapté au contexte du choix de localisation résidentielle (autocorrélation spatiale, grand nombre d'alternatives de choix), et à la définition de l'échelle de désagrégation appropriée pour l'estimation.

Partant de ces constats, l'objectif de cette thèse était d'enrichir le corpus des travaux sur les déterminants de choix de localisation, en investiguant les développements récents des modèles de choix discrets en présence d'un nombre élevé d'alternatives pour analyser empiriquement les choix de localisation résidentielle de l'aire urbaine de Lyon pour comprendre les arbitrages faits par les ménages en matière de choix d'une localisation.

L'analyse des travaux empiriques traitant les déterminants des choix de localisation résidentielle (tableaux 2.1 à 2.3, chapitre 2) a mis en avant la nécessité d'intégrer dans le modèle de choix de localisation des variables qui permettent de mesurer les effets liés à quatre dimensions de choix de localisation, à savoir l'arbitrage du marché entre coût de transport et

distance au centre, les aménités spatiales, l'environnement social et les caractéristiques des ménages.

En s'appuyant constamment sur cette typologie dans nos analyses empiriques, nous avons commencé par la construction d'une base de données riche en information caractérisant le ménage et la zone de localisation à travers une liste de variables permettant de mesurer directement les effets liés à ces quatre dimensions.

La structure simple du logit multinomial nous a permis d'estimer un modèle de base en 1999. Par ailleurs, la propriété d'indépendance par rapport aux choix non retenus fondamentale pour l'estimation d'un modèle logit multinomial a nécessité le test de plusieurs échelles spatiales afin de choisir celle qui permet de traiter partiellement le problème d'autocorrélation spatiale et de constituer des entités géographiques relativement distinctes entre elles et homogènes pour chacune d'elles afin de respecter en moins en partie cette propriété. A l'issue de ces tests, nous avons retenu les Grands quartiers comme échelle d'estimation. Il s'agit d'un zonage intermédiaire entre un zonage administratif (Commune) et statistique (IRIS).

En termes de résultats, la modélisation des choix de localisation résidentielle des ménages en 1999 a montré que la qualité de l'environnement social est le facteur le plus déterminant, mais ne remet pas en cause le rôle des contraintes du marché (accessibilité à l'emploi et prix de l'immobilier). A l'inverse du contexte américain où la préférence des ménages à haut revenu pour les aménités sociales est plus importante que l'accessibilité à l'emploi, les ménages de l'aire urbaine de Lyon cherchent à être en soi tout en restant dans des zones accessibles à l'emploi et aux aménités centrales. L'importance accordée simultanément aux facteurs d'accessibilité et à l'environnement social par les ménages peut être une opportunité pour les acteurs publics permettant d'articuler politiques de transport et politiques de logement et renouveler l'attractivité des quartiers les moins favorisés. La ségrégation socio-spatiale est à la fois conséquence et cause des choix de localisations résidentielles (Figure 2).

Au plan méthodologique, le questionnement des différentes structures dans les chapitres 3 et 4 nous a conduit à conclure que théoriquement le modèle emboîté permet de traiter les problèmes liées à l'aspect géographique de ces choix, à savoir le problème d'autocorrélation spatiale et la contrainte de la multitude des alternatives. Toutefois, l'application de la nouvelle méthodologie développée par Guevara et Ben-Akiva (2013a) pour corriger les biais liés à l'échantillonnage d'alternatives pour analyser empiriquement les choix de localisation des ménages de l'aire urbaine de Lyon a montré des problèmes de consistances de ce type de modèle. Plusieurs structures hiérarchiques ont été testées pour regrouper sous le même nid les

Grand quartiers qui se ressemblent, sans pouvoir pour autant résoudre ce problème de consistance. Cela est liée en grande partie à la difficulté de l'application de cette méthode sur des bases de données complexes (Daly et al, 2014)

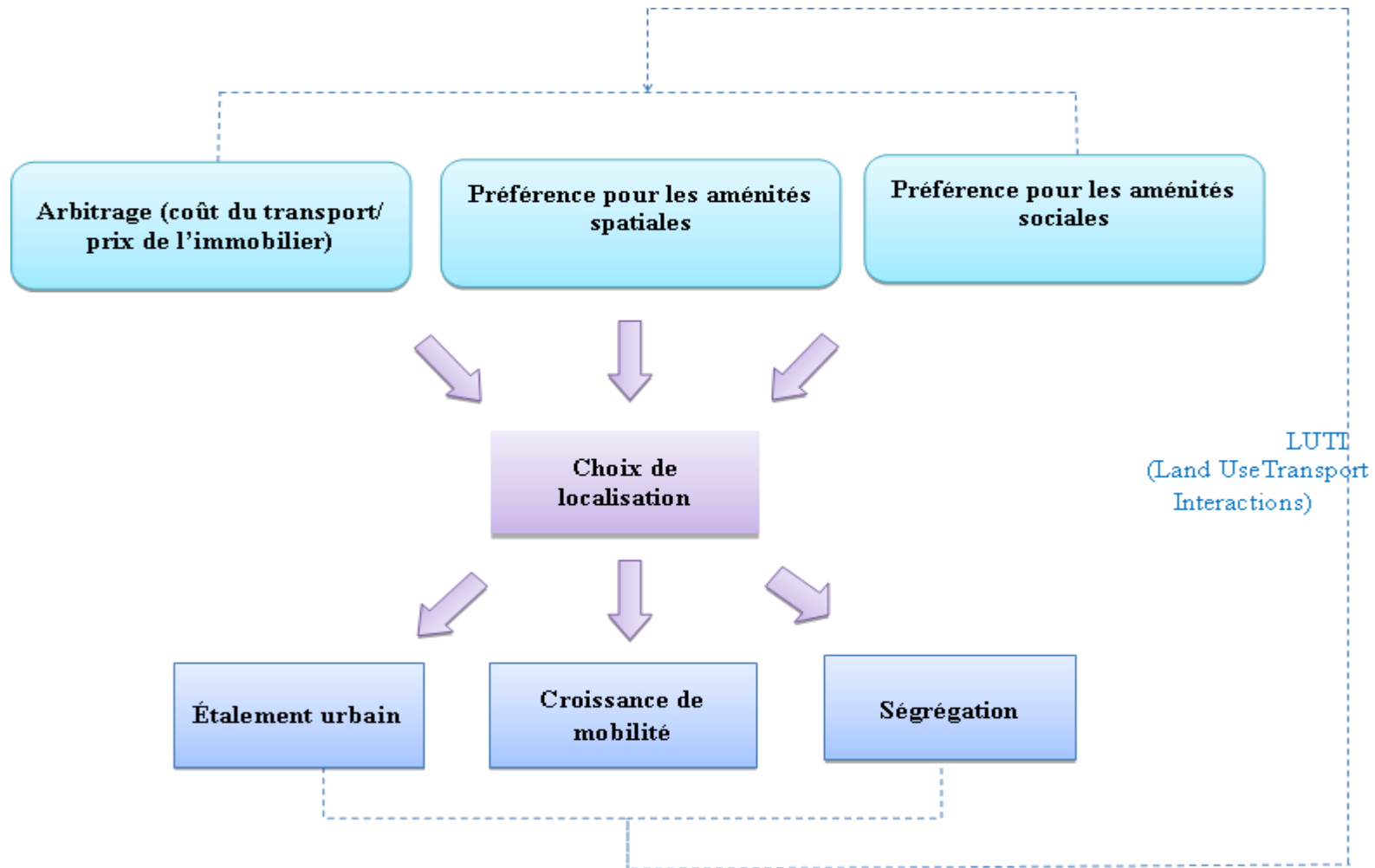
Pour ce qui est de l'évolution temporelle, le calage du modèle de choix de localisation retenu en 1999 sur les données de 2007 a confirmé l'impact fort et stable dans le temps de l'accessibilité et de l'environnement social. Il a montré également que l'effet du prix de l'immobilier et du statut d'occupation croisé à la localisation (central/périphérique) et plus particulièrement pour la catégorie des propriétaires de la périphérie a pris de l'importance. Toutefois, les résultats des tests de transférabilité ont montré que notre modèle n'est pas transférable dans le temps.

En termes d'approfondissement possibles, le travail sur seulement deux dates n'a permis de répondre définitivement sur la transférabilité de notre modèle. Le travail sur d'autres dates dans le futur pourra apporter plus d'éléments de réponse.

La comparaison de plusieurs villes dans le chapitre 2 a montré la variation des effets des variables explicatives de choix de localisation résidentielle en fonction de la forme urbaine, cela mérite une analyse exploratoire du lien entre la structure urbaine et le comportement des ménages en matière de choix d'une zone de localisation.

Le questionnement de différentes structures de modèles de choix discrets (chapitre3) a montré que les modèles développés par Bhat et Guo (2004) permettent de traiter le problème des autocorrélations spatiales. Toutefois, l'absence d'une méthode qui permet de corriger les biais liés à l'échantillonnage dans le cas de ces modèles limitent leur utilisation dans les analyses empiriques. Le développement d'une méthode, à partir de celle développée par Guevara et Ben-Akiva (2013a, 2013b) est aussi à approfondir.

Figure 2 : Articulation entre les choix de localisation et modèles LUTI



Source : *Élaboration propre*

Bibliographie

- Abraham, J., & Hunt, J. (1997). Specification and estimation of nested logit model of home, workplaces, and commuter mode choices by multiple-worker households. *Journal of the Transportation Research Board*, 1606, 17-24.
- Aissaoui, H., Bouzouina, L., & Bonnel, P. (2015). Choix de localisation résidentielle, entre contraintes du marché et préférences individuelles : Application à l'aire urbaine de Lyon (1999). *Revue d'Économie Régionale et Urbaine*, 4, 629-656.
- Alonso, W. (1964). *Location and land use : Toward a general theory of land rent*. Harvard Univ Press, Cambridge.
- Anas, A. (1982). *Residential location markets and urban transportation : Economic theory, econometrics, and policy analysis with discrete choice models*. Academic Press, New York.
- Anas, A. (1990). Taste Heterogeneity and Urban Spatial Structure : The Logit Model and Monocentric Theory Reconciled. *Journal of Urban Economics*, 28(3), 318–335.
- Anas, A. (1994). METROSIM : A unified economic model of transportation and land use. Unpublished system description, Willamsville, New York.
- Anas, A., Arnott, R.J., & Small, K.A. (1998). Urban Spatial Structure. *Journal of Economic Literature*, 36, 1426-1464.
- Andrew, M., & Meen, G. (2006). Population structure and location choice : A study of London and South East England. *Regional Science*, 85(3), 401-419.
- Badoe, D. A., Miller, E. J. (1995). Analysis of the temporal transferability of disaggregate work trip mode choice models. *Journal of the Transportation Research Board*, 1493, 1–11.
- Bayoh, I., Irwin, E.G., Haab, T. (2006). Determinants of Residential Location Choice : How Important Are Local Public Goods in Attracting Homeowners to Central City Locations?. *Journal of Regional Science*, 46, 97–120.

- Ben-Akiva, M. (1973). Structure of passenger travel demand models, PhD thesis, Massachusetts Institute of Technology, Cambridge.
- Ben-Akiva, M. (1974). Structure of passenger travel demand models. *Transportation Research Record*, 526, 26-42.
- Ben-Akiva, M., & Bierlaire, M. (1999). Discrete Choice Methods and Their Applications to Short Term Travel Decisions. In : Randolph W. Hall (Ed.), *Handbook of Transportation Science* (pp. 5-33). Springer, US.
- Ben-Akiva, M., & Bowman, J.L. (1998). Integration of an activity-bases model system and a residential location model. *Urban Studies*, 35(7), 1131–1153.
- Ben-Akiva, M., & Lerman, S.R. (1985). *Discrete choice analysis : Theory and application to travel demand*. MIT Press, Cambridge.
- Bhat, C.R., & Guo, J. (2004). A mixed spatially correlated logit model : Formulation and application to residential choice modeling, 2004. *Transportation Research Part B*, 38, 147–168.
- Bolduc, D., & Ben-Akiva, M. (1991). A multinomial probit formulation for large choice sets. *Proceedings of 6th international conference on travel behavior*, 2, 243-258.
- Bolduc, D., & Kaci, M. (1993). Estimation des modèles probit polytomiques : Un survol des techniques. *L'Actualité économique*, 69(3), 161-191.
- Bonnafous, A. (1989). *Le siècle des ténèbres de l'économie*, Economica, Paris.
- Bonnel, P. (2004). *Prévoir la demande de transport*. Presses de l'école nationale des ponts et chaussées, Paris.
- Bonvalet, C. (2010). Les logiques des choix résidentiels des franciliens. In : Authier J.Y., Bonvalet, C., & Lévy, J.P. (dir.), *Elire domicile* (pp. 55-76). Presses Universitaires, Lyon.
- Bouzouina, L. (2008). *Ségrégation spatiale et dynamiques métropolitaines*. Thèse pour le doctorat en sciences économiques, Université Lumière Lyon 2.
- Brownstone, D., & Train, K. (1999). Forecasting new product penetration with flexible substitution Patterns. *Journal of Econometrics*, 89, 109-129.
- Brueckner, J., Thisse, J.F., & Zenou, Y. (1999). Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review*, 43, 91-107.

- Bürgle, M. (2006). Residential location choice model for the Greater Zurich area. Paper presented at the 6th Swiss Transport Research Conference, Monte Verità / Ascona
- Cabrera Delgado, J. (2013). *Quelle Prise en Compte des Dynamiques Urbaines dans la Prévion de la Demande de Transport ?* Thèse pour le doctorat en Sciences Économiques, Université Lumière Lyon 2.
- Chattopadhyay, S. (2000). The effectiveness of McFaddens's nested logit model in valuing amenity improvement. *Regional Science and Urban Economics*, 30, 23–43.
- Chen, Y.C., Duann, L.S., & Hu, W.P. (2005). The estimation of discrete choice models with large choice set. *Journal of the Eastern Asia Society for Transportation Studies*, 6, 1724-1739.
- Chen, J., Chen, C., & Timmermans, H. (2008). Accessibility trade-offs in household residential location decisions. *Transportation Research Record*, 2077, 71–79.
- Clark, W.A.V. (1986). Residential segregation in American cities : A review and interpretation. *Population Research and Policy Review*, 5, 95-129.
- Clark, W.A.V., Huang, Y., & Withers, S. (2003). Does commuting distance matter? Commuting tolerance and residential change. *Regional Science and Urban Economics*, 33, 199-221.
- Clark, W.A.V, Deurloo, M.C., & Dieleman, F.M. (2006). Residential Mobility and Neighbourhood Outcoms. *Housing Studies*, 21(3), 323-342.
- Cornelis, E., Barthelemy, J., Pauly, X., & Walle, F. (2012). Modélisation de la mobilité résidentielle en vue d'une micro-simulation des évolutions de population. *Les Cahiers Scientifiques Du Transport* , 62, 65-84.
- Cornuel, D. (2010). Choix résidentiel et analyse économique. In : Authier, J.Y., Bonvalet, C., & Lévy, J.P. (dir.), *Elire domicile* (pp. 15-34). Presses Universitaires, Lyon.
- Daly, A., & Zachary, S. (1979). Improved multiple choice models. In : Hensher, D., & Dalvi, Q. (Eds.), *Identifying and measuring the determinants of mode choice* (pp. 335-357). Teakfield Ltd, London.
- Daly, A., Hess, S., & Dekker, T. (2014). Practical solutions for sampling alternatives in large scale models. *Journal of the Transportation Research Board*, 2429, 148-156.

- Dantan, S., & Picard, N. (2009). Rôle des préférences et des contraintes dans le phénomène de ségrégation résidentielle : Cas de l'Ile-de-France. Un papier présenté au 26ème journée de microéconomie, Dijon.
- Debreu, G. (1960). Review of R.D. Luce, Individual Choice Behavior : A theoretical Analysis. *American Economic Review*, 50, 186-188.
- de Palma, A., & Thisse, J.F. (1987). Les modèles de choix discrets. *Annales d'Économie et de statistiques*, 9, 151-190.
- de Palma, A., Motamedi, K., Picard, N., & Waddell, P. (2007). Accessibility and environmental quality : Inequality in the Paris housing market. *European transport*, 36, 47-74.
- de Palma, A., Picard, N., & Waddell, P. (2007). Discrete choice models with capacity constraints : An empirical analysis of the housing market of the greater Paris region. *Journal of urban economics*, 62, 204-230.
- DeSalvo, J.S., & Huq, M. (2005). Mode Choice, Commuting Cost, and Urban Household Behavior. *Journal of Regional Science*, 45(3), 493-517.
- Ettema, D. (2010). The impact of telecommuting on residential relocation and residential preferences. *The journal of transport and land use*, 3, 7-24.
- Feridhanusetyawan, T., & Kilkenny, M. (1996). Rural/Urban residence location choice. Working Paper, n°157, Center for Agricultural and Rural Development, Iowa State University.
- Filion, P., Bunting, T., & Warriner, K. (1999). The Entrenchment of Urban Dispersion : Residential Preferences and Location Patterns in the Dispersed City. *Urban Studies*, 36(8), 1317-1347.
- Fujita, M. (1989). *Urban Economic Theory : Land use and city size*. Cambridge University Press, Cambridge.
- Gabriel, S.A., & Rosenthal, S.S. (1989). Household Location and Race: Estimates of a Multinomial Logit Model. *The Review of Economics and Statistics*, 71(2), 240-249.
- Galster, G. (1988). Residential segregation in American cities : A contrary review. *Population Research and Policy Review*, 7, 113-121.

- Gaschet, F., & Le Gallo, J. (2005). The spatial dimension of segregation : A case study in four urban areas. Cahiers du Groupement de recherches économiques et sociales, 12-24.
- Gayda, S.(1998). Stated preference survey on residential location choice and modal choice in Brussels. Paper presented at the 8th World Conference on Transportation Research - Antwerp, Belgium.
- Gérardin, B., Patier, D., Routhier, J.L., & Ségalou, E. (2000). Diagnostic du transport de marchandises dans une agglomération. Rapport pour le compte du programme national marchandises en ville, LET pour le compte de la DRAST, Lyon.
- Gilbert, F. (2012). Tarification logit dans un réseau. Thèse pour le doctorat en Informatiques, Université de Montréal.
- Glaeser, E.L., & Kahn, M.E. (2004). Sprawl and urban growth. In : Henderson, V.J., & Thisse, J.F. (Eds.), Handbook of regional and urban economics (pp. 2481-2527). Elsevier, Amsterdam.
- Glaeser, E.L., Kahn, M.E., & Rappaport, J. (2008). Why do the poor live in cities? The role of public transportation. Journal of Urban Economics, 63, 1–24.
- Godinot, C., Bonnel, P., & Nicolas, J.P. (2008). Phase d'affectation : Mise en forme du réseau de routier sur l'Aire Urbaine de Lyon. Rapport intermédiaire n°8, LET, Lyon.
- Goffette-Nagot, F., Schaeffer, Y. (2013). Accessibilité ou voisinage ? Une analyse des sources de la ségrégation résidentielle au sein des aires urbaines françaises. Revue économique, 64, 857-882.
- Guevara, A.C. (2010). Endogeneity and Sampling of Alternatives in Spatial Choice Models. PhD thesis, Massachusetts Institute of Technology, Cambridge.
- Guevara, A.C., & Ben-Akiva, M. (2013a). Sampling of alternatives in multivariate extreme value (MEV) models, Transportation Research Part B, 48(1), 35-52.
- Guevara, A.C., & Ben-Akiva, M. (2013b). Sampling of alternatives in Logit Mixture models, Transportation Research Part B, 58(1), 185-198.
- Gueymard, S. (2006). Facteurs environnementaux de proximité et choix résidentiels Le rôle de l'ancrage communal, des représentations et des pratiques des espaces verts. Développement durable et territoires (En ligne), URL :

<http://developpementdurable.revues.org/2716>.

- Guo, J.Y., & Bhat, C.R. (2001). Residential Location Choice Modeling : Accommodating Sociodemographic, School Quality and Accessibility Effects. Working paper, University of Texas, Austin.
- Guo, J.Y. (2004). Addressing spatial complexities in residential location choice models. PhD thesis, University of Texas.
- Guo, J.Y., & Bhat C.R. (2007). Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography*, 15, 31–45.
- Habib, M.A., & Miller, E.J. (2007). Modeling Residential and Spatial Search Behaviour : Evidence from the Greater Toronto Area. Paper presented at the sixth Triennial Symposium on Transportation Analysis, Phuket Island-Thailand.
- Hamilton, B.W. (1982). Wasteful Commuting. *Journal of Political Economy*, 90(5), 1035-1053.
- Hansen, W.G. (1959). How accessibility shapes land use. *Journal of the American institute of planners*, 25, 73-76.
- Hausman, J., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*, 52(2), 1219-1240.
- Henderson, V., & Slade, E. (1993). Development Games in Non-monocentric Cities. *Journal of Urban Economics*, 34, 207-229.
- Herbert, D.J., & Stevens, B.H. (1960). A model for the distribution of residential activity in urban areas, *Journal of Regional Science*, 2, 21-36.
- Hoang Huu, P., & Wakely, P. (2000). Status, Quality and the Other Trade-off : Towards a New Theory of Urban Residential Location. *Urban studies*, 37, 7–35.
- Holland, D.S., & Sutinen, J.G. (2000). Location Choice in New England Trawl Fisheries : Old Habits Die Hard. *Land Economics*, 76(1), 133-149.
- Homocianu, G. (2009). Modélisation de l'interaction transport-urbanisme - choix résidentiels des ménages dans l'aire urbaine de Lyon. Thèse pour le doctorat en Sciences Économiques, Université Lumière Lyon 2.
- Hunt, J.D. (2010). Stated Preference Examination of factors influencing residential attraction. Residential location choice, Springer. In : Pagliara, F., Preston, J.,

- Simmonds, D. (Eds.), *Residential Location Choice* (pp 21-59). Springer, Berlin Heidelberg.
- Hunt, L.M., Boots, B., & Kanaroglou, P.S. (2004). Spatial choice modelling : New opportunities to incorporate space into substitution patterns. *Progress in Human Geography*, 28(6), 746-766.
- Hunt, J.D., Kriger, D.S., & Miller, E.J. (2005). Current operational urban land-use-transport modeling frameworks: A review. *Transport Reviews*, 25(3), 329-376.
- Hurtubia, R., Flötteröd, G., & Bierlaire, M. (2010). Estimation techniques for MEV models with sampling of alternatives. Paper presented at the European Transport Conference, Glasgow, Scotland.
- Ibeas, A., Cordera, R., dell'Olio, L., & Coppola, P. (2013). Modelling the spatial interactions between workplace and residential location. *Transportation Research Part A*, 49, 110-122.
- Karasmaa, N. (2003). *The Transferability of Travel Demand Models: An analysis of transfer methods, data quality and model estimation*, PhD thesis, Helsinki University of Technology.
- Kling, C.L., & Herriges, J.A. (1995). An Empirical Investigation of the Consistency of Nested Logit Models with Utility Maximization. *American Journal of Agricultural Economics*, 77(4), 875-884.
- Kim, J.H. (2010). *Land Use, Spatial Structure, and Regional Economic Performance : Assessing the Economic Effects of Land Use Planning and Regulation*, PhD thesis. University of Illinois.
- Kim, J.H., Pagliara, F., & Preston, J. (2005). The Intention to Move and Residential Location Choice Behaviour. *Urban Studies*, 42, 1621–1636.
- Koppelman, F.S., & Wen, C.H. (1998). Alternative nested logit models : Structure, properties and estimation. *Transportation Research Part B*, 32(5), 289-298.
- Koppelman, F.S., & Wen, C.H. (2000). The paired combinatorial logit model : Properties, estimation and application. *Transportation Research Part B*, 34, 75-89.
- Kryvobokov, M., & Bouzouina, L. (2014). Willingness to pay for accessibility under the conditions of residential segregation. *International Journal of Strategic Property Management*, 18(2), 101-115.

- Le Corre, R. (1983). Localisation résidentielle et mobilité intra-urbaine : Esquisse d'un cadre théorique. *Cahiers québécois de démographie*, 12(1), 141-154.
- Lee, B.H.J., Beamish, J.O., & Goss, R.C. (2008). Location preferences of multifamily housing residents, *Housing and Society*. 35(1), 41-58.
- Lee, B.H.L., & Waddell, P. (2010). Residential Mobility and Location Choice : A Nested Logit Model with Sampling of Alternatives. *Transportation*, 37, 587-601.
- Lee, B.H.L., Waddell, P., Wang, L., & Pendyala, R.M. (2010a). Reexamining the influence of work and nonwork accessibility on residential location choices with a microanalytic framework. *Environment and Planning A*, 42, 913-930.
- Lee, B.H.Y., Waddell, P., Wang, L., & Pendyala, R.M. (2010b). Operationalizing Time-Space Prism Accessibility in a Building-Level Residential Choice Model : Empirical Results from the Puget Sound Region. Paper presented at the 88st Annual Meeting of the transportation Research Board (TRB), Washington DC, US.
- Le Gallo, J. (2000). *Économétrie spatiale*, 1. Autocorrélation spatiale. Document de travail, n°2000-05, LATEC, Université de Bourgogne.
- Lerman, S. (1976). Location, housing, automobile ownership, and mode to work : A joint choice model. *Transportation Research Record*, 610, 6-11.
- Leroy, S.F., & Sonstelie, J. (1983). Paradise lost and regained : Transportation innovation, income, and residential location. *Journal of Urban Economics*, 13, 67-89.
- Li, J., Lai, X., & Yu, Z. (2013). A paired combinatorial logit route choice model with probit-based equivalent impedance. *Journal of transportation systems engineering and information technology*, 13(4), 100-105.
- Lowry, I. (1964). *A Model of Metropolis*. The Rand Corporation, Santa Monica.
- Luce, R.D. (1959). *Individual Choice Behavior : A Theoretical Analysis*, Wiley, New York.
- Manski, C.F. (1977). The Structure of random utility models. *Theory and decision*, 8, 229-254.
- Marschak, J. (1960). Binary choice constraints and random utility indicators. In : Arrow, K.J., Karlin, S., & Suppes, P. (Eds.), *Mathematical methods in the social science* (pp. 312-329), Stanford University Press, Stanford.

- Masson, S. (1998). Interactions entre système de transport et système de localisation : De l'héritage des modèles traditionnels à l'apport des modèles interactifs de transport et d'occupation des sols. *Les Cahiers Scientifiques du Transport*, 33, 79-108.
- Maurin, E. (2004). *Le ghetto français : Enquête sur le séparatisme social*. La République des idées / Seuil, Paris.
- McFadden, D. (1978). Modelling the choice of residential location. In : Karlqvist, A., Lundqvist, L., Snickars, F., & Weibull, J (Eds.). *Spatial Interaction Theory and Planning Models* (pp. 75-96). North- Holland, Amsterdam.
- Molin, E., & Timmermans, H. (2002). Accessibility considerations in residential choice decisions : Accumulated evidence from the Benelux. Paper presented at the Annual transportation research board meeting, Washington D.C.
- Nerella, S., & Bhat, C.R. (2004). A Numerical Analysis of the Effect of Sampling of Alternatives in Discrete Choice Models. *Journal of the Transportation Research Board*, 1894, 11-19.
- Ng, C.F. (2008). Commuting distances in a household location choice model with amenities. *Journal of Urban Economics*, 63, 116-129.
- Nicolas, J.P.(2010). SIMBAD : Un outil pour intégrer le développement durable dans les politiques publiques. In : Antoni, J.P. (Ed.), *Modéliser la ville : Formes urbaines et politiques de transport* (pp. 119-164). Economica, Paris.
- Oates, W.E. (1969). The effects of Property taxes and Local Public Spending on Property Values : An Empirical Study of Tax Capitalization and the Tiebout Hypothesis, *Journal of Political Economy*, 77(6), 957-971.
- Openshaw, S. (1977). Optimal zoning systems for spatial interaction models. *Environment and Planning A*, 9(2), 169–184.
- Openshaw, S. (1984). *The Modifiable Areal Unit Problem : Concepts and Techniques in Modern Geography*. Geo Books, Norwich, Royaume-Uni.
- Parsons, G.R., & Kealy, M.J. (1992). Randomly drawn opportunity sets in a random utility model of lake recreation. *Land Economics*, 68(1), 93-106.
- Ozturk, E., & Irwin, E.G. (2001). Explaining Household Location Choices Using a Spatial Probit Model. Paper presented at the American Agricultural Economics Association meeting Chicago, IL.

- Pagliara, F., & Wilson, A. (2010). The state-of-the-art in building residential location models. In : Pagliara, F., Preston, J., Simmonds, D. (Eds.), *Residential Location Choice* (pp. 1–20). Springer, Berlin Heidelberg.
- Pinçon, M., Pinçon-Charlot, M. (2004). *Sociologie de Paris. La découverte*, Paris.
- Pravinvongvuth, S., & Chen, A. (2005). Adaptation of the paires combinatorial logit model to the route choice problem. *Transportmetrica*, 1(3), 223-240.
- Putman, S.H. (2010). DRAM Residential Location and Land Use Model : 40 Years of Development and Application. In : Pagliara, F., Preston, J., Simmonds, D. (Eds.), *Residential Location Choice* (pp. 61–76). Springer, Berlin Heidelberg.
- Quigley, J.M. (1985). Consumer choice of dwelling, neighborhood and public services. *Regional Science and Urban Economics*, 15, 41–63.
- Quigley, J.M., & Weinberg, D.H. (1977). Intra- Urban Residential Mobility : A Review and Synthesis. *International Regional Science Review*, 2, 41–66.
- Rakotomalala, R. (2014). *Pratique de la régression logistique. Régression logistique binaire et polytomique. Note de cours (Version 2.0)*, Université Lumière Lyon 2.
- Richardson, H. W. (1977). A Generalization of Residential Location Theory. *Regional Science and Urban Economics*, 7, 251-66.
- Rivera, M.A.I, Tiglao, N.C.C. (2005). Modeling residential location choice, workplace location choice and mode choice of two-worker households in metro Manila. *Proceedings of Eastern Asia Society for Transportation Studies*, 5, 1167-1178.
- Rosen, S. (1974). Hedonic prices and implicit markets : Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55.
- Schelling, T.C. (1969). Models of segregation. *American Economic Review*, 59(2), 488-493.
- Sener, I.N., Pendyala, R.M., & Bhat, C.R. (2011). Accommodating spatial correlation across choice alternatives in discrete choice models : An application to modeling residential location choice behavior. *Journal of Transport Geography*, 19, 294–303.
- Senior, M. L. (1973). Approaches to residential location modelling 1 : Urban ecological and spatial interaction models (A review). *Environment and Planning*, 5, 165-19.

- Sermons, M.W., & Koppelman, F.S., 2001. Representing the differences between female and male commute behavior in residential location choice models. *Journal of Transport Geography*, 9, 101-110.
- Simmonds, D., Echenique, M., Bates, J., & Oosterhaven, J. (1999). Review of land-use/transport interaction models. Reports to The Standing Advisory Committee on Trunk Road Assessment, Department of the Environment, Transport and the Regions, London.
- Schirmer, P.M., Eggermond, M.A.B., & Axhausen, K.W. (2014). The role of location in residential location choice models : A review of literature. *The Journal of Transport and Land Use*, 7(2), 3-21.
- Small, K.A. (1987). A Discrete Choice Model for Ordered Alternatives. *Econometrica*, 55(2), 409-424.
- Small, K.A., & Song, S. (1992). "Wasteful" Commuting : A Resolution. *Journal of Political Economy*, 100(4), 888-898.
- Srour, I.M., Kockelman, K.M., & Dunn, T.P. (2002). Accessibility Indices : A Connection to Residential Land Prices and Location Choices. *Journal of the Transportation Research Board*, 1805, 25-34.
- Swait, J. (2001). Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B*, 35, 643-666.
- Tabourin, E., Andan, O., & Routhier, J.L. (1995). Les formes de la croissance urbaine. Modèle de Bussière appliqué à l'agglomération lyonnaise. Rapport pour le PIR-Villes CNRS, LET, Lyon.
- Tayyar, M.R., Khan, A.M, & Anderson, D.A. (2003). Impact of telecommuting and intelligent transportation systems on residential location choice. *Transportation Planning and Technology*, 26(2), 171–193.
- Thill, J.C. (1992). Choice set formation for destination choice modelling. *Progress in Human Geography*, 16(3), 361-382.
- Tillema, T., Wee, B.V., & Ettema, D. (2010). Road Pricing and Relocation Decisions of Dutch Households. *Urban Studies*, 47(14), 3013–3033.
- Tobler, W. (1970). A computer model simulating urban growth in the Detroit region. *Economic Geography*, 46(2), 234-240.

- Train, K.E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, second Edition, Cambridge.
- Train, K.E., Ben-Akiva, M., & Atherton, T. (1989). Consumption Patterns and Self-selecting Tariffs. *The Review of Economics and Statistics*, 71(1), 62-73.
- Van Der Haegen, H., Van Hecke, E., Juchtmans, G. (1996). Les régions urbaines belges en 1991. *Études Statistiques n ° 104*, Institut National de Statistique, Bruxelles.
- Vovsha, P. (1997). Application of the Cross-Nested Logit Model to Mode Choice in the Tel-Aviv, Israel, Metropolitan Area. *Transportation Research Record*, 1607, 6-15.
- Waddell, P. (1993). Exogenous Workplace Choice in Residential Location Models : Is the Assumption Valid?. *Geographical Analysis*, 25, 65-82.
- Waddell, P. (1996). Accessibility and Residential Location : The Interaction of Workplace, Residential Mobility, Tenure, and Location Choices. Paper presented at the Lincoln Land Institute TRED conference, Cambridge.
- Waddell, P. (2000). A Behavioral Simulation Model for Metropolitan Policy Analysis and Planning : Residential Location and Housing Market Components Of Urbansim. *Environment and planning B : Planning and Design*, 27, 247-263.
- Waddell, P., Bhat, C.R., Ruiters, E., Bekhor, S., Outwater, M., & Loy Schroer, E. (2001). Land Use and Travel Demand Forecasting Models : Review of the Literature and Operational Models. Final Report to the Puget Sound Regional Council, Seattle, WA.
- Walker, J.L., & Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, 9(1), 77-101.
- Wang, D., & Li, S.M. (2004). Housing preferences in a transitional housing system : The case of Beijing, China. *Environment and Planning A*, 36, 69-87.
- Wegener, W., & Fürst, F. (1999). Land use transport interaction : State of the art. IRPUD report, University of Dortmund.
- Weisbrod, G., Ben-Akiva, M., & Lerman, S. (1980). Tradeoffs in residential location decisions : Transportation versus other factors. *Transportation Policy and Decision-Making*, 1, 13-26.
- Wen, C.H., & Koppelman, F.S. (2001). The generalized nested logit model. *Transportation Research Part B*, 35, 627-641.

- Wheaton, W.C. (1974). A comparative static analysis of urban spatial structure. *Journal of Economic Theory*, 9, 223–237.
- Wilson, A. (1967). A Statistical Theory of Spatial Distribution Models. *Transportation Research*, 1, 253–69.
- Wilson, A. (1970). Interregional Commodity Flows : Entropy-Maximising Procedures. *Geographical Analysis*, 2, 255–82.
- Wilson, A.G. (1998). Land-use/Transport Interaction Models. *Journal of Transport Economics and Policy*, 32(1), 3-26.
- Wong, G.K.M. (2002). A conceptual model of the household's housing decision-making process : The economic perspective. *Review of Urban & Regional Development Studies*, 14 (3), 217-234.
- Wonnacott, T.H., & Wonnacott, R.J. (1991). *Statistiques* (traduction française de l'ouvrage *Introductory statistics for business and economics*, 4ème édition, John Wiley et Sons, New York), Economica, Paris.
- Yao, E., Morikawa, T., Kurauchi, S., & Tokida, T. (2002). A Study on Nested Logit Mode Choice Model for Intercity High-Speed Rail System with Combined RP/SP Data. *Proceedings of the International Conference on Traffic and Transportation Studies*, 612-619.
- Zolfaghari, A., Sivakumar, A., & Polak, J.W. (2012). Choice set formation in residential location choice modelling : Empirical comparison of alternative approaches. Paper presented at the 91st Annual Meeting of the transportation Research Board (TRB), Washington DC, US.
- Zolfaghari, A. (2013). *Methodological and empirical challenges in modelling residential location choices*. PhD thesis, Imperial College London.
- Zondag, B., Pieters, M. (2005). Influence of accessibility on residential location choice. *Transportation Research Record*, 1902, 63-70.

Annexes

Annexe 1 : Les établissements de type G1 et G2

	TYP015	Libellé TYP015
Emploi G1	AD	Administration, associations
	EC	Industrie des biens de consommation
	EM	Activités immobilières et financières
	EN1	Services aux entreprises (Front office)
	EN2	Services aux entreprises (Back Office)
	EP	Services aux particuliers
	EQ	Education, santé, action sociale
	J3	Commerce de détail, réparations
Emploi G2	AG	Agriculture
	EF	Industrie des biens intermédiaires
	EG	Energie
	EH	Construction
	EK	Transports
	EK2	Logistique
	J2	Commerce de gros, intermédiaires

Annexe 2 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, en croisant le prix de l'immobilier avec les quintiles du revenu (avec un échantillonnage aléatoire de 7 alternatives)

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0101	0,00015	65,3
	prix_imm*rev1	-0,0008	0,00005	-16,4
	prix_imm*rev2	-0,0001	0,00006	-1,3
	prix_imm*rev3	0,0001	0,00005	2
	prix_imm*rev4	0,0006	0,00005	12,4
	prix_imm*rev5	0,0017	0,00005	33,2
Aménité spatiale	prox_serv	0,8013	0,01908	42
	etb_prim	0,1	0,00102	97,8
	etb_sec	0,0575	0,00333	17,3
	acc_jeux	0,0995	0,00981	10,2
Environnement social	taux1*rev1	-0,0114	0,00115	-9,9
	taux2*rev2	0,0098	0,00086	11,3
	taux3*rev3	0,0071	0,00036	19,8

	taux4*rev4	0,0033	0,00039	8,5
	taux5*rev5	0,008	0,00055	14,5
	%hlm*rev1	1,5365	0,04327	35,5
	%hlm*rev2	0,7836	0,0557	14,1
	%hlm*rev3	0,3097	0,04515	6,9
	%hlm*rev4	-0,1137	0,05097	-2,2
	%hlm*rev5	-1,3558	0,06535	-20,7
Caractéristiques des ménages	peri*2VP+	0,7507	0,02051	36,6
	centre*VP0	0,8737	0,02109	41,4
	TC*VP0	0,3388	0,01912	17,7
	TC*VP1	0,1874	0,01491	12,6
	peri*loc	-0,7799	0,01602	-48,7
	peri*prop	-0,6481	0,02374	-27,3
	centre*loc	0,0926	0,01467	6,3
	centre*propr	-0,4362	0,02156	-20,2
	peri*tail >=3	0,2558	0,01967	13
	centre*Etud	2,1023	0,03666	57,3

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-154 557</i>
<i>Log likelihood zero</i>	<i>-218 160</i>
<i>R² ajusté</i>	<i>0,292</i>

Annexe 3 : Résultats de l'estimation du modèle de localisation des ménages, avec l'ensemble des variables

Catégorie	Variabes	Coef	Se	t_test
Aménité spatiale	prox_serv	0,8936	0,01904	46,9
	acc_jeux	0,1419	0,00976	14,5
	prox_etb_prim	0,1078	0,00102	105,4
	prox_etb_sec	0,0431	0,00335	12,9
Environnement social	%hlm*rev1	1,3825	0,03835	36,1
	%hlm*rev2	0,7223	0,05136	14,1
	%hlm*rev3	-0,0007	0,04193	-0,0
	%hlm*rev4	-0,6795	0,04822	-14,1
	%hlm*rev5	-2,3961	0,06247	-38,4
Caractéristiques des ménages	TC*VP0	0,4687	0,01880	24,9
	TC*VP1	0,2309	0,01486	15,5
	centre*Etud	1,7199	1 174,99	0,0

centre*retraité	-0,4525	1 174,99	-0,0
centre*actif	0,1193	1 174,99	0,0
centre*autre	-0,1311	1 174,99	-0,0
periph*Etudiant	-0,8758	1 291,50	-0,0
periph*retraité	-0,0370	1 291,50	-0,0
periph*actif	-0,1004	1 291,50	-0,0
periph*autre	-0,2838	1 291,50	-0,0
centre*taille_men1	0,6899	273,797	0,0
centre*taille_men2	0,6087	273,797	0,0
centre*taille_men3	0,2045	273,797	0,0
centre*taille_men4	-0,0652	273,797	-0,0
centre*taille_men5	-0,1822	273,797	-0,0
periph*taille_men1	-0,2351	136,899	-0,0
periph*taille_men2	-0,2642	136,899	-0,0
periph*taille_men3	-0,2796	136,899	-0,0
periph*taille_men4	-0,1963	136,899	-0,0
periph*taille_men5	-0,3217	136,899	-0,0
centre*VP0	0,9358	1 468,07	0,0
centre*1VP	0,4077	1 468,076	0,0
centre*2VP+	-0,0879	1 468,076	-0,0
periph*non_motorisé	-0,8685	1 387,12	-0,0
periph*1VP	-0,4675	1 387,12	-0,0
peri*2VP+	0,0389	1 387,12	0,0
peri*loc	0,0960	0,04417	2,2
peri*prop	0,2685	0,04744	5,7
centre*loc	0,0766	0,03398	2,3
centre*propr	-0,2385	0,03911	-6,1
cour1_est*rev1	0,4782	0,03083	15,5
cour1_est*rev2	0,0716	0,03702	1,9
cour1_est*rev3	0,3637	0,03097	11,7
cour1_est*rev4	0,0724	0,03385	2,1
cour1_est*rev5	-0,4786	0,04332	-11,0
cour1_ouest*rev1	0,0487	0,03033	1,6
cour1_ouest*rev2	0,0227	0,03211	0,7
cour1_ouest*rev3	0,3351	0,02640	12,7
cour1_ouest*rev4	0,1789	0,02702	6,6
cour1_ouest*rev5	0,3228	0,02664	12,1

Annexe 4 : Résultats du modèle MNL avec un échantillonnage aléatoire de 54 alternatives

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0120	0,00013	91,5
	prix_imm	-0,0005	0,00003	-16,7
Aménité spatiale	prox_serv	0,7658	0,01676	45,7
	etb_prim	0,0961	0,00076	126,5
	etb_sec	0,0426	0,00253	16,8
	acc_jeux	0,0784	0,00849	9,2
Environnement social	taux3*rev3	6,6655	0,16619	40,1
	taux4*rev4	8,3141	0,23798	34,9
	taux5*rev5	3,7632	0,07750	48,6
	%hlm*rev1	1,2854	0,03132	41,0
	%hlm*rev2	0,4587	0,04161	11,0
	%hlm*rev3	-0,1300	0,03765	-3,5
	%hlm*rev4	0,1916	0,05147	3,7
	%hlm*rev5	-0,8921	0,06164	-14,5
Caractéristiques des ménages	peri*2VP+	0,7102	0,01957	36,3
	centre*VP0	0,7592	0,01709	44,4
	TC*VP0	0,4228	0,01435	29,5
	TC*VP1	0,1425	0,01060	13,4
	peri*loc	-0,6348	0,01610	-39,4
	peri*prop	-0,5362	0,02260	-23,7
	centre*loc	0,0749	0,01221	6,1
	centre*propr	-0,4325	0,01847	-23,4
	peri*tail >=3	0,3232	0,01782	18,1
	centre*Etud	2,0478	0,03029	67,6

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-368 710</i>
<i>Log likelihood zero</i>	<i>-447 213</i>
<i>R² ajusté</i>	<i>0,175</i>

Annexe 5 : Résultats du modèle MNL avec un échantillonnage aléatoire de 4 alternatives

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0107	0,00019	57,1
	prix_imm	-0,0001	0,00004	-3,2
Aménité spatiale	prox_serv	0,8332	0,02150	38,8
	etb_prim	0,1038	0,00123	84,3
	etb_sec	0,0786	0,00410	19,2
	acc_jeux	0,1134	0,01102	10,3
Environnement social	taux3*rev3	7,9589	0,22626	35,2
	taux4*rev4	8,2606	0,31189	26,5
	taux5*rev5	4,2432	0,11707	36,2
	%hlm*rev1	1,9674	0,04854	40,5
	%hlm*rev2	0,8615	0,05889	14,6
	%hlm*rev3	0,2411	0,05052	4,8
	%hlm*rev4	0,5792	0,06798	8,5
Caractéristiques des ménages	%hlm*rev5	-0,4917	0,08096	-6,1
	peri*2VP+	0,6925	0,02412	28,7
	centre*VP0	0,8897	0,02513	35,4
	TC*VP0	0,4045	0,02162	18,7
	TC*VP1	0,1541	0,01443	10,7
	peri*loc	-0,5958	0,01999	-29,8
	peri*prop	-0,4245	0,02778	-15,3
	centre*loc	0,1240	0,01701	7,3
	centre*propr	-0,3844	0,02471	-15,6
	peri*tail >=3	0,2401	0,02192	11,0
centre*Etud	2,0983	0,04509	46,5	

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-101 544</i>
<i>Log likelihood zero</i>	<i>-155 420</i>
<i>R² ajusté</i>	<i>0.347</i>

Annexe 6 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, en croisant le prix de l'immobilier avec les quintiles du revenu (avec un échantillonnage aléatoire de 35 alternatives)

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0111	0,00014	80,7
	prix_imm*rev1	-0,0016	0,00004	-44,7
	prix_imm*rev2	-0,0011	0,00004	-26,2
	prix_imm*rev3	-0,0012	0,00003	-34,0
	prix_imm*rev4	-0,0004	0,00004	-11,1
	prix_imm*rev5	0,0008	0,00007	10,9
Aménité spatiale	prox_serv	0,4768	0,01474	32,3
	etb_prim	0,0939	0,00079	118,7
	etb_sec	0,0455	0,00266	17,1
	acc_jeux	0,0996	0,00873	11,4
Environnement social	taux1*rev1	1,4293	0,09473	15,1
	taux2*rev2	-0,2776	0,20500	-1,4
	taux3*rev3	2,5355	0,14307	17,7
	taux4*rev4	3,1837	0,21468	14,8
	taux5*rev5	0,9403	0,14554	6,5
	%hlm*rev1	-0,1014	0,04908	-2,1
	%hlm*rev2	0,0627	0,04475	1,4
	%hlm*rev3	-0,6222	0,03804	-16,4
	%hlm*rev4	-0,4947	0,04909	-10,1
	%hlm*rev5	-1,5916	0,06658	-23,9
Caractéristiques des ménages	peri*2VP+	0,7907	0,01984	39,8
	centre*VP0	0,8499	0,01767	48,1
	TC*VP0	0,3338	0,01469	22,7
	TC*VP1	0,0900	0,01068	8,4
	peri*loc	-0,9126	0,01546	-59,0
	peri*prop	-0,7820	0,02228	-35,1
	centre*loc	0,0603	0,01239	4,9
	centre*propr	-0,4569	0,01873	-24,4
	peri*tail >=3	0,2811	0,01819	15,5
	centre*Etud	2,0161	0,02959	68,1

<i>Nombre d'observations</i>	112 112
Log likelihood	-323 688
Log likelihood zero	-398 597
R ² ajusté	0,188

Annexe 7 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, en croisant le prix de l'immobilier avec les 1^{er} et 5^{ème} quintiles du revenu (avec un échantillonnage aléatoire de 35 alternatives)

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0091	0,00012	73,0
	prix_imm*rev1	-0,0006	0,00004	-14,0
	prix_imm*rev5	0,0026	0,00008	34,3
Aménité spatiale	prox_serv	0,7807	0,01672	46,7
	etb_prim	0,0954	0,00075	127,2
	etb_sec	0,0279	0,00259	10,8
	acc_jeux	0,1018	0,00863	11,8
Environnement social	taux1*rev1	3,3619	0,10637	31,6
	taux5*rev5	-1,1082	0,14783	-7,5
	%hlm*rev1	-0,3547	0,05166	-6,9
	%hlm*rev5	-1,2373	0,06718	-18,4
Caractéristiques des ménages	peri*2VP+	0,7810	0,01994	39,2
	centre*VP0	0,8240	0,01778	46,4
	TC*VP0	0,3730	0,01479	25,2
	TC*VP1	0,1438	0,01072	13,4
	peri*loc	-0,6543	0,01540	-42,5
	peri*prop	-0,5195	0,02222	-23,4
	centre*loc	0,0849	0,01240	6,8
	centre*propr	-0,4309	0,01875	-23,0
	peri*tail >=3	0,2653	0,01802	14,7
centre*Etud	2,0592	0,03061	67,3	

<i>Nombre d'observations</i>	112 112
Log likelihood	-321 554
Log likelihood zero	-398 597
R ² ajusté	0,193

Annexe 8 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, en croisant le prix de l'immobilier avec les 1^{er}, 2^{ème} et 3^{ème} quintiles du revenu (avec un échantillonnage de 35 alternatives)

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0123	0,00013	96,6
	prix_imm*rev1	-0,0012	0,00004	-26,9
	prix_imm*rev2	-0,0002	0,00006	-3,4
	prix_imm*rev3	-0,0001	0,00007	-1,4
Aménité spatiale	prox_serv	0,7994	0,01672	47,8
	etb_prim	0,0857	0,00075	113,8
	etb_sec	0,0569	0,00260	21,9
	acc_jeux	0,1266	0,00853	14,8
Environnement social	taux1*rev1	2,5157	0,10718	23,5
	taux2*rev2	2,7162	0,13451	20,2
	taux3*rev3	-2,4299	0,14136	-17,2
	%hlm*rev1	-0,2050	0,05154	-4,0
	%hlm*rev2	0,8836	0,05428	16,3
	%hlm*rev3	-0,7223	0,04468	-16,2
Caractéristiques des ménages	peri*2VP+	0,7951	0,01985	40,1
	centre*VP0	0,8047	0,01771	45,4
	TC*VP0	0,3441	0,01476	23,3
	TC*VP1	0,1002	0,01072	9,3
	peri*loc	-0,6592	0,01591	-41,4
	peri*prop	-0,5370	0,02248	-23,9
	centre*loc	0,0220	0,01229	1,8
	centre*propr	-0,4767	0,01859	-25,6
	peri*tail >=3	0,2609	0,01825	14,3
	centre*Etud	2,0241	0,03054	66,3

<i>Nombre d'observations</i>	112 112
<i>Log likelihood</i>	-323 451
<i>Log likelihood zero</i>	-398 597
<i>R² ajusté</i>	0,188

Annexe 9 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, en croisant le prix de l'immobilier avec les 1^{er} et 5^{ème} quintiles du revenu, sans les ménages appartenant au 3^{ème} quintile (avec un échantillonnage aléatoire de 35 alternatives)

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0100	0,00012	81,4
	prix_imm*rev5	0,0024	0,00008	31,3
	prix_imm*rev1	-0,0012	0,00004	-30,0
Aménité spatiale	prox_serv	0,7525	0,01669	45,1
	etb_prim	0,0952	0,00077	124,2
	etb_sec	0,0341	0,00258	13,2
	acc_jeux	0,0917	0,00861	10,7
Environnement social	taux4*rev4	7,6196	0,23861	31,9
	taux5*rev5	-0,7720	0,14756	-5,2
	%hlm*rev1	0,8206	0,03591	22,8
	%hlm*rev2	0,7511	0,03823	19,6
	%hlm*rev4	0,4451	0,04956	9,0
	%hlm*rev5	-1,1997	0,06704	-17,9
Caractéristiques des ménages	peri*2VP+	0,8098	0,01995	40,6
	centre*VP0	0,8625	0,01768	48,8
	TC*VP0	0,4304	0,01465	29,4
	TC*VP1	0,1518	0,01069	14,2
	peri*loc	-0,6250	0,01556	-40,2
	peri*prop	-0,4851	0,02233	-21,7
	centre*loc	0,0659	0,01240	5,3
	centre*propr	-0,4559	0,01875	-24,3
	peri*tail >=3	0,2518	0,01803	14,0
	centre*Etud	2,0511	0,03057	67,1

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-321 251</i>
<i>Log likelihood zero</i>	<i>-398 597</i>
<i>R² ajusté</i>	<i>0,194</i>

Annexe 10 : Résultats du modèle de l'annexe 9, sans la variable mesurant l'effet croisé du taux hlm avec les ménages appartenant au 4^{ème} quintile du revenu

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0099	0,00012	81,0
	prix_imm*rev5	0,0024	0,00008	31,3
	prix_imm*rev1	-0,0012	0,00004	-30,0
Aménité spatiale	prox_serv	0,7553	0,01670	45,2
	etb_prim	0,0962	0,00076	126,8
	etb_sec	0,0326	0,00257	12,6
	acc_jeux	0,0905	0,00861	10,5
Environnement social	taux4*rev4	6,3879	0,19591	32,6
	taux5*rev5	-0,7692	0,14756	-5,2
	%hlm*rev1	0,8066	0,03591	22,5
	%hlm*rev2	0,7368	0,03823	19,3
	%hlm*rev5	-1,2210	0,06707	-18,2
Caractéristiques des ménages	peri*2VP+	0,8074	0,01995	40,5
	centre*VP0	0,8654	0,01768	48,9
	TC*VP0	0,4274	0,01465	29,2
	TC*VP1	0,1506	0,01070	14,1
	peri*loc	-0,6391	0,01548	-41,3
	peri*prop	-0,4995	0,02228	-22,4
	centre*loc	0,0689	0,01240	5,6
	centre*propr	-0,4529	0,01875	-24,1
	peri*tail >=3	0,2515	0,01804	13,9
	centre*Etud	2,0562	0,03060	67,2

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-321 291</i>
<i>Log likelihood zero</i>	<i>-398 597</i>
<i>R² ajusté</i>	<i>0,194</i>

Annexe 11 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, en croisant le prix de l'immobilier avec les 1^{er} et 5^{ème} quintiles du revenu et le taux hlm avec les 1^{er} et 2^{ème} quintiles du revenu (avec 35 alternatives)

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0100	0,00012	82,1
	prix_imm*rev5	0,0022	0,00008	29,1
	prix_imm*rev1	-0,0012	0,00004	-29,8
Aménité spatiale	prox_serv	0,7536	0,01671	45,1
	etb_prim	0,0948	0,00075	125,8
	etb_sec	0,0332	0,00256	13,0
	acc_jeux	0,0929	0,00862	10,8
Environnement social	taux4*rev4	6,3399	0,19594	32,4
	taux5*rev5	0,4085	0,13171	3,1
	%hlm*rev1	0,8308	0,03584	23,2
	%hlm*rev2	0,7609	0,03815	19,9
Caractéristiques des ménages	peri*2VP+	0,8121	0,01998	40,7
	centre*VP0	0,8629	0,01767	48,8
	TC*VP0	0,4370	0,01464	29,9
	TC*VP1	0,1577	0,01068	14,8
	peri*loc	-0,6189	0,01546	-40,0
	peri*prop	-0,4739	0,02226	-21,3
	centre*loc	0,0665	0,01239	5,4
	centre*propr	-0,4544	0,01874	-24,3
	peri*tail >=3	0,2558	0,01806	14,2
	centre*Etud	2,0535	0,03061	67,1

<i>Nombre d'observations</i>	112 112
<i>Log likelihood</i>	-321 466
<i>Log likelihood zero</i>	-398 597
<i>R² ajusté</i>	0,194

Annexe 12 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, en croisant le prix de l'immobilier avec les quintiles du revenu et le taux hlm avec les 1^{er} et 5^{ème} quintiles (avec un échantillonnage aléatoire de 35 alternatives)

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0094	0,00013	69,6
	prix_imm*rev1	-0,0007	0,00005	-15,4
	prix_imm*rev2	-0,0007	0,00004	-17,2
	prix_imm*rev3	-0,0005	0,00003	-16,0
	prix_imm*rev4	0,0006	0,00003	18,6
	prix_imm*rev5	0,0025	0,00008	31,7
Aménité spatiale	prox_serv	0,7869	0,01674	47,0
	etb_prim	0,0942	0,00077	123,2
	etb_sec	0,0298	0,00262	11,4
	acc_jeux	0,1046	0,00864	12,1
Environnement social	taux1*rev1	3,2467	0,10769	30,1
	taux5*rev5	-0,9623	0,14974	-6,4
	%hlm*rev1	-0,3420	0,05168	-6,6
	%hlm*rev5	-1,2311	0,06707	-18,4
Caractéristiques des ménages	peri*2VP+	0,7929	0,01997	39,7
	centre*VP0	0,8339	0,01780	46,8
	TC*VP0	0,3714	0,01481	25,1
	TC*VP1	0,1428	0,01073	13,3
	peri*loc	-0,6705	0,01550	-43,3
	peri*prop	-0,5320	0,02230	-23,9
	centre*loc	0,0843	0,01242	6,8
	centre*propr	-0,4364	0,01878	-23,2
	peri*tail >=3	0,2658	0,01813	14,7
	centre*Etud	2,0547	0,03063	67,1

<i>Nombre d'observations</i>	112 112
<i>Log likelihood</i>	-321 036
<i>Log likelihood zero</i>	-398 597
<i>R² ajusté</i>	0,195

Annexe 13 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, sans les ménages appartenant au 3^{ème} quintile du revenu (avec un échantillonnage aléatoire de 35 alternatives)

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0115	0,00013	87,7
	prix_imm	-0,0006	0,00002	-24,5
Aménité spatiale	prox_serv	0,7769	0,01676	46,4
	etb_prim	0,0943	0,00077	122,8
	etb_sec	0,0448	0,00259	17,3
	acc_jeux	0,0915	0,00864	10,6
Environnement social	taux4*rev4	8,1776	0,23993	34,1
	taux5*rev5	4,1242	0,07728	53,4
	%hlm*rev1	1,2262	0,03026	40,5
	%hlm*rev2	0,3909	0,04082	9,6
	%hlm*rev4	0,1396	0,05110	2,7
	%hlm*rev5	-0,8240	0,06185	-13,3
Caractéristiques des ménages	peri*2VP+	0,7117	0,01971	36,1
	centre*VP0	0,7743	0,01736	44,6
	TC*VP0	0,4126	0,01463	28,2
	TC*VP1	0,1556	0,01073	14,5
	peri*loc	-0,6582	0,01572	-41,9
	peri*prop	-0,5492	0,02245	-24,5
	centre*loc	0,0704	0,01237	5,7
	centre*propr	-0,4311	0,01868	-23,1
	peri*tail >=3	0,2813	0,01791	15,7
	centre*Etud	2,0505	0,03055	67,1

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-322 002</i>
<i>Log likelihood zero</i>	<i>-398 597</i>
<i>R² ajusté</i>	<i>0,192</i>

Annexe 14 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, en croisant le taux hlm avec les ménages appartenant aux 1^{er} et 5^{ème} quintiles et sans les ménages de la classe moyenne (avec un échantillonnage aléatoire de 35 alternatives)

Catégorie	Variables	Coef	Se	t_test
Arbitrage du marché	ac_g_VP	0,0115	0,00013	87,9
	prix_imm	-0,0006	0,00002	-25,4
Aménité spatiale	prox_serv	0,7709	0,01674	46,1
	etb_prim	0,0933	0,00076	123,4
	etb_sec	0,0446	0,00259	17,2
	acc_jeux	0,0923	0,00865	10,7
Environnement social	taux4*rev4	7,7410	0,20108	38,5
	taux5*rev5	4,7245	0,06391	73,9
	%hlm*rev1	1,2334	0,02989	41,3
	%hlm*rev2	0,4055	0,04043	10,0
Caractéristiques des ménages	peri*2VP+	0,7162	0,01971	36,3
	centre*VP0	0,7744	0,01736	44,6
	TC*VP0	0,4178	0,01461	28,6
	TC*VP1	0,1592	0,01072	14,9
	peri*loc	-0,6440	0,01552	-41,5
	peri*prop	-0,5310	0,02229	-23,8
	centre*loc	0,0711	0,01236	5,8
	centre*propr	-0,4307	0,01867	-23,1
	peri*tail >=3	0,2843	0,01791	15,9
	centre*Etud	2,0484	0,03055	67,1

<i>Nombre d'observations</i>	<i>112 112</i>
<i>Log likelihood</i>	<i>-322 101</i>
<i>Log likelihood zero</i>	<i>-398 597</i>
<i>R² ajusté</i>	<i>0,192</i>

Annexe 15 : Description et corrélation des variables testées

Variables	Description	1999		2007	
		Moyenne	écart_type	Moyenne	écart_type
ac_g_VP	Accessibilité gravitaire à l'emploi en voiture particulière	35556	39216,3	35318	37958
ac_g_TC	Accessibilité gravitaire à l'emploi en transport en commun	8628	15513,7	8644	15407
acc_grav_G1	Accessibilité gravitaire à l'emploi aux établissements de service			-	-
acc_grav_G2	Accessibilité gravitaire à l'emploi aux établissements industriels			-	-
acc_vp_g1	Accessibilité gravitaire à l'emploi aux établissements de service en voiture particulière			-	-
acc_vp_g2	Accessibilité gravitaire à l'emploi aux établissements de service en voiture particulière			-	-
acc_tc_g1	Accessibilité gravitaire à l'emploi aux établissements de service en transport en commun			-	-
acc_tc_g2	Accessibilité gravitaire à l'emploi aux établissements industriels en transport en commun			-	-
acc_eng	Accessibilité gravitaire à l'emploi aux établissements d'enseignement			-	-
prix_imm	Prix de l'immobilier estimé à partir de l'enquête logement et du fichier du marché immobilier des notaires	747,7	210,243	2398	269,8
prix_imm*rev1	rev1 : Une variable dichotomique qui prend la valeur 1 si le ménage appartient au premier quintile et 0 sinon			-	-
prix_imm*rev2	rev2 : Une variable dichotomique qui prend la valeur 1 si le ménage appartient au deuxième quintile et 0 sinon			-	-
prix_imm*rev3	rev3 : Une variable dichotomique qui prend la valeur 1 si le ménage appartient au troisième quintile et 0 sinon			-	-
prix_imm*rev4	rev4 : Une variable dichotomique qui prend la valeur 1 si le ménage appartient au quatrième quintile et 0 sinon			-	-
prix_imm*rev5	rev5 : Une variable dichotomique qui prend la valeur 1 si le ménage appartient au cinquième quintile et 0			-	-

	sinon				
prox_metro	Une variable dichotomique qui prend la valeur 1 s'il y a une station de métro (funiculaire inclus)	0,100	0,300	-	-
prox_bus	Une variable dichotomique qui prend la valeur 1 s'il y a un arrêt de bus d'une fréquence supérieur ou égale à un passage de bus tous les 5 minutes	0,243	0,429	-	-
<i>TC</i>	Une variable dichotomique qui prend la valeur 1 s'il y a une station de métro (funiculaire inclus) ou un arrêt de bus d'une fréquence supérieur ou égale à 30 (un bus tous les 5 minutes)	0,472	0,500	0,285	0,452
café_tabac	Une variable dichotomique qui prend la valeur 1 s'il y a un café ou un bureau de tabac	0,199	0,400	-	-
boulang	Une variable dichotomique qui prend la valeur 1 s'il y a une pâtisserie ou une boulangerie	0,711	0,454	-	-
supermarché	Une variable dichotomique qui prend la valeur 1 s'il y a un supermarché	0,718	0,451	-	-
magasin	Une variable dichotomique qui prend la valeur 1 s'il y a un magasin	0,620	0,486	-	-
serv_base1	xboupat+xtaba	0,734	0,442	-	-
serv_base2	binaire(xsup+xmags)	0,782	0,413	-	-
prox_serv	Variable binaire qui vaut 1 s'il y a un service de base dans la macro-zone (boulangerie, superette...)	0,857	0,351	0,852	0,356
serv_part	Nombre des établissements dédiés aux services particuliers dans la macro-zone	38,780	51,050	-	-
serv_santé	Nombre des étab dans la macro-zone lissements de fonction médicale ou paramédicale dans la macro-zone	18,120	25,556	-	-
banque_poste	Nombre des banques et postes dans la macro-zone	2,481	3,830	-	-
ciné	Nombre des cinémas dans la macro-zone	0,097	0,420	-	-
resto_trad	Nombre des restaurants de type traditionnels dans la macro-zone	6,713	13,550	-	-
rest_rapide	Nombre des restaurants de type rapide dans la macro-zone	1,208	2,990	-	-
disco	Nombre des discothèques dans la macro-zone	0,310	1,034	-	-

biblio	Nombre des bibliothèques dans la macro-zone	0,035	0,196	-	-
autre_cult	Nombre des établissements dédiés à d'autres activités culturelles comme les musées dans la macro-zone	0,081	0,334	-	-
salle_sport	Nombre des salles de sports dans la macro-zone	0,303	0,623		
spot_autre	Nombre des établissements pour d'autres activités de sports dans la macro-zone	3,259	3,296		
acc_jeux	Nombre de parcs de jeu, de manèges et de parcs d'attraction dans le Grand quartier	0,093	0,368	0,090	0,332
equ_recr	C'est la somme des variables : xcine, xdisco, xresttra et xrestrapide	8,329	17,151		
equ_cult	C'est la somme des variables : xbib et xculautre)	0,116	0,409		
sport	C'est la somme des variables : salle_sport et spot_autre	3,562	3,521		
etb_sec	Nombre des établissements secondaires	0,539	0,965	0,537	0,921
etb_prim	Nombre des établissements primaires et crèches	3,785	3,515	3,785	3,431
taux1*rev1	Variable croisée entre l'appartenance du ménage au quintile du revenu (1 oui, 0 sinon) et le taux de ménages du même quintile de revenu dans le quartier (%)	0,1731	0,09407	0,1683	0,101
taux2*rev2		0,18782	0,0463	0,2127	0,071
taux3*rev3		0,2334	0,04229	0,2445	0,07
taux4*rev4		0,199	0,03971	0,1858	0,062
taux5*rev5		0,2067	0,09987	0,1887	0,121
age1*taux1	Variable croisée entre l'appartenance du ménage à une classe d'âge (1 oui, 0 sinon) et le taux de ménages du même classe d'âge dans le quartier (%)				
age2*taux2					
age3*taux3					
age4*taux4					
age5*taux5					
taux_person1*person1	Variable croisée de la taille du ménage (1 oui, 0 sinon) et le taux de ménages de la même taille dans le quartier (%)				
taux_person2*person2					
taux_person3*person3					
taux_person4*person4					
taux_person5*person5					
taux_sup3*person_sup3	Variable croisée des ménages composés de 3 personnes ou plus (1 oui, 0 sinon) et le taux de ménages de la				

	même taille dans le quartier (%)				
hlm	le pourcentage des logements sociaux HLM	0,113	0,178	0,118	0,168
%hlm*rev1	Variable croisée entre l'appartenance du ménage au quintile du revenu (1 oui, 0 sinon) et le pourcentage des logements sociaux HLM				
%hlm*rev2					
%hlm*rev3					
%hlm*rev4					
%hlm*rev5					
centre	Variable dichotomique qui vaut 1 si la zone de localisation est dans le centre (Lyon et Villeurbanne) et 0 sinon	0,183	0,387	0,183	0,387
peri	Variable binaire qui vaut 1 si la zone de localisation est dans la couronne périurbaine de Lyon et 0 sinon	0,563	0,497	0,563	0,497
TC*VP0	VP0 : une variable dichotomique qui vaut 1 si le ménage est non motorisé et 0 sinon				
TC*VP1	VP1 : une variable dichotomique qui vaut 1 si le ménage a au moins une voiture et 0 sinon				
peri*2VP+	2VP+ : une variable dichotomique qui vaut 1 si le ménage a au moins deux voitures et 0 sinon				
centre*VP0					
peri*loc	loc : une variable binaire qui vaut 1 si le ménage se est locataire et 0 sinon				
peri*prop	prop : une variable binaire qui vaut 1 si le ménage se est propriétaire et 0 sinon				
centre*loc					
centre*prop					
peri*tail >=3	tail>=3 : une variable binaire qui vaut 1 si le ménage se compose de 3 personnes ou plus et 0 sinon				
centre*Etud	Etud : une variable dichotomique qui vaut 1 si le chef du ménage est étudiant et 0 sinon				
retseurb*person5	person5 : une variable binaire qui vaut 1 si le ménage se compose de 5 personnes et 0 sinon				

Liste des illustrations

Liste des tableaux

Tableau 2.1 : Synthèse des travaux empiriques mettant en avant le rôle de l'arbitrage du marché.....	36
Tableau 2.2 : Synthèse des travaux empiriques mettant en avant le rôle des aménités spatiales et/ou sociales	43
Tableau 2.3 : Synthèse des travaux empiriques mettant en avant le rôle des caractéristiques des ménages	47
Tableau 5.1 : pourcentage des ménages ayant emménagé entre 2006 et 2009	91
Tableau 5.2 : Résultats de l'estimation du modèle de localisation des ménages à l'échelle de la cellule	95
Tableau 5.3 : Résultats de l'estimation du modèle de localisation des ménages à l'échelle de la commune	96
Tableau 5.4 : Résultats de l'estimation du modèle de localisation des ménages à l'échelle de la commune, sans variable d'accessibilité	97
Tableau 5.5 : Résultats de l'estimation d'un modèle de localisation des ménages à l'échelle de la commune avec seulement le prix de l'immobilier	98
Tableau 5.6 : Résultats de l'estimation d'un modèle de localisation des ménages à l'échelle de la commune avec seulement les caractéristiques des ménages	98
Tableau 5.7 : Résultats de l'estimation du modèle de localisation des ménages à l'échelle de l'IRIS	99
Tableau 6.1 : Résultats de l'estimation du modèle de localisation des ménages, avec seulement les variables relatives à l'arbitrage du marché.....	109
Tableau 6.2 : Résultats de l'estimation du modèle de localisation des ménages, avec les accessibilités gravitaires TC et VP et le prix de l'immobilier.....	110
Tableau 6.3 : Résultats de l'estimation du modèle de localisation des ménages, avec l'accessibilité gravitaire VP et le prix de l'immobilier	110
Tableau 6.4 : Résultats de l'estimation du modèle de localisation des ménages, avec l'accessibilité gravitaire TC et le prix de l'immobilier	111
Tableau 6.5 : Résultats de l'estimation du modèle de localisation des ménages, avec les variables relatives aux aménités spatiales.....	112
Tableau 6.6 : Résultats de l'estimation du modèle de localisation des ménages avec les variables relatives aux aménités spatiales, en remplaçant les services de bases par des indicateurs agrégés ...	113
Tableau 6.7 : Résultats de l'estimation du modèle de localisation des ménages avec les variables relatives aux aménités spatiales, en remplaçant les services de bases par des indicateurs agrégés et en enlevant les variables moins significatives	114
Tableau 6.8 : Résultats de l'estimation du modèle de localisation des ménages avec les variables relatives aux aménités spatiales retenues	114
Tableau 6.9 : Résultats de l'estimation du modèle de localisation des ménages, avec les variables relatives aux aménités sociales.....	116
Tableau 6.10 : Résultats de l'estimation du modèle de localisation des ménages avec les variables relatives aux aménités sociales, en regroupant les ménages d'une taille supérieure ou égale à 3 dans une seule variable	117
Tableau 6.11 : Résultats de l'estimation du modèle de localisation des ménages avec les variables relatives aux aménités sociales retenues	117
Tableau 6.12 : Résultats de l'estimation du modèle de localisation des ménages, avec certaines caractéristiques de la zone et la catégorie socioprofessionnelle du ménage	118

Tableau 6.13 : Résultats de l'estimation du modèle de localisation des ménages, avec certaines caractéristiques de la zone et la taille du ménage.....	119
Tableau 6.14 : Résultats de l'estimation du modèle de localisation des ménages, avec certaines caractéristiques de la zone et le statut d'occupation du logement.....	119
Tableau 6.15 : Résultats de l'estimation du modèle de localisation des ménages, avec certaines caractéristiques de la zone et du ménage	120
Tableau 6.16 : Résultats de l'estimation du modèle de localisation des ménages avec certaines caractéristiques de la zone et du ménage, en remplaçant la proximité aux métros par celle aux arrêts bien desservis	121
Tableau 6.17 : Résultats de l'estimation du modèle final de localisation des ménages en 1999	123
Tableau 6.18 : Résultats de l'estimation du modèle final de localisation des ménages en 1999, sans le prix de l'immobilier	124
Tableau 6.19 : Résultats du test du ratio de vraisemblance du modèle avec et sans le prix de l'immobilier.....	125
Tableau 6.20 : Variation de la valeur de log-vraisemblance en fonction de la taille de l'échantillon d'alternatives	127
Tableau 7.1 : Résultats du modèle emboîté selon la structure centrée sur la répartition des revenus (4 zones), avec les variables du modèle multinomial	134
Tableau 7.2 : Résultats du modèle emboîté selon la structure centrée sur la répartition des revenus (4 zones) avec les variables du modèle multinomial, mais sans l'environnement social.....	135
Tableau 7.3 : Résultats du modèle emboîté selon la structure centrée sur la répartition des revenus (4 zones) avec les variables du modèle multinomial, sans les variables liées à l'environnement social et aux caractéristiques des ménages	136
Tableau 7.4 : Résultats du modèle emboîté selon la structure centrée sur la répartition des revenus (4 zones) avec les variables du modèle multinomial, sans les variables liées à l'environnement social, aux caractéristiques des ménages et la proximité des services de base	137
Tableau 7.5 : Résultats de l'estimation du modèle de Bussière	138
Tableau 7.6 : Résultats du modèle emboîté selon une structure centre-périphérie définie à partir de la densité de population et de la distance, avec les variables du modèle multinomial	140
Tableau 7.7 : Résultats du modèle emboîté selon une structure centre-périphérie définie à partir de la densité de population et de la distance, en enlevant les variables de localisation (centre, périphérie)	142
Tableau 7.8 : Résultats du modèle emboîté selon la structure multicritère sur la base d'une matrice de contiguïté (4 classes), avec les variables du modèle multinomial.....	145
Tableau 7.9 : Résultats du modèle emboîté selon la structure multicritère sur la base d'une matrice de contiguïté (5 classes), avec les variables du modèle multinomial.....	147
Tableau 7.10 : Résultats du modèle MNL avec un échantillonnage aléatoire de 35 alternatives.....	149
Tableau 7.11 : La correction des biais liés à l'échantillonnage des alternatives.....	150
Tableau 7.12 : Résultats du modèle MNL avec un échantillonnage stratifié et correctif selon la méthode proposée par McFadden (1978).....	152
Tableau 7.13 : Résultats du modèle MNL avec un échantillonnage stratifié (emboîté avec logsum=1, estimation selon la méthode itérative de Guevara (2010)).....	153
Tableau 8.1 : Résultats de l'estimation du modèle MNL en 2007	167
Tableau 8.2 : Résultats de l'estimation du modèle emboîté en 2007	170

Liste des figures

Figure 1 : Architecture du système modélisé par SIMBAD	10
Figure 1.1 : Classification des approches de la modélisation des choix résidentiels	17

Figure 3.1 : Exemple illustrant l'équivalence entre un logit emboîté et un PCL.....	62
Figure 3.2 : Relation entre complexité des calculs et restriction des hypothèses des modèles.....	68
de choix discrets avec un grand nombre d'alternatives.....	68
Figure 5.1 : Évolution de la population de l'aire urbaine de Lyon dans le centre et les deux couronnes urbaine et périurbaine (1968-2007).....	84
Figure 6.1 : Processus d'estimation d'un MNL avec un échantillon d'alternatives	105
Figure 6.2 : Différence d'utilités entre le 5ème et le 95ème percentile des valeurs des variables du modèle retenu	125
Figure 7.1 : Résultats du modèle Bussière appliqué à l'aire urbaine de Lyon.....	139
Figure 7.2 : Dendrogramme de la classification hiérarchique.....	143
Figure 8.1 : Différence d'utilités entre le 5ème et le 95ème percentile des valeurs des variables du modèle 1999 et 2007	168
Figure 2 : Articulation entre les choix de localisation et modèles LUTI.....	178

Liste des graphiques

Graphique 3.1 : Représentation graphique de l'utilité en fonction de la probabilité du choix.	57
Graphique 8.1 : Variation du prix de l'immobilier (en fonction de la distance au centre (1999 et 2007))	158
Graphique 8.2 : Variation de l'accessibilité gravitaire à l'emploi en VP (en fonction de la distance au centre (1999 et 2007))	159
Graphique 8.3 : Nombre d'établissements primaires et crèches (en fonction de la distance au centre (1999 et 2007)).....	160
Graphique 8.4 : Nombre d'établissements secondaires (en fonction de la distance au centre (1999 et 2007))	160
Graphique 8.5 : Nombre de parcs de jeux (en fonction de la distance au centre (1999 et 2007))	161
Graphique 8.6 : Concentration des logements sociaux (en fonction de la distance au centre (1999 et 2007))	161
Graphique 8.7 : Variation des taux des ménages de chaque quintile de revenu (en fonction de la distance au centre (1999 et 2007))	163
Graphique 8.8 : Variation des taux de motorisation (en fonction de la distance au centre (1999 et 2007))	164
Graphique 8.9 : Concentration des étudiants (en fonction de la distance au centre (1999 et 2007)).....	165
Graphique 8.10 : Variation des taux de locataires et de propriétaires (en fonction de la distance au centre (1999 et 2007))	165
Graphique 8.11 : Variation des taux de ménages composés de 3 personnes ou plus (en fonction de la distance au centre (1999 et 2007))	166

Liste des cartes

Carte 5.1 : Un exemple de l'analyse spatiale illustrant la différence entre les différents zonages testés	100
Carte 7.1 : Structure centrée sur la répartition des revenus (4 zones)	133
Carte 7.2 : Distinction centre-périphérie à partir de la densité de population et la distance au centre (2 zones)	140
Carte 7.3 : Structure multicritère sur la base d'une matrice de contiguïté (4 classes)	144
Carte 7.4 : Structure multicritère sur la base d'une matrice de contiguïté (5 classes)	146

Table des matières

Sommaire	5
Introduction	8
Partie I - Choix de localisation résidentielle : revue de littérature	14
Introduction	15
Chapitre 1 - Cadres théoriques du choix de localisation résidentielle	16
1.1. Le modèle standard de l'économie urbaine et ses extensions.....	18
1.1.1. Le modèle standard d'Alonso.....	18
1.1.2. Extensions du modèle standard.....	20
1.2. L'apport de l'approche des interactions spatiales.....	22
1.3. La théorie de l'utilité aléatoire.....	24
L'approche d'utilité aléatoire comme cadre d'analyse de la localisation résidentielle.....	26
Chapitre 2 - Les déterminants des choix de localisation résidentielle : travaux empiriques	27
2.1. Arbitrage du marché : accessibilité vs prix de l'immobilier	27
2.1.1. Accessibilité à l'emploi	28
2.1.2. Prix de l'immobilier	31
2.2. Préférence pour les aménités spatiales et sociales	37
2.2.1. Préférence pour les aménités spatiales	37
2.2.2. Préférence pour les aménités sociales.....	39
2.3. Caractéristiques des ménages : cycle de vie et accès à la propriété	44
Comparaison entre différents contextes : quels enseignements par rapport aux déterminants du choix de localisation ?	48
Conclusion	50
Partie II : Les modèles de choix discrets comme outil pour appréhender les choix de localisation résidentielle.....	52
Chapitre 3 - Les familles de modèles de choix discrets : structures et limites	54

3.1.	Spécificités générales des modèles de choix discrets.....	55
3.2.	La famille des modèles GEV	56
3.2.1.	Modèle logit Multinomial	57
3.2.2.	Modèle emboîté.....	59
3.2.3.	D'autres familles de modèles GEV	60
3.3.	Modèle probit	64
3.4.	Familles des modèles mixtes.....	65
3.5.	D'autres structures.....	65
	Synthèse : Modèles développés pour modéliser les choix de localisation.....	66

Chapitre 4 – Quels apports des développements récents pour les choix de localisation 69

4.1.	La constitution de l'ensemble des choix.....	70
4.2.	Échantillonnage des alternatives	71
4.2.1.	MNL.....	72
4.2.2.	GEV.....	74
4.3.	Taille de l'échantillon	76
	Quelle méthode pour réduire le nombre élevé d'alternatives.....	77

Conclusion 79

Partie III : Choix de localisation résidentielle dans l'aire urbaine de Lyon (1999, 2007) 80

Chapitre 5 – Aire urbaine de Lyon : Échelles et données 82

5.1.	Données et constructions des variables	82
5.1.1.	Arbitrage du marché : accessibilité et prix de l'immobilier	84
5.1.2.	Aménité spatiale : services quotidiens et équipements publics.....	87
5.1.3.	Environnement social : niveau de vie et logements sociaux	89
5.1.4.	Caractéristiques des ménages : catégorie socioprofessionnelle, revenu... ..	91
5.2.	Échelles spatiales	94
5.2.1.	Le zonage le plus fin : carroyage	94
5.2.2.	Le zonage administratif : Commune	96

5.2.3.	Le zonage statistique : IRIS.....	99
5.2.4.	Le Grand quartier comme alternative intermédiaire.....	100
	Analyse empirique des choix de localisation : Données et échelles.....	101

Chapitre 6 - Analyse empirique des choix de localisation des ménages : un modèle logit Multinomial..... 103

6.1.	Processus de l'estimation	104
6.2.	Les critères de l'évaluation de la qualité du modèle	106
6.2.1.	Test de Wald	106
6.2.2.	Ratio de vraisemblance.....	106
6.2.3.	L'élasticité	107
6.2.4.	Test de la propriété d'IIA.....	107
6.3.	Estimation des modèles en 1999	108
6.4.	Résultats et discussions.....	122
6.4.1.	Présentation et interprétation du modèle retenu.....	122
6.4.2.	Effet de la taille de l'échantillon.....	127
	Un modèle logit multinomial avec échantillonnage aléatoire : Quelles variables et quelle taille d'échantillon ?.....	129

Chapitre 7 - Une structure hiérarchique pour modéliser les choix de localisation : Quel apport ?..... 131

7.1.	Définition des nids et estimation d'un modèle emboîté	132
7.1.1.	Structure centrée sur la répartition des revenus.....	133
7.1.2.	Structure centrée sur la densité de population et la distance au centre	138
7.1.3.	Structure multicritère sur la base d'une matrice de contiguïté.....	143
7.2.	Comparaison avec d'autres structures de modèle	148
7.2.1.	Un MNL avec un échantillonnage aléatoire ou un MNL avec un échantillonnage stratifié.....	149
7.2.2.	Un NL ou un MNL avec un échantillonnage stratifié	153
	Estimation d'un logit emboîté : Quel apport ?	156

Chapitre 8 - Analyse diachronique des déterminants de localisation des ménages (1999, 2007)..... 157

8.1.	Analyse descriptive	158
------	---------------------------	-----

8.1.1.	Arbitrage du marché	158
8.1.2.	Aménité spatiale	159
8.1.3.	Aménité sociale	161
8.1.4.	Caractéristiques des ménages	163
8.2.	Résultats des estimations et comparaison	166
8.3.	Analyse de la transférabilité temporelle.....	171
8.3.1.	Les tests de transférabilité.....	171
8.3.2.	Résultats des tests.....	172
	Le modèle estimé en 1999 est-il transférable ?.....	173
	Conclusion	173
	Conclusion générale	175
	Bibliographie	179
	Annexes.....	192
	Liste des illustrations	212
	Liste des tableaux.....	212
	Liste des figures.....	213
	Liste des graphiques.....	214
	Liste des cartes.....	215
	Table des matières	216