

Université de Liège
Département de Géographie
Sciences Géographiques

Approche hédonique du marché des terrains à bâtir

***Modélisation des prix dans les bassins d'emploi de Liège
et de Charleroi, comparaison interbassin
et évolution temporelle liégeoise***

Mémoire présenté par Hubert Maldague

Sous la direction de M. Jean-Marie Halleux

Dans le cadre de l'obtention du diplôme de Master en Sciences Géographiques

Orientation générale

Finalité Développement Territorial et Géomatique

Année académique 2013 – 2014

Approche hédonique du marché des terrains à bâtir

***Modélisation des prix dans les bassins d'emploi de Liège
et de Charleroi, comparaison interbassin
et évolution temporelle liégeoise***

Mémoire présenté par Hubert Maldague

Sous la direction de M. Jean-Marie Halleux

Dans le cadre de l'obtention du diplôme de Master en Sciences Géographiques

Orientation générale

Finalité Développement Territorial et Géomatique

Année académique 2013 – 2014

Avant tout, je tiens à remercier Monsieur J.-M. Halleux, promoteur de ce mémoire, pour son encadrement, sa disponibilité et les précieux conseils promulgués tout au long de la réalisation du présent travail.

Mes remerciements vont également à Madame B. Mérenne-Schoumaker et Monsieur R. Billen qui ont accepté de lire ce mémoire.

Je remercie par ailleurs toutes les personnes qui, de près ou de loin, ont contribué à l'élaboration de ce travail par leurs conseils et leur aide précieuse, notamment Madame M. Trotta et Messieurs Y. Cornet et J.-P. Kasprzyk.

J'adresse aussi mes remerciements à ma famille, mes condisciples et amis pour leurs encouragements et avis.

Enfin, j'exprime ma complète gratitude à ma mère, qui a relu attentivement l'ensemble du travail, et à ma compagne pour leur soutien inconditionnel durant cette dernière année.

*Hubert Maldague
Août 2014*

Table des matières

TABLE DES MATIÈRES	2
TABLE DES FIGURES	5
TABLE DES TABLEAUX	6
INTRODUCTION GÉNÉRALE	7
THÉMATIQUE ET PROBLÉMATIQUE	7
HYPOTHÈSES DE TRAVAIL	8
1. ETAT DE L'ART	11
1.1. QUELQUES GRANDES DÉFINITIONS	11
1.1.1. Le foncier	11
1.1.2. Valeur et prix	11
1.2. LE FONCTIONNEMENT DES MARCHÉS FONCIERS ET IMMOBILIERS	12
1.2.1. Valeur foncière et prix foncier	12
1.2.2. La segmentation typologique des marchés fonciers	12
1.2.3. Un marché des biens fonciers en concurrence imparfaite	13
1.2.4. Un marché inefficace	14
1.2.5. L'évaluation foncière et immobilière	14
1.2.6. Marginalité décroissante des prix en fonction de la superficie	15
1.3. L'APPROCHE HÉDONIQUE	15
1.3.1. Le terrain comme ensemble de caractéristiques	15
1.3.2. Un marché implicite	16
1.3.3. L'approche hédonique	16
1.3.4. Apports de l'approche hédonique à l'étude des marchés fonciers et immobiliers	18
2. MÉTHODOLOGIE	22
2.1. PRÉSENTATION DES MODÈLES DÉVELOPPÉS	22
2.2. PRÉPARATION ET GÉOLOCALISATION DES DONNÉES LIÉES AUX TERRAINS	22
2.2.1. Immoweb	23
2.2.2. Listing papier des transactions foncières des années nonante	25
2.2.3. Géolocalisation des observations	25
2.2.4. Préparation des données provenant d'Immoweb	27
2.2.5. Préparation des données provenant du listing papier	28
2.3. PRÉPARATION DES VARIABLES	28
2.3.1. Variable dépendante	29
2.3.2. Variables structurelles	29
2.3.3. Variables d'accessibilité	30
2.3.4. Variables d'environnement physique	34
2.3.5. Variables relatives à la valorisation sociale	41
2.4. LES ÉTAPES DE MODÉLISATION	42
2.4.1. Sources logicielles utilisées pour la modélisation	42
2.4.2. Paramètres utilisés lors des régressions	42
2.4.3. Statistiques descriptives	44
2.4.4. Analyse des corrélations entre variables explicatives	44
2.4.5. Régression linéaire multiple par MCO	44

2.4.6. Correction de l'hétéroscédasticité	45
2.4.7. Retrait des observations à résidus extrêmes et ajustement du modèle	46
2.4.8. Contrôle de l'absence d'autocorrélation spatiale	46
3. MODÉLISATION DU PRIX DES TERRAINS À BÂTIR DANS LE BASSIN D'EMPLOI DE LIÈGE	48
3.1. SÉLECTION DES OBSERVATIONS	48
3.1.1. Prise en compte du bassin d'emploi	48
3.1.2. Isolement du marché des terrains à bâtir	48
3.1.3. Isolement du marché des terrains juridiquement urbanisables	49
3.2. MODÉLISATION DES PRIX POUR L'ENSEMBLE DU BASSIN D'EMPLOI DE LIÈGE EN 2013-2014	50
3.2.1. Développement du modèle	50
3.2.1.1. Statistiques descriptives	50
3.2.1.2. Analyse des corrélations entre variables explicatives	53
3.2.1.3. Tentative de développement d'un premier modèle	53
3.2.1.4. Modélisation par MCO	55
3.2.1.5. Correction de l'hétéroscédasticité	57
3.2.1.6. Suppression des observations à résidus extrêmes et ajustement du modèle	59
3.2.1.7. Analyse de l'autocorrélation spatiale des résidus	62
3.2.2. Présentation des résultats	63
3.2.2.1. Variables structurelles	64
3.2.2.2. Variables d'accessibilité	66
3.2.2.3. Variables d'environnement physique	68
3.2.2.4. Variables liées à la valorisation sociale	71
3.3. MODÉLISATION DES PRIX POUR LA PÉRIPHÉRIE DU BASSIN D'EMPLOI DE LIÈGE EN 2013-2014	71
3.3.1. Développement du modèle	72
3.3.1.1. Statistiques descriptives	72
3.3.1.2. Analyse des corrélations entre variables explicatives	73
3.3.1.3. Modélisation par MCO	73
3.3.1.4. Correction de l'hétéroscédasticité	74
3.3.1.5. Suppression des observations à résidus extrêmes et ajustement du modèle	75
3.3.1.6. Analyse de l'autocorrélation spatiale des résidus	78
3.3.2. Présentation des résultats	79
3.3.2.1. Variables structurelles	80
3.3.2.2. Variables d'accessibilité	80
3.3.2.3. Variables d'environnement physique	81
3.3.2.4. Variables de valorisation sociale	82
3.4. CONCLUSION SUR LES MODÈLES LIÉGEOIS	82
4. MODÉLISATION DU PRIX DES TERRAINS À BÂTIR DANS LE BASSIN D'EMPLOI DE CHARLEROI	84
4.1. SÉLECTION DES OBSERVATIONS	84
4.2. MODÉLISATION DES PRIX POUR L'ENSEMBLE DU BASSIN D'EMPLOI DE CHARLEROI EN 2013-2014	85
4.2.1. Développement du modèle	85
4.2.1.1. Statistiques descriptives	85
4.2.1.2. Analyse des corrélations entre variables explicatives	87
4.2.1.3. Modélisation par MCO	88
4.2.1.4. Correction de l'hétéroscédasticité	89
4.2.1.5. Suppression des observations à résidus extrêmes et ajustement du modèle	90
4.2.1.6. Analyse de l'autocorrélation spatiale	92
4.2.2. Présentation des résultats	93
4.2.2.1. Variables structurelles	94
4.2.2.2. Variables d'accessibilité	95

4.2.2.3. Variables d'environnement physique	97
4.2.2.4. Variables de valorisation sociale.....	99
4.3. CONCLUSION SUR LE MODÈLE CAROLORÉGIEN.....	99
5. COMPARAISON ENTRE LES MODÈLES LIÉGEOIS ET LE MODÈLE CAROLORÉGIEN	100
5.1. VARIABLES COMPARABLES	101
5.2. VARIABLES COMPARABLES AVEC PRUDENCE.....	102
5.3. VARIABLES COMPARABLES PAR SEUILS DIFFÉRENTS	102
5.4. VARIABLES DIFFICILEMENT COMPARABLES.....	103
5.5. CONCLUSION	105
6. MODÉLISATION DU PRIX DES TERRAINS À BÂTIR DE LA FIN DES ANNÉES NONANTE	106
6.1. SÉLECTION DES OBSERVATIONS	106
6.2. MODÉLISATION DU PRIX DES TERRAINS À BÂTIR DE LA FIN DES ANNÉES NONANTE	106
6.2.1. Développement du modèle	106
6.2.1.1. Statistiques descriptives.....	107
6.2.1.2. Analyse des corrélations entre variables explicatives	109
6.2.1.3. Modélisation par MCO	109
6.2.1.4. Correction de l'hétéroscédasticité.....	110
6.2.1.5. Suppression des observations à résidus extrêmes et ajustement du modèle.....	110
6.2.1.6. Analyse de l'autocorrélation spatiale des résidus.....	111
6.2.2. Présentation des résultats.....	112
6.2.2.1. Variables structurelles.....	113
6.2.2.2. Variables d'accessibilité.....	114
6.2.2.3. Variables d'environnement physique	115
6.2.2.4. Variables de valorisation sociale.....	116
6.3. CONCLUSION	116
7. COMPARAISON TEMPORELLE LIÉGEOISE	117
CONCLUSION GÉNÉRALE ET PERSPECTIVES.....	120
BIBLIOGRAPHIE	123
ANNEXES	

Table des figures

FIGURE 1 – DIAGRAMME D'ORGANISATION DES CATÉGORIES ET SOUS-CATÉGORIES DE VARIABLES.	16
FIGURE 2 – RÉSIDUS EN FONCTION DE LA SUPERFICIE ISSUS DU MODÈLE 0 – 2400 M ²	50
FIGURE 3 – RÉPARTITION DES OBSERVATIONS RETENUES AU SEIN DE LA ZONE D'ÉTUDE.	51
FIGURE 4 – DISTRIBUTION DES VALEURS DE PRIX.	52
FIGURE 5 – DISTRIBUTION DES VALEURS DE SUPERFICIE.	52
FIGURE 6 – DISTRIBUTION DES VALEURS DE PENTE.	52
FIGURE 7 – CLUSTERS LISA POUR LE PREMIER MODÈLE.	54
FIGURE 8 – DISTRIBUTION DES RÉSIDUS EN FONCTION DE LA SUPERFICIE.	57
FIGURE 9 – DISTRIBUTION DES RÉSIDUS EN FONCTION DE LA SUPERFICIE APRÈS CORRECTION DE L'HÉTÉROSCÉDASTICITÉ.	58
FIGURE 10 – DISTRIBUTION DES RÉSIDUS APRÈS ÉPURATION.	60
FIGURE 11 – DIAGRAMME QUANTILE-QUANTILE DES RÉSIDUS.	61
FIGURE 12 – DISTRIBUTION DES RÉSIDUS EN FONCTION DU PRIX.	61
FIGURE 13 – CLUSTERS LISA POUR LE MODÈLE GÉNÉRAL RETENU.	62
FIGURE 14 – DISTRIBUTION DES VALEURS DE PRIX.	72
FIGURE 15 – DISTRIBUTION DES VALEURS DE SUPERFICIE.	72
FIGURE 16 – DISTRIBUTION DES VALEURS DE PENTE.	73
FIGURE 17 – DISTRIBUTION DES RÉSIDUS EN FONCTION DE LA SUPERFICIE.	74
FIGURE 18 – DISTRIBUTION DES RÉSIDUS EN FONCTION DE LA SUPERFICIE APRÈS CORRECTION DE L'HÉTÉROSCÉDASTICITÉ.	75
FIGURE 19 – DISTRIBUTION DES RÉSIDUS APRÈS ÉPURATION.	76
FIGURE 20 – DIAGRAMME QUANTILE-QUANTILE DES RÉSIDUS.	77
FIGURE 21 – DISTRIBUTION DES RÉSIDUS EN FONCTION DU PRIX.	77
FIGURE 22 – CLUSTERS LISA POUR LE MODÈLE PÉRIPHÉRIQUE.	78
FIGURE 23 – RÉSIDUS DE LA VARIABLE SUPERFICIE ISSUS DU MODÈLE 0 – 2400 M ²	85
FIGURE 24 – RÉPARTITION DES OBSERVATIONS RETENUES AU SEIN DE LA ZONE D'ÉTUDE.	86
FIGURE 25 – DISTRIBUTION DES VALEURS DE PRIX.	86
FIGURE 26 – DISTRIBUTION DES VALEURS DE SUPERFICIE.	87
FIGURE 27 – DISTRIBUTION DES VALEURS DE PENTE.	87
FIGURE 28 – DISTRIBUTION DES RÉSIDUS EN FONCTION DE LA SUPERFICIE.	89
FIGURE 29 – DISTRIBUTION DES RÉSIDUS EN FONCTION DE LA SUPERFICIE APRÈS CORRECTION DE L'HÉTÉROSCÉDASTICITÉ.	90
FIGURE 30 – DISTRIBUTION DES RÉSIDUS APRÈS ÉPURATION.	91
FIGURE 31 – DIAGRAMME QUANTILE-QUANTILE DES RÉSIDUS.	91
FIGURE 32 – DISTRIBUTION DES RÉSIDUS EN FONCTION DU PRIX.	92
FIGURE 33 – CLUSTERS LISA POUR LE MODÈLE CAROLORÉGIEN.	93
FIGURE 34 – TAUX DE DISPONIBILITÉ FONCIÈRE EN 2008.	101
FIGURE 35 – DISTRIBUTION DES VALEURS DE PRIX AVANT ÉCRÉMAGE.	107
FIGURE 36 – DISTRIBUTION DES VALEURS DE PRIX APRÈS ÉCRÉMAGE.	108
FIGURE 37 – DISTRIBUTION DES VALEURS DE SUPERFICIE.	108
FIGURE 38 – DISTRIBUTION DES VALEURS DE PENTE.	108
FIGURE 39 – RÉPARTITION DES BIENS RETENUS AU SEIN DE LA ZONE D'ÉTUDE.	109
FIGURE 40 – DISTRIBUTION DES RÉSIDUS APRÈS ÉPURATION.	111
FIGURE 41 – DIAGRAMME QUANTILE-QUANTILE DES RÉSIDUS.	111
FIGURE 42 – CLUSTERS LISA POUR LE MODÈLE DES ANNÉES NONANTE.	112

Table des tableaux

TABLEAU 1 – STATISTIQUES DESCRIPTIVES APPLIQUÉES AUX SUPERFICIES DE L'ENSEMBLE DE L'ÉCHANTILLON LIÉGEOIS.	49
TABLEAU 2 – STATISTIQUES DESCRIPTIVES POUR LE PRIX ET LA SUPERFICIE.	51
TABLEAU 3 – PARAMÈTRES DU MODÈLE RETENU DÉVELOPPÉ PAR MCO.	56
TABLEAU 4 – PARAMÈTRES DU MODÈLE CORRIGEANT L'HÉTÉROSCÉDASTICITÉ.	58
TABLEAU 5 – PARAMÈTRES DU MODÈLE CORRIGEANT L'HÉTÉROSCÉDASTICITÉ APRÈS SUPPRESSION DES RÉSIDUS DÉLINQUANTS.	59
TABLEAU 6 – RÉSULTATS DU MODÈLE GÉNÉRAL LIÉGEOIS.	63
TABLEAU 7 – CLASSEMENT DES VARIABLES EXPLICATIVES PAR ORDRE D'IMPORTANCE.	64
TABLEAU 8 – ILLUSTRATION DE LA MARGINALITÉ DÉCROISSANTE DES PRIX SELON LA SUPERFICIE.	65
TABLEAU 9 – PARAMÈTRES DU MODÈLE RETENU DÉVELOPPÉ PAR MCO.	73
TABLEAU 10 – PARAMÈTRES DU MODÈLE CORRIGEANT L'HÉTÉROSCÉDASTICITÉ.	74
TABLEAU 11 – PARAMÈTRES DU MODÈLE CORRIGEANT L'HÉTÉROSCÉDASTICITÉ APRÈS SUPPRESSION DES RÉSIDUS DÉLINQUANTS.	76
TABLEAU 12 – RÉSULTATS DU MODÈLE PÉRIPHÉRIQUE LIÉGEOIS.	79
TABLEAU 13 – CLASSEMENT DES VARIABLES EXPLICATIVES PAR ORDRE D'IMPORTANCE.	79
TABLEAU 14 – STATISTIQUES DESCRIPTIVES APPLIQUÉES AUX SUPERFICIES DE L'ENSEMBLE DE L'ÉCHANTILLON CAROLORÉGIEN.	84
TABLEAU 15 – STATISTIQUES DESCRIPTIVES POUR LE PRIX ET LA SUPERFICIE.	85
TABLEAU 16 – PARAMÈTRES DU MODÈLE RETENU DÉVELOPPÉ PAR MCO.	88
TABLEAU 17 – PARAMÈTRES DU MODÈLE CORRIGEANT L'HÉTÉROSCÉDASTICITÉ.	89
TABLEAU 18 – PARAMÈTRES DU MODÈLE CORRIGEANT L'HÉTÉROSCÉDASTICITÉ APRÈS SUPPRESSION DES RÉSIDUS DÉLINQUANTS.	90
TABLEAU 19 – RÉSULTATS DU MODÈLE CAROLORÉGIEN.	94
TABLEAU 20 – CLASSEMENT DES VARIABLES EXPLICATIVES PAR ORDRE D'IMPORTANCE.	94
TABLEAU 21 – ILLUSTRATION DE LA MARGINALITÉ DÉCROISSANTE DES PRIX SELON LA SUPERFICIE.	95
TABLEAU 22 – COMPARAISON ENTRE LE MODÈLE CAROLORÉGIEN ET LES DEUX MODÈLES LIÉGEOIS.	100
TABLEAU 23 – STATISTIQUES DESCRIPTIVES POUR LE PRIX ET LA SUPERFICIE.	109
TABLEAU 24 – PARAMÈTRES DU MODÈLE RETENU DÉVELOPPÉ PAR MCO.	110
TABLEAU 25 – PARAMÈTRES DU MODÈLE CORRIGEANT L'HÉTÉROSCÉDASTICITÉ.	110
TABLEAU 26 – PARAMÈTRES DU MODÈLE CORRIGEANT L'HÉTÉROSCÉDASTICITÉ APRÈS SUPPRESSION DES RÉSIDUS DÉLINQUANTS.	110
TABLEAU 27 – RÉSULTATS DU MODÈLE LIÉGEOIS DE LA FIN DES ANNÉES NONANTE.	112
TABLEAU 28 – CLASSEMENT DES VARIABLES EXPLICATIVES PAR ORDRE D'IMPORTANCE.	113
TABLEAU 29 – CORRECTION DES VALEURS DES COEFFICIENTS DE RÉGRESSION.	113
TABLEAU 30 – ILLUSTRATION DE LA MARGINALITÉ DÉCROISSANTE DES PRIX SELON LA SUPERFICIE.	114
TABLEAU 31 – COMPARAISON TEMPORELLE ENTRE LES MODÈLES LIÉGEOIS.	117

Introduction générale

Thématique et problématique

Au sein du vaste domaine qu'est l'aménagement du territoire, l'occupation du sol prend une place importante et la parcimonie de son utilisation devient au fil du temps un défi majeur face aux enjeux environnementaux avec lesquels nos sociétés modernes doivent désormais compter. Ceci est d'autant plus vrai en Wallonie où la consommation importante d'espace par habitant et une aspiration à vivre en milieu calme, vert et rural ont conduit à un étalement poussé de l'urbanisation (J.-M. Halleux, 2012). Les problèmes environnementaux et économiques présents et futurs appellent à un changement de mentalité et à une nouvelle conception de l'aménagement du territoire. Ceux-ci doivent aboutir à une gestion parcimonieuse de la ressource foncière, permettant un équilibre entre les différentes affectations s'y exerçant et une minimisation des coûts pour la collectivité et l'environnement. Cependant, cette nouvelle conception ne peut se faire sans la compréhension du fonctionnement actuel des marchés fonciers et immobiliers. Ce n'est que sur base de la connaissance des logiques actuelles que nos sociétés pourront envisager les solutions aux défis qui se présentent à elles. Or, en Belgique, le mécanisme de fonctionnement des marchés reste relativement mal connu (J.-M. Halleux, 2005), malgré un développement certain de ce champ d'étude dans notre pays. Le développement des SIG et des technologies associées ainsi qu'une utilisation de plus en plus aisée des logiciels de traitements statistiques amènent pourtant de nouveaux moyens d'étude des marchés fonciers et immobiliers mais aussi une facilité offerte dans la génération des données nécessaires à cette étude.

Parmi les logiques de fonctionnement des marchés fonciers et immobiliers, l'étude de la formation des prix des habitations et des terrains à bâtir revêt un intérêt tout particulier. Elle doit ainsi permettre de connaître la contribution de la localisation, de l'accessibilité et de l'environnement du bien à cette formation. En effet, les seules caractéristiques structurelles de ce bien (superficie du terrain, nombre d'étages de la maison...) sont nécessaires mais insuffisantes pour expliquer son prix. Néanmoins, le cadre belge reste peu investigué. Citons toutefois, dans le domaine du foncier, une contribution de C. Vandermotten *et al.* (1996) présentant un modèle permettant d'établir les prix des terrains à bâtir sur base d'un certain nombre de variables explicatives et significatives. Un autre travail s'est également attaché à l'influence de différentes variables sur le prix des terrains à bâtir en utilisant des données agrégées au niveau des communes (F. Goffette-Nagot *et al.*, 2011). Il a mis en évidence le rôle important de l'accessibilité à l'emploi, alors que l'influence de l'environnement semble plus réduite. Les auteurs précisent néanmoins que le rôle des variables environnementales est sans doute bien plus important à l'échelle locale.

Cette dernière étude appelle à la nécessité d'analyser le comportement des marchés fonciers et immobiliers en utilisant non pas des données agrégées, mais un échantillon de terrains à bâtir considérés individuellement. La discipline de l'approche hédonique est bien adaptée à ce type d'étude. Cette pratique est développée à l'étranger, avec notamment une littérature scientifique abondante à ce sujet en Extrême-Orient, en Amérique et, dans une moindre mesure, en Europe. En Belgique, ce champ de recherche n'en est encore qu'au stade de l'émergence. Durant les dernières années, différents mémoires réalisés à l'Université de Liège se sont intéressés à

l'approche hédonique appliquée aux marchés immobiliers, et plus précisément le marché des maisons unifamiliales dans les régions de Bruxelles (C. Delbar, 2012) et de Liège (Y. Grignet, 2011 ; P. Dethier, 2012). Les modèles développés montrent une grande robustesse statistique de l'impact des variables structurelles¹ des habitations sur leur prix de vente. L'influence de certaines variables de localisation est par contre nettement moins claire. Certains déterminants ne présentent pas de significativité statistique, ou l'explication qu'ils procurent paraît contradictoire avec les résultats d'autres études. La revue de la littérature associée à ce mémoire n'a fait que renforcer cette impression.

L'existence même des variables structurelles expliquerait la médiocre mise en évidence de l'influence de l'environnement sur les prix. La suppression des caractéristiques structurelles du bâti, via l'étude des terrains, pourrait permettre de mieux cerner la contribution des variables de localisation, parmi lesquelles les variables liées à l'environnement (C. J. Fausold et R. J. Lilieholm, 1999 ; L. Tyrvaïnen et A. Miettinen, 2000 ; V. R. Sharma, 2013). De plus, les variables de localisation jouent un rôle important dans l'établissement du prix des terrains à bâtir (J. Comby, 2003).

Plusieurs études ont montré la faisabilité de l'application de l'approche hédonique aux terrains vierges de construction en Europe Occidentale (R. Paris, 2003 ; G. Ahlfeldt et W. Maennig, 2008 ; A. Donzel *et al.*, 2008), en Asie orientale (K. Andoh et M. Ohta, 1997 ; X. Gao et Y. Asami, 2007 ; Y. Liu *et al.*, 2010 ; Y. Fumihiko *et al.*, 2011 ; K. Matsui et M. Fukushige, 2012), en Amérique du Nord (C. G. Leggett et N. E. Bockstael, 1999 ; V. R. Sharma, 2013) ou encore en Océanie (C. J. Barnett, 1985).

Par ce travail, nous souhaitons contribuer au développement du champ d'étude de l'approche hédonique en Belgique, plus particulièrement dans le cadre des marchés fonciers.

Hypothèses de travail

Les différentes lectures effectuées ont abouti au questionnement à la base de ce mémoire. Premièrement, il n'existe sans doute pas d'étude sur les terrains à bâtir utilisant l'approche hédonique en Belgique. De plus, le sujet a été entamé sur base des constatations mises en évidence lors de l'étude du marché des maisons unifamiliales dans la région de Liège. Nous souhaitons pouvoir appliquer l'approche hédonique et la méthodologie développée par Y. Grignet (2011) aux marchés fonciers liégeois et mettre davantage en évidence l'influence des variables de localisation. Nous voulons ensuite vérifier et comparer les résultats avec ceux obtenus dans un autre bassin d'emploi belge, celui de Charleroi. Dès lors, la question de recherche se construit comme suit : les variables de localisation expliquent-elles la formation des prix des terrains à bâtir dans les régions de Liège et de Charleroi ?

La bibliographie présentée montre qu'il est possible d'appliquer l'approche hédonique aux marchés fonciers et d'étudier l'influence des variables de localisation sur la formation des prix. Les trois mémoires réalisés à l'Université de Liège ont, de plus, mis en évidence qu'il était possible de développer un modèle hédonique décrivant le marché des maisons unifamiliales sur base d'un échantillon constitué à partir des annonces immobilières obtenues sur le site Immoweb, et ce sur des zones d'études différentes. Nous pouvons supposer qu'il en va de même pour les terrains à

¹ Les différentes catégories de variables sont présentées dans l'état de l'art. Voir point 1.3.1.

bâtir. C'est pourquoi ce mémoire propose comme premier objectif la vérification de l'hypothèse suivante :

L'influence des variables de localisation sur la formation des prix des terrains à bâtir des bassins d'emplois de Liège et de Charleroi est identifiable à partir de données collectées sur un site d'annonces immobilières.

Cette hypothèse en suppose une autre, à savoir qu'il est possible de construire un modèle hédonique robuste pour les deux zones d'étude précitées en utilisant les annonces de vente de terrains à bâtir du site Immoweb. Une autre hypothèse est proposée :

L'influence des variables est comparable entre les bassins d'emplois de Liège et de Charleroi.

Le choix du bassin d'emploi n'est pas anodin. Il s'agit bien souvent de l'aire de fonctionnement des marchés (J.-M. Halleux, 2012b). Dès lors, la limite de ces bassins correspond aux limites de la segmentation spatiale des marchés. Le critère de la distance au lieu de travail est en effet prépondérant dans les décisions de localisation des candidats acheteurs d'un bien foncier ou immobilier. La deuxième justification est l'arbitrage entre la nécessité d'obtenir un nombre d'observations suffisamment important pour permettre la construction de modèles statistiquement robustes et l'obligation de considérer une aire ne recouvrant pas différentes réalités. Ainsi, il est illusoire de vouloir travailler à l'échelle de la commune ou de quelques communes en particulier pour la simple et bonne raison que le nombre d'observations collectées sera nettement insuffisant pour permettre une régression robuste. Il est en effet généralement admis qu'il faut un minimum strict de trente observations pour autoriser l'utilisation des statistiques paramétriques et qu'il est souhaitable d'avoir à disposition un minimum de cent observations pour permettre la construction d'analyses statistiques solides. Enfin, à l'autre extrême, travailler à l'échelle de la Wallonie n'aurait que peu de sens compte tenu de l'ensemble des différents bassins d'emplois et bassins de vie recouverts (J.-M. Halleux, 2012b). C'est ainsi que les données ont été collectées en tenant compte des bassins d'emplois tels que définis par l'IWEPS² pour l'année 2008. Ceux-ci sont construits suivant la méthodologie MIRABEL de l'INSEE³ : elle consiste à utiliser des matrices origine-destination des travailleurs, élaborée dans le cas présent sur base des données fournies par l'ONSS⁴. Nous renvoyons à l'article de la CPDT⁵ pour davantage d'informations concernant la construction des bassins d'emploi (CPDT, 2011). La liste des communes appartenant aux bassins de Liège et de Charleroi est consultable aux annexes I et II.

L'opportunité d'étudier les évolutions temporelles entre les années nonante et la situation actuelle s'est présentée suite à une proposition de J.-M. Halleux de valoriser une liste de terrains à bâtir dont les transactions de vente ont eu lieu entre les années 1991 et 1998 dans l'arrondissement de Liège. Ce listing de transactions a été transmis en 1999 à J.-M. Halleux par la Chambre des notaires de Liège dans le cadre de sa recherche doctorale. Plusieurs études consultées se sont par ailleurs intéressées à l'évolution temporelle de l'influence des variables structurelles et de localisation propres aux terrains à bâtir et aux biens immobiliers. Elles montrent que cette influence est rarement fixe dans le temps. Dès lors, une nouvelle question de

² Institut wallon pour l'évaluation statistique et prospective.

³ Institut National des Statistiques français.

⁴ Office national de sécurité sociale.

⁵ Conférence permanente du développement territorial.

recherche a pu être émise : des changements de l'influence des variables de localisation entre les années nonante et la situation actuelle peuvent-ils être mis en évidence ? Ainsi, nous pouvons émettre l'hypothèse suivante :

L'influence des variables de localisation a évolué entre les années nonante et la période actuelle dans la région de Liège.

Nous sommes ici tenus par la répartition géographique des données historiques, dont l'aire n'occupe pas entièrement le bassin d'emploi. Le terme « région » recouvre deux arrondissements, ceux de Liège et de Waremme. La liste des communes faisant partie de la zone d'étude est reprise à l'annexe III.

Dans un premier temps, nous nous attarderons sur quelques concepts et explications théoriques liés à l'approche hédonique et au fonctionnement des marchés fonciers et immobiliers, nécessaires pour la bonne compréhension des matières traitées au cours de ce mémoire.

Deuxièmement, nous présenterons la méthodologie employée, déclinée en différentes étapes. Nous expliquerons les démarches liées à la collecte de données relatives aux terrains ainsi que leur localisation. Nous traiterons ensuite des variables retenues et leur création. Nous nous arrêterons également sur les étapes théoriques de modélisation proprement dite.

Nous passerons alors au développement du modèle actuel liégeois et à la présentation des résultats que nous discuterons. Nous verrons également comment nous avons pu améliorer les performances de la régression en ne tenant pas compte des biens localisés au sein de l'agglomération opérationnelle liégeoise.

Nous présenterons ensuite le développement du modèle carolorégien et de ses résultats, avant de mettre en parallèle les deux modèles généraux ainsi constitués.

Enfin, nous développerons le modèle liégeois des années nonante, ainsi qu'un modèle actuel de comparaison, basé sur une zone d'étude identique. Nous confronterons brièvement quelques variables.

1. Etat de l'art

Nous présentons ici quelques grands concepts nécessaires à la compréhension des développements élaborés dans le cadre de ce mémoire. A la suite de définitions principales, nous caractérisons le fonctionnement des marchés fonciers et immobiliers et les apports de l'approche hédonique et de ses variables.

1.1. Quelques grandes définitions

1.1.1. Le foncier

J. Comby (2013) donne au terme *foncier* la définition suivante :

« Dans son acception première, le foncier désigne la terre (et parfois, par extension, les constructions et aménagements qui sont dessus), non pour elle-même, mais en tant que fonds d'une exploitation ou d'une rente. Il n'y a pas de foncier sans une notion d'appropriation et de valorisation. Dans l'étude du foncier, le juridique et l'économique sont intimement liés. Sans droit d'appropriation (individuelle ou collective), une terre n'a aucune valeur particulière : lorsqu'elle sort du champ juridique, la terre sort également du champ économique. » (J. Comby, 2013)

Le bien foncier est un bien complexe, doublement approprié, durable et hétérogène (J.-M. Halleux, 2012b) :

- complexe car le foncier est un champ de connaissances mal identifié : il est situé à la croisée entre différentes disciplines telles que l'économie, le droit, l'urbanisme et la géographie. Les acteurs agissant dans ce domaine sont diversifiés : promoteurs, constructeurs, financiers, notaires, mais aussi des particuliers vendant et achetant les biens. C'est de plus un sujet administré à différentes échelles territoriales ;
- doublement approprié car la parcelle appartient à la fois à un propriétaire et à une collectivité qui s'approprie le territoire dans lequel se trouve le bien foncier ;
- durable en raison de sa durée de vie économiquement infinie ;
- hétérogène, car en raison de son caractère localisé et au-delà des différentes catégories de terrains, un terrain d'une même catégorie ne sera jamais identique à un autre terrain de cette même catégorie.

Il convient dès lors de faire la distinction avec la définition populaire du terme foncier, qui tend à l'approcher d'une notion générale de l'immobilier. Dans le langage courant, l'immobilier fait référence aux biens qui ne peuvent être déplacés, aussi bien des constructions que des terrains nus. Or, l'objet de ce mémoire est le terrain à bâtir, donc un bien foncier, à savoir un terrain nu pouvant supporter une construction (J.-M. Halleux, 2012b).

1.1.2. Valeur et prix

F. Des Rosiers (2001) stipule que « l'établissement de la valeur d'un bien est essentiellement issu du concept de marché : c'est le résultat de l'interaction de l'offre et de la demande, la première reflétant les conditions et contraintes de production de ce bien (coût des intrants, salaires, profit d'entreprise) et la seconde les caractéristiques des consommateurs (goûts et préférences, niveau des revenus, modes, etc.) » (F. Des Rosiers, 2001). Nous pouvons ajouter à cela la définition

suivante : « La valeur marchande d'un bien est définie comme le prix de transaction le plus probable auquel en arriveront un acheteur et un vendeur qui disposent de toute l'information nécessaire pour juger des caractéristiques du produit devant faire l'objet de la transaction (marché de concurrence parfaite) et qui agissent en toute indépendance et en toute liberté, sans que ni l'un ni l'autre ne soit en mesure d'exercer quelque forme de contrôle que ce soit sur le marché du bien en question (marché atomistique). » (F. Des Rosiers, 2001).

Il convient dès lors de bien différencier les notions de prix et de valeur :

- le prix est la somme d'argent que le vendeur reçoit en échange d'un bien ou d'un service ;
- la valeur n'est pas directement observable.

1.2. Le fonctionnement des marchés fonciers et immobiliers

1.2.1. Valeur foncière et prix foncier

Le prix d'un bien n'est égal à sa valeur qu'en de très rares circonstances de marché en concurrence parfaite où l'information circule de manière absolument libre. La valeur vénale d'un bien foncier ou immobilier n'est qu'une approximation du prix sous lequel le bien changera de propriétaire et passera ainsi du vendeur à l'acheteur. Elle est une tendance centrale au sein d'une distribution de probabilités constituées par les prix de vente observés. Selon le postulat qui prétend que les individus basent leurs choix de localisation et de quantité sur le principe de la maximisation du bien-être (S. Rosen, 1974), le prix d'équilibre auquel s'échangera le bien sera plutôt le reflet du consensus trouvé entre l'acheteur et le vendeur, ceux-ci ayant un pouvoir de négociation et d'information parfois fort différent. L'acteur le mieux informé pourra augmenter ou diminuer le prix à son avantage et en tirera ainsi profit (J.-M. Halleux, 2009). Encore ne faut-il pas perdre de vue qu'une compétition entre différents acheteurs potentiels pour l'obtention d'un même bien foncier peut mener à une surenchère. Autrement dit, le propriétaire vendra son bien à l'acheteur qui lui proposera le meilleur prix (J. Comby, 2003).

Dans une région définie, la localisation d'un terrain par rapport à différents pôles attractifs influence grandement la valeur de ce dernier : « D'après la théorie économique urbaine classique, les valeurs des sols des zones les plus accessibles à l'emploi, aux loisirs et aux autres lieux d'attraction sont les plus élevées, toutes choses égales par ailleurs » (G. Deymier, 2003). Une modification portée à l'environnement du bien foncier ou immobilier peut entraîner une variation du prix de ce bien. La dimension temporelle n'est pas à oublier non plus. Soit un laps de temps suffisamment important, la valeur d'un bien pourra évoluer suite aux modifications portées à l'influence des variables explicatives.

1.2.2. La segmentation typologique des marchés fonciers

Le marché foncier est, tel qu'il est défini plus haut, constitué de biens hétérogènes. Il est dès lors possible de catégoriser les différents biens qui le composent.

J.-M. Halleux (2012b) fait la différence entre deux types de marchés fonciers :

- le marché des gisements dans lequel le bien est du foncier brut à aménager, offert par des propriétaires initiaux et demandé par des promoteurs⁶ fonciers ;
- le marché des terrains à bâtir pour lequel le bien est un lot directement constructible, offert par un promoteur foncier et demandé par un promoteur immobilier, qu'il soit professionnel ou particulier.

Outre cela, il existe une autre segmentation entre le marché des terrains « neufs », c'est-à-dire n'ayant jamais reçu de construction auparavant, et le marché des terrains à recycler pour lequel une construction existante sied sur le terrain (J. Comby, 2003). Les constructions existantes sur le deuxième type de terrain seront démolies pour laisser place à une nouvelle opération immobilière.

1.2.3. Un marché des biens fonciers en concurrence imparfaite

Selon J.-M. Halleux (2005), le marché en concurrence parfaite est défini selon le respect de quatre conditions :

- l'atomicité de l'offre et de la demande. Les offreurs et les demandeurs sont suffisamment nombreux pour que les actions de l'un d'entre eux n'engendrent pas de changements de comportement chez les autres acteurs ;
- l'homogénéité du bien échangé. Les biens mis en vente dans leur ensemble peuvent être considérés comme identiques entre eux ;
- la transparence du marché : tous les offreurs et les demandeurs sont parfaitement au courant des conditions de prix qui régissent le marché. Ils sont également au courant des quantités de biens disponibles et des caractéristiques propres de ces biens. Ils ont dès lors une connaissance complète du marché et de ses tenants et aboutissants ;
- la mobilité parfaite des offreurs et des demandeurs : l'entrée dans le marché est caractérisée par une totale liberté tandis que la sortie est exempte de toute entrave quelle qu'elle soit.

Par ses caractéristiques – immobilité physique, indivisibilité, hétérogénéité et information partielle – le marché immobilier (et également foncier) ne répond pas à cette logique (V. Renard, 1975 ; cité par J.-M. Halleux, 2005). La seule loi de l'offre et de la demande, qui établit le prix d'équilibre dans de nombreux marchés économiques, n'est d'aucun secours pour expliquer les prix atteints par les biens fonciers (J. Comby, 2003). En premier lieu, le terrain à bâtir – et plus généralement, tout bien immobilier ou foncier – impose une immobilité à son propriétaire souhaitant le vendre. En deuxième lieu, le fait que, régulièrement, un terrain soit mis en vente par un seul offreur et intéresse un nombre variable, parfois très petit, de potentiels acheteurs va à l'encontre du principe d'atomicité de l'offre et de la demande. Troisièmement, les biens fonciers sont hétérogènes, non seulement à cause de leur localisation, mais aussi en raison de toute une série d'éléments comme la superficie, l'activité que le terrain permet de remplir.... Nous sommes en présence d'une situation de concurrence monopolistique : l'offreur est le seul à offrir un bien différencié, mais les potentiels acheteurs ont la liberté de pouvoir se rabattre vers des produits substitués si le premier bien proposé ne les intéresse pas.

⁶ Promoteur : acteur prenant l'initiative de la réalisation d'une opération foncière ou immobilière (J.-M. Halleux, 2012b).

1.2.4. Un marché inefficace

Selon A. W. Evans (1995), les marchés immobiliers sont inefficaces dans la mesure où il existe un écart entre les évaluations et le niveau réel des transactions dont l'importance peut excéder en moyenne les 10 %, quelles que soient les méthodes d'évaluation, par l'intermédiaire d'experts ou via des statistiques (J.-M. Halleux, 2009). Cela est notamment dû au caractère très hétérogène des biens immobiliers et fonciers qui vont à l'encontre d'un prix qui soit admis à la fois par les offreurs et les demandeurs dans leur entièreté. Le pouvoir de négociation de ces derniers, plus ou moins important selon les cas, peut aboutir à un prix de transaction variable d'un cas à l'autre. A. W. Evans souligne que cette situation expliquerait pourquoi certaines personnes ont fait fortune en investissant dans l'immobilier. L'auteur déclare également à propos des marchés immobiliers et fonciers qu'il n'existe pas de valeur précise pour caractériser le prix de vente d'un bien, mais uniquement une étendue de valeurs possibles. Ces constatations ont été précédemment exposées au point 1.2.1.

1.2.5. L'évaluation foncière et immobilière

Selon Y. Grignet (2011), l'évaluation immobilière peut être définie comme « l'ensemble des méthodes qui permettent d'estimer la valeur d'un bien immobilier ». Son équivalent appliqué aux terrains à bâtir est l'évaluation foncière. En d'autres termes, elle permet de définir, avec plus ou moins de précision, la quantité d'argent que le possesseur peut s'attendre à recevoir lorsqu'il vend son bien.

L'expertise est loin d'être une science exacte. B. de Polignac et J.-P. Monceau (2002, cités par J.-M. Halleux, 2009) signalent que « l'expert conclut à une estimation qui se trouve sensiblement au centre d'une fourchette à l'intérieur de laquelle il pense, compte tenu de ses connaissances du marché, que le bien pourrait se vendre. Même s'il est tabou dans le monde des experts de parler de précision ou d'imprécision d'une expertise, et même si les chiffres qui vont suivre peuvent perturber certains clients, donnons avec modestie quelques indications :

- pour les biens courants, relevant d'un large marché, rester constamment dans une fourchette de + 5 % à - 5 % de la valeur vénale est un objectif très ambitieux. De plus, aucun expert ne peut prétendre s'être toujours situé dans un écart de + 10 % à - 10 % ;
- pour les biens relevant d'un marché étroit, tant du fait de leur nature (châteaux, usines) que de leur situation, il y a des cas où un écart de 30 % voire plus ne signifie pas une erreur de l'expert.

Cela doit être rapproché du fonctionnement imparfait, au sens économique du terme, du marché de l'immobilier. » (B de Polignac et J.-P. Monceau, 2002).

Pour estimer la valeur des biens, les experts immobiliers ont notamment recours à la technique des points de comparaison, l'approche hédonique étant la formalisation de cette méthode (J.-M. Halleux, 2005). Elle consiste à établir le prix d'un bien particulier sur base des prix de transactions récentes et pour lesquelles le bien associé est le plus semblable possible en terme de consistance, d'état et de situation à celui qui est mis en vente. La contribution de chaque attribut à la valeur marchande du bien est isolée à partir d'un échantillon de biens mis en vente et jugés similaires et comparables (J.-M. Halleux, 2009).

1.2.6. Marginalité décroissante des prix en fonction de la superficie

Les prix fonciers sont caractérisés par une marginalité décroissante (J.-M. Halleux, 2005). Cela signifie qu'une augmentation de la superficie d'un terrain à vendre engendre une décroissance du prix moyen par mètre carré, et ce alors que le prix de la transaction en elle-même augmente. Une part de cette décroissance provient de l'existence des gisements fonciers, terrains à prix par mètre carré généralement moins élevé. En retirant ces biens, nous pouvons isoler les deux autres composantes de la marginalité décroissante.

La première provient du fait que l'utilité marginale – soit l'utilité par unité de superficie – que retire un ménage d'une parcelle plus importante diminue, et ce en parallèle du fait qu'un ménage retirera une utilité générale plus importante d'une parcelle plus grande (J.-M. Halleux, 2005). Ainsi, un mètre carré de terrain supplémentaire sera davantage désiré par les acheteurs pour un petit terrain. Au contraire, ce même mètre carré de terrain pour un terrain plus grand procure une utilité marginale moins importante.

La deuxième est due aux coûts de production à charge des promoteurs fonciers, ceux-ci n'étant pas proportionnels avec la superficie. Le promoteur reporte ainsi les coûts qu'il supporte pour la viabilisation du terrain sur le prix de vente qu'il propose aux promoteurs immobiliers (J.-M. Halleux, 2005). Ainsi, pour deux terrains viabilisés, nous observerons un coût par mètre carré plus élevé pour le terrain le plus petit.

1.3. L'approche hédonique

1.3.1. Le terrain comme ensemble de caractéristiques

L'approche hédonique repose sur un principe fondamental énoncé par K. J. Lancaster (1966) et selon lequel les consommateurs tirent leur utilité des caractéristiques des biens et non du bien lui-même. Ceci met en avant le fait que le bien ne doit pas être considéré comme homogène, mais comme un ensemble de caractéristiques. Ce principe est applicable au cas d'une parcelle de terrain : celle-ci peut être décomposée en un ensemble de caractéristiques, chacune d'entre elles pouvant être considérée comme un argument distinct de la fonction d'utilité procurée à son acquéreur (J.-M. Halleux, 2005). Ces caractéristiques du même bien ne peuvent être vendues séparément car elles font partie du même ensemble. Aussi, et compte tenu de ces différentes caractéristiques, le terrain doit être considéré comme un bien hétérogène : il n'existe pas deux terrains à bâtir identiques.

Ainsi, un bien quelconque est caractérisé par un vecteur X possédant k caractéristiques tel que

$$X = (x_1, x_2, \dots, x_k)$$

Avec x_i la valeur d'une caractéristique particulière.

Le fait que le bien immobilier – ou foncier dans le cas présent – soit, comme son nom l'indique, non déplaçable, implique la prise en compte de caractéristiques liées à sa situation géographique et à son environnement (P. Srikhum, 2012) en plus de ses caractéristiques propres. Les différents types de caractéristiques sont détaillés ci-après :

- les caractéristiques intrinsèques ou structurelles, qui sont propres à la structure du bien : superficie du terrain pour les biens fonciers, nombre de pièces, présence d'une cave... pour les biens immobiliers ;
- les caractéristiques extrinsèques ou de localisation. Elles tiennent compte de l'environnement et de la distance à différents pôles, affectations et services. Elles peuvent être divisées en deux groupes : d'une part les variables d'accessibilité au bien et d'autre part les variables liées à l'environnement et qui caractérisent le voisinage de la position du bien. Ce dernier groupe peut à son tour être divisé en deux : celui des variables de qualité physique (ou d'environnement physique) et celui des variables de valorisation sociale.

Le schéma ci-dessous (Figure 1) résume les catégories et sous-catégories des différentes variables.

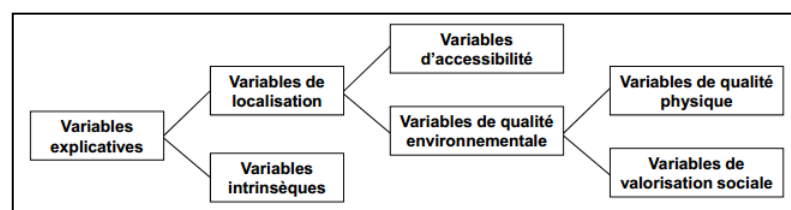


Figure 1 – *Diagramme d'organisation des catégories et sous-catégories de variables (Source : J.-M. Halleux, 2012b).*

1.3.2. Un marché implicite

Le marché des caractéristiques composant le bien foncier est implicite (J.-M. Halleux, 2009). Cette théorie stipule qu'est implicite un marché où s'échangent des caractéristiques dont les prix ne sont pas directement observables (S. Rosen, 1974). Ainsi, un bien immobilier – ou foncier dans le cas présent – est constitué d'un ensemble de caractéristiques qui ne peuvent être vendues séparément et dont le prix est implicite (A. W. Evans, 1995). Si le prix de cet ensemble est connu, celui de chacune de ses composantes ne l'est pas : nous ne connaissons pas le prix du mètre carré de terrain seul, ni le prix de la distance au centre-ville. Il est cependant possible, moyennant un échantillon suffisamment important de biens, de déterminer statistiquement le prix des différentes caractéristiques.

1.3.3. L'approche hédonique

La méthode hédonique a été initiée en 1974 par S. Rosen qui démontre comment l'équilibre d'un marché de biens différenciés, avec un continuum de biens du côté de l'offre et un continuum de consommateurs hétérogènes du côté de la demande, aboutit à une relation fonctionnelle entre les caractéristiques des biens et leur prix, appelée fonction hédonique (M. Baudry et M. Maslianskaïa-Pautrel, 2011). A toute combinaison de caractéristiques est associé le prix qu'un futur propriétaire devra déboursier pour obtenir un bien immobilier ou foncier (J.-M. Halleux, 2009). La fonction hédonique dépend des positions respectives de la totalité des courbes d'enchères et d'acceptation (J.-M. Halleux, 2005). La fonction d'enchère E représente le prix maximum que chaque ménage est prêt à payer pour tout X, le vecteur des k caractéristiques du bien analysé. Ce prix maximum est défini selon le fait que le ménage cherche à maximiser son utilité sous la contrainte de son revenu (J.-M. Halleux, 2005). A contrario, la fonction d'acceptation A représente le prix minimum exigé par le vendeur pour mettre sur le marché l'ensemble des k caractéristiques du bien, ce prix étant lié à la recherche de profit sous contrainte

des coûts de production (J.-M. Halleux, 2005). Les fonctions E et A définies pour un bien sont tangentes en un point précis représentant l'échange. La combinaison de ces points de tangence pour toute une série de biens définit la fonction hédonique. Sa forme est variable selon les marchés car elle dépend de la population des fonctions d'enchère et d'acceptation, elles-mêmes variables selon ces mêmes marchés. Le changement de la fonction hédonique d'un marché à l'autre relève du niveau inter-marché. Cette fonction se définit donc uniquement au sein du marché qu'elle représente : c'est le niveau intra-marché, ici délimité par les frontières du bassin d'emploi. Ajoutons à cela que l'approche hédonique nécessiterait l'utilisation d'un échantillon de biens individuels auxquels sont associés les valeurs individuelles des caractéristiques, et non des données agrégées. En effet, en procédant selon la deuxième manière, l'analyse porte simultanément sur des déterminants de niveau inter-marché et intra-marché (J.-M. Halleux, 2005). Néanmoins, nous ne pourrions éviter l'utilisation de données agrégées telles que le revenu ou le taux de chômage.

Les prix hédoniques peuvent également être définis comme les prix implicites des caractéristiques composant le bien. Ces prix peuvent être déterminés sur base du coût explicite du bien considéré dans son entièreté et des différentes caractéristiques le composant. Ainsi, la fonction hédonique peut être révélée implicitement par les prix des marchés (S. Rosen, 1974), ces prix correspondant à la somme des prix implicites propres à chacune des caractéristiques du bien en question (P. Srikhum, 2012).

C'est à partir des années quatre-vingt que l'approche hédonique a commencé à être appliquée à l'étude des marchés immobiliers et fonciers. La théorie n'est en effet pas propre à ces domaines. Elle a également trouvé des débouchés dans d'autres domaines comme l'environnement, l'agronomie, l'épidémiologie, la météorologie et les sciences de la terre (P. Srikhum, 2012).

La théorie hédonique est idéale pour l'analyse d'éléments constitutifs d'un marché en concurrence monopolistique, ceci étant le cas du marché des terrains à bâtir et de celui des maisons d'habitation (J.-M. Halleux, 2009). Dès lors, il est possible d'envisager la participation de des caractéristiques à la formation du prix des biens considérés (J.-M. Halleux, 2005). La fonction de prix modélisée décrit ainsi la relation entre le prix d'équilibre d'un terrain ou d'une maison, bien hétérogène, et le vecteur de ses caractéristiques (F. Des Rosiers, 2001). Dès lors, en considérant P comme le prix et $x_{1, \dots, k}$ les différentes caractéristiques, nous pouvons écrire :

$$P = f(X) = f(x_1, x_2, \dots, x_k)$$

Si nous adaptons la formule au cas du marché des maisons tout en tenant compte des différentes catégories de variables, nous obtenons l'équation suivante :

$$P = f(V, X, Y, Z, W) + E$$

où :

- P est le prix du bien ;
- V représente l'ensemble des caractéristiques structurelles du terrain ;
- X représente l'accessibilité du bien ;
- Y représente l'ensemble des caractéristiques de valorisation sociale du lieu dans lequel se trouve le bien ;
- Z représente l'ensemble des caractéristiques de l'environnement physique ;

- W représente l'ensemble des caractéristiques de la structure bâtie qui sied sur le terrain ;
- E représente un terme d'erreur aléatoire.

Or, comme nous l'avons énoncé dans l'introduction, l'explication du prix procurée par les caractéristiques de la structure bâtie peut mener à une occultation de l'influence de l'environnement. Dès lors, il devient malaisé de mettre en évidence l'influence des variables composant le vecteur Z. Ce problème peut être supprimé en s'intéressant uniquement aux terrains sans structure bâtie (V. R. Sharma, 2013). Ainsi, l'équation s'appliquant à ces terrains s'exprimera comme suit :

$$P = f(V, X, Y, Z) + E$$

La base de la théorie hédonique repose sur la technique économétrique de la régression multiple à travers les relations existantes entre le prix du bien et les différentes caractéristiques via l'importance de leurs valeurs contributives et le sens de cette influence, positive ou négative. Pour rappel, la fonction de régression linéaire multiple peut s'écrire de la façon suivante :

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n$$

avec n variables explicatives $X_{1,\dots,n}$ de coefficients de régression non-standardisés $B_{1,\dots,n}$ qui expriment le prix Y. B_0 est l'ordonnée à l'origine.

L'approche hédonique comporte certaines limites inhérentes à l'utilisation de la régression multiple. Ainsi, certaines hypothèses théoriques doivent être respectées (F. Des Rosiers, 2001) :

- existence de données complètes et fiables ;
- représentativité de l'échantillon ;
- linéarité de la relation découlant de la modélisation statistique ;
- caractère additif des termes de l'équation ;
- indépendance des variables explicatives ;
- normalité dans la distribution des résidus ;
- constante dans la variation des résidus ;
- indépendance des termes d'erreur.

En présence d'un nombre suffisamment important de biens dans l'échantillon utilisé et un bon choix de variables pertinentes et importantes, le degré de précision sera très proche voire identique à celui d'une évaluation de terrain réalisée par des professionnels (J. S. Dodgson et N. Topham, 1990 ; A. W. Evans, 1995, cités par J.-M. Halleux, 2009). Néanmoins, compte tenu de l'inefficacité des marchés fonciers et immobiliers, l'approche hédonique ne permet pas de modéliser le prix exact d'un bien.

1.3.4. Apports de l'approche hédonique à l'étude des marchés fonciers et immobiliers

De manière générale, les études en immobilier cherchent à mettre en évidence l'influence d'une ou de quelques variables en particulier. Les études généralistes sont plus rares. Néanmoins, les modèles construits utilisent des variables qui reviennent régulièrement au travers de la littérature consultée et qui sont indispensables à la construction d'un modèle hédonique solide. Ainsi, les déterminants structurels et d'accessibilité sont régulièrement repris.

Les études concernant l'application de l'approche hédonique aux marchés fonciers sont moins nombreuses, mais existent néanmoins. A l'instar des études concernant l'immobilier, elles se penchent généralement sur un impact particulier, les travaux plus généralistes restant relativement rares. Toutefois, comme dans le cas des marchés immobiliers, l'usage de variables dites de contrôle est nécessaire pour une modélisation correcte. Ces variables de contrôle, très diverses, ne font pas réellement partie du champ de recherche que l'étude investigate.

En France, citons l'étude de R. Paris (2003) qui étudie l'impact d'une opération d'aménagement sur le prix des terrains à Colombes-Garenne, en banlieue parisienne. Malgré le faible effectif de son échantillon, l'auteur montre l'impact négatif de la proximité de l'opération d'aménagement sur le prix des terrains.

Toujours en France, A. Donzel *et al.* (2008) développent plusieurs modèles hédoniques à partir des données relatives aux terrains à bâtir. A travers cette étude, les auteurs mettent en évidence les effets fixes de la localisation à différentes échelles (région, département, commune) en utilisant des déterminants propres à chacun de ces niveaux d'organisation territoriale. Ils intègrent pour ce faire des données de transactions de terrains à bâtir concernant l'ensemble de la France. Ils mettent en évidence l'impact de la distance à de nombreux services (écoles et commerces...) ainsi qu'à différentes villes selon des seuils de population. L'environnement social et bâti est également étudié. Les résultats montrent que l'impact de la distance aux différents commerces et écoles est négatif, signe de leur attraction. Seules les écoles maternelles semblent avoir un effet répulsif que les auteurs n'ont pas encore pu expliquer. L'étude montre également un effet répulsif d'une grande présence d'immeubles HLM sur le prix des terrains, de même que les quartiers habités par une importante proportion d'étrangers.

Un impact positif des arènes sportives sur le prix des terrains est mis en évidence à Berlin, et ce dans un rayon de 3000 mètres autour desdites arènes (G. Ahlfeldt et W. Maennig). Outre ce paramètre, les auteurs étudient l'influence de la proximité des stations de métro et de RER. Ainsi, ils montrent un impact positif différentiel entre l'ouest et l'est de la ville. Ils imputent cela au développement des réseaux, celui du métro est davantage développé à l'ouest qu'à l'est. Le développement du RER est quant à lui plus poussé à l'est de la ville. Une influence négative d'une part d'étrangers croissante, d'une densité population croissante et de la distance aux corps hydrographiques est également mise en évidence. A l'inverse, la distance aux lignes de chemin de fer a une influence positive.

A l'instar de l'étude des marchés immobiliers, l'application de l'approche hédonique aux terrains est davantage exploitée en Amérique et en Extrême-Orient où il existe une plus grande transparence de l'information (R. Paris, 2003). Les données peuvent ainsi être obtenues plus facilement.

Au Japon, K. Andoh et M. Ohta (1997) s'intéressent à l'impact de la proximité de Tokyo et de Kofu sur les terrains à bâtir de la préfecture de Yamanashi et étudient l'évolution de cet impact entre 1985 et 1995. Ils utilisent le temps d'accès en train aux gares de Tokyo et de Kofu, considérées comme le centre des deux pôles exerçant une attraction sur la préfecture. Le temps d'accès à ces deux pôles a une influence négative, mais cette influence diminue au cours du temps. L'étude montre également l'influence négative d'un éloignement aux échangeurs autoroutiers. Celle-ci

s'explique par l'utilité que procure la proximité de l'échangeur : celui-ci permet d'accéder au réseau express.

Toujours au Japon, X. Gao et Y. Asami (2007) étudient l'influence du paysage sur le prix des terrains à bâtir. Une enquête paysagère a été réalisée au préalable pour construire différents indices représentant la compatibilité du bâti ainsi que le caractère vert de l'environnement. Ces deux indices, en augmentant, entraîne une plus-value. Ceci représente l'aspiration à vivre dans un lieu vert et harmonieux. L'étude montre aussi l'impact positif de la proximité d'une gare et d'un arrêt de bus. La forme du terrain joue également, puisque la largeur à front de rue a un impact positif sur les prix. A l'inverse, un terrain de forme irrégulière voit son prix diminuer. La morphologie des rues dans lesquelles se trouvent les terrains est également importante puisque une rue plus large entraîne une plus-value, tandis qu'une rue en cul-de-sac entraîne une moins-value. Les auteurs expliquent cela par les possibilités et les entraves que représentent les voiries pour l'évacuation en cas d'incendie dans un quartier dense. L'altitude moyenne du quartier fait augmenter les prix lorsque celle-ci s'élève, en raison de la vue et du moindre risque d'inondation engendré par un emplacement à altitude plus élevée. Une densité de population croissante entraîne une moins-value.

Y. Liu *et al.* (2010) étudient l'influence de nombreuses variables liées à l'environnement physique ainsi qu'à l'accessibilité de terrains à bâtir de Danyang, en Chine. L'impact négatif d'une distance croissante aux espaces verts, aux gares, aux bus et aux commerces est mis en évidence. L'influence négative de la proximité d'une industrie est également démontrée.

Y. Fumihiko *et al.* (2011) mettent en évidence l'influence d'un complexe nucléaire japonais sur le prix des terrains des environs à travers le temps. Ils montrent notamment que l'annonce de la construction d'une centrale nucléaire a entraîné une baisse des prix des terrains environnants, phénomène que les auteurs expliquent par l'anxiété générée par le risque nucléaire. D'autres variables sont utilisées comme la distance à la gare la plus proche (impact négatif), la largeur de la rue desservant le terrain (impact positif) et la présence d'industries (impact négatif).

K. Matsui et M. Fukushige (2012) s'intéressent aux prix des terrains à Tokyo. Ils montrent l'influence négative de l'éloignement des gares et d'une densité de population croissante. Un terrain de forme irrégulière ou trapézoïdale connaît une moins-value par rapport au terrain de même superficie mais de forme régulière.

En Amérique, C. G. Leggett et N. E. Bockstael (1999) montrent l'impact positif de la proximité d'un corps hydrographique sur le prix des terrains. Ils notent également une influence positive de la distance aux industries et une influence négative de la distance au centre-ville.

Dans le Colorado, V. R. Sharma (2013) montre l'influence négative de la proximité des forêts, phénomène qu'elle attribue au risque d'incendie. Une distance croissante au centre-ville entraîne une diminution du prix des terrains, toutes choses égales par ailleurs. Il en va de même pour la distance aux stations de ski.

C. J. Barnett (1985) applique l'approche hédonique aux terrains à bâtir de la ville de Perth, en Australie. Cette étude, au contraire des précédentes, est plutôt généraliste. Différentes variables d'accessibilité, d'environnement physique et de valorisation sociale sont utilisées. Citons ainsi la plus-value portée par une localisation dans un cul-de-sac, la proximité d'un espace public et une

distance croissante aux industries. A l'inverse, une distance croissante au centre-ville entraîne une moins-value. Il en va de même pour la distance au littoral et aux écoles.

Les études consultées illustrent bien la diversité des investigations permises par l'application de l'approche hédonique sur les marchés fonciers, mais aussi immobiliers. Elles permettent ainsi de mettre en évidence le rôle de l'accessibilité au centre-ville et à différents services et établissements scolaires mais aussi de connaître l'influence des dessertes des transports en commun sur les prix des maisons et des terrains. Les variables liées à l'environnement physique sont utilisées assez régulièrement pour l'étude de l'influence des aménités naturelles telles que les forêts, les espaces verts, la diversité de l'occupation du sol et les cours d'eau. Les nuisances sont également étudiées au travers des travaux consultés, notamment celles liées aux aéroports, aux lignes de chemin de fer et aux routes à grand trafic. L'approche hédonique contribue ainsi à une meilleure compréhension de l'influence des différentes caractéristiques des biens immobiliers et fonciers. Elle apporte un regard complémentaire à celui de l'expert, dont la connaissance et le savoir-faire ne pourront jamais être remplacés (C. Delbar, 2012).

Ces travaux ainsi que diverses études concernant les marchés immobiliers nous ont permis de déterminer quelles variables pourraient être intégrées dans nos modèles. Nous reviendrons sur la création de ces variables au point 2.3.

2. Méthodologie

Cette section traite de tous les aspects méthodologiques liés à la modélisation. Nous rappelons ici les différents modèles développés dans le cadre de ce mémoire. Ensuite, nous détaillerons l'ensemble des étapes de collecte de données ainsi que les traitements menant à la construction des différentes variables utilisées pour les régressions et choisies sur base de la littérature. Enfin, nous nous attarderons sur le canevas théorique des étapes de modélisation.

2.1. Présentation des modèles développés

Il importe de présenter, à cette étape, les différents modèles qui vont être développés dans le cadre de ce mémoire, compte tenu des questionnements soulevés et des hypothèses avancées au commencement du présent travail.

Le premier modèle développé porte sur les terrains à bâtir appartenant au bassin d'emploi de Liège⁷ tel que défini par l'IWEPS. Nous présentons les différentes phases de son élaboration selon le canevas théorique retenu. Par la suite, nous détaillons les résultats obtenus. Nous verrons comment nous avons mis en évidence la marginalité décroissante du prix en fonction de la superficie, concept abordé au cours de l'état de l'art et concernant les terrains à bâtir. Compte tenu des résultats de ce modèle général, nous tentons de développer des sous-modèles spatiaux afin d'améliorer la mise en évidence de la contribution des différentes variables utilisées sur la formation des prix.

Nous procédons ensuite, selon le même canevas théorique, à l'élaboration du modèle portant sur les terrains à bâtir localisés au sein du bassin d'emploi de Charleroi tel que défini par l'IWEPS⁸. Nous nous attachons ensuite aux résultats de ce modèle, que nous comparons par la suite aux modèles liégeois.

Enfin, nous construisons un modèle pour les terrains liégeois vendus entre 1996 et 1998. La zone d'étude étant réduite à l'arrondissement de Liège et à une partie de celui de Waremme en raison de la disponibilité des données, nous élaborons en parallèle un modèle avec les observations actuelles se situant à l'intérieur de cette zone d'étude. Nous présentons les résultats des deux modèles et les comparons.

2.2. Préparation et géolocalisation des données liées aux terrains

Nous avons jugé important de détailler les sources utilisées et la manière dont les données ont été collectées, préparées et traitées. L'ensemble de ces étapes prend en effet une place considérable dans le déroulement du mémoire, notamment en termes de temps.

Les variables utilisées dans nos modèles proviennent de différentes sources. Certaines d'entre elles sont immédiatement disponibles et utilisables :

- pour les données actuelles, la variable dépendante – le prix – et certaines variables explicatives structurelles et de localisation sont immédiatement affichées sur les

⁷ Voir annexe I pour la liste des communes concernées.

⁸ Voir annexe II pour la liste des communes concernées.

annonces immobilières du site Immoweb. Il s'agit de la superficie, des dimensions, de la présence de raccordements aux impétrants, de l'appartenance à un lotissement et de la présence d'un bâtiment ;

- pour les données des années nonante, nous avons considéré le prix, la superficie du terrain et la date de la transaction, informations reprises dans le listing papier fourni par J.-M. Halleux.

Les deux sources citées ci-dessus ne fournissent pas suffisamment de variables permettant l'élaboration des modèles. De nombreux traitements ont été effectués pour générer les données relatives à l'accessibilité, à l'environnement physique et à l'environnement socio-économique de chaque bien. Nous y viendrons au point 2.3. Compte tenu de son utilisation inévitable pour la génération des valeurs de ces variables de localisation et de deux variables structurelles (pente et orientation), l'adresse est également considérée dans les deux cas.

2.2.1. Immoweb

Les travaux effectués sur les marchés des maisons unifamiliales dans les régions de Liège et de Charleroi se sont intéressés au site d'annonces Immoweb, créé en 1996 (Y. Grignet, 2011 ; C. Delbar, 2012 ; P. Dethier, 2012), et ce à défaut de pouvoir utiliser d'autres sources, notamment notariales. Il se définit en effet comme étant le site d'annonces immobilières en ligne le plus complet et le plus visité de Belgique avec environ 140 000 visiteurs par jour (Y. Grignet, 2011). Compte tenu des différentes remarques et du succès des études de 2011 et de 2012, nous avons choisi de travailler avec ce même site. En effet, il propose une catégorie « terrains à bâtir » qui s'applique très bien au cadre de ce mémoire. Cette catégorie propose des informations primordiales comme la superficie du terrain, son prix et l'adresse. Cependant, cette dernière n'est pas obligatoirement indiquée pour chaque bien.

Un problème de propriété intellectuelle des informations diffusées sur le site Immoweb a été mis en évidence (Y. Grignet, 2011). Le règlement général du site⁹ le stipule très bien :

«Tout le contenu présent sur le site Immoweb, y compris notamment les textes, graphiques, logos, boutons, images, code HTML, base de données et icônes, est la propriété intellectuelle de la société anonyme Immoweb SA et est protégé par des droits de propriété intellectuelle tel que le droit d'auteur et par les droits du producteur d'une base de données.»

La législation belge¹⁰ autorise toutefois l'utilisation de ce type de données à des fins de recherche :

«Art. 7. L'utilisateur légitime d'une base de données qui est licitement mise à la disposition du public de quelque manière que ce soit, peut, sans l'autorisation du producteur :

1° extraire une partie substantielle du contenu d'une base de données non électronique lorsque cette extraction est effectuée dans un but strictement privé ;

⁹ Accessible à l'adresse suivante :

http://www.immoweb.be/fr/Pages/Page.cfm?Page=LegalNotices.TermsOfUse.htm&mycurrent_section=Global&

¹⁰ Loi du 31 août 1998 transposant en droit belge la directive européenne du 11 mars 1996 concernant la protection juridique des bases de données, consultable à l'adresse suivante :

http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=fr&la=F&cn=1998083141&table_name=loi

2° extraire une partie substantielle du contenu de la base de données lorsque cette extraction est effectuée à des fins d'illustration de l'enseignement ou de recherche scientifique pour autant que cette extraction soit justifiée par le but non lucratif poursuivi ;

3° extraire et/ou réutiliser une partie substantielle du contenu de la base de données à des fins de sécurité publique ou aux fins d'une procédure administrative ou juridictionnelle.

Le nom du producteur et le titre de la base de données dont le contenu est extrait à des fins d'illustration de l'enseignement ou de recherche scientifique, doivent être mentionnés. »

La collecte des données d'Immoweb est une opération très coûteuse en termes de durée. Pour accélérer quelque peu la procédure, nous nous sommes intéressés à la méthodologie de collecte développée par Y. Grignet (2011) et reprise par C. Delbar (2012) et P. Dethier (2012). Un script Java a été spécialement développé à cet effet. Cependant, un essai a vite mis en évidence l'inadéquation de ce programme pour le cadre précis de ce mémoire. En effet, il a été conçu dans le but de récolter les informations issues du canevas des annonces de maisons à vendre. Or, ce canevas est très différent de celui des terrains à bâtir, les informations étant formatées différemment. Il a dès lors été nécessaire de reconstruire un programme pour la collecte des informations concernant les terrains à bâtir. Toutefois, nos compétences limitées en programmation Java nous ont poussé à obtenir une aide extérieure précieuse en la personne de Jean-Paul Kasprzyck¹¹ pour construire un code Java performant et adapté au cadre du présent travail. Le code est disponible à l'annexe IV.

La collecte automatisée nécessite en input une liste d'adresses URL successives et séparées par une virgule. Lors de la demande de recherche sur Immoweb, les codes postaux des communes concernées sont entrés dans le champ réservé. Le site limite le nombre de codes à 61 par requête. Il est donc nécessaire d'effectuer plusieurs fois la démarche au cours d'une même séance de collecte. Il est indispensable de demander au site d'afficher les résultats en liste de biens. Les résultats sont affichés les uns après les autres, de manière verticale. Pour collecter rapidement les URL cachées derrière les liens, l'extension *Multi Links* de *Mozilla Firefox* a été utilisée. La sélection est ensuite collée dans Notepad pour ordonnancement, de manière à obtenir les adresses URL alignées les unes à la suite des autres et séparées par une virgule.

Le script Java tourne sous le programme Eclipse, mais peut être utilisé sous n'importe quel logiciel de programmation en Java. Lors de l'exécution, il lit tour à tour chaque fiche et extrait, à partir du site web, les informations pour chaque bien correspondant. Ces informations sont compilées dans un document *comma separated value* ou *CSV* où chaque ligne correspond à un bien. A chacun de ces biens sont associées les différentes valeurs des variables disponibles sur le site Immoweb. Les résultats de chaque exécution du programme sont exportés dans un document Excel complété au fur et à mesure des collectes. Le programme permet également de copier dans un dossier les pages internet correspondant à chacun des biens exportés.

Un autre problème déjà identifié auparavant est la difficulté de récupérer l'adresse de chaque bien (Y. Grignet, 2011). En effet, le script Java ne peut l'exporter, celle-ci étant codée. Constatant cela, il a été décidé d'encoder l'adresse manuellement pour chaque bien. Cette opération, bien que grandement consommatrice en temps, est toutefois utile car elle permet de vérifier en

¹¹ Jean-Paul Kasprzyck est doctorant à l'Unité de Géomatique.

parallèle l'orthographe des noms de rues. Elle a été réalisée en utilisant le site Internet *Google Maps*.

Comme nous l'évoquons plus haut, nous avons constaté que l'information individuelle des biens était rarement complète, particulièrement en ce qui concerne l'adresse. Ceci est problématique dans la mesure où la localisation de chaque terrain à bâtir doit être la plus précise possible pour une attribution adéquate des valeurs des différentes variables créées. La seule utilisation des biens possédant une adresse complète présentait un risque de non-représentativité statistique de l'échantillon, c'est pourquoi il a été décidé de collecter également les biens ne possédant pas d'adresse complète pour une utilisation ultérieure.

Cette collecte a commencé le 1^{er} juillet 2013 et devait initialement se terminer fin janvier 2014. Cependant, le nombre de biens mis en vente étant faible, il a été décidé de prolonger la collecte de données jusque la fin du mois d'avril 2014. Une séance de collecte a eu lieu tous les quinze jours. La première d'entre elles a porté sur tous les biens disponibles au moment de son exécution le 1^{er} juillet 2013. Les autres collectes ont été effectuées en ne demandant en résultat que les biens pour lesquels les annonces ont été mises en ligne durant les quinze jours précédant chaque séance. Ceci permettait d'éviter au maximum les risques de doublons. Néanmoins, chaque fois qu'un bien était collecté, nous avons pris soin de vérifier que celui-ci ne se trouvait pas déjà dans le fichier Excel collecteur en recherchant l'identifiant de chaque annonce tel qu'attribué sur Immoweb. La restriction aux biens de moins de quinze jours permet également d'afficher des biens plus vieux, mais pour lesquels le prix de vente a été modifié. Chaque fois qu'un changement de prix vers le bas était constaté pour un bien, l'ancien prix indiqué dans le fichier Excel était remplacé par ce nouveau prix. Ceci se justifie dans la mesure où ces abaissements surviennent suite à l'absence de candidats acheteurs intéressés par le terrain.

2.2.2. Listing papier des transactions foncières des années nonante

Ce listing papier de 207 pages, transmis à J.-M. Halleux en 1999, reprend un très grand nombre de transactions concernant des terrains à bâtir dans les arrondissements de Liège et de Waremme. Ces transactions sont classées par ancienne commune, puis par rue. Dans quelques très rares cas, les biens possèdent un numéro permettant de les situer dans la rue.

Compte tenu du temps d'encodage conséquent, il a été convenu avec J.-M. Halleux de ne retenir que les biens pour lesquels les transactions ont eu lieu en 1992, 1994, 1996 et 1998. L'ensemble des informations ont été copiées dans un document Excel. Cette opération s'est déroulée durant les trois premières semaines du mois de mars 2014.

Au final, seules les transactions des années 1996 et 1998 ont été utilisées, en raison d'un manque de données de localisation pour le début de la décennie nonante, permettant de générer un nombre suffisant de variables.

2.2.3. Géolocalisation des observations

Les biens possédant une adresse correcte semblaient pouvoir être localisés sans grande difficulté, mais tel n'a pas été le cas. Y. Grignet (2011) a utilisé la base de données du Cadastre

mise à disposition de l'ULg et le PICC¹² de la Région Wallonne pour localiser les biens et vérifier l'exactitude de certaines variables. Il a cependant rencontré des problèmes de correspondances qui lui ont fait perdre un certain nombre de biens. Ceci était notamment le cas lorsque l'orthographe des adresses du PICC et de son échantillon n'étaient pas similaires. Nos propres échantillons ayant une taille qui nous semblait fort restreinte, nous ne pouvions nous permettre de perdre des biens. De plus, nous avons rapidement pris la décision de grossir les échantillons en localisant manuellement les biens pour lesquels nous possédions au moins la rue et la commune. Nous verrons plus loin comment nous avons procédé. Enfin, une maîtrise trop superficielle de SQL¹³, utilisé par Y. Grignet, nous a poussé à chercher d'autres solutions.

Nous souhaitons néanmoins procéder à une vérification de la superficie en confrontant les données d'Immoweb avec celles du Cadastre. Nous nous sommes intéressés à l'application en ligne CadGIS, disponible librement et gratuitement sur Internet. Cette application mise à disposition par les services du Cadastre propose des plans du parcellaire et les bâtiments existants. Il est possible d'atteindre une localisation via une fenêtre de recherche où sont demandées la rue et la commune. Cependant, il est apparu que celle-ci ne pouvait être utilisée. En effet, nous avons mis en évidence qu'une proportion non négligeable d'adresses encodées au sein d'Immoweb correspondait à des parcelles existantes et supportant un bâtiment, celles-ci étant souvent situées à côté d'une parcelle vierge que nous présumons être la parcelle concernée par l'annonce immobilière disponible sur le site Immoweb. Il peut être supposé que le vendeur, ne connaissant pas l'adresse de sa parcelle, la réfère avec celle d'une habitation voisine pour que les acheteurs potentiels puissent la localiser. Il est également apparu que certaines adresses renvoyaient à de grandes parcelles porteuses d'un bâtiment. Compte tenu du fait que la superficie renseignée sur les annonces immobilières se trouvait être très inférieure à la superficie renseignée par l'application CadGIS, nous avons supposé que de tels biens provenaient de la division de la parcelle de base et de la vente de la partie non bâtie en tant que terrain à bâtir. Dans de tels cas, la division de la parcelle n'est généralement pas encore opérée ou les données cadastrales n'ont pas encore été mises à jour. Nous devons dès lors abandonner la procédure de vérification des superficies via le Cadastre.

Compte tenu des difficultés risquant d'être rencontrées pour la géolocalisation de l'ensemble des biens et l'impossibilité de vérifier les superficies, l'utilisation des données cadastrales pour la localisation des terrains à bâtir de notre échantillon est remise en question. Les recherches de solutions alternatives pour la géolocalisation se sont vite arrêtées sur un batch géocodeur utilisé notamment par le département de Géomatique. Ce batch utilise le même algorithme que Google Maps, Goog Earth et leur extension commune Street View. Il est disponible sur Internet, à l'adresse suivante : http://www.findlatitudeandlongitude.com/batch-geocode/#.U9bPt-N_tQf

Ce petit programme demande en entrée les adresses ordonnées de manière bien précise. Cet ordre doit être strictement respecté, sous peine d'empêcher l'application de localiser le bien correspondant : *rue, n°, code postal, commune*.

L'application accepte l'import de données Excel. Ceci se révèle intéressant dans la mesure où les résultats de la collecte ont été stockés au sein d'un fichier de ce type.

¹² Projet informatique de cartographie continue.

¹³ Structured Query Language.

Ce programme fournit en sortie la liste des adresses accompagnées de leur localisation en latitude et en longitude. Les adresses n'ayant pu être localisées sont accompagnées d'un code d'erreur justifiant la cause de la non-géolocalisation. L'explication des codes peut être trouvée à l'annexe V. Il est ainsi possible de vérifier les adresses concernées et de les localiser manuellement par après si le programme ne peut le faire lui-même (code d'erreur 3 ou 5). Si malgré les recherches, la localisation ne pouvait être effectuée manuellement, le bien était supprimé du fichier de données.

La table Excel ainsi récupérée en sortie de la géolocalisation peut être importée sous ArcGIS pour créer un shapefile de points sur base des coordonnées en longitude et en latitude fournies par le batch géocodeur.

2.2.4. Préparation des données provenant d'Immoweb

Comme nous l'avons signalé précédemment, certains biens ne présentent pas d'adresse complète. En parallèle, le développement des modèles uniquement sur base des terrains à bâtir possédant le nom de la rue et le numéro risquait d'amener quelques problèmes de représentativité. Il a donc été décidé de conserver les terrains pour lesquels cette complétude n'était pas rencontrée. Les annonces dont au moins le nom de la rue, le code postal et la commune étaient connus ont été intégrées à l'échantillon initial. Leur localisation s'est effectuée au moyen des photos disponibles pour chacun des biens. Ces photos ont été confrontées manuellement au système de visualisation *Street View* disponible sous *Google Maps* ou *Google Earth*. Le numéro attribué manuellement correspond à celui qu'attribue l'algorithme utilisé par Google Maps de telle sorte que ces terrains sans numéro se retrouvent correctement localisés.

Nous avons constaté, en milieu rural, que l'algorithme ne localisait pas toujours les biens au bon endroit. Dès lors, nous avons choisi de procéder à la vérification de tous les biens possédant une adresse complète sur le site Immoweb et de leur réattribuer le numéro de Street View si celui-ci n'était pas identique au numéro renseigné sous Immoweb. Une triple confrontation a été effectuée entre l'annonce Immoweb (adresse et photos), l'application CadGIS et Google Maps. S'il s'avérait que les numéros d'adresse ne correspondaient pas, c'est celui donné par Google Maps qui a été utilisé pour la géolocalisation, pour la raison déjà expliquée que Google Maps et le batch géocodeur fonctionnent sur base du même algorithme. Le numéro de l'algorithme de Google Maps permet la géolocalisation adéquate des biens.

Toujours en milieu rural, nous nous sommes rendu compte que certaines rues n'étaient pas couvertes par l'algorithme : celui-ci ne fournissait aucun numéro lorsque nous cherchions une adresse à l'aide de Google Maps et de Street View. Lorsque ce problème concernait un bien, nous prenions soin de le noter et celui-ci était relocalisé manuellement sous ArcGIS une fois la géolocalisation des autres biens effectuée par le batch géocodeur.

Les quelques biens ne possédant pas de photos ont été supprimés, sauf dans un cas précis : certaines annonces fournissent, en commentaire, le numéro d'identification cadastral de la parcelle. Celui-ci peut ainsi être encodé sous CadGIS afin de retrouver ou de vérifier la localisation exacte du bien.

Une fois la géolocalisation effectuée, nous avons contrôlé la correspondance des adresses en entrée et en sortie et prêté attention au code d'erreur. Lorsque, pour un bien, la localisation a

échoué, nous avons noté son identifiant permettant ensuite une relocalisation manuelle sous ArcGIS, via modification du shapefile de points.

Ces étapes de vérification et de localisation des biens sans adresse complète sont extrêmement lourdes en termes de temps. La durée de la recherche est fonction de la longueur de la rue : les rues courtes, de quelques dizaines de mètres, permettent une localisation pratiquement immédiate, tandis que les rues de grande longueur engendrent une recherche de localisation pouvant excéder le quart d'heure par bien. Par ailleurs, cette démarche n'aboutit pas toujours à une localisation complète des biens. Dans ce cas, nous n'avons guère d'autre choix que de les retirer de nos échantillons.

2.2.5. Préparation des données provenant du listing papier

La géolocalisation de ces biens étant nécessaire, nous avons choisi de réutiliser le Batch Géocodeur. Celui-ci localise par défaut les biens n'ayant pas de numéro au milieu de la rue correspondante, ce qui est ici le cas pour la très grande majorité des terrains à bâtir renseignés dans la liste papier. Contrairement à Immoweb, il n'existe pas, dans le cas présent, de sources alternatives (notamment des photographies des biens) permettant d'employer d'autres méthodes pour la géolocalisation. Ainsi, il a été choisi de créer des catégories de longueur de rue, dans l'optique d'établir des seuils pour les modèles à créer, et de déterminer le meilleur compromis entre le pouvoir explicatif et la robustesse statistique. Cette étape a également occupé un temps conséquent suite à la vérification individuelle de la longueur de chaque rue via Google Maps et l'outil de mesure associé.

2.3. Préparation des variables

Le choix et le nombre des variables à sélectionner pour calibrer le modèle est très important, notamment parce qu'un trop grand nombre de ces variables n'apporte guère davantage de précision au modèle. En effet, « *au-delà de 5 à 10 critères, l'amélioration de l'estimation n'est plus en proportion de l'effort déployé pour produire et traiter de nouvelles données* » (J.-M. Halleux, 2009). S. Rosen (1974) abondait en ce sens. De plus, multiplier les variables risque de mener à un problème de multi-colinéarité (P. Srikhum, 2012). A l'inverse, P. Srikhum cite plusieurs études indiquant qu'il faut un grand nombre de variables explicatives pour capter toutes les valeurs liées à la localisation (R. K. Pace *et al.*, 1998 ; R. K. Pace *et al.*, 1998b ; J. Valente *et al.*, 2005). Le nombre de variables choisies relève de l'arbitrage de l'analyste. Dans notre cas, nous avons choisi de générer un nombre important de variables de localisation, par prudence. Nous pensons que, parmi nos variables, seul un nombre relativement réduit va effectivement expliquer le prix de manière significative. Cette prudence est d'autant plus légitime que les études concernant le marché des maisons unifamiliales ont débouché sur un faible nombre de variables significatives, qui plus est peu explicatives en regard des variables structurelles (Y. Grignet, 2011 ; C. Delbar, 2012 ; P. Dethier, 2012).

Pour rappel, certaines variables sont utilisables directement car mentionnées sur les annonces immobilières du site Immoweb pour les observations actuelles ou dans le listing papier pour les observations des années nonante. D'autres doivent au contraire être collectées ou produites à partir de données sources. Nous avons ainsi dû générer nous-mêmes certaines données, soit parce que nous ne disposions pas de données directement spatialisables, soit parce qu'il fallait

modifier les informations pour tenir compte de la situation dans les années nonante. C'est notamment le cas du raster de temps de parcours par pixel *Cosstreets* fourni par le SEGEFA, nécessaire pour calculer les temps d'accès. Nous émettons le postulat que le réseau routier liégeois n'a pratiquement pas évolué, et que les vitesses autorisées étaient les mêmes qu'actuellement. Nous avons toutefois supprimé le tunnel de Cointe pour tenir compte de la situation des années nonante.

Ce point présente les différentes variables retenues sur base de la littérature et des trois mémoires concernant l'étude du marché des maisons unifamiliales ainsi que les sources utilisées et les traitements nécessaires à leur génération. Nous détaillerons ces variables selon leur catégorie :

- la variable dépendante ;
- les variables structurelles ;
- les variables d'accessibilité ;
- les variables d'environnement physique ;
- les variables de valorisation sociale.

Le lecteur trouvera aux annexes VI et VII les tableaux reprenant l'ensemble des variables créées.

2.3.1. Variable dépendante

Il s'agit du prix. Celui-ci est obtenu directement à partir des annonces immobilières Immoweb et du listing papier.

2.3.2. Variables structurelles

Superficie du terrain

Le choix de cette variable est essentiel, celle-ci étant utilisée dans l'ensemble des études retenues pour la bibliographie du présent travail. Elle constitue la première source d'explication du prix, celui-ci évoluant de manière positive avec l'aire, mais suivant une décroissance marginale (J.-M. Halleux, 2005). Comme mentionné au point 2.2., la superficie a été extraite des annonces en ligne du site Immoweb via le script Java conçu à cet effet. Pour les données des années nonante, cette information est reprise au sein du listing papier.

Largeur à front de rue et ratio largeur-profondeur

La largeur à front de rue peut être employée comme variable (X. Gao et Y. Asami, 2007 ; K. Matsui et M. Fukushige, 2012). Ceci nous pousse à utiliser cette variable, mais aussi à envisager un ratio largeur-profondeur. En effet, d'après J. Comby (2013), un terrain plus profond sera moins cher qu'un terrain de même superficie mais possédant une largeur à front de rue plus importante. Ces données ont été collectées via le script Java. Il est également possible de mettre en évidence l'influence de la forme du terrain : le fait d'avoir un terrain trapézoïdal ou irrégulier peut influencer le prix (K. Matsui et M. Fukushige, 2012), mais étant donné que nous ne connaissons pas cette morphologie, nous ne pouvons tester cette variable.

Présence préalable d'un bâtiment sur le terrain

La présence d'une construction préalable de type abri de jardin ou garage peut engendrer une influence sur le prix. Cette information est mentionnée en commentaire ou dans les photos associées sur Immoweb. Nous créons une variable binaire dans nos tableaux Excel.

Présence d'un raccordement à l'eau, au gaz et/ou à l'électricité et présence d'un raccordement à l'égout

Ces mentions apparaissant sur les annonces actuelles, nous les avons reprises au sein de nos variables.

Pente

Nous pouvons supposer qu'un terrain en pente aura moins de valeur qu'un terrain plat. Le fait que l'information d'un terrain plat soit parfois reprise dans les annonces Immoweb consultées pousse à l'investigation de cette variable. La génération de l'information est rendue très facile par la fonction *Slope* du logiciel-SIG ArcGIS appliquée à un Modèle Numérique de Terrain de la DGO3. La fonction *Extract Multi Value to Points* permet ensuite d'extraire les données au niveau de chaque terrain à bâtir.

Orientation au sud

L'exposition est parfois mentionnée dans les annonces consultées sur Immoweb, principalement lorsque le terrain est orienté au sud. Ceci pousse à l'inclusion de cette variable dans le modèle. L'exécution de la fonction *Aspect* de ArcGIS sur le MNT permet d'obtenir un raster où l'orientation du terrain est représentée en degrés, dans le sens horloger. La fonction *Extract Multi Value to Points* est ensuite utilisée. Pour obtenir la variable binaire, nous utilisons la fonction SI sous Excel accordant le chiffre 1 à tous les biens dont l'orientation est comprise entre 135° et 225°.

2.3.3 Variables d'accessibilité

Distance aux arrêts de bus

La distance à l'arrêt de bus le plus proche est fréquemment étudiée en milieu urbain, pour les terrains (X. Gao et Y. Asami, 2007 ; Y. Liu *et al.*, 2010) et pour les biens immobiliers (S. Zheng et M. E. Kahn, 2008 ; R. Cervero et C. D. Kang, 2011 ; I.-H. Lin *et al.*, 2013). Dans tous les cas, la proximité à un arrêt de bus entraîne une plus-value sur les biens. Un parasitage par les nuisances sonores peut apparaître durant les toutes premières dizaines de mètres d'éloignement (R. Cervero et C. D. Kang, 2011). Il aurait été utile d'envisager différents pas de distance très resserrés, mais la taille limitée de l'échantillon contraint ce type d'analyse.

La distance euclidienne aux arrêts de bus a été utilisée comme variable pour l'étude du marché des maisons unifamiliales à Liège : une moins-value est mise en évidence lorsque l'arrêt se trouve à moins de 700 mètres. Il est envisagé une contamination de la variable par des nuisances liées à un milieu urbain dense et à la proximité d'axes à fort trafic (P. Dethier, 2012).

Nous exécutons la fonction *Euclidian Distance* sur un shapefile des arrêts de bus wallons généré par le TEC et extrayons les données du raster ainsi généré par la fonction *Extract Multi Value to Points*. De cette distance euclidienne, nous déclinons, via la fonction SI, des seuils de 100, 500 et

700 mètres tels qu'utilisés pour l'étude du marché liégeois des maisons unifamiliales (P. Dethier, 2012). Après concertation avec J.-M. Halleux, nous avons utilisé le même shapefile pour générer la distance aux arrêts de bus dans les années nonante, le réseau liégeois n'ayant pratiquement pas évolué depuis.

Distance et temps d'accès aux gares

La distance à la gare de train, de RER ou de métro la plus proche fait l'objet d'un grand nombre d'études qui mettent en évidence l'importante plus-value engendrée par la proximité de ces infrastructures, aussi bien pour les terrains (X. Gao et Y. Asami, 2007 ; G. Ahlfeldt et W. Maennig, 2008 ; Y. Liu *et al.*, 2010 ; K. Matsui et M. Fukushige, 2012) que pour les maisons (D. R. Bowes et K. R. Ihlanfeldt, 2001 ; C.-H. C. Bae *et al.*, 2003 ; S. Zheng et M. E. Kahn, 2008 ; D. E. Andersson *et al.*, 2010 ; F. De Vor et H. L. F. De Groot, 2011). Des pas de distance sont souvent employés pour mettre en évidence l'atténuation de cette influence avec la distance mais aussi le parasitage dû aux nuisances dans les dizaines de mètres les plus proches de l'infrastructure (D. R. Bowes et K. R. Ihlanfeldt, 2001). Une nouvelle fois, la taille limitée de l'échantillon ne permet pas ce type d'analyse par pas de distance resserrés.

Un précédent mémoire a mis en évidence une moins-value pour les biens situés à moins de 1 km des gares et ce pour la région urbaine de Bruxelles. Il est avancé que l'environnement socio-économique des gares et les nuisances en sont responsables (C. Delbar, 2012).

Nous avons choisi de décliner la proximité des gares sous deux formes : la distance euclidienne à la gare la plus proche et le temps d'accès en voiture le plus court. En effet, les gares sont davantage rejointes en voiture que les arrêts de bus, bien que la proportion de trajets à pied reste importante. Ainsi, le cheminement piéton est représenté par la distance euclidienne. Nous utilisons un shapefile des gares et points d'arrêts SNCB pour l'année 2000, générés par J. Charlier, professeur à l'ULB. Nous avons vérifié son adaptabilité à la situation actuelle en vérifiant si les gares renseignées étaient toujours desservies grâce aux horaires de la SNCB. Nous exécutons par la suite la fonction *Euclidian Distance*. Les temps d'accès en voiture ont été déterminés à partir du fichier *Cosstreets* du SEGEFA. Ils sont générés par la fonction *Cost Distance*. Les valeurs des deux rasters pour chaque terrain sont extraites par la fonction *Extract Multi Value to Points*.

La distance euclidienne a été déclinée en seuils de 100, 500, 700, 1000 et 1500 mètres. Au-delà de 1500 mètres, nous pensons que l'influence d'une gare n'existe plus.

De plus, pour la situation actuelle, les gares ont été déclinées en deux catégories selon le type de train qui les dessert : des gares IC/IR et des gares L¹⁴. Pour déterminer les types de gares, nous avons utilisés les horaires de la SNCB. L'hypothèse sous-jacente est que les gares desservies par les trains IC/IR exercent une influence plus importante sur le prix des biens que les gares desservies par les autres types de train. Nous avons exécuté *Euclidian Distance* et *Cost Distance*, puis *Extract Multi Value to Points*. La distance euclidienne a été déclinée en seuils de 500, 1000 et 1500 mètres autour des gares IC/IR et 100, 500 et 700 mètres autour des autres gares.

¹⁴ Trains IC/IR : trains desservant les gares les plus importantes – Trains L : trains desservant toutes les gares.

Temps d'accès en voiture à différents pôles

La décroissance des prix des terrains en fonction de la distance au centre-ville, fournisseur de services, de biens et d'emplois, est un phénomène très connu et largement étudié (P. F. Colwell et H. J. Munneke, 1997). La littérature consultée envisage dans pratiquement chaque modèle la distance au centre-ville – ou au Central Business District dans les métropoles.

L'étude de l'influence engendrée par des pôles secondaires, en plus du pôle central, a déjà été menée précédemment pour le marché des maisons unifamiliales en Belgique (Y. Grignet, 2011 ; C. Delbar, 2012).

C. Delbar (2012) s'appuie sur une étude réalisée en 2009 (E. Van Hecke *et al.*, 2009) pour déterminer les localisations à choisir et construire les variables de distance aux centralités à partir de celles-ci. Elle préconise l'utilisation des communes classées en A (communes centrales des principales agglomérations) et A1 (communes morphologiquement urbaines et à urbanisation fonctionnelle forte). Nous proposons de suivre une démarche identique. Sont obtenues les communes suivantes :

- pour le bassin d'emploi de Liège : Liège et Verviers. Les localisations précises au sein de ces agglomérations sont respectivement la Place Saint-Lambert et la Place du Marché ;
- pour le bassin d'emploi de Charleroi : Charleroi, avec la Place Charles II prise comme centre.

Pendant, nous nous sommes rendu compte, en procédant de la sorte, que des localisations intéressantes n'étaient pas prises en compte. Nous parlons ici de communes générant un nombre relativement important d'emplois et bien peuplées, mais aussi de pôles jugés importants à l'intérieur même des communes considérées. Nous retenons ainsi les localisations supplémentaires suivantes :

- Herstal, Place Jean Jaurès ;
- Seraing, Place Communale ;
- Ans, Esplanade de l'Hôtel Communal ;
- Sart-Tilman, Avenue du Bois Saint-Jean ;
- l'Aéropôle de Charleroi, terminal de l'aéroport ;
- Châtelet, Place de l'Hôtel de Ville ;
- Fleurus, Place Charles Gailly.

Outre ces pôles, des variables similaires ont été développées pour étudier l'influence du temps d'accès à Bruxelles, métropole à l'influence importante et attractive pour l'emploi (C. Vandermotten *et al.*, 1996 ; CPDT, 2011), et à Namur, capitale régionale wallonne. Les adresses choisies sont respectivement la Grand-Place et la Place de la Station.

Nous avons donc exécuté la fonction *Cost Distance* sur le raster *Cosstreets*, pour chacune des localisations retenues, et exporté les données pour chaque bien via la fonction *Extract Multi Value to Points*.

Distance et temps d'accès aux commerces et aux nodules commerciaux

La distance des terrains à bâtir aux commerces est régulièrement discutée (C. J. Barnett, 1985 ; X. Gao et Y. Asami, 2007 ; Y. Liu *et al.*, 2010), et a dans la plupart des cas un effet négatif. Il en va

de même pour étude sur les maisons (L. Tyrvaïnen et A. Miettinen, 2000). Une étude a montré une influence positive de la distance sur le prix (C. J. Barnett, 1985).

Il nous a semblé pertinent de décliner les commerces en deux volets, à l'instar de ce qui a été fait pour l'étude des maisons unifamiliales en région de Liège (P. Dethier, 2012) : d'une part, les commerces alimentaires rejoints pour des achats fréquents, et d'autre part, les nodules commerciaux tels que définis par le SEGEFA¹⁵. Nous utiliserons la distance euclidienne aux commerces alimentaires et le temps d'accès aux nodules commerciaux. Les fonctions *Euclidian Distance* et *Cost Distance* ont été utilisées. Les données pour chaque terrain sont exportées depuis les rasters par la fonction *Extract Multi Value to Points*. A l'instar de P. Dethier, nous avons décliné la distance euclidienne aux commerces alimentaires en seuils de 100, 500 et 700 mètres. Les shapefiles des commerces alimentaires et des nodules proviennent du SEGEFA.

Temps d'accès et distance aux sorties d'autoroute

La distance aux sorties d'autoroute est fréquemment étudiée et est responsable d'une influence négative sur le prix des terrains à bâtir (J. S. Shonkwiler et J. E. Reynolds, 1986 ; K. Andoh et M. Ohta, 1997) et des maisons d'habitation (D. R. Bowes et K. R. Ihlanfeldt, 2001 ; D. E. Andersson *et al.*, 2010 ; F. De Vor et H. L. F. De Groot, 2011). Une étude montre une influence positive de la distance, probablement en raison des nuisances du trafic routier qui masquent l'utilité procurée par la proximité de l'accès autoroutier (D. E. Andersson *et al.*, 2010). Nous avons choisi d'utiliser le temps d'accès en voiture le plus court aux accès à l'autoroute. La distance euclidienne a également été calculée, extraite et déclinée en seuils de 100, 500, 700, 1000 et 1500 mètres. Le shapefile des sorties d'autoroute est fourni par le LEPUR¹⁶. Nous l'avons également utilisé pour obtenir la situation en 1998, après vérification effectuée grâce à la carte routière de 1998¹⁷.

Distance au réseau RAVeL

Bien que ni la littérature, ni les mémoires ne l'aient prise en compte, cette variable a été choisie sous l'hypothèse que la présence d'un RAVeL à proximité des biens se fait de plus en plus attractive pour les loisirs notamment. Nous avons une nouvelle fois utilisé les fonctions *Euclidian Distance* et *Extract Multi Value to Points*. Cette distance a été déclinée en seuils de 100 et 500 mètres. Le shapefile du réseau provient de la DGO1. Nous l'avons modifié sur base d'une carte papier¹⁸ pour également obtenir le réseau en 1999.

Accessibilité en mode de transports alternatifs

Les variables de part modale du train, du bus et des modes lents ont été utilisées dans les précédents mémoires concernant les maisons d'habitations de la région de Liège (Y. Grignet, 2011 ; P. Dethier, 2012). C'est pourquoi il semblait intéressant d'effectuer une analyse similaire pour les terrains à bâtir. Pour davantage d'informations concernant la conception de ces indicateurs, nous renvoyons à l'article accompagnant leur création (Y. Cornet *et al.*, 2005). Les mémoires mettent en évidence l'influence positive d'une part modale des modes lents croissante

¹⁵ SEGEFA : Service d'Etude en Géographie Economique Fondamentale et Appliquée.

¹⁶ LEPUR : Centre de Recherche sur la Ville, le Territoire et le Milieu rural.

¹⁷ MICHELIN (1998), *Carte routière et touristique, Belgique-Luxembourg*, 1/350000^e, Zellik.

¹⁸ DGATLP (1999), *Réseau autonome de Voies Lentes*, 1/300000^e, 3^{ème} édition, Namur.

sur le prix des biens, mais aucune influence de la part modale du train et du bus (Y. Grignet, 2011 ; P. Dethier, 2012).

Les variables sont directement utilisables : les valeurs au niveau de chaque bien peuvent être directement exportées des rasters reprenant, pour chacun d'entre eux, la valeur d'un indice par pixel. La fonction *Extract Multi Value to Points* est de nouveau utilisée.

Temps d'accès aux écoles et universités

La proximité d'écoles est régulièrement étudiée et déclinée sous différentes formes selon le type d'établissement : école maternelle, école primaire, école secondaire, établissement d'enseignement supérieur ou universitaire. Il est ainsi montré une influence négative de la distance aux écoles, tous types confondus, aussi bien pour les terrains (C. J. Barnett, 1985 ; R. Paris, 2003) que pour les maisons (L. Tyrvaïnen et A. Miettinen, 2000). Il en va de même, pour les terrains à bâtir, lorsque les différents types d'écoles sont séparés en écoles secondaires, écoles primaires, écoles maternelles (A. Donzel *et al.*, 2008). Une étude portant toujours sur les terrains ne met pas en évidence d'influence significative (Y. Liu *et al.*, 2010). Un dernier travail montre que la proximité d'une université engendre une plus-value sur le prix des maisons (S. Zheng et M. E. Kahn, 2008).

Nous avons choisi d'étudier le temps d'accès en voiture le plus court à chaque type d'établissement : écoles maternelles et primaires, écoles secondaires, enseignement supérieur. Ceci a été obtenu grâce à la fonction *Cost Distance* appliquée sur le raster *Cosstreets*. Les valeurs sont exportées via la fonction *Extract Multi Value to Points*. Le temps d'accès a été décliné en seuils de 30 et 60 secondes pour les écoles maternelles et primaires, de 60 et 120 secondes pour les écoles secondaires et de 60 et 300 secondes pour l'enseignement supérieur. Les shapefiles reprenant la localisation des établissements proviennent du LEPUR.

2.3.4. Variables d'environnement physique

Localisation dans un lotissement

Le choix de cette variable est motivé par l'hypothèse que, pour un terrain de même superficie et identiquement localisé, le prix de vente peut être différent selon que le bien appartienne à un lotissement ou non. Cette information est présentée sur Immoweb en commentaire ou à travers les photos associées. Une variable binaire est ainsi créée dans les tableaux Excel contenant les biens.

Distance aux SAED¹⁹

Ces sites sont caractérisés par un état de délabrement avancé. Leur proximité peut influencer les prix des maisons et des terrains à bâtir. C'est pourquoi il a été jugé intéressant d'intégrer cette variable dans les modèles. Elle a par ailleurs été utilisée pour l'approche hédonique du marché des maisons unifamiliales en région de Liège. Aucune influence n'a cependant été mise en évidence, quel que soit le seuil de proximité utilisé, à savoir 100, 200 et 500 mètres (P. Dethier, 2012). Le shapefile des SAED provient de la DGO4. Dans le cadre de ce travail, nous avons utilisé la fonction *Euclidian Distance*, puis *Extract Multi Values to Points*. Nous avons ensuite transformé

¹⁹ Site d'activités économiques désaffecté.

cette distance selon les mêmes seuils utilisés par P. Dethier (2012) sous Excel en utilisant la fonction SI. Trois variables binaires sont ainsi générées.

Distance aux industries

La proximité d'une industrie engendre une moins-value sur les prix, aussi bien pour les terrains à bâtir (C. J. Barnett, 1985 ; C. G. Leggett et N. E. Bockstael, 1999 ; Y. Liu *et al.*, 2010) que pour les maisons d'habitation (M. Tavieres *et al.*, 2009 ; F. De Vor et H. L. F. De Groot, 2011). F. De Vor et H. L. F. De Groot (2011) mettent en évidence, dans leur étude, la décroissance progressive de cette influence avec la distance, avant sa disparition à plus de 2 km. En région urbaine de Liège, une moins-value sur les maisons unifamiliales situées à moins de 200 mètres d'un site industriel est mise en évidence. Les autres seuils (100 et 500 mètres) ne sont pas significatifs (P. Dethier, 2012). Nous choisissons, outre la distance euclidienne, de tester ces mêmes seuils, auxquels nous ajoutons un seuil à 1000 mètres. Une fois les sites industriels extraits de la carte d'occupation des sols de la Région Wallonne (source : SPW-DGO4), nous avons créé le raster de distance via la fonction *Euclidian Distance*. Les valeurs obtenues sont exportées grâce à la fonction *Extract Multi Values to Points*. Ensuite, la fonction SI de Excel permet de générer les seuils souhaités. Pour la situation des années nonante, nous avons utilisé le Corine Land Cover de 2000 et 2006. Les rasters sont téléchargeables sur le site de l'Agence Européenne pour l'Environnement (European Environment Agency en anglais).

Distance aux carrières

Cette variable a été utilisée pour l'étude du marché liégeois des maisons unifamiliales. Cependant, aucune influence de la proximité de ces carrières n'est mise en évidence, sans doute par manque de biens concernés (Dethier, 2012). Nous réutilisons les mêmes seuils : 100, 200 et 500 mètres. Nous avons extrait les carrières du Corine Land Cover pour 2000 et 2006. Nous avons ensuite exécuté les fonctions *Euclidian Distance* et *Extract Multi Value to Points* avant de dériver les rasters obtenus en variables binaires selon la fonction SI sous Excel.

Bruit

Une variable d'intensité du bruit généré par le trafic automobile, ferroviaire ou aérien est fréquemment utilisée, essentiellement pour des études sur les maisons d'habitation (I. R. Lake *et al.*, 1998 ; J. E. C. Dekkers et J. Willemijn Van Der Straaten, 2009 ; M. Tavieres *et al.*, 2009). Cependant, les données de bruit sont souvent incomplètes, voire inexistantes. Ceci est notamment le cas de la Wallonie : il existe bien quelques données de bruit, mais l'extension de l'information est incomplète et bien trop réduite pour pouvoir l'utiliser dans une approche hédonique. Nous avons cependant pris contact auprès de la DGRANE²⁰ pour obtenir un exemplaire des données en format raster, sans résultat.

Pour remédier au manque de données de bruit, certains auteurs proposent de calculer des distances euclidiennes autour des voies ferrées, des grands axes de circulation routière et des aéroports (I. R. Lake *et al.*, 1998 ; G. Ahlfeldt et W. Maennig, 2008 ; S.-H. Cho *et al.*, 2008 ; M. Tavieres *et al.*, 2009 ; Y. Liu *et al.*, 2010 ; I.-H. Lin *et al.*, 2013). Dans tous les cas, une augmentation du niveau de bruit (ou un rapprochement de l'élément générateur de nuisances sonores) entraîne une moins-value de plus en plus importante sur les biens. Une étude

²⁰ Direction Générale Opérationnelle Agriculture, Ressources naturelles et Environnement.

s'intéressant à l'effet de différentes sources de bruit montre que ce sont les nuisances dues au trafic aérien qui entraînent les plus grandes moins-values (J. E. C. Dekkers et J. Willemijn Van Der Straaten, 2009). Les données des plans d'exposition au bruit du trafic aérien disponibles auprès de la SOWAER ont par ailleurs été utilisées pour l'étude du marché des maisons unifamiliales dans la région urbaine de Bruxelles où les biens concernés par des niveaux de bruit de plus de 60 dB étaient victimes d'une moins-value (C. Delbar, 2012). Pour l'étude de l'influence du trafic aérien, nous choisissons de réutiliser ce seuil, en plus du seuil des 55 dB. Nous n'avons pas envisagé les autres seuils de bruit plus importants étant donné le très faible nombre voire l'absence de biens concernés. Nous avons créé une variable binaire d'appartenance à chacune de ces zones dans Excel sur base des cartes et shapefiles de la SOWAER.

En région liégeoise, un précédent mémoire n'a pas démontré l'existence de moins-value significative portée aux biens se situant à moins de 100 mètres du réseau routier principal²¹ (P. Dethier, 2012). Une autre approche consiste à prendre pour variable le trafic des routes génératrices de bruit (K. S. Kim *et al.*, *et al.*, 2007 ; Y. Liu *et al.*, 2010). Cependant, ces chiffres ne sont pas disponibles pour l'ensemble du réseau wallon, et il n'est dès lors pas possible de développer ce type d'étude. Une variante qui sera utilisée consiste à tester l'influence de la distance euclidienne au réseau routier dont le trafic est supérieur à différents seuils de nombre de véhicules par jour. Nous choisissons deux seuils : 10000 et 20000 véhicules par jour. Sur base des comptages de l'année 2009 (SPF Mobilité-Transport), nous avons modifié un shapefile des routes (source : Open Street Map) en ne gardant que les axes concernés. Nous avons fait de même avec les comptages de l'année 2000 (SPF Mobilité-Transport) pour représenter la situation de la fin des années nonante. Nous avons ensuite appliqué la fonction *Euclidian Distance* sur les deux shapefiles ainsi créés. Les valeurs ont été exportées via la fonction *Extract Multi Value to Points*. Les données ont ensuite été dérivées en variables binaires selon trois seuils de 100, 200 et 500 mètres. Cette opération a été effectuée dans Excel grâce à la fonction SI.

Nous procédons de la même manière avec les lignes de chemin de fer. Nous avons fusionné différents shapefiles provenant de J. Charlier, chercheur à l'ULB, et de la CPDT pour reconstituer le réseau des chemins de fer en 2000 et en 2014. Nous avons ensuite exécuté les fonctions *Euclidian Distance* et *Extract Multi Value to Points*, puis dérivé cette variable en seuils de 100, 500 et 700 mètres.

Distance à l'eau

La proximité immédiate d'un cours d'eau peut avoir une influence positive significative sur le prix des terrains à bâtir (C. G. Leggett et N. E. Bockstael, 1999 ; G. Ahlfeldt et W. Maennig, 2008) et des maisons d'habitations (J. Luttik, 2000 ; S.-T. Anderson et S.-E. West, 2006 ; H. A. Sander et R. G. Haight, 2012). De même, il peut être intéressant de considérer la proximité d'autres types de corps hydrographiques. Ainsi, deux études montrent une influence négative de la distance à un lac, et ce de manière très significative (S.-T. Anderson et S.-E. West, 2006 ; H. A. Sander et R. G. Haight, 2012). Une étude menée sur les maisons dans l'estuaire de la Seine ne montre cependant pas d'influence significative de la distance à un cours d'eau (M. Tavieres *et al.*, 2009). Il semble dès lors que l'influence de la présence d'un corps hydrographique puisse varier d'un cas à l'autre.

²¹ Ensemble des autoroutes et routes nationales où la vitesse maximale autorisée est supérieure à 50 km/h.

Pour le bassin d'emploi de Charleroi, nous choisissons de considérer les lacs de l'Eau d'Heure en raison de leur étendue relativement importante et de l'essor touristique qui leur est associé. N'ayant pas de shapefile de ces lacs, nous avons choisi d'en créer un en prenant comme modèle les cartes topographiques de l'IGN à l'échelle 1/10000^{ème}, fournies par le LEPUR en format raster. Nous avons ensuite utilisé les fonctions *Euclidian Distance* puis *Extract Multi Value to Points* et dérivé la variable obtenue selon deux seuils : 2000 et 5000 mètres.

Pour le réseau hydrographique, nous avons utilisé deux shapefiles : le premier provient de l'Atlas Numérique de Belgique et représente les cours d'eau importants²². Le second est plus complet et reprend également les cours d'eau de moindre importance, y compris les ruisseaux. Il nous a été fourni par le LEPUR. Afin d'approcher la visibilité procurée par la proximité immédiate d'un cours d'eau, nous avons choisi de tester, pour les deux cas, un seuil de 100 mètres autour des deux types de réseaux hydrographiques étudiés. Nous calculons également des seuils à 200 et à 500 mètres. Ceci a été rendu possible via la fonction SI sous Excel, au départ du raster généré par la fonction *Euclidian Distance*. Les données au niveau de chaque bien sont exportées via la fonction *Extract Multi Value to Points*.

Distance aux lignes à haute tension

Cette variable n'a pas été étudiée dans les études consultées. Nous supposons cependant que la proximité de ces infrastructures peut entraîner une moins-value sur les terrains à bâtir. Outre la distance euclidienne générée par la fonction *Euclidian Distance* et extraite au niveau de chaque bien par la fonction *Extract Multi Value to Points*, des seuils de 100, 200 et 500 mètres générés grâce à la fonction SI sous Excel seront également testés. Nous avons utilisé un shapefile créé par la CPDT.

Part de bâti/d'artificialisation et diversité de l'environnement

L'influence de la part de bâti résidentiel et de la diversité de l'environnement a été étudiée par J. Geoghegan *et al.* (1997) pour des buffers de 100 mètres et de 1 km de rayon. Le seuil des 100 mètres est une approximation de ce qui est directement observable à partir du point d'observation. Le seuil des 1000 mètres représente l'environnement accessible à pied en 15 – 20 minutes. Une fonction hédonique est ainsi construite pour une zone située à moins de 30 miles de Washington sur base d'une régression par moindres carrés ordinaires. Cette étude porte sur les maisons.

La part de bâti résidentiel est simplement calculée comme suit :

$$B = \frac{\sum_b P_b}{\text{aire du buffer}}$$

P_b correspond à l'aire de chaque superficie de bâti résidentiel. A moins de 100 mètres, aucune influence significative n'est mise en exergue. Au sein du buffer de 1000 mètres, une part de bâti plus importante génère une moins-value sur les biens (J. Geoghegan *et al.*, 1997).

²² Le shapefile correspondant reprend les cours d'eau suivants : Meuse, Ourthe, Amblève, Vesdre, Berwinne, Geer, Hoegne, Mehaigne et Hoyoux pour le bassin d'emploi de Liège ; Sambre, Haine et Eau Noire pour le bassin d'emploi de Charleroi.

La diversité de l'environnement est mesurée au travers de l'indice de Shannon (J. Choumert et M. Travers, 2010). Il se calcule comme suit :

$$H = - \sum_{i=1}^k P_i \ln P_i$$

- P_i correspond à la proportion de la surface d'un type d'affectation par rapport à la surface totale, toutes affectations confondues ;
- K représente quand à lui le nombre d'affectations différentes.

Lorsqu'une seule affectation est présente, l'indice est égal à 0. Une élévation de l'indice correspond à un environnement de plus en plus diversifié. Une diversité croissante engendre une élévation du prix, toutes choses égales par ailleurs (J. Geoghegan *et al.*, 1997). J. Choumert et M. Travers (2010) sont arrivés à la même conclusion.

Y. Grignet (2011) a repris ces formules et les a appliquées au cadre de son mémoire portant sur l'application de l'approche hédonique aux maisons unifamiliales de la région urbaine de Liège. Il a adapté la première formule et étudié l'influence de la part des surfaces artificialisées sur le prix des maisons unifamiliales de la région urbaine de Liège. Ce choix provient des catégories d'affectations du raster Corine Land Cover utilisé. Les surfaces artificialisées correspondent aux codes 1 à 11 dans la légende du raster. Les seuils retenus sont respectivement 200 mètres et 2 km, le premier étant choisi en fonction de la résolution du Corine Land Cover (cellules de 100 mètres de côté). Il montre qu'une part artificialisée grandissante tend à générer une influence négative sur le prix des maisons, mais compte tenu de la p-value associée (légèrement supérieure à 0,10), il n'est pas possible de conclure à sa significativité (Y. Grignet, 2011). Un autre mémoire réalisé sur le même échantillon ne montre aucune influence (P. Dethier, 2012). En région urbaine de Bruxelles, il n'est pas non plus mis en évidence d'influence de la part artificialisée sur le prix des maisons unifamiliales pour les deux seuils retenus, à savoir 200 et 1000 mètres (C. Delbar, 2012).

Aucune influence significative de la diversité de l'environnement sur le prix des maisons de la région urbaine de Bruxelles n'est mise en évidence (C. Delbar, 2012). En région liégeoise, une relation positive est mise en évidence entre une diversité grandissante et le prix des biens, mais uniquement au seuil de 100 mètres. L'explication avancée est qu'une faible diversité correspond, en région liégeoise, à de l'urbanisation. Il n'y a pas d'influence au seuil de 1000 mètres (P. Dethier, 2012). Une autre étude menée à Liège ne montre aucune influence, quel que soit le seuil choisi (Y. Grignet, 2011).

Compte des différents modes opératoires énoncés ci-dessus, nous choisissons d'étudier l'influence de la part de bâti selon deux approches :

- la part de bâti, soit l'ensemble des maisons et bâtiments ayant des fonctions résidentielles et de service. Il s'agit de la dénomination « Urban Fabric » du Corine Land Cover (valeur 1 et 2). Cet indice s'approche le plus de la part de bâti résidentiel étudié par J. Geoghegan (1997) ;
- la part d'artificialisation, c'est-à-dire l'ensemble des superficies construites par l'homme, soit l'ensemble des « Artificial Surfaces » (valeurs 1 à 11) du Corine Land Cover, à l'instar des développements de Y. Grignet (2011).

Nous adaptons dès lors la formule permettant de calculer l'indice B à ces deux cas de figure. Nous retenons les seuils choisis par C. Delbar (2012), à savoir 200 mètres et 1000 mètres, ceux-ci tenant compte à la fois des seuils retenus par J. Geoghegan *et al.* (1997) et de la résolution du Corine Land Cover. Nous utilisons ces mêmes seuils pour la diversité de l'environnement, calculée sur base de l'indice de Shannon.

Pour construire ces variables, nous avons utilisé les rasters du Corine Land Cover pour 2000 et 2006. La fonction *Tabulate Area* disponible sous ArcGIS permet de calculer la superficie de différentes affectations du sol contenues à l'intérieur d'une zone bien définie, dans ce cas-ci des buffers de respectivement 200 et 1000 mètres. La répétition manuelle d'une même opération sur plusieurs centaines de points représentatifs de chaque bien serait une opération extrêmement coûteuse en termes de temps. Pour éviter cela, il a été choisi de procéder à une automatisation des tâches comme effectué par Y. Grignet (2011), C. Delbar (2012) et P. Dethier (2012). Un script Python créé à l'époque par Y. Grignet (2011) a été repris et modifié pour les besoins du présent travail (Annexe VIII). Il exécute la fonction *Tabulate Area* autant de fois que demandé dans le script et crée une table reprenant les aires de chaque affectation pour chacun des buffers. Ensuite, l'outil *Merge* permet de fusionner les tables. Les calculs peuvent ensuite être effectués sous Excel, selon les formules exposées plus haut, avant que la table finale ne soit jointe aux fichiers des points-terrains via la fonction *Join*, par l'intermédiaire d'un identifiant.

Densité de la population par secteur statistique

Cette variable est fréquemment utilisée dans les modèles. La croissance de la densité est généralement responsable d'une moins-value, aussi bien pour les terrains (X. Gao et Y. Asami, 2007 ; G. Ahlfeldt et W. Maennig, 2008 ; K. Matsui et M. Fukushige, 2012) que pour les maisons (C.-H. C. Bae *et al.*, 2003 ; Cho, S.-H. *et al.*, 2008 ; F. De Vor et H. L. F. De Groot, 2011).

Elle a également été utilisée pour caractériser le marché des maisons unifamiliales dans la région de Liège, où une relation négative sur les prix est mise en évidence (P. Dethier, 2012).

Les données utilisées proviennent de l'INS. Les chiffres de l'année 2001 sont utilisés pour représenter la situation de la fin des années nonante, tandis que les chiffres de l'année 2009 sont utilisés pour représenter la situation actuelle. Nous avons lié ces données à un shapefile des secteurs statistiques via la fonction *Join*. Ensuite, l'outil *Intersect* permet d'associer les données aux observations.

Aléas dus aux inondations

A notre connaissance, cette variable n'a jamais été intégrée dans une analyse appliquant l'approche hédonique aux marchés immobiliers. Nous estimons cependant que le fait qu'un bien se trouve en zone inondable peut entraîner une moins-value sur son prix. Nous avons considéré deux seuils d'aléas : faible et fort. Nous avons repris le shapefile des aléas d'inondation généré par la DGO3. La fonction *Intersect* exécutée entre la couche des points-terrains et la couche des aléas d'inondation permet de déterminer quels points sont concernés. La valeur de l'aléa est ensuite transformée en variable binaire.

Distance aux forêts et aux espaces verts

Certains travaux étudient l'influence de la proximité d'une forêt sur le prix des maisons d'habitations (L. Tyrvainen et A. Miettinen, 2000 ; S.-H. Cho *et al.* 2008 ; J. Choumert et M. Travers, 2010 ; F. Waltert et F. Schöpfer, 2010) ou des terrains (V. R. Sharma, 2013). Les effets peuvent varier d'un cas à l'autre. Ainsi, si un article met en évidence l'influence positive de la proximité d'une zone forestière (L. Tyrvainen et A. Miettinen, 2000), certaines études montrent au contraire une influence positive de l'éloignement vis-à-vis des forêts (S.-H. Cho *et al.*, 2008 ; V.R. Sharma, 2013). Dans ce cas, il est supposé que le risque d'incendie (S.-H. Cho *et al.*, 2008) ou que la perte des feuilles en basse saison (V. R. Sharma, 2013) en est responsable. Une étude réalisée à Angers ne met pas d'influence en évidence (J. Choumert et M. Travers, 2010).

L'application de l'approche hédonique au marché des maisons unifamiliales de la région urbaine de Liège n'a montré aucune influence de la proximité des forêts et des espaces verts urbains (P. Dethier, 2012).

Nous avons choisi d'étudier la distance euclidienne aux forêts et aux espaces verts, à l'instar de P. Dethier (2012). Nous avons premièrement extrait ces affectations de la Carte d'Occupation des Sols pour les modèles actuels et du Corine Land Cover 2000 et 2006 pour le modèle de la fin des années nonante et le modèle de comparaison actuel lié. Nous avons utilisé les fonctions *Euclidian Distance* et *Extract Multi Value to Points*. Nous avons ensuite créé deux variables binaires en utilisant deux seuils : 100 et 500 mètres.

Distance à la centrale nucléaire de Tihange

Eu égard au risque d'incident dommageable pour l'environnement et la société humaine, la proximité d'une centrale nucléaire est supposée avoir une influence très négative sur le prix des terrains. Y. Fumihiko *et al.* (2011) ont montré l'impact négatif qu'avait l'annonce de la construction d'une centrale nucléaire sur le prix des terrains à bâtir. Nous nous intéressons dès lors au cas de la centrale nucléaire de Tihange. Outre cette distance, des seuils de 2000, 5000 et 10000 mètres seront utilisés. Nous avons créé un shapefile de point à l'emplacement du site de la centrale, puis exécuté les fonctions *Euclidian Distance* et *Extract Multi Value to Points* avant de produire les variables binaires sous Excel.

Proportion de maisons à quatre façades par secteurs statistiques

Cette variable a été testée dans une étude française. Il est mis en évidence une évolution positive entre le pourcentage de maisons individuelles et le prix des terrains à bâtir (A. Donzel *et al.*, 2008). Nous choisissons de tester une variable semblable, mais pour l'ensemble du bassin d'emploi de Liège et de Charleroi. A partir des données de l'INS pour 2001, nous avons calculé, pour chaque secteur statistique, la part de maisons à quatre façades. Nous avons ensuite joint la table Excel ainsi créée au shapefile des secteurs statistiques via la fonction *Join*. La fonction *Intersect* est ensuite exécutée.

Cul-de-sac

Une influence négative sur le prix des terrains à bâtir situés dans un cul-de-sac est mise en évidence dans une étude japonaise (X. Gao et Y. Asami, 2007). A l'inverse, une plus-value est mise en évidence sur le prix des terrains suite au faible trafic de la rue (C. J. Barnett, 1985). Il est apparu

intéressant de vérifier si de telles constatations peuvent être faites en Belgique pour les terrains à bâtir. Cette variable a été générée en même temps que la vérification et la correction des localisations de chaque bien, grâce au site internet Google Maps. Le chiffre « 1 » a été attribué aux biens situés dans une rue en cul-de-sac.

2.3.5. Variables relatives à la valorisation sociale

Revenu médian par secteurs statistiques

Le revenu est fréquemment utilisé pour caractériser l'environnement socio-économique des biens. Plus ces valeurs croissent, plus le prix des maisons et des terrains s'élève. Il est donc mis en évidence une influence positive souvent très significative, pour les maisons (K. R. Ihlanfeldt, 2006 ; S.-H. Cho *et al.*, 2008 ; D. E. Andersson *et al.*, 2010) et les terrains à bâtir (A. Donzel *et al.*, 2008).

Compte tenu de l'agencement des données obtenues, nous utiliserons le revenu médian par ménage, décliné par secteurs statistiques. S.-H. Cho *et al.* (2008) utilisent notamment cette variable. Cette variable a par ailleurs été étudiée précédemment pour les maisons unifamiliales à Bruxelles, où une relation positive avec les prix était mise en évidence (C. Delbar, 2012). A Liège, deux études similaires mettent en évidence une influence semblable (Y. Grignet, 2011 ; P. Dethier, 2012).

Les données proviennent de l'INS et sont disponibles de 1994 à 2011. Nous avons lié ces données à un shapefile des secteurs statistiques via la fonction *Join*. Les données étaient ensuite associées aux observations via la fonction *Intersect*.

Taux de chômage par secteurs statistiques et par commune

Le taux de chômage est également utilisé pour caractériser l'environnement socio-économique des biens. Il tend à engendrer une moins-value de plus en plus importante lorsqu'il croît (S.-H. Cho *et al.*, 2008). Il a de ce fait été choisi d'intégrer le taux de chômage dans notre travail.

Nous disposons des données de chômage par commune pour 1997 et 2010 (source : IWEPS) et par secteurs statistiques pour 2001 (source : INS). Nous joignons les données aux shapefiles des secteurs statistiques et des communes via la fonction *Join*, puis nous exécutons *Intersect*.

Proportion d'individus diplômés du supérieur par commune

Une relation négative significative entre la proportion d'individus non diplômés et le prix des biens en zone non rurale française est mise en évidence (A. Donzel *et al.*, 2008). En zone rurale, toujours en France, le phénomène inverse est constaté, mais la relation n'est pas significative au seuil de 5 %. P. Dethier (2012) a utilisé le taux d'individus diplômés du supérieur pour illustrer le niveau d'instruction. Une variable similaire a été retenue pour ce mémoire, pour l'ensemble des bassins d'emplois de Liège et de Charleroi. Nous disposons de ces données, pour 1991 et 2001, à l'échelle des communes. Elles proviennent de l'INS. Nous associons ces données au shapefile des communes grâce à la fonction *Join*, puis nous exécutons l'outil *Intersect*.

Fiscalité des communes

Cette variable n'a pas été testée parmi les études consultées. Il semble cependant intéressant de vérifier s'il n'existe pas une relation significative entre le taux d'IPP de chaque commune et le prix

des terrains à bâtir s’y trouvant. Nous avons utilisé les données fournies par le LEPUR. Nous joignons ces données au shapefile des communes via la fonction *Join*, puis nous exécutons l’outil *Intersect*.

2.4. Les étapes de modélisation

La modélisation suivra dans les grandes lignes la méthodologie proposée par F. Des Rosiers (2001). Nous avons également suivi les développements de P. Dethier (2012) tenant compte notamment de la correction de l’hétéroscédasticité. Ce chapitre réalise la synthèse théorique des étapes de modélisation utilisées pour chacun de nos modèles.

2.4.1. Sources logicielles utilisées pour la modélisation

Outre ArcGIS (version 10.2), nécessaire pour la construction des différentes variables de localisation, trois logiciels de traitements statistiques ont été utilisés :

- le premier, Statistica (version 10), est un logiciel développé par Statsoft. Ce logiciel propose de nombreux modules pour l’analyse statistique, le traitement et la représentation des données. Dans le cas présent, il permet de générer des statistiques descriptives pour les variables employées. Ce logiciel est payant. Il est cependant offert gratuitement aux étudiants de l’ULg via une licence accordée à l’université. Il peut être téléchargé sur le site Gitan, à l’adresse suivante :
<http://www.gitan.ulg.ac.be/Intranet/intranet-statistica.html> ;
- le deuxième logiciel utilisé est Gretl pour Gnu Regression, Econometrics and Time-series Library (version 1.9.90), logiciel libre développé par Allin Cottrell (P. Dethier, 2012). Bien que l’interface et l’ergonomie soient moins poussées que celles de Statistica, Gretl possède certaines fonctionnalités qui ne sont pas disponibles sous le premier logiciel telles que les tests d’hétéroscédasticité et de colinéarité ainsi que le développement de modèles débarrassés de l’hétéroscédasticité. Son grand avantage est qu’il peut, comme son nom l’indique, être téléchargé gratuitement et librement à l’adresse suivante :
<http://gretl.sourceforge.net/win32/> ;
- le troisième est GeoDa, développé par Luc Anselin, spécialiste de l’analyse spatiale, en économétrie notamment. Ce logiciel calcule notamment les indices général ou local de Moran et cartographie les clusters LISA. Il sera ainsi possible de déterminer de possibles variables d’influence locale ayant été oubliées dans le modèle. Il est libre, gratuit et peut être téléchargé à l’adresse suivante : <https://geodacenter.asu.edu/software/downloads>. Nous avons choisi d’utiliser ce programme suite à une incompatibilité supposée²³ entre l’outil similaire disponible sous ArcGIS et Windows Vista, le système d’exploitation installé sur l’ordinateur supportant les traitements.

2.4.2. Paramètres utilisés lors des régressions

Les régressions ont été développées dans le logiciel Gretl. Outre la sortie d’une synthèse de la régression présentant les différentes variables et le coefficient de régression linéaire associé, le programme fournit plusieurs paramètres qui peuvent être utilisés pour la sélection du modèle jugé adéquat.

²³ Cette remarque est issue d’une discussion avec M. Binard et M. Trotta, suivant la découverte du non-fonctionnement de l’outil de calcul de l’indice de Moran sous ArcGIS 10.2.

L'erreur-type de l'estimation permet de juger la performance prédictive du modèle. Il s'agit d'une représentation de l'erreur engendrée par le modèle dans l'estimation de la variable dépendante (C. Delbar, 2012). Elle s'exprime selon l'unité de la variable dépendante, dans le cas présent, l'euro. Elle peut être relativisée par rapport au prix moyen de l'échantillon. Une erreur-type de régression égale ou inférieure à 15 % du prix moyen caractérise un modèle performant. Une valeur inférieure à 10 % est cependant souhaitée (F. Des Rosiers, 2001). Le logiciel Gretl ne fournissant pas cette donnée, nous la recalculerons manuellement.

Le coefficient de détermination R^2 exprime la part de variance de la variable dépendante qui est expliquée par les différentes variables indépendantes entrées dans la régression (F. Des Rosiers, 2001).

Davantage que le R^2 , c'est le comportement du coefficient de détermination ajusté R^2 ajusté qui doit être observé. En effet, une augmentation du R^2 n'indique pas forcément une augmentation du pouvoir explicatif du modèle. Ceci est notamment vrai lorsque de nombreuses variables sont intégrées dans la régression : l'ajout d'une variable ne permet pas de compenser la perte d'un degré de liberté, le nombre de ces degrés étant fonction de la différence entre la taille de l'échantillon et le nombre de variables entrées dans la régression (F. Des Rosiers, 2001).

Le test de Fischer F est un test d'hypothèse permettant de vérifier que l'ensemble des variables indépendantes entrées dans le modèle explique de manière significative les valeurs de la variable dépendante. L'hypothèse H_0 correspond à une explication non-significative. Aussi, il est indispensable que la valeur de F générée par le logiciel soit supérieure à celle qui peut être déterminée à partir des tables statistiques, en veillant au degré de liberté du numérateur (égal à k , le nombre de variables explicatives entrées dans la régression) et à celui du dénominateur (égal à $N - k - 1$, N correspondant à la taille de l'échantillon). Une telle situation permet de rejeter l'hypothèse nulle et de conclure à la robustesse statistique du modèle ainsi développé.

La p-value exprime la probabilité que la valeur du coefficient déterminée pour chaque variable soit due au hasard. Il est préférable qu'elle soit la plus petite possible.

Le logiciel Gretl ne fournit pas les coefficients de régression standardisés nécessaires pour mener la comparaison entre les variables. Ils peuvent néanmoins être recalculés manuellement pour chaque variable en utilisant les coefficients de régression B , l'écart-type lié X et l'écart-type de la variable dépendante Y . Ces données sont consultables dans les statistiques descriptives. La formule à employer est la suivante :

$$B' = \frac{B(x) * \text{écart} - \text{type } X}{\text{écart} - \text{type } Y}$$

Le critère d'Akaike (ou AIC) permet de sélectionner le modèle adapté – présentant la valeur minimale de cet indice – entre différents autres (R. Lancelot et M. Lesnoff, 2005 ; cités par P. Dethier, 2012). Deux autres critères de même utilité sont fournis : le critère de Schwarz (ou BIC) et l'indice Hannan-Quinn. Le premier tient compte, en plus du nombre de variables, de la taille de l'échantillon. Le second est un intermédiaire entre le critère AIC et celui du BIC. Il prend également en compte le nombre d'observations, mais de manière moins importante que le BIC. Ces indices doivent également être les plus petits possibles (R. Lancelot et M. Lesnoff, 2005 ; cités par P. Dethier, 2012).

2.4.3. Statistiques descriptives

Une première étape de la modélisation, avant même l'exécution des régressions proprement dite, consiste à analyser les statistiques descriptives des différentes variables. Ceci permet notamment de vérifier qu'il n'existe pas de valeurs aberrantes vis-à-vis du domaine d'existence mathématique des variables retenues, ou de valeurs extrêmes par rapport au reste de l'échantillon.

Ensuite, à l'instar de Y. Grignet (2011), C. Delbar (2012) et P. Dethier (2012), nous prêtons attention à la distribution des variables structurelles quantitatives, dans notre cas la superficie et la pente du terrain. Cette étape est même vivement recommandée pour la variable dépendante, le prix. Cette dernière doit en effet présenter une distribution s'approchant au maximum de la loi normale (F. Des Rosiers, 2001). A l'inverse, l'obtention d'une distribution normale pour la superficie n'est pas obligatoire, mais il est préférable que l'histogramme soit continu (Y. Grignet, 2011). Ceci permet de plus de repérer des valeurs extrêmes.

2.4.4. Analyse des corrélations entre variables explicatives

Une autre étape consiste à analyser une matrice de corrélation entre toutes les variables explicatives et d'y identifier les couples présentant une corrélation dite excessive. Il est indispensable de ne pas intégrer dans les régressions des variables présentant une telle corrélation avec une ou plusieurs autres variables. L'inverse peut entraîner des problèmes de colinéarité et biaiser les coefficients du modèle. La colinéarité parfaite entre les variables explicatives d'une régression multiple survient lorsque une de ces variables – ou plusieurs – est une combinaison linéaire exacte des autres variables du modèle (P. Dethier, 2012). La littérature définit habituellement la valeur 0,8 en valeur absolue en tant que seuil. Nous avons choisi d'utiliser la valeur de +/-0,75 afin d'éviter autant que possible les problèmes de multicollinéarité pouvant survenir lorsque de nombreuses variables sont utilisées (P. Srikhum, 2012). Ce problème avait été rencontré pour l'étude du marché des maisons unifamiliales en région de Liège alors qu'un seuil de 0,8 en valeur absolue avait été choisi (P. Dethier, 2012).

2.4.5. Régression linéaire multiple par MCO

Un temps certain est consacré à de nombreux essais de combinaisons de variables générant autant de modèles différents. Ceci permet de sélectionner les variables ayant le plus grand pouvoir explicatif sur le prix, mais aussi de tester diverses transformations de variables en logarithme, inverse ou racine carrée. Nous cherchons à obtenir le modèle au plus grand pouvoir explicatif (R^2 ajusté maximisé) et à la plus grande robustesse statistique (F maximisé). Nous utilisons pour ce faire la régression par moindres carrés ordinaires.

Lorsqu'un modèle nous convenant a été défini, nous passons à une étape d'écrémage du nombre de variables par retrait des paramètres ne présentant pas une p-value significative. Y. Grignet (2011) et P. Dethier (2012) ont retenu 0,10 comme seuil critique. F. Des Rosiers (2001) propose une épuration en deux phases : à 0,15 puis à 0,05. Soucieux d'obtenir un modèle avec un maximum de variables significatives, nous choisissons de procéder à une épuration en deux phases, en éliminant premièrement toutes les variables ne présentant pas de p-value inférieure à 0,20. A la deuxième étape de l'écrémage, nous retirons les variables pour lesquelles la p-value reste supérieure à 0,10. Le premier seuil (p-value = 0,20) peut paraître élevé, mais il est motivé par le constat d'une grande variabilité des p-values de certaines variables de localisation. Cette constatation a été faite lors d'un travail concernant l'approche hédonique réalisé dans le cadre

des Travaux Pratiques de Marchés fonciers et immobiliers. En effet, quelques p-values se réajustent lors de la diminution du nombre de variables. Le risque est ainsi de voir une variable potentiellement intéressante et significative retirée trop tôt de la modélisation. Cette procédure peut entraîner une légère baisse du R^2 ajusté. Ceci ne porte pas préjudice à l'analyse étant donné que les risques de multicollinéarité se retrouvent réduits et que, surtout, la robustesse statistique du test en ressort renforcée (élévation marquée du F).

Une fois cette double phase d'épuration effectuée, il est indispensable de contrôler l'absence de colinéarité entre les variables retenues. Cette étude de colinéarité se fait au moyen des indices VIF ou facteurs d'inflation de variance. Une valeur de VIF trop élevée peut indiquer un problème de colinéarité engendré par la variable qu'il mesure, et donc le fait que cette variable est expliquée en grande partie par l'ensemble des autres variables du modèle. Le logiciel Gretl signale que toute valeur supérieure à 10 est suspecte. Une telle valeur indique que la variable est expliquée à 90 % par les autres variables (F. Des Rosiers, 2001). P. Dethier (2012) considère le seuil de 7,5. Dans tous les cas, il s'agit de se montrer vigilant lorsqu'une valeur excédant 5 est repérée, les variables binaires étant moins sensibles à cet indice que les variables quantitatives (F. Des Rosiers, 2001). Si cela est rencontré, la variable doit être retirée du modèle. Le VIF correspondant à une variable i se calcule comme suit :

$$\text{VIF}(i) = \frac{1}{1 - R^2(i)}$$

avec $R^2(j)$ le coefficient de corrélation multiple de la variable i avec l'ensemble des autres variables.

Nous nous assurons ensuite que l'hypothèse d'homoscédasticité des résidus est bien vérifiée. L'homoscédasticité est la situation pour laquelle la variance des résidus est constante et n'est pas fonction des variables explicatives. Son opposé, l'hétéroscédasticité, est à éviter.

La détection de l'hétéroscédasticité peut se faire à l'observation d'une distribution en cône ou en sablier des résidus dans un graphique les associant à la variable dépendante. Il est cependant préférable d'avoir recours à une solution statistique pour confirmer ou infirmer l'existence du problème, surtout lorsque la taille de l'échantillon est réduite, rendant l'observation et l'analyse complexes. Le test de Breusch-Pagan fourni dans Gretl permet justement d'identifier les variables concernées. Cette statistique est construite en régressant les résidus au carré sur les variables explicatives du modèle (J. Dubé, 2010).

Si aucun problème d'hétéroscédasticité n'est présent, la modélisation continue au point 2.4.7.

2.4.6. Correction de l'hétéroscédasticité

S'il s'avère que certaines variables sont bien concernées par de l'hétéroscédasticité, plusieurs solutions s'offrent à l'analyste :

- dans un premier temps, une réflexion sur l'éventuel oubli de variables explicatives importantes doit être menée ;
- la deuxième étape consiste à transformer les variables concernées et à vérifier si la robustesse et le pouvoir explicatif du modèle en ressortent renforcés. Si ce n'est pas le cas, il est préférable d'avoir recours à la dernière solution ;

- cette solution consiste en l'utilisation d'une régression corrigeant *de facto* l'hétéroscédasticité. Le logiciel Gretl propose un tel module utilisant l'ajustement par moindres carrés généralisés. Ce module commence par réaliser une régression classique par moindres carrés ordinaires. Il régresse ensuite le logarithme des carrés des résidus en fonction des variables initiales et de leur carré. Enfin, la solution utilise la méthode des moindres carrés pondérés, exploitant l'inverse de la variance estimée. Cette régression MCG permet d'accorder davantage d'importance aux observations à variance faible et de diminuer celle des observations possédant une variance importante (P. Dethier, 2012).

Une fois le nouveau modèle développé, une augmentation du F et du R^2 ajusté est généralement observée, tandis que les critères AIC, BIC et Hannan-Quinn voient leur valeur fortement diminuée. Enfin, un nouveau test de colinéarité est souhaité pour vérifier l'absence de ce problème.

2.4.7. Retrait des observations à résidus extrêmes et ajustement du modèle

L'opération suivante consiste à supprimer les résidus importants du modèle. Le logiciel Gretl identifie les résidus supérieurs à 2,5 fois l'erreur standard d'estimation. Il convient de retirer de tels résidus afin d'essayer d'augmenter la robustesse statistique et le pouvoir explicatif du modèle. Le modèle est ensuite réajusté sur base de l'échantillon épuré.

Il est indispensable que les résidus soient distribués normalement. Le logiciel Gretl réalise un test d'hypothèse sur base du Chi carré. L'hypothèse nulle étant la distribution normale des résidus, la p-value associée au test doit être supérieure au seuil critique 0,05.

Le graphique de la droite de Henry, nommé aussi QQ-plot ou diagramme quantile-quantile, permet de repérer des variations dans la distribution des résidus en fonction de leur valeur, et donc des intervalles d'ajustement moins performant du modèle. L'abscisse correspond aux quantiles de la loi théorique et l'ordonnée présente les quantiles réellement observés. Les résidus doivent se distribuer au plus près de la droite.

L'analyse de la distribution des résidus des variables indépendantes en fonction de la variable dépendante permet d'identifier les raisons de l'ajustement imparfait du modèle et les valeurs qui sont concernées.

2.4.8. Contrôle de l'absence d'autocorrélation spatiale

Une dernière étape consiste à vérifier l'absence d'autocorrélation spatiale entre les résidus. Cette vérification a été effectuée grâce au logiciel GeoDa. Les résidus standardisés sont exportés depuis le logiciel Gretl et sont joints au shapefile des observations sur base de l'identifiant attribué à la fin de la période d'encodage. Une fois le shapefile prêt, celui-ci est importé sous GeoDa. Il est ainsi possible d'obtenir l'indice général de Moran. Celui-ci doit présenter une valeur proche de 0, caractéristique d'une absence totale d'autocorrélation spatiale. Il doit de plus ne pas être significativement différent de l'indice de Moran attendu. Nous veillerons ainsi à la p-value du test, qui doit être supérieure au seuil critique (0,05). A l'indice est associé un diagramme de dispersion de Moran. Le graphique est décomposé en quatre quadrants, dont la signification est reprise dans le tableau situé à l'annexe IX.

La présence d'autocorrélation peut indiquer l'oubli d'une variable explicative importante ou une contribution non linéaire d'une des variables à l'explication de la variable dépendante. Il est

également possible de générer une analyse spatiale LISA permettant d'identifier des clusters d'autocorrélation locale des résidus, signe d'un ajustement imparfait du modèle pour les lieux où se présentent ces clusters. Ceci permet d'éventuellement identifier des variables qui auraient été oubliées lors de la construction des modèles. Dans ce cas, il est nécessaire de recommencer la procédure de modélisation.

3. Modélisation du prix des terrains à bâtir dans le bassin d'emploi de Liège

Ce chapitre porte sur le développement et la présentation des résultats des modèles développés sur base des terrains à bâtir localisés au sein du bassin d'emploi de Liège. Une première section présente la méthodologie employée pour isoler le marché des terrains à bâtir. Nous présentons ensuite le développement et les résultats du modèle général conformément aux démarches explicitées dans le chapitre précédent. Nous constatons une prédiction médiocre de notre modèle au niveau de l'agglomération opérationnelle liégeoise de 1971. C'est pourquoi, dans une troisième section, nous réalisons un deuxième modèle, cette fois basé sur les terrains à bâtir localisés en dehors de l'agglomération opérationnelle afin de nuancer certains résultats du modèle général. Nous avons également tenté de développer un modèle pour l'agglomération opérationnelle, sans succès, en raison d'un nombre trop faible de biens concernés, de problèmes de colinéarité et d'une distribution non normale des résidus.

3.1. Sélection des observations

3.1.1. Prise en compte du bassin d'emploi

Comme nous l'avons déterminé plus haut, les limites de la zone d'étude pour le développement des modèles liégeois ont été fixées aux frontières du bassin d'emploi tel que défini par l'IWEPS. Nous considérons cette zone d'étude car nous pensons qu'elle délimite un marché homogène et qu'elle est donc pertinente pour l'établissement de la fonction hédonique. Celle-ci ne peut porter que sur un seul marché²⁴. Il est à noter que, malgré leur appartenance au bassin d'emploi de Liège, les Fourons n'ont pas été repris dans la zone d'étude. La couverture spatiale d'un bon nombre de données, limitée à la Wallonie, a poussé à cette exclusion. En effet, des documents tels que le modèle numérique de terrain, le réseau hydrographique, les shapefiles des arrêts de bus, des gares ou encore certaines données socio-économiques sont limités aux frontières de la Wallonie.

La listes des communes faisant partie de la zone d'étude est disponible à l'annexe I. Toutes les entités font partie de la province de Liège, à l'exception de Durbuy, Erezée et Manhay, communes du nord de la province du Luxembourg.

3.1.2. Isolement du marché des terrains à bâtir

Une première étape a scindé, au sein de l'échantillon, le marché des terrains à bâtir et le marché des gisements fonciers. Leur distinction a été exposée au sein de l'état de l'art²⁵. Ces deux marchés répondent à des logiques différentes : nous supposons que les variables explicatives et leurs impacts sur les prix ne sont pas toujours identiques. Par exemple, la superficie a une contribution marginale au prix plus élevée dans le cas d'un terrain à bâtir que dans celui d'un gisement foncier. Dans la mesure où la fonction hédonique est représentative d'un seul marché, cette segmentation est cohérente.

²⁴ Voir point 1.3.3.

²⁵ Voir point 1.2.2.

La séparation entre les marchés peut être opérée sur base de la superficie. Les gisements fonciers ont des superficies généralement plus importantes que les terrains à bâtir, toutes choses égales par ailleurs. Cependant, la limite n'est pas nette, les deux marchés s'imbriquant l'un dans l'autre. Il existe ainsi un continuum entre le marché des terrains à bâtir et le marché des gisements où les représentants du premier diminuent progressivement en nombre au profit de celui des représentants du deuxième. Il est dès lors difficile de séparer distinctement les deux marchés. Par exemple, Stadim²⁶ place arbitrairement la limite entre les deux marchés à 1500 m². Il est cependant fort probable que, dans le contexte wallon, la limite puisse être placée plus haut, la Wallonie étant caractérisée par une moyenne de superficie des terrains plus importante (J.-M. Halleux, 2012b).

Les statistiques descriptives appliquées aux superficies des terrains de l'échantillon montrent clairement une large gamme d'étendues, comprises entre 156 m² et 6300 m² (Tableau 1). La moyenne s'établit à 1117 m². Ceci signifie que la plupart des observations ont une superficie réduite ou moyenne, mais que quelques observations possèdent néanmoins des surfaces importantes. L'observation possédant la superficie minimale est un terrain à bâtir, l'observation présentant la superficie maximale est un gisement foncier.

	Nombre	Moyenne	Ecart-Type	Minimum	Maximum
Superficie	471	1117,238	691,9213	156,000	6300,000

Tableau 1 – Statistiques descriptives appliquées aux superficies de l'ensemble de l'échantillon.

Pour isoler le marché des terrains à bâtir, nous avons utilisé un modèle²⁷ reposant sur les terrains de superficie comprise entre 0 et 2400 m². Plusieurs seuils ont été testés, mais l'intervalle ici retenu met, de manière optimale, les deux marchés en évidence sur base de l'analyse des résidus en fonction de la superficie. Le premier est représenté par une concentration de résidus légèrement croissante dans la partie gauche du graphique ci-dessous (Figure 2), entre 0 et approximativement 1600 m². Ce marché est celui des terrains à bâtir. Passé ce seuil, une grande partie des résidus se trouve systématiquement sous-évaluée par le modèle ainsi développé. Ces résidus proviennent de terrains à plus faible prix par mètre carré et peuvent être considérés comme appartenant à un marché différent du premier. Ce marché est défini comme étant celui des gisements fonciers. Connaissant d'une part la limite arbitraire des 1500 m² définie par Stadim et le fait que les terrains wallons ont tendance à avoir une superficie moyenne plus importante, la valeur des 1600 m² est tout à fait cohérente et peut être interprétée comme le seuil au-delà duquel le marché des gisements prend progressivement et de manière significative le relais du marché des terrains à bâtir.

3.1.3. Isolement du marché des terrains juridiquement urbanisables

Pour nous assurer que l'échantillon de biens sur lequel vont porter les différentes analyses représente bien un marché homogène, nous avons croisé le shapefile des observations avec un shapefile des zones d'habitat et d'habitat à caractère rural du Plan de Secteur. Ceci permet d'éliminer d'éventuelles annonces encodées dans la catégorie « terrains à bâtir » du site Immoweb par des propriétaires ignorant que leur terrain n'est en réalité pas constructible. Ceci

²⁶ Consultant anversois en immobilier.

²⁷ Ce modèle a été développé de manière similaire à la procédure détaillée au point 2.4.

évite également d'utiliser les quelques biens situés dans des zones telles que les ZAE²⁸, qui sont susceptibles de répondre à d'autres logiques et d'appartenir dès lors à un marché distinct.

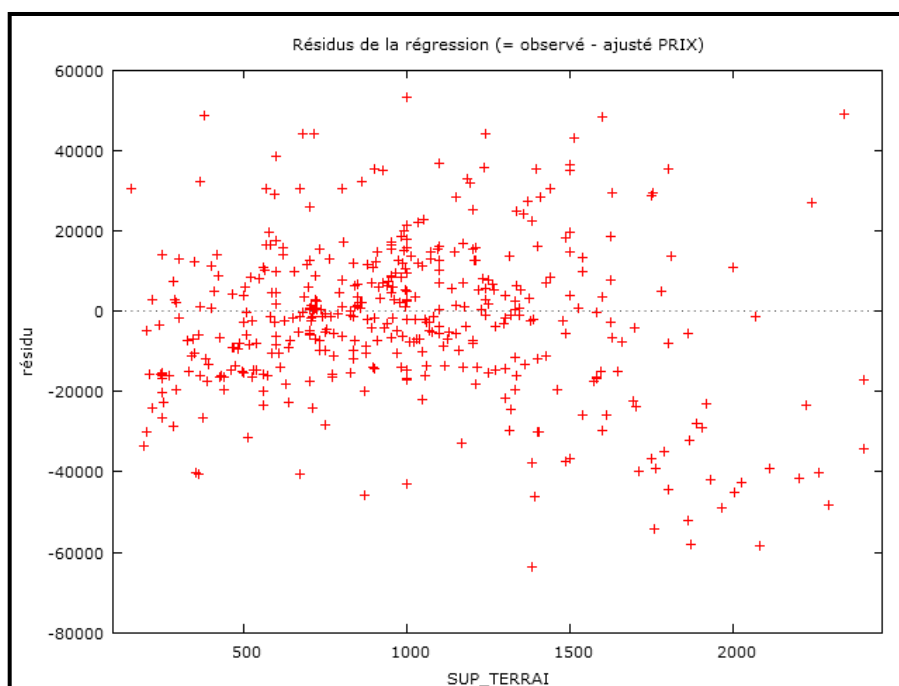


Figure 2 – Résidus en fonction de la superficie issus du modèle 0 – 2400 m².

3.2. Modélisation des prix pour l'ensemble du bassin d'emploi de Liège en 2013-2014

3.2.1. Développement du modèle

3.2.1.1. Statistiques descriptives

En premier lieu, nous analysons les statistiques descriptives de la variable dépendante, le prix. Il en ressort que, si la plupart des valeurs sont distribuées de manière pratiquement normale, certains terrains présentent des prix très élevés. Nous supprimons les biens de plus de 175 000 euros, car nous considérons qu'ils ne sont pas réellement représentatifs du marché des terrains à bâtir étudié. Il s'agit probablement de biens exceptionnels. De plus, cette opération permet de rapprocher davantage la distribution des prix d'une loi normale, cette condition devant être respectée pour la variable dépendante (F. Des Rosiers, 2001).

En étudiant les statistiques descriptives des autres variables, nous repérons deux valeurs négatives de la part modale des modes lents. Or, ceci est incohérent avec le domaine d'existence mathématique de la variable, compris entre 0 et 1. Nous supprimons donc ces deux observations.

L'échantillon final comporte 356 terrains à bâtir répartis à travers le bassin d'emploi de Liège (Figure 3). La densité des observations est plus importante à proximité des pôles urbains de Liège et de Huy, ainsi qu'en Hesbaye. A l'inverse, les mises en vente sont moins nombreuses dans la moitié sud du bassin d'emploi, plus rurale.

²⁸ Zone d'activités économiques.

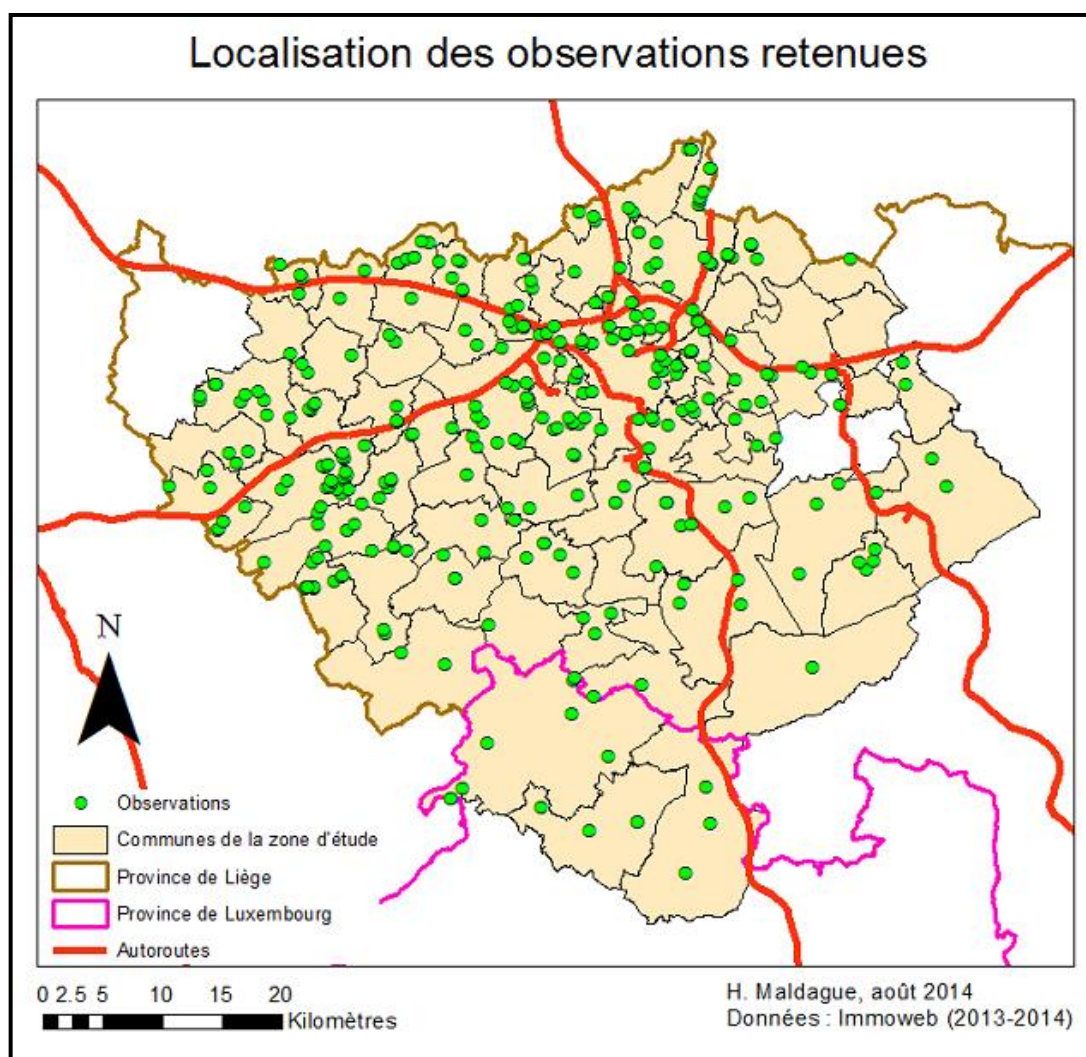


Figure 3 – Répartition des observations retenues au sein de la zone d'étude.

L'échantillon final sur lequel les régressions multiples vont être appliquées présente donc les caractéristiques suivantes :

- superficies comprises entre 156 et 1600 m², jugées représentatives du marché des terrains à bâtir ;
- prix compris entre 12 000 et 175 000 euros, jugés représentatifs du marché des terrains à bâtir standard ;
- biens situés dans les zones d'habitat et d'habitat à caractère rural, et donc juridiquement urbanisables.

Variabes	Minimum	Maximum	Moyenne	Ecart-type
Prix (euros)	12000	175000	79026,25	32058,59
Superficie (m ²)	156	1600	889,34	352,43

Tableau 2 – Statistiques descriptives pour le prix et la superficie.

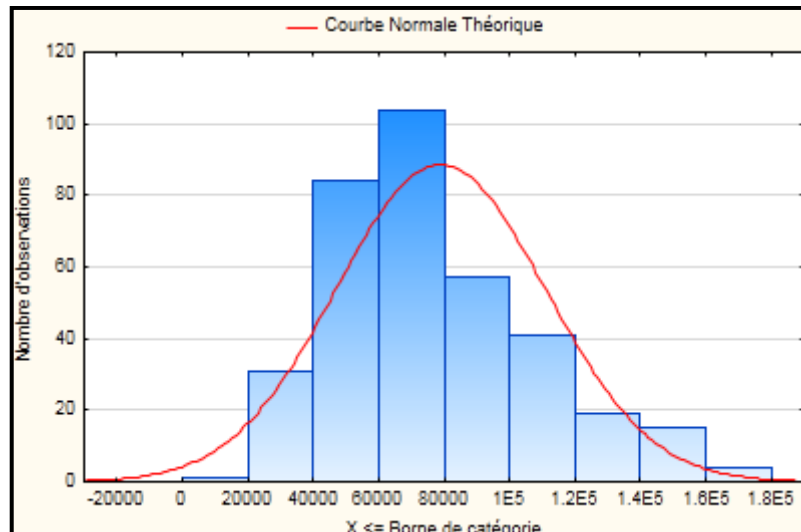


Figure 4 – Distribution des valeurs de prix.

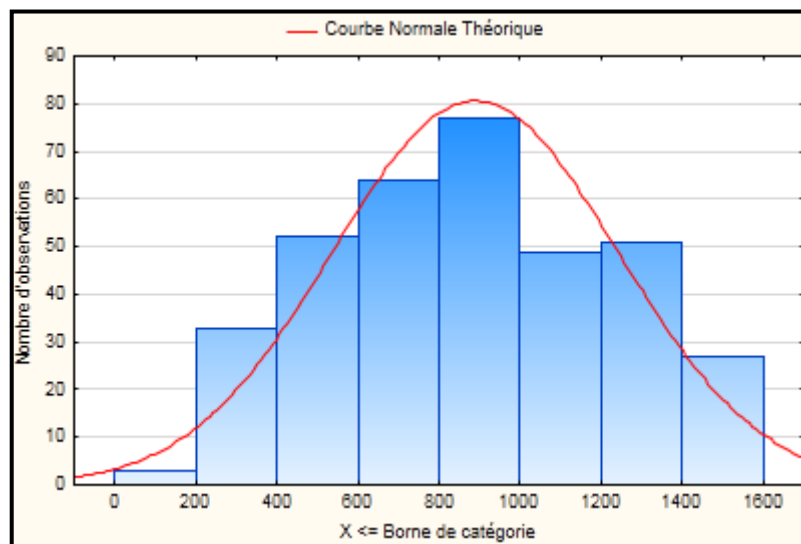


Figure 5 – Distribution des valeurs de superficie.

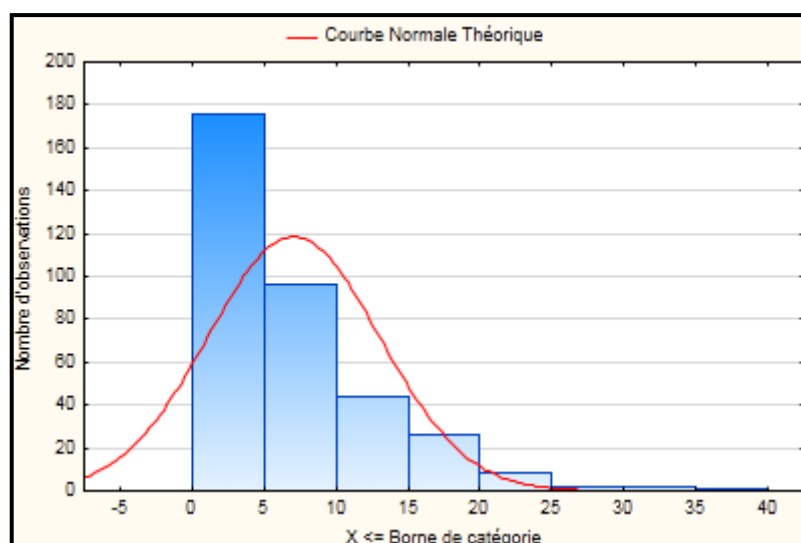


Figure 6 – Distribution des valeurs de pente

La figure 4 montre que les valeurs de prix sont distribuées de manière pratiquement normale, l'histogramme étant de plus continu, une fois les prix de plus de 175 000 euros retirés de l'échantillon.

Comme Y. Grignet (2011), C. Delbar (2012) et P. Dethier (2012), nous prêtons attention à la distribution des variables structurelles quantitatives, à savoir la superficie et la pente. La figure 5 montre un histogramme continu pour la superficie.

Nous pouvons constater qu'il en va de même pour la pente au niveau du terrain (Figure 6).

3.2.1.2. Analyse des corrélations entre variables explicatives

Bon nombre de variables d'accessibilité présentent des corrélations excessives entre elles. Ainsi, les variables de temps d'accès à Liège, Seraing, Herstal, Ans, Verviers, Huy et au Sart-Tilman de même qu'aux écoles secondaires et aux établissements de l'enseignement supérieur sont systématiquement concernées par des corrélations excessives avec une ou plusieurs autres variables. Cela est sans doute dû à la proximité des pôles étudiés par rapport à l'espace couvert par le bassin d'emploi. Cependant, la décomposition des variables de temps d'accès aux écoles en variables binaires permet de contourner ces corrélations excessives.

Il existe également une corrélation entre certaines variables de proximité comme, par exemple, la distance euclidienne aux lignes de chemin de fer et la distance euclidienne aux gares. Une fois encore, la décomposition de ces variables quantitatives en variables binaires permet de solutionner le problème.

Le lecteur trouvera à l'annexe X la liste des couples de variables concernées par une colinéarité excessive (corrélation supérieure à 0,75).

Ces corrélations excessives forcent l'abandon d'un certain nombre de variables jugées intéressantes. Il est décidé de conserver le temps d'accès à Liège, celui restant le pôle central de l'agglomération (E. Van Hecke *et al.*, 2009). Il existe également une corrélation excessive entre le temps d'accès à Verviers, à Huy et la distance euclidienne à la centrale nucléaire de Tihange. Etant donné que cette dernière variable a été créée sur base de la littérature, elle est conservée. Ceci est d'autant plus pertinent que l'attraction exercée par Huy et Verviers est marginale pour le premier pôle et limitée à l'environnement immédiat pour le second (CDPT, 2011). Nous choisissons le temps d'accès à Bruxelles au lieu du temps d'accès à Namur. La première, en tant que première métropole belge, est une destination importante pour l'emploi et exerce son influence sur une grande partie du pays (F. Goffette-Nagot *et al.*, 2011).

3.2.1.3. Tentative de développement d'un premier modèle

Un premier modèle est développé selon la méthodologie théorique expliquée au point 2.4. Ce modèle présente des paramètres témoignant de sa robustesse statistique ainsi que de son pouvoir explicatif et prédictif. Cependant, l'étude des résidus met en évidence la présence d'une autocorrélation spatiale légère mais significative de ceux-ci. L'analyse de l'autocorrélation spatiale locale de ces résidus met en évidence trois concentrations de prédiction médiocre du modèle (Figure 7).

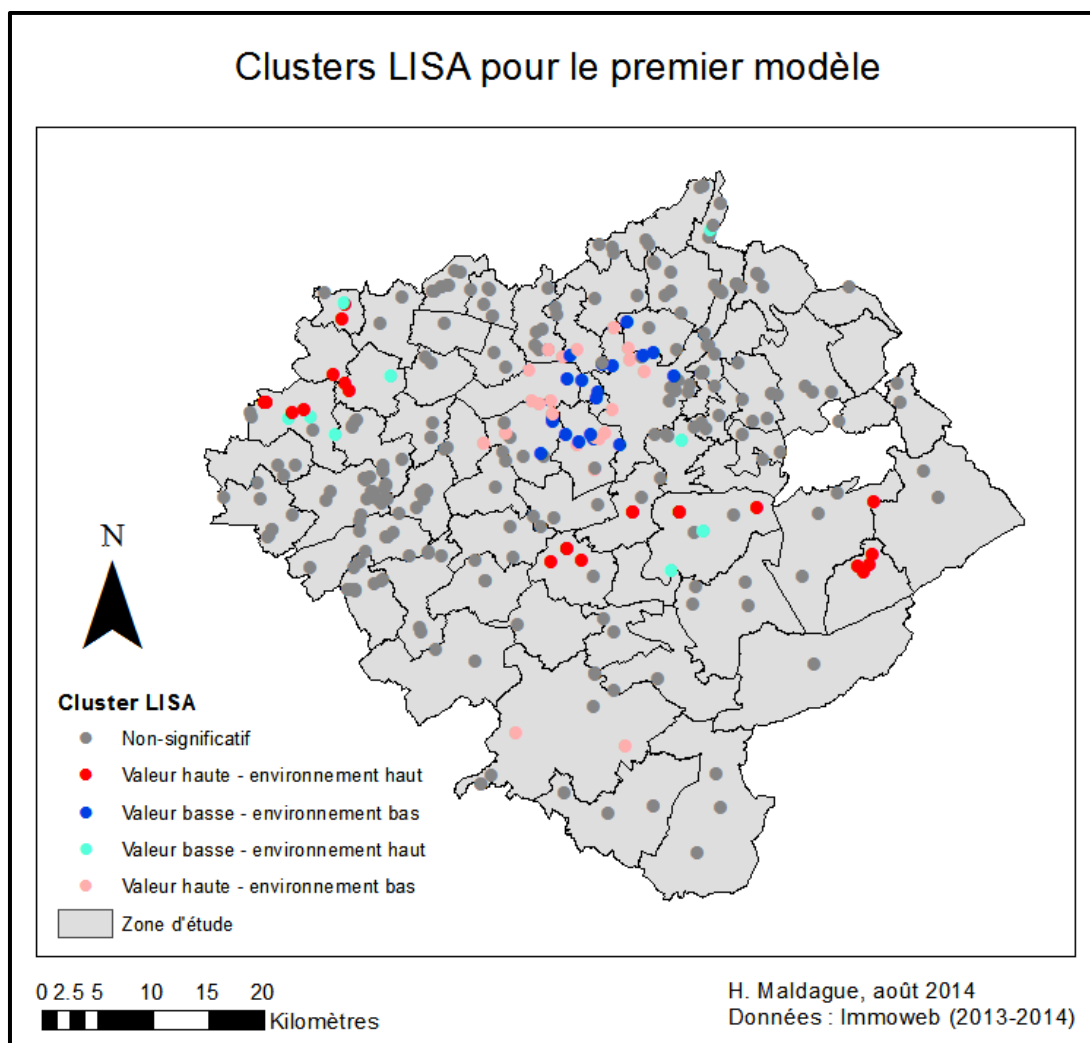


Figure 7 – Cluster LISA pour le premier modèle.

La première d'entre elles correspond au cœur de l'agglomération de Liège, soit la commune de Liège et les communes limitrophes. Une concentration de valeurs basses d'autocorrélation positive s'y retrouve, ainsi que quelques valeurs d'autocorrélation négative. L'environnement bas dénote d'une surestimation des prix par le modèle.

Une hypothèse avancée est qu'il existerait, pour l'agglomération de Liège, un « effet de valorisation » lié à l'appartenance de chaque terrain au quartier dans lequel il se trouve. Cette perception serait très variable d'un quartier à l'autre, et ce sur des distances parfois courtes. Les valeurs basses d'autocorrélation spatiale positive tendent à se concentrer dans l'axe de la vallée de la Meuse, tandis que les valeurs hautes d'autocorrélation spatiale négative se retrouvent principalement sur le haut des versants et le plateau en surplomb de la vallée. Cette perception du quartier est difficilement modélisable et ne peut donc être aisément appréhendée.

Une série de valeurs d'environnement haut se rencontre dans le nord-ouest du bassin d'emploi de Liège. Dans cette zone, le modèle sous-estime les valeurs de prix observées. Ce phénomène peut être attribué à l'influence du bassin d'emploi de Bruxelles. Ce phénomène a été clairement mis en évidence avec une appartenance alternée des communes du nord-ouest de la province de Liège entre le bassin d'emploi éponyme et celui de Bruxelles (CPDT, 2011). Il peut être supposé qu'une pression immobilière s'exerçant sur cette zone entraîne une augmentation des prix en réponse

aux lois de l'offre et de la demande. Ce secteur est en effet une localisation recherchée par les individus travaillant à Bruxelles (J.-M. Halleux, 2005 ; Compagnie des notaires de Liège, 2012). Nous pouvons également supposer que les hautes valeurs de prix observées dans cette région sont dues à la proximité, dans le bassin d'emploi de Bruxelles, de prix fonciers de plus en plus élevés, s'étendant au cours du temps de plus en plus loin de la capitale, et donc sur le nord-ouest du bassin d'emploi de Liège.

La dernière concentration de valeurs hautes est observée aux alentours de Spa. Au niveau de la commune, le modèle sous-estime les prix observés. Les hautes valeurs de prix spadoises pourraient s'expliquer par une offre foncière de plus en plus réduite face à une demande relativement soutenue et en partie due à un sentiment d'appartenance à la commune de la part des habitants qui préféreraient demeurer au sein de celle-ci (J.-B. Jehain, 2014).

Compte tenu des trois zones de prédiction imparfaite du modèle, nous avons cherché à l'améliorer et à supprimer l'autocorrélation spatiale. Il est apparu que la concentration liégeoise correspondait aux communes de tradition industrielle, qui peut être approchée par l'agglomération opérationnelle de 1971 définie sur base des anciennes communes par H. Van Der Haegen et M. Pattyn (1979). Nous pensons que l'appartenance d'un terrain à bâtir à cette agglomération entraîne une moins-value sur son prix. Celle-ci pourrait provenir d'un ensemble de nuisances sonores, visuelles et environnementales liées à un milieu urbain dense à caractère industriel. Dès lors, une localisation au sein de l'agglomération serait moins recherchée qu'une localisation plus périphérique. Nous avons dès lors créé une variable binaire d'appartenance à l'agglomération opérationnelle des années septante nommée *AGOP*. Nous avons également créé une variable binaire *BEBRUX* qui reprend les communes concernées par une mauvaise prédiction du modèle dans le nord-ouest du bassin d'emploi. Ces communes sont Berloz, Geer, Braives et Faimés. Enfin, nous isolons la commune de Spa en créant également une variable binaire, *SPA*. Les trois variables ajoutées ne présentent pas de corrélation excessive entre elles ou avec les autres variables utilisées.

Enfin, nous avons tenté d'isoler la vallée de la Meuse via l'altitude en raison de son caractère industriel et de l'impact qu'elle semble exercer sur les prix des terrains. Néanmoins, il était relativement difficile de définir un seuil d'altitude cohérent car nous prenions alors en compte d'autres vallées où le caractère industriel est beaucoup moins présent. Nous avons dès lors renoncé à explorer cette voie.

3.2.1.4. Modélisation par MCO

De nombreuses régressions par moindres carrés ordinaires ont été effectuées pour déterminer la combinaison de variables la plus intéressante, la plus robuste et la plus explicative des prix des terrains. Ces tests visent ainsi à :

- tester l'influence des différentes variables quantitatives ;
- si celles-ci ne se révèlent pas intéressantes, les remplacer par des variables binaires ;
- trouver, parmi ces variables binaires, le seuil qui apporte le plus d'explication.

Nous avons également testé des transformations pour certaines variables explicatives. Cependant, les paramètres R^2 ajusté et F se trouvaient légèrement diminués. C'est pourquoi les variables non-transformées ont été conservées pour la modélisation. Ces variables sont :

- la superficie, déclinée en logarithme en base 10, en logarithme népérien et en racine carrée ;
- le revenu, décliné en logarithme en base 10.

Nous avons finalement choisi de ne pas intégrer la variable de proximité aux sorties d'autoroutes, malgré sa significativité. Cette variable a en effet une influence sur les paramètres des autres variables et semble perturber ceux-ci. Les modèles développés avec cette variable étaient systématiquement moins performants. Etant donné qu'elle semblait mettre en évidence une moins-value, nous lui avons préféré la variable de proximité des axes routiers ayant un trafic de plus de 20 000 véhicules par jour. A nos yeux, celle-ci est en effet plus explicite.

Lorsque nous avons trouvé ce qui nous semble être la combinaison de variables indépendantes la plus intéressante sur base des différents paramètres fournis par Gretl, nous avons procédé à l'épuration du nombre de variables. A la première étape, nous retirons toutes les caractéristiques présentant une p-value supérieure à 0,20. A la seconde étape, nous abaissons ce seuil critique à 0,10. Le modèle ainsi obtenu présente les paramètres suivants (Tableau 3) :

Moy. Var. dep	79026,25	Ec. typ. var. dep.	32058,59
Somme carré résidus	1,19e+11	Ec. typ. de régression	18748,90
R ²	0,673386	R ² ajusté	0,657971
F (16,339)	43,68274	p. critique (F)	2,56e-72
Log de vraisemblance	-3999,077	Critère d'Akaike	8032,155
Critère de Schwarz	8098,029	Hannan-Quinn	8058,359

Tableau 3 – Paramètres du modèle retenu développé par MCO.

Nous constatons que le modèle développé selon la régression par moindres carrés ordinaires est relativement robuste d'un point de vue statistique : le *F* de Fischer présente une valeur bien supérieure au seuil critique qui peut être trouvé dans les tables statistiques. La p-value critique associée est très petite, confirmant la robustesse du modèle. Le *R² ajusté* n'est par contre pas très élevé. Le *R²* indique que les variables indépendantes intégrées dans le modèle n'expliquent que 67 % de la variance de la variable dépendante. Cependant, il présente déjà un net progrès par rapport aux premiers modèles développés pour tester les différentes variables, qui présentaient un *R²* généralement compris entre 0,45 et 0,55.

Le test de la colinéarité (annexe XI) indique clairement qu'aucune des variables retenues ne présente ce problème. En effet, aucune valeur de VIF²⁹ supérieure à 2,56 n'est présente, la variable de temps d'accès à Liège étant le paramètre concerné. Cette valeur est bien loin du seuil d'attention 5 défini par F. Des Rosiers (2001), et davantage des seuils 7,5 et 10.

Le test de Breusch-Pagan (annexe XII) indique que notre modèle présente de l'hétéroscédasticité de manière significative, et ce sur quatre variables :

- la présence d'un raccordement à l'eau, au gaz ou à l'électricité ;
- l'appartenance à un lotissement ;
- la présence d'une gare à moins de 700 mètres ;
- la superficie du terrain.

²⁹ Voir point 2.4.5.

La présence d'hétéroscédasticité peut se repérer, pour les variables quantitatives, sur un graphique de distribution des résidus de la variable concernée par rapport aux valeurs de cette variable. Le graphique ci-dessous (Figure 8) présente la distribution des résidus selon la superficie. Cependant, le repérage d'une forme en cône ou en sablier typique de la présence de l'hétéroscédasticité est relativement malaisé dans le cas présent, justifiant pleinement le recours au test de Breusch-Pagan. Cette morphologie de distribution des résidus serait sans doute apparue de manière beaucoup plus nette si nous étions en présence d'un échantillon plus fourni et surtout d'une hétéroscédasticité bien plus forte. Il est cependant possible de repérer une légère morphologie en forme de cône : il existe des résidus extrêmes dont le nombre croît légèrement avec la superficie. Ce phénomène est plus rapidement marqué pour les résidus positifs que pour les résidus négatifs.

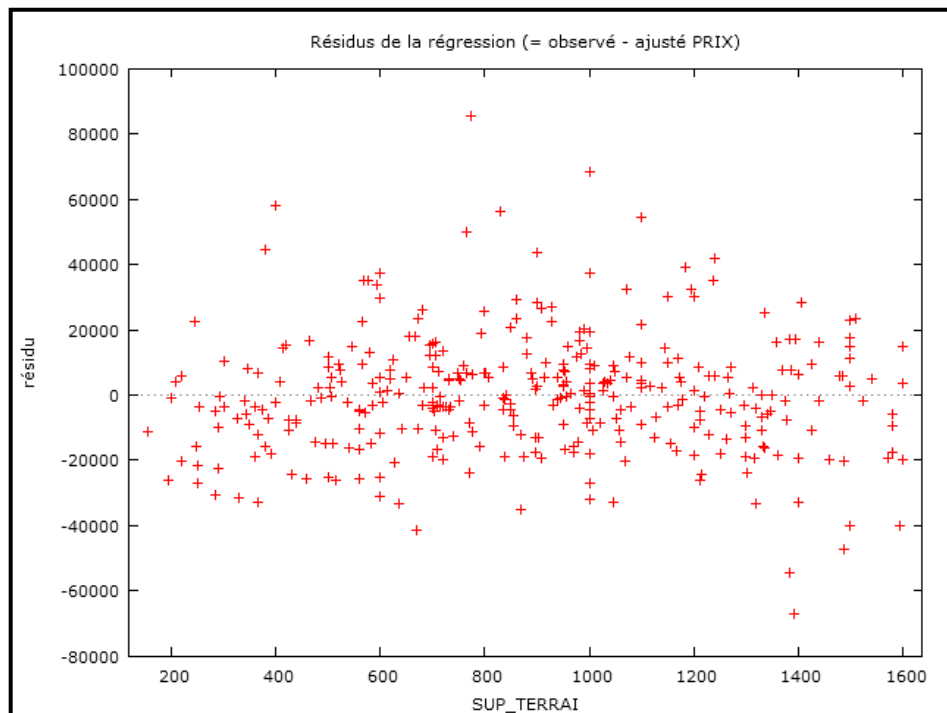


Figure 8 – Distribution des résidus en fonction de la superficie.

3.2.1.5. Correction de l'hétéroscédasticité

La première manipulation pouvant permettre de corriger le modèle de l'hétéroscédasticité est la transformation des variables concernées. Ces manipulations ne sont cependant d'aucun secours dans le cas présent, et ce pour les raisons suivantes :

- il n'est pas possible de transformer les variables binaires concernées que sont l'appartenance à un lotissement, la présence d'un raccordement et la présence d'une gare à moins de 700 mètres ;
- la transformation de la superficie a déjà été expérimentée dès les premières étapes de la modélisation, mais la dégradation du F et du R^2 ajusté nous a poussé à renoncer à cette manipulation.

La deuxième étape consiste à redévelopper un modèle correcteur d'hétéroscédasticité. Les paramètres de ce modèle MCG sont présentés ci-dessous (Tableau 4) :

Moy. Var. dep	79026,25	Ec. typ. var. dep.	32058,59
Somme carré résidus	1,36e+11	Ec. typ. de régression	20065,97
R ²	0,829727	R ² ajusté	0,820632
F (18,337)	91,23197	p. critique (F)	1,3e-117
Log de vraisemblance	-729,8484	Critère d'Akaike	1497,697
Critère de Schwarz	1571,320	Hannan-Quinn	1526,983

Tableau 4 – Paramètres du modèle corrigeant l'hétéroscédasticité.

La correction de l'hétéroscédasticité a amélioré les performances du modèle. Le F connaît une augmentation importante par rapport au modèle MCO, indiquant une très grande robustesse statistique. Il en va de même pour le pouvoir explicatif du modèle, représenté par le R^2 ajusté qui connaît également une augmentation conséquente de sa valeur d'environ 16 points. Les critères d'Akaike et de Schwarz ainsi que l'indice Hannan-Quinn voient leurs valeurs respectives fortement abaissées.

Les résidus en fonction de la superficie présentent une distribution moins conique que celle issue de la régression MCO (Figure 9). Le nombre de résidus extrêmes semble en effet plus constant avec l'augmentation des valeurs de la variable. De plus, les résidus sont davantage concentrés autour de la droite horizontale de valeur nulle, confirmant de manière visuelle une amélioration du modèle.

Une nouvelle analyse de colinéarité (annexe XI) sur les variables explicatives montre que ce problème n'est pas présent de manière significative. Le VIF associé à la variable de temps d'accès à Liège s'est toutefois élevé à 3,53.

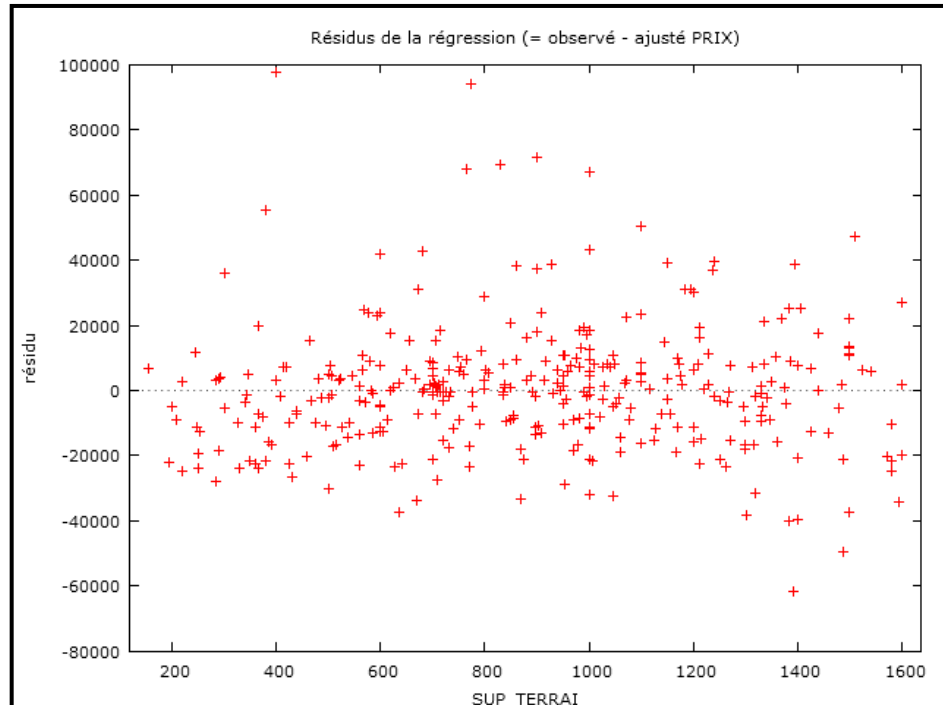


Figure 9 – Distribution des résidus en fonction de la superficie après correction de l'hétéroscédasticité.

3.2.1.6. Suppression des observations à résidus extrêmes et ajustement du modèle

Afin de vérifier le bon ajustement du modèle, nous procédons à l'analyse des résidus. Le logiciel Gretl permet d'identifier les résidus importants, dits aussi résidus délinquants. Les observations qui y sont associées doivent être retirées de l'analyse. Ceci doit mener à un meilleur ajustement du modèle. Onze résidus supérieurs à 2,5 fois l'erreur standard sont identifiés. Huit d'entre eux sont liés à des terrains présentant des prix très élevés, supérieurs à 129 000 euros, sans doute liés à une situation exceptionnelle. Deux autres terrains ont des prix plus classiques, mais sont situés en milieu très dense. Le dernier d'entre eux est un terrain de 55 000 euros pour une superficie de 1393 m², assez mal situé (terrain étroit, proximité du chemin de fer et d'une ligne à haute tension).

Nous retirons ces résidus de l'échantillon et ajustons notre modèle. Celui-ci présente les paramètres suivants (Tableau 5) :

Moy. Var. dep	77439,55	Ec. typ. var. dep.	30233,17
Somme carré résidus	8,29e+10	Ec. typ. de régression	15873,43
R ²	0,818380	R ² ajusté	0,810099
F (15,329)	98,83153	p. critique (F)	5,8e-112
Log de vraisemblance	-702,9368	Critère d'Akaike	1437,874
Critère de Schwarz	1499,370	Hannan-Quinn	1462,364

Tableau 5 – Paramètres du modèle corrigeant l'hétéroscédasticité après suppression des résidus délinquants.

La performance prédictive du modèle, quoiqu'acceptable, n'est pas optimale. L'erreur-type d'estimation recalculée sur base des résidus du modèle vaut 15 568,81 euros. Relativisée par rapport à la moyenne des prix, ceci correspond à une erreur de 20,10 %. Cette erreur est supérieure au seuil de 15 % retenu par F. Des Rosiers (2001). Ceci est à mettre en parallèle avec l'inefficacité des marchés immobiliers : pour chaque bien, « il n'existe pas un prix de marché arrêté, seulement une étendue de valeurs possibles » (A. W. Evans, 1995 ; cité par Y. Grignet, 2011). Le marché foncier étant également un marché inefficace, une réflexion identique peut être appliquée à notre travail. Tout porte cependant à croire que l'erreur peut être diminuée. Celle-ci est en effet probablement due à la faible taille de l'échantillon qui n'a pas permis de construire un modèle au plus grand pouvoir prédictif.

Le F élevé fait état d'un modèle robuste, robustesse qui s'est améliorée avec la suppression des neuf observations à résidus délinquants. Il a en effet gagné pratiquement neuf points par rapport au précédent modèle. A l'inverse, le R^2 ajusté a légèrement diminué, mais reste très satisfaisant. Les variables indépendantes retenues expliquent 82 % de la variance du prix.

La distribution des résidus ne peut être considérée comme normale, la p-value associée au test du Chi carré étant légèrement inférieure au seuil de 0,05 (Figure 10). L'hypothèse nulle, pour rappel la distribution normale des résidus, doit donc être rejetée. Nous constatons la présence de quelques résidus de valeur élevée. De même, la décroissance des occurrences est irrégulière pour les résidus positifs autour de la valeur 10 000. Cependant, le problème peut être relativisé, la distribution des résidus restant très proche de la normale.

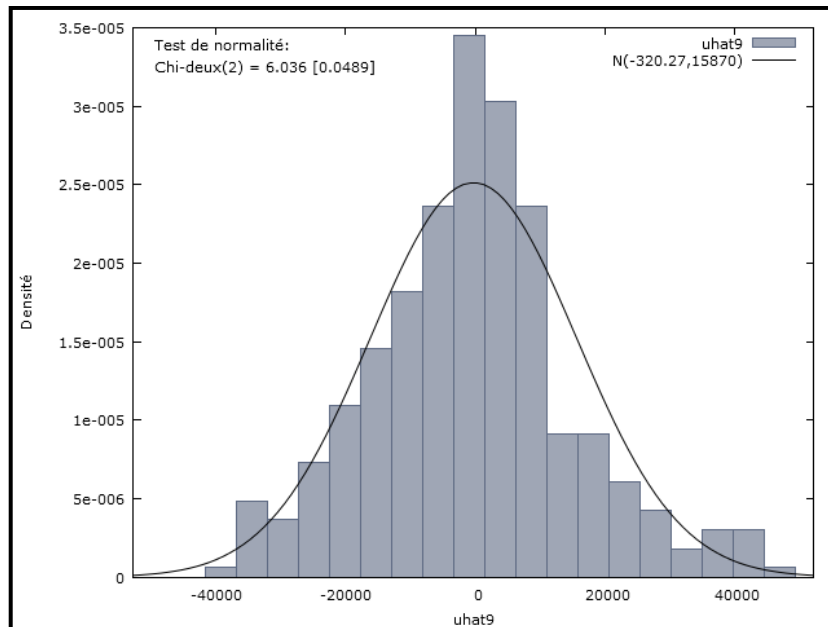


Figure 10 – Distribution des résidus après épuration.

Le diagramme quantile-quantile (Figure 11) montre que les résidus suivent globalement la droite de Henry. Ils ont une distribution proche de la normale pour l'intervalle compris entre -35 000 euros et +30 000 euros. En dehors de cet intervalle, l'ajustement du modèle est imparfait. Les résidus en-dessous de -35 000 euros sont moins négatifs que ce que prévoit le modèle. Au-dessus de 30 000 euros, ils sont plus importants que la prévision.

L'analyse du diagramme de distribution des résidus en fonction du prix (Figure 12) montre que le modèle voit ses performances diminuées au-delà de 110 000 euros : son pouvoir prédictif décline au-delà de cette valeur. Le phénomène s'aggrave une fois passé les 130 000 euros, l'ensemble des résidus devenant positifs. Ce problème a déjà été rencontré auparavant lors de l'étude du marché des maisons unifamiliales à Bruxelles (C. Delbar, 2012) et à Liège (Y. Grignet, 2011, P. Dethier, 2012). Il s'agit d'un phénomène couramment observé en évaluation immobilière (F. Des Rosiers, 2001). Les constatations du présent travail semblent indiquer que ce problème existe aussi pour l'évaluation foncière. Enfin, les valeurs très négatives de résidus entre 50 000 et 80 000 euros sont peut-être représentatives des premiers gisements fonciers qui, pour rappel, se distinguent du marché des terrains à bâtir. Il semble ainsi que notre échantillon couvre différents sous-marchés dont la distinction est relativement malaisée.

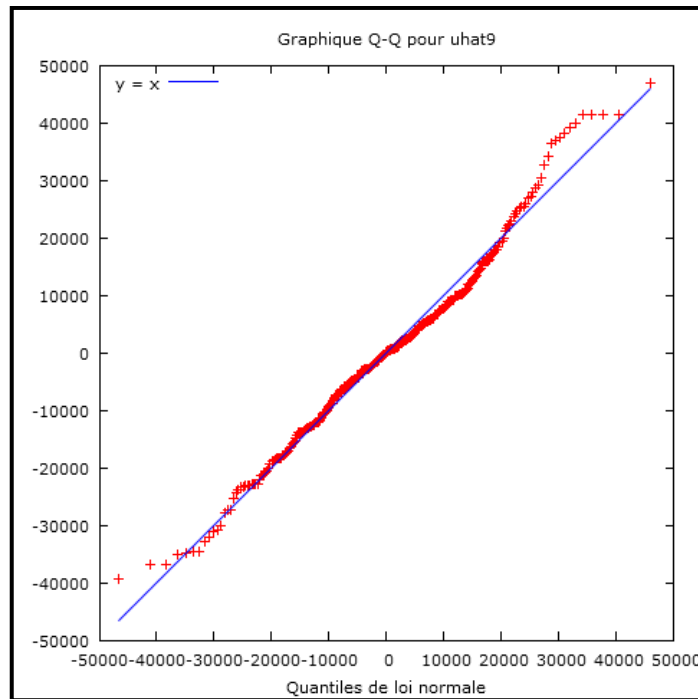


Figure 11 – Diagramme quantile-quantile des résidus.

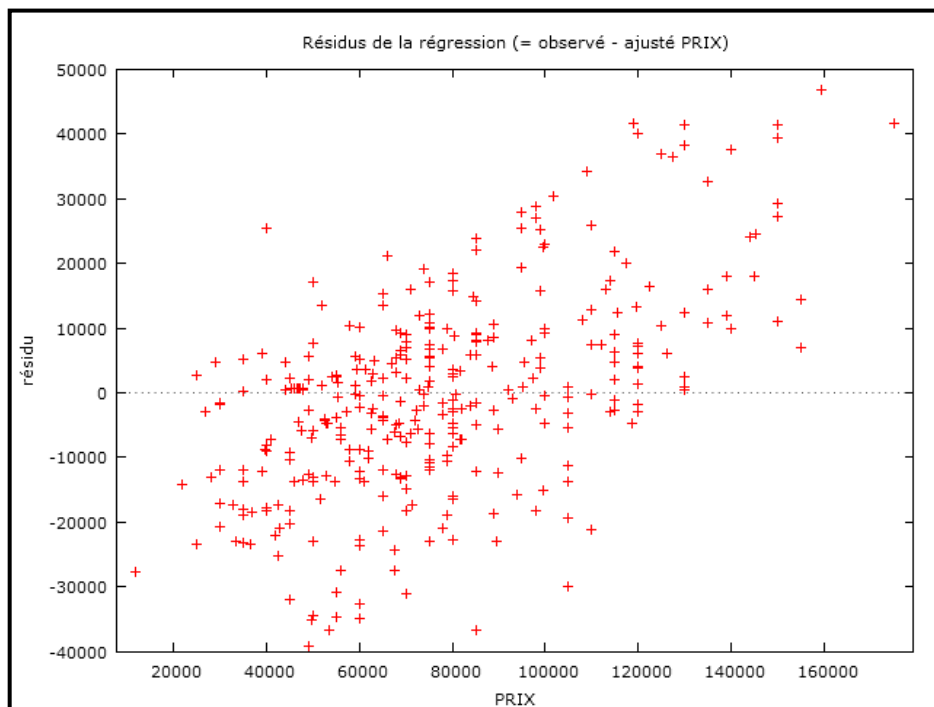


Figure 12 – Distribution des résidus en fonction du prix.

Nous aurions pu essayer de développer de nouveaux modèles en resserrant l'intervalle des prix. Cependant, nous avons écarté cette option en raison de la diminution d'un échantillon déjà fort réduit à la base. De plus, nous avons déjà procédé à un écrémage des observations après bon nombre d'essais, en enlevant tous les biens supérieurs à 175 000 euros. Enfin, si nous supposons une influence de quelques représentants du marché des gisements fonciers, nous ne pouvons décerner avec certitude quels sont les biens qui en font effectivement partie. Cela complique l'écrémage en ce sens.

3.2.1.7. Analyse de l'autocorrélation spatiale des résidus

L'analyse de l'autocorrélation spatiale des résidus montre que celle-ci est absente. L'indice général de Moran est proche de 0 (valeur exacte : 0,011009), et n'est de plus pas statistiquement différent (p-value : 0,18200) de l'indice attendu (-0,0029). Le diagramme de dispersion de Moran est consultable à l'annexe XIII. L'ajout des variables *AGOP*, *BEBRUX* et *SPA* a permis le retrait de l'autocorrélation. Au niveau local, nous devons cependant relativiser ce constat (Figure 13) :

- l'effet d'appartenance à l'agglomération opérationnelle n'a été que partiellement pris en compte par la variable *AGOP*, bien que celle-ci soit significative, signe de l'existence d'un sous-marché. L'effet de valorisation variable entre quartiers, supposé plus haut, n'a pas été correctement cerné. Il est, comme pressenti, difficilement modélisable ;
- la variable *BEBRUX* n'est pas significative. Ceci semble indiquer que l'effet de la proximité du bassin d'emploi de Bruxelles ne s'exerce pas de manière généralisée sur l'ensemble des quatre communes concernées ;
- le mauvais ajustement de Spa existe toujours, mais il a néanmoins été diminué.

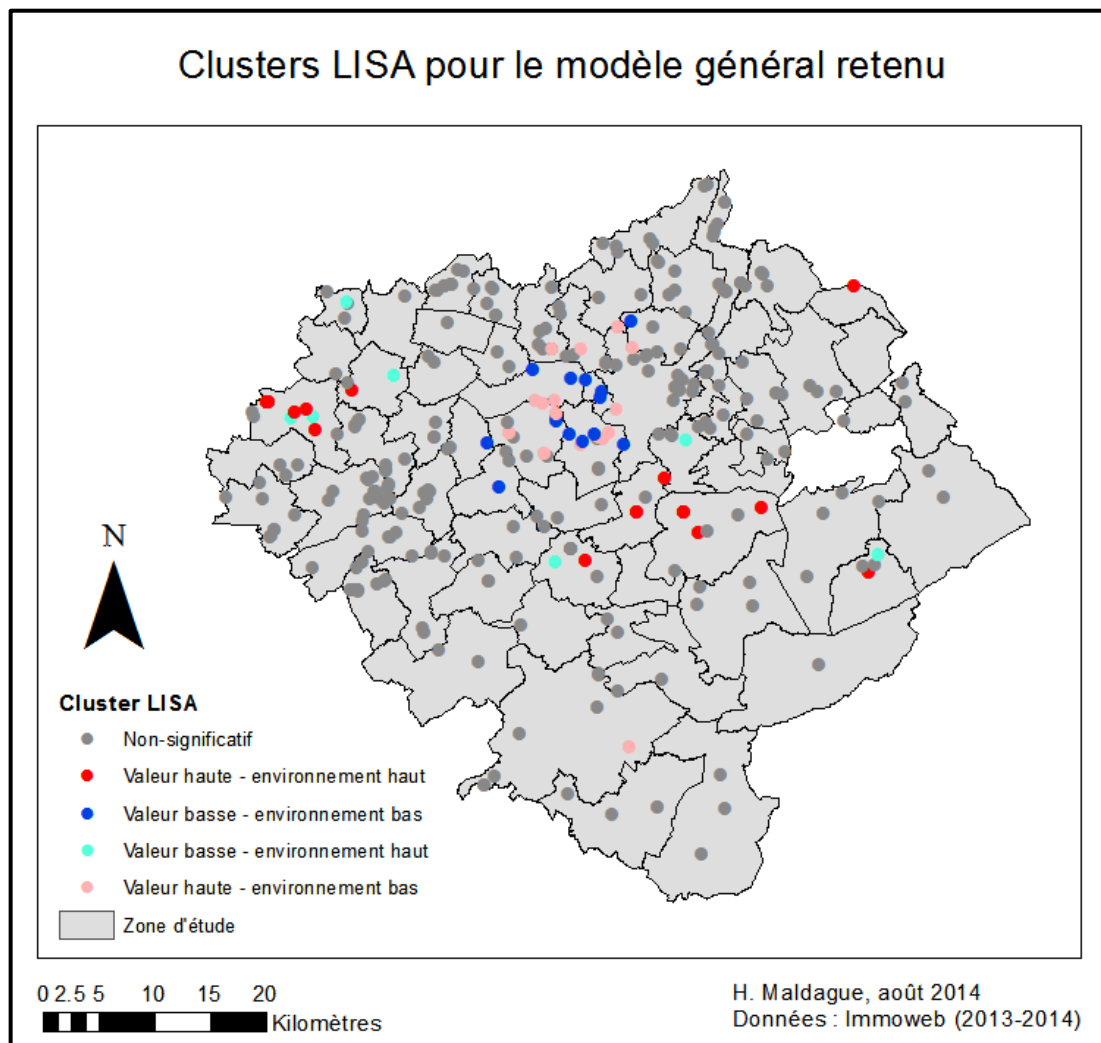


Figure 13 – Clusters LISA pour le modèle général retenu.

Nous aurions pu nous attendre à observer un mauvais ajustement du modèle autour des pôles urbains de Huy et de Verviers, l'influence du temps d'accès à ceux-ci n'ayant pu être appréhendée, faute d'une trop forte corrélation avec d'autres variables menaçant les modèles

développés de problèmes de colinéarité excessive. Le modèle semble pourtant prédire correctement les prix des terrains se trouvant près de ces pôles. L'attraction qu'exercent Huy et Verviers sur leur environnement n'est sans doute pas suffisante pour les faire émerger du rayon d'attraction de Liège. Ceci semble montrer que ni Huy ni Verviers ne contribuent à structurer la géographie des valeurs foncières du bassin d'emploi de Liège :

- le bassin d'emploi de Huy s'est dilué dans celui de Liège entre 1991 et 2001. La situation reste inchangée depuis lors (CPDT, 2011).
- le bassin d'emploi de Verviers, encore consistant en 1991, a fondu depuis pour ne plus concerner que Verviers et Pepinster en 2008. Verviers, qui est de plus proche de Liège, émerge de moins en moins comme un pôle secondaire et est progressivement assimilé au bassin d'emploi liégeois (CPDT, 2011).

3.2.2. Présentation des résultats

Nous présentons dans cette section les résultats du modèle retenu pour représenter le marché des terrains à bâtir du bassin d'emploi de Liège (Tableau 6). Gretl ne fournit pas les coefficients standardisés. Nous les avons recalculés manuellement (Tableau 7).

Sur base des coefficients standardisés, il est possible de dégager un ordre général des types de variables selon leur contribution à l'explication du prix. Le modèle montre l'importance de la superficie dans cette explication. Viennent ensuite les variables d'accessibilité et de valorisation sociale. Les variables d'environnement physique contribuent plus faiblement à la formation des prix. Ceci pourrait indiquer que les ménages, lors d'un achat, prêtent avant tout attention à la superficie et à l'accessibilité du terrain ainsi qu'à la valorisation sociale avant d'envisager l'environnement physique et les éventuelles nuisances générées. Cet ordre est assez similaire à celui mis en évidence par Y. Grignet (2011), bien que le nombre de variables d'environnement physique utilisées ait été plus faible.

	<i>Coefficient</i>	<i>Erreur Std</i>	<i>t de Student</i>	<i>p. critique</i>	
Constante	33949,5	7640,3	4,4435	0,00001	***
Appartenance à la commune de Spa	26109,5	6976,68	3,7424	0,00022	***
Superficie	54,1631	2,35953	22,9551	<0,00001	***
Localisation dans un lotissement	-5028,83	1418,79	-3,5444	0,00045	***
Bruit aérien > 60 dB	-11912,5	3800,66	-3,1343	0,00188	***
Revenus médians	1,11851	0,208633	5,3612	<0,00001	***
Cours d'eau à moins de 100 m	-6105,1	1746,83	-3,4950	0,00054	***
Distance à Tihange	0,570395	0,0853388	6,6839	<0,00001	***
Distance aux arrêts de bus	5,20125	2,00973	2,5880	0,01008	**
Ligne HT à moins de 500 m	-3909,92	1666,21	-2,3466	0,01954	**
Temps d'accès à Liège	-14,5478	1,82042	-7,9915	<0,00001	***
Temps d'accès aux nodules commerciaux	-30,5355	5,40504	-5,6495	<0,00001	***
Route > 20000 veh/j à moins de 200 m	-21218,4	7751,24	-2,7374	0,00653	***
Temps d'accès à Bruxelles	-2,65997	1,16904	-2,2753	0,02353	**
Appartenance à l'agglomération opérat.	-6089,74	2313,23	-2,6326	0,00887	***
Localisation dans un cul-de-sac	4738,63	2272,12	2,0856	0,03779	**
* <i>p-value inférieure à 0,10</i> ** <i>p-value inférieure à 0,05</i> *** <i>p-value inférieure à 0,01</i>					

Tableau 6 – Résultats du modèle général liégeois.

Variable	B*	B
Superficie	0,63080	54,1631
Temps d'accès à Liège	-0,29436	-14,5478
Distance à Tihange	0,21611	0,570395
Temps d'accès aux nodules commerciaux	-0,18202	-30,5355
Revenu médian	0,16280	1,11851
Route > 20000 véhicules/jour à moins de 200 m	-0,09909	-21218,4
Appartenance à la commune de Spa	0,09258	26109,5
Appartenance à l'agglomération opérationnelle	-0,08476	-6089,74
Cours d'eau à moins de 100 m	0,07861	-6105,1
Localisation dans un lotissement	0,07489	-5028,83
Bruit du trafic aérien supérieur à 60 dB	-0,06620	-11912,5
Temps d'accès à Bruxelles	-0,06194	-2,65997
Localisation dans un cul-de-sac	0,06101	4738,63
Ligne à haute tension à moins de 500 m	-0,06036	-3909,92
Distance aux arrêts de bus	0,06035	5,20125

Tableau 7 – Classement des variables explicatives par ordre d'importance.

3.2.2.1. Variables structurelles

La variable la plus significative est la superficie. C'est le premier facteur explicatif du prix en termes d'importance. Pour chaque mètre carré de terrain gagné, une augmentation de 54,16 euros est constatée. Cependant, la contribution de chaque mètre carré ajouté au prix n'est pas constante avec l'augmentation de la superficie du terrain. Ainsi, le prix du mètre carré tend à diminuer au fur et à mesure que la superficie du terrain augmente. Cette marginalité décroissante du prix en fonction de la superficie est notamment expliquée par l'utilité marginale décroissante. Le propriétaire d'un terrain retire en effet davantage d'utilité d'un mètre carré pour un petit terrain que ce même mètre carré pour un terrain plus grand. Nous pouvons également y ajouter le coût marginal, plus important pour les petites superficies³⁰. Ce concept avait notamment été clairement mis en évidence pour le marché des maisons unifamiliales de la région urbaine de Liège (Y. Grignat, 2011 ; P. Dethier, 2012) et de Bruxelles (C. Delbar, 2012). A noter que P. Dethier (2012) mettait en évidence, à l'aide de l'approche hédonique appliquée sur la région urbaine de Liège, une décroissance de 9484 euros lorsque la superficie du terrain passait de 700 à 300 m². Dans notre modèle, la décroissance est plus importante, avec une baisse de 21 600 euros. Ceci peut s'expliquer par deux phénomènes. Premièrement, nous avons développé notre modèle sur un intervalle déterminé de superficies, à savoir 0 – 1600 m², alors que les modèles réalisés sur les maisons liégeoises incluaient des terrains de superficie bien plus importante, cette superficie étant concernée par la marginalité décroissante. Deuxièmement, la valeur d'un mètre carré de terrain nu est plus importante que ce même mètre carré bâti. L'utilité procurée par un mètre carré de terrain nu est ainsi différente de celle procurée par ce même mètre carré de terrain bâti.

Il nous a paru intéressant de tenter de mettre en évidence la marginalité décroissante du prix en fonction de la superficie. P. Dethier (2012) avait pour ce faire développé un modèle comprenant le logarithme de la superficie du terrain des maisons unifamiliales. Nous avons effectué la même démarche en développant un modèle similaire pour les terrains à bâtir. Nous avons cependant constaté que, outre une robustesse statistique et un pouvoir explicatif moindre, ce modèle s'ajustait relativement mal, prédisant des valeurs de 200 000 euros pour un terrain de 1000 m², ce

³⁰ Voir point 1.2.6.

qui est bien trop important. Nous avons alors choisi d'approcher ce phénomène par une autre méthode appliquée par Y. Grignet (2011) et C. Delbar (2012), mais aussi en complément par P. Dethier (2012). Nous avons scindé notre échantillon sur base des quartiles. Les trois seuils séparateurs sont respectivement 600 m², 899 m² et 1168 m². Sur chacun des quatre sous-échantillons, nous avons développé des modèles sur base de la procédure standard appliquée dans ce mémoire. Les résultats sont repris ci-dessous (Tableau 8). Nous attirons l'attention sur le fait que ces modèles sont développés sur base d'échantillons de petite taille.

Intervalle de superficie	coefficient	p-value
0 – 600	83,48	<0,00001
601 – 898	50,59	0,01986
899 – 1167	33,98	0,03000
1168 – 1600	25,41	0,11691

Tableau 8 – Illustration de la marginalité décroissante des prix selon la superficie.

Sur base des quatre modèles développés, nous mettons en évidence l'existence de la marginalité décroissante du prix en fonction de la superficie. Nous constatons une augmentation progressive de la p-value avec la superficie, celle-ci devenant non-significative au seuil retenu de 0,10 pour le dernier intervalle. Ceci semble cohérent si nous envisageons le fait qu'un acheteur potentiel sera beaucoup plus attentif à la superficie pour un petit terrain que pour un plus grand. Une même augmentation de la superficie pour un petit terrain est mieux mise en évidence et prise en compte que cette même augmentation pour un terrain plus important.

Le programme codé en Java ayant été conçu pour récolter un maximum d'informations, nous avons choisi de mettre en évidence une éventuelle influence de la forme des terrains sur leur prix. Nous avons ainsi exporté de l'échantillon principal les terrains présentant des données de largeur à front de rue et de profondeur. De là, nous avons calculé un ratio largeur – profondeur. Nous avons par la suite défini différents seuils au-delà desquels les terrains peuvent être considérés comme trop étroits ou trop large. Dans le premier cas, deux seuils ont été choisis : 0,15 et 0,30. Dans le second cas, nous avons choisi de prendre les terrains plus larges que longs, avec un ratio largeur – profondeur supérieur à 1, ceci suite au faible nombre de biens de ce type. Nous avons également repris une variable testée par X. Gao et Y. Asami (2007), à savoir la largeur à front de rue pondérée par la superficie (l/S). Le sous-échantillon ainsi constitué comporte 140 biens. Le modèle³¹ est développé en suivant la même procédure que celle utilisée tout au long de ce mémoire. Nous parvenons à mettre en évidence un impact sur les prix de la part de terrains présentant un ratio largeur – profondeur inférieur à 0,30. La moins-value est de 8321,56 euros. La p-value est néanmoins élevée (0,05576). Ceci semble indiquer qu'un terrain étroit sera moins attractif.

La présence d'un bâtiment n'a pas d'influence significative sur le prix. La raison envisagée est que les bâtiments édifiés sur les terrains concernés de notre échantillon sont des garages ou des abris de jardin qui peuvent au choix être valorisés ou détruits.

Enfin, les paramètres liés au relief, la pente au niveau du terrain et l'orientation au sud, ne sont pas significatifs. Pour la première variable, ceci peut s'expliquer par le faible nombre de terrains en forte pente dans notre échantillon : seuls treize biens ont une pente de plus de 20 %, trois à

³¹ Paramètres du modèle final : F = 64,15, R² ajusté = 0,653.

peine ont une pente de plus de 30 %. Il n'est guère difficile d'adapter la maison que l'on souhaite construire à une pente importante ; le coût du terrassement nécessaire peut de plus être contrebalancé par le coût implicite de la vue qu'offre une situation en pente.

Nous nous rendons compte que l'orientation du terrain est une notion imprécise, ce qui explique l'absence de significativité. En effet, l'influence de l'orientation peut varier selon la pente : un terrain fort en pente vers le sud a peut-être une plus-value par rapport à ce même terrain sur la même pente forte mais orientée au nord, mais cette plus-value peut disparaître pour des terrains à pente très faible.

3.2.2.2. Variables d'accessibilité

La distance euclidienne aux arrêts de bus est contre-intuitive et engendre une plus-value. Ceci est contraire aux résultats des études consultées qui prédisent une plus-value portée aux terrains engendrée par la proximité d'un arrêt de bus (X. Gao et Y. Asami, 2007 ; Y. Liu *et al.*, 2010). Chaque mètre d'éloignement entraîne une élévation du prix de 5,20 euros. La p-value, bien que significative, est assez élevée. Pour le bien le plus éloigné d'un arrêt de bus, à 2783 mètres, cela représente une plus-value de 14 475,08 euros. La moyenne de distance des biens aux arrêts, 322,24 mètres, montre qu'une grande partie de ces biens est située à proximité des arrêts. Les variables binaires au seuil de 500 et 700 mètres se sont également montrées significatives, mais de manière moins prononcée que la variable de distance euclidienne. Ce phénomène de moins-value engendrée par la présence d'un arrêt de bus à moins de 700 mètres avait déjà été mis en évidence pour le marché des maisons unifamiliales (P. Dethier, 2012). L'explication alors avancée était la densité du milieu desservi par les transports en commun. Ceux-ci ont de plus tendance à se concentrer sur les grands axes. Les seules nuisances, qui s'exercent sur des distances plus courtes (R. Cervero et C. D. Kang, 2011), ne peuvent expliquer à elles seules cette moins-value. Un raisonnement similaire peut être appliqué à notre travail. Les biens les plus éloignés sont caractéristiques de milieux moins denses et moins bruyants. Ils connaissent donc une plus-value. Ceci indique que l'utilité procurée par la proximité d'un arrêt de bus reste faible comparée aux nuisances du lieu d'implantation. L'existence de corrélations significatives et positives entre la distance aux arrêts de bus d'une part, le revenu et le pourcentage de maisons quatre façades d'autre part, confirme cette constatation. Il en va de même pour les corrélations significatives mais opposées entre la distance aux arrêts de bus d'une part, la densité et la part de bâti d'autre part. Les ménages semblent en premier lieu prendre en compte les nuisances du milieu dense dans lequel l'arrêt se trouve. La non-significativité de la part modale du bus vient apporter un éclairage supplémentaire. Le bus ne semble pas apparaître comme un moyen de transport suffisamment intéressant pour que son utilité puisse être ressentie sur le prix des terrains bien desservis.

Le temps d'accès à Liège est une variable très fortement significative. Les prix vont en décroissant au fur et à mesure que l'on s'éloigne de la Place Saint-Lambert, prise pour centre de l'agglomération. Ceci est en accord avec la plupart des études concernant l'approche hédonique des marchés fonciers et immobiliers (C. J. Barnett, 1985 ; C. G. Leggett et N. E. Bockstael, 1999 ; L. Tyrvaïnen et A. Miettinen, 2000 ; D. R. Bowes et K. R. Ihlanfeldt, 2001 ; S. Zheng et M. E. Kahn, 2008 ; M. Taviens *et al.*, 2009 ; D. E. Andersson *et al.*, 2010 ; F. De Vor et H. L. F. De Groot, 2011 ; H. A. Sander et R. G. Haight, 2012). La décroissance des prix avec la distance au centre-ville répond à la concentration des services, des emplois et des facilités qu'une ville peut offrir, en

comparaison avec sa périphérie et le milieu rural (P. F. Colwell et H. J. Munneke, 1997). Ainsi, pour chaque seconde d'éloignement à Liège, le prix diminue de 14,55 euros. Ceci correspond à 872,87 euros par minute. Ainsi, le bien le plus éloigné en termes de temps, à 50 minutes et 20 secondes de la Place Saint-Lambert, est moins cher de 43 934,36 euros par rapport à un bien identique qui serait localisé sur la Place même. A titre de comparaison, P. Dethier (2012) met en évidence, pour le marché des maisons unifamiliales liégeoises, une décroissance de 1165 euros par minute. Cette valeur est proche du coefficient de notre modèle, généré sur une étendue plus importante et où la contribution d'une minute supplémentaire éloignée de Liège est moins importante que la contribution d'une minute à proximité du centre.

Il en va de même pour le temps d'accès aux nodules commerciaux, variable très significative. Ainsi, chaque seconde d'éloignement vis-à-vis d'un nodule entraîne une baisse de 30,54 euros, ce qui correspond à une moins-value de 34 047,08 euros pour le bien le plus éloigné, toutes choses égales par ailleurs, à 18 minutes et 35 secondes d'un nodule commercial. Cette influence négative de l'éloignement aux commerces est cohérente avec deux études sur les quatre consultées dans le cadre de ce travail (X. Gao et Y. Asami, 2007 ; Y. Liu *et al.*, 2010). Ces nodules commerciaux sont nombreux et répartis à travers le bassin d'emploi de Liège, avec une plus grande concentration au niveau de l'agglomération centrale, constat confirmé par l'existence de corrélations modérées négatives entre le temps d'accès aux nodules et les variables de part de bâti et de densité. Une corrélation modérée positive existe entre le temps d'accès aux nodules et le temps d'accès à Liège (0,55). Ceci signifie qu'une part de l'explication du prix donnée par le temps d'accès aux nodules provient du temps d'accès à Liège.

Le comportement de la variable de temps d'accès à Bruxelles montre qu'il existe une polarisation de la métropole et capitale fédérale au sein du bassin d'emploi de Liège. Cette attraction de Bruxelles, premier pôle d'emploi de la Belgique, s'étend sur l'ensemble du pays (J.-F. Thisse et I. Thomas, 2007 ; cités par F. Goffette-Nagot *et al.*, 2011). La proportion de Liégeois travaillant à Bruxelles est en effet non négligeable et en constante augmentation. D'autre part, la pression foncière et immobilière de plus en plus forte en périphérie de Bruxelles force les ménages à trouver un logement ou un terrain à bâtir à plus grande distance, là où les prix sont pour eux acceptables (CPDT, 2011). Pour chaque seconde d'éloignement par rapport à Bruxelles, le prix diminue de 2,66 euros, ce qui représente 159,60 euros par minute. Ainsi, entre le bien le plus proche de Bruxelles (à 3143 secondes, soit 52 minutes et 20 secondes) et le bien le plus éloigné (à 6649 secondes, soit 1 heure 50 minutes et 50 secondes) s'établit une différence de 9325,85 euros. Le rôle prépondérant de Bruxelles sur les prix des terrains à bâtir a déjà été mis en évidence par le passé (C. Vandermotten *et al.*, 1996 ; F. Goffette-Nagot *et al.*, 2011).

La proximité d'un commerce alimentaire n'a pas d'influence significative. Ceci est peut être révélateur d'un effet de préférence pour un commerce d'une chaîne particulière qui n'est pas nécessairement le plus proche. Ainsi, certains ménages souhaiteront faire leurs courses chez Colruyt plutôt que chez Delhaize, par exemple, pour des raisons liées notamment à leurs finances et à la préférence de certaines marques. La situation est relativement différente de celle des nodules commerciaux qui concentrent généralement une offre très diversifiée de biens de consommation. Le ménage a dès lors davantage de chance de trouver ce qu'il cherche dans le nodule le plus proche.

La proximité d'une gare ne semble pas engendrer d'influence sur les prix. Ceci peut s'expliquer par le fait que l'utilité procurée par le train peut être parasitée par différentes nuisances comme le bruit du trafic ferroviaire ou la vue de l'emprise des voies. Il semble cependant que les ménages tendent à rechercher la proximité d'un arrêt de train qu'ils considèrent comme une utilité (Compagnie des notaires de Liège, 2014). Nous ne pouvons dès lors confirmer ou infirmer l'existence d'une plus-value apportée par la proximité des gares. Ajoutons également que la part modale du train n'est pas significative, ce même résultat ayant été obtenu par Y. Grignet (2011) et P. Dethier (2012) pour le marché des maisons unifamiliales.

La proximité des différents types d'établissements scolaires ne semble pas avoir d'influence sur les prix. Leur forte présence, notamment dans l'agglomération liégeoise, peut expliquer une certaine indifférence des ménages quant aux distances. D'autre part, de nombreux biens concernés par cette proximité sont situés à proximité d'industries, d'où une perturbation de la variable.

La part modale des modes lents ne semble pas être un critère qui soit pris en compte par les acheteurs de terrains à bâtir. Les déplacements à pied ou à vélo restent encore très minoritaires en Wallonie. Ceci est à rapprocher de l'absence d'influence sur les prix engendrée par la proximité d'un RAVeL.

3.2.2.3. Variables d'environnement physique

L'appartenance à l'agglomération opérationnelle telle que définie en 1971 entraîne une moins-value de 6089,74 euros. Ceci semble indiquer des prix moins importants au sein de cette agglomération de tradition industrielle et de bâti ancien, par rapport à sa périphérie et au milieu rural. Néanmoins, comme nous l'avons mis en évidence au cours de l'analyse de l'autocorrélation, cette variable ne semble pas cerner de manière efficace l'effet de perception et de valorisation. Une étude basée sur l'approche par contingence réalisée à Wavre a mis en évidence, pour les maisons situées dans des quartiers anciens, une moins-value de 33 000 euros en considérant un bien de référence de 300 000 euros (J. Strée et J.-M. Halleux, 2013). Ceci vient confirmer la moins-value observée pour les biens situés au sein de l'agglomération opérationnelle de 1971.

La localisation dans un lotissement entraîne une moins-value de 5028,83 euros. Ceci pourrait s'expliquer par le fait que, pour un bien identique, les acheteurs potentiels préfèrent qu'il n'appartienne pas à un lotissement. Ceci est probablement dû aux règles urbanistiques qui peuvent contraindre les demandeurs. Ceux-ci préféreront avoir davantage de liberté dans le choix des matériaux, dans l'agencement de leur future maison et dans la date d'exécution des travaux (Construire facile, 2012 ; Compagnie des notaires de Liège, 2013). Il semble également qu'il existe une offre trop importante de terrains en grands lotissements par rapport à la demande, ceci entraînant une baisse des prix des terrains situés dans ces lotissements (Compagnie des notaires de Liège, 2013).

La localisation dans un cul de sac entraîne une plus-value de 4738,63 euros pouvant s'expliquer par le faible trafic automobile empruntant ces voiries destinées à une circulation exclusivement locale, ce qui rejoint les propos de C. J. Barnett (1985). Ceci est le reflet d'une aspiration pour le calme. De plus, l'analyse des corrélations montre qu'une partie de ces cul-de-sac sont situés dans des secteurs à revenus élevés et à habitat dispersé.

Les biens concernés par un bruit du trafic aérien de plus de 60 dB connaissent une moins-value de 11 912,50 euros. De toutes les nuisances sonores générées, le bruit du trafic aérien semble être le plus mal vécu (J. E. C. Dekkers et J. Willemijn Van Der Straaten, 2009). Ceci est d'autant plus important que l'aéroport de Liège-Bierset accueille un trafic nocturne qui peut être plus dérangeant que le trafic diurne. Ce paramètre est justement pris en compte dans l'établissement du Plan d'Exposition au Bruit de l'aéroport de Bierset (SOWAER, n.d.). L'impact des nuisances est assez fort pour mettre en évidence la moins-value sur base des seuls dix biens soumis à un bruit supérieur à 60 dB. Lors de nos tests de l'influence des différentes variables au début de la modélisation, nous avons constaté que les biens soumis à un bruit inférieur à 55 dB étaient également concernés par une moins-value significative. Cependant, celle-ci était moins importante et la significativité de la variable était moins prononcée.

Une centrale nucléaire est vue comme un élément menaçant et dangereux en cas d'incident et peut donc être responsable d'une moins-value sur les prix des terrains à bâtir. C'est pourquoi chaque mètre d'éloignement de la centrale entraîne une plus-value de 0,57 euro, associé à une p-value très petite. Pour le bien le plus éloigné de la centrale, à 49,003 km, cela représente une plus-value de 27 951,07 euros par rapport au même bien situé sur le site même. Les transformations en variables binaires (seuils à 2000, 5000 et 10000 mètres) sont également significatives, mais de manière moins importante que la distance euclidienne. Il semble ainsi que l'influence de la centrale nucléaire de Tihange s'exerce sur une large superficie.

La présence, à moins de 200 mètres, d'une route ayant un trafic de plus de 20 000 véhicules par jour entraîne une importante moins-value de 21 218,40 euros. Ceci est compréhensible dans la mesure où ces voies sont essentiellement des autoroutes et des routes nationales. Le bruit intense lié à la vitesse des véhicules et à leur grand nombre est réellement dérangeant, ces nuisances se répercutant sur le prix des maisons (I. R. Lake *et al.*, 1998 ; K. S. Kim *et al.*, 2007 ; S.-H. Cho *et al.*, 2008 ; J. E. C. Dekkers et J. Willemijn Van Der Straaten, 2009 ; M. Tavieres *et al.*, 2009 ; I.-H. Lin *et al.*, 2013) et des terrains à bâtir. Les autres variables binaires se montraient elles aussi significatives, mais c'est pour le seuil des 200 mètres que la moins-value est la plus importante et la p-value la plus petite. Ainsi, nous pouvons estimer que l'influence du trafic routier est fortement néfaste jusqu'à 200 mètres, puis commence petit à petit à décroître.

La variable de proximité d'une ligne à haute tension (500 mètres) apparaît significative, avec une moins-value engendrée de 3909,92 euros. Cet élément est donc pris en compte par les ménages lors de l'achat de leur terrain. Les autres seuils ne sont pas significatifs. Ceci est sans doute dû à un manque de biens concernés.

Enfin, la dernière variable d'environnement physique significative a une influence en opposition avec le comportement attendu sur base des études consultées (C. G. Leggett et N. E. Bockstael, 1999 ; G. Ahlfeldt et W. Maennig, 2008). La proximité immédiate d'un cours d'eau (seuil des 100 mètres) induit une moins-value si nous considérons le réseau hydrographique dans sa totalité. La variable similaire développée sur base du réseau hydrographique principal n'est quant à elle pas significative. Les 64 biens concernés connaissent une moins-value moyenne de 6105,10 euros. Nous pensons que d'autres paramètres viennent polluer cette variable. Aussi, l'éventuel avantage procuré par la proximité immédiate d'un cours d'eau est fortement contrebalancé par des facteurs de risque et de localisation. Premièrement, le risque d'inondation que présentent les cours d'eau peut entraîner une baisse des prix. Cependant, la variable binaire

expressément utilisée, la localisation en zone inondable, ne s'est pas révélée significative. Ceci pourrait provenir du fait que les vendeurs et acheteurs ne connaissent pas réellement l'étendue de ces zones inondables et se font donc une idée plus ou moins floue de l'importance des débordements en cas de crue. D'autres facteurs tels que la localisation en fond de vallée réduisant le champ de vision, la présence d'espaces industriels et/ou portuaires pour les biens situés dans l'agglomération de Liège et le trafic fluvial jouent certainement un rôle.

Aucune variables de diversité, de part de bâti et de part d'artificialisation n'est significative, ceci contredisant les résultats de l'étude de J. Geoghegan *et al.* (1997). La diversité des affectations laisse donc les ménages acheteurs indifférents, d'autant plus qu'il peut s'agir d'un élément relativement difficile à juger (Y. Grignet, 2011). Il en va de même pour la part de bâti, qui ne semble pas affecter les prix, constatations similaires à celles mises en évidence par P. Dethier (2012). Le fait que les deux variables de densité et de proportion de maisons quatre façades soient également non significatives abonde dans le sens que les ménages ne sont que peu regardants à la structure du bâti dans le voisinage du bien qu'ils souhaitent acheter. P. Dethier (2012) avait mis en évidence une influence négative de la densité, avec une p-value significative mais relativement élevée (0,04). Cependant, au vu des corrélations modérées existant entre les variables de densités et de proportion de maisons quatre façades d'une part et du revenu d'autre part, nous pensons que l'influence des premières variables est implicitement reprise par la seconde. Une autre explication est qu'au sein des secteurs statistiques, les variations de densité et de proportion de maisons quatre façades sont parfois suffisamment importantes pour avoir une différence marquée entre la valeur du secteur statistique et la situation réellement observée localement. Ceci est notamment vrai pour les grands secteurs statistiques en milieu rural.

Un certain nombre de variables censées tenir compte des nuisances visuelles et sonores ne sont pas significatives, quels que soient les seuils utilisés. Il s'agit de la proximité d'une ligne de chemin de fer, d'une industrie, d'une carrière, d'un site d'activité économique désaffecté et d'une ligne à haute tension. Ceci peut s'expliquer de plusieurs manières :

- un trop faible nombre de biens au plus proche de l'élément pour rendre compte de l'influence ;
- le nombre de biens concernés augmente avec des seuils plus éloignés, mais les valeurs de ces biens sont de moins en moins impactées avec la distance ;
- il peut exister des barrières au bruit et à la vue dans le voisinage du bien : bâtiments, bosquets et dispositifs antibruit peuvent réduire le bruit et limiter la vue. Les nuisances ainsi atténuées ne sont plus suffisamment dérangeantes pour entraîner une influence sur les prix. P. Dethier (2012) a avancé cette explication pour expliquer la différence de moins-value engendrée par la proximité des industries entre le modèle hédonique et l'approche par contingence développés sur les maisons liégeoises. La moins-value mise en évidence par l'approche hédonique est moins importante que celle déterminée sur base de l'approche par contingence. C'est essentiellement la vue sur les industries qui est responsable de la moins-value, vue limitée par des barrières visuelles qui ne peuvent être appréhendées au travers de l'approche hédonique ;
- les représentants d'un type d'élément n'ont sans doute pas tous le même impact. Ceci est notamment le cas pour les industries : une industrie bruyante et polluante n'a pas le même effet qu'une industrie davantage respectueuse de l'environnement. Ceci rejoint les

constats de l'étude de Y. Liu *et al.* (2010). Une ligne de chemin de fer ne voit pas passer les mêmes types et fréquences de trafic qu'une autre.

La proximité d'une forêt, d'un bois ou d'un espace vert ne joue pas de manière significative sur les prix. Ceci est sans doute dû au fait qu'elle apparaît comme un élément intéressant pour certains acheteurs et comme une barrière visuelle pour d'autres. Nous pouvons supposer que les biens les plus proches peuvent être menacés par des chutes de branche ou d'arbres en cas de tempête. De plus, la perte du feuillage en hiver lui fait perdre de manière saisonnière son attrait esthétique et peut paraître dérangeante par le dépôt des feuilles (S.-H. Cho *et al.*, 2008).

3.2.2.4. Variables liées à la valorisation sociale

L'appartenance à la commune de Spa engendre une plus-value de 26 109,50 euros. Comme nous l'avons expliqué auparavant, un effet de valorisation et un sentiment d'appartenance pourrait être responsable de cette importante plus-value.

Le revenu médian par secteur statistique est une autre variable expliquant pour une grande part le prix des terrains à bâtir. Ainsi, une augmentation du revenu de un euro entraîne une augmentation du prix de 1,12 euros. Ceci montre l'aspiration des ménages capables financièrement de faire bâtir leur habitation à vivre dans des quartiers aisés. Ainsi, le prix varie de 29 470,50 euros entre les deux extrêmes de revenus (respectivement 13 091 euros et 39 439 euros). La contribution du revenu au prix des maisons unifamiliales liégeoises est du même ordre de grandeur, avec un coefficient de 3,56 euros (P. Dethier, 2012). Si nous considérons que l'investissement peut être divisé entre l'achat du terrain et la construction de la maison à raison d'un rapport de un tiers – deux tiers, le coefficient obtenu par notre modèle est cohérent.

Le taux de chômage et le taux de diplômés du supérieur ne sont pas significatifs. Il est cependant délicat de conclure au fait que les ménages ne prennent pas ces paramètres en compte. L'ancienneté des données (2001) peut expliquer la non-significativité. Les taux ne reflètent plus la situation actuelle.

Le taux IPP n'a pas d'influence significative sur les prix. Ceci semble indiquer que cette donnée n'a pas d'impact sur les valorisations foncières.

3.3. Modélisation des prix pour la périphérie du bassin d'emploi de Liège en 2013-2014

Notre modèle général rencontrant des difficultés de prédiction au niveau de l'agglomération opérationnelle de Liège, nous choisissons de scinder le bassin d'emploi et d'étudier séparément le comportement de nos différentes variables explicatives dans chacun des deux sous-espaces créés. Il semble en effet que nous soyons en présence de sous-marchés dont l'étude peut être approchée via une segmentation spatiale de l'échantillon (A. Goodman et T. Thibodeau, 1998 ; cités par J. Dubé, 2010). La scission s'est opérée sur base de la variable *AGOP* : nous développons ainsi un modèle pour la périphérie. Nous reprenons les variables *SPA* et *BEBRUX* afin de confirmer ou d'infirmer la présence de sous-marchés en ces lieux. Nous avons également tenté de développer un modèle pour l'agglomération opérationnelle de 1971, sans succès.

3.3.1. Développement du modèle

3.3.1.1. Statistiques descriptives

L'échantillon utilisé pour le développement de ce modèle comporte 270 biens. Il répond aux critères utilisés depuis le début de ce mémoire :

- appartenance aux zones d'habitat ou d'habitat à caractère rural du Plan de Secteur ;
- biens de moins de 175 000 euros ;
- superficies comprises entre 156 et 1600 m².

L'échantillon ayant une taille différente de l'original, il est utile de vérifier la distribution des valeurs de certaines variables, notamment la variable dépendante. Les valeurs de celle-ci suivent une distribution relativement proche de la normale (Figure 14).

L'histogramme des valeurs de superficie présente une distribution continue (Figure 15). Il en va de même pour l'histogramme des valeurs de pente (Figure 16).

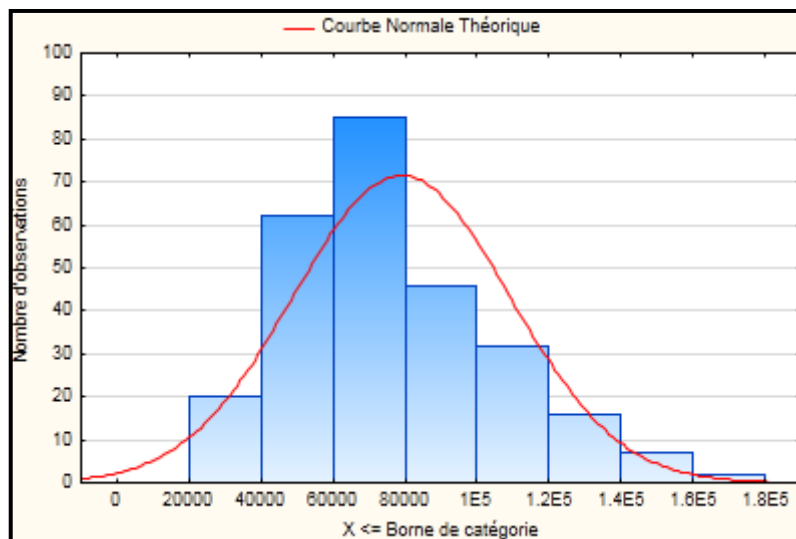


Figure 14 – Distribution des valeurs de prix.

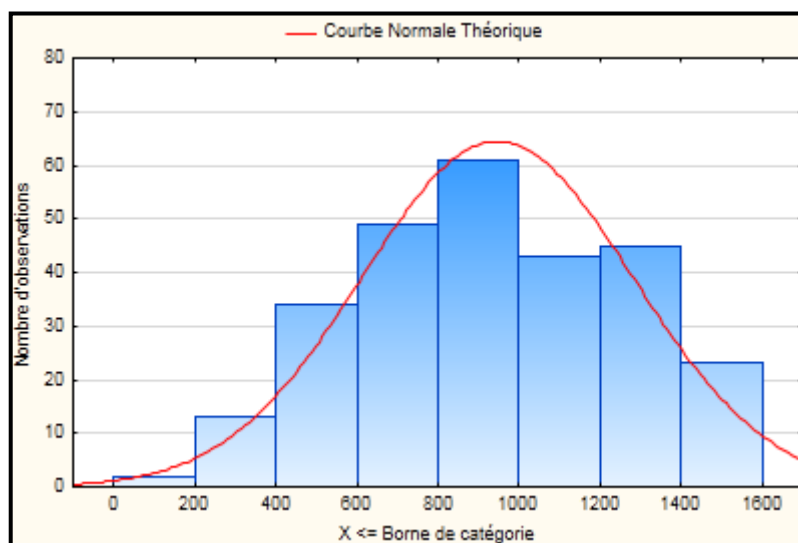


Figure 15 – Distribution des valeurs de superficie.

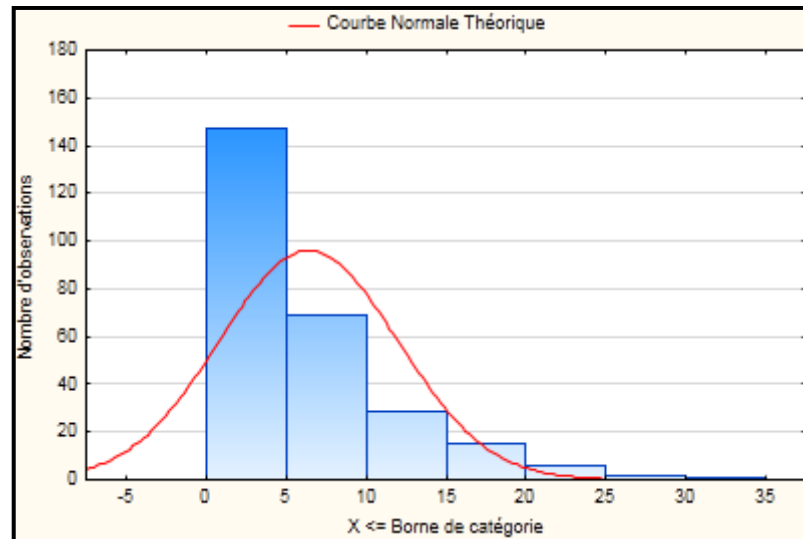


Figure 16 – Distribution des valeurs de pente.

3.3.1.2. Analyse des corrélations entre variables explicatives

L'échantillon étant différent, il est impératif de vérifier à nouveau les corrélations entre les variables explicatives. Le lecteur trouvera, à l'annexe XIV, la liste des couples de variables explicatives concernés par des corrélations supérieures à 0,75 en valeur absolue. Nous avons choisi, sur base des corrélations, de retenir les mêmes variables que celles qui avaient été choisies pour l'élaboration du modèle général.

3.3.1.3. Modélisation par MCO

De la même manière que celle ayant servi au développement du modèle général, nous avons effectué une série de tests afin de conserver la combinaison de variables qui nous semblait la plus optimale sur base de l'observation des différents paramètres fournis par Gretl. Une fois que cette combinaison a été trouvée, nous avons procédé à l'épuration du modèle par suppression des variables non significatives, et ce en deux fois selon les seuils de la p-value (0,20 puis 0,10). Le modèle ainsi épuré présente les paramètres suivants (Tableau 9) :

Moy. Var. dep	79025,92	Ec. typ. var. dep.	30125,80
Somme carré résidus	5,38e+10	Ec. typ. de régression	14695,21
R ²	0,779747	R ² ajusté	0,762056
F (20,249)	44,07593	p. critique (F)	4,79e-70
Log de vraisemblance	-2962,907	Critère d'Akaike	5967,815
Critère de Schwarz	6043,382	Hannan-Quinn	5998,159

Tableau 9 – Paramètres du modèle retenu développé par MCO.

Le modèle développé par MCO présente des caractéristiques intéressantes. Le pouvoir explicatif est déjà relativement élevé avec un R^2 ajusté de 0,762, soit environ dix points au-dessus de celui du modèle général à ce même stade. La robustesse statistique est également mieux marquée, avec un F de 44,08, légèrement supérieur à celui du modèle général.

Le test de colinéarité (annexe XV) ne montre pas de coefficients inquiétants. C'est toujours le temps d'accès à Liège qui possède le VIF le plus élevé, avec une valeur de 2,85. Il n'y a donc pas de problème de colinéarité excessive entre les variables indépendantes du modèle.

L'étude de la présence d'hétéroscédasticité via le test de Breusch-Pagan (annexe XVI) montre que deux variables explicatives sont concernées par ce problème de manière significative, mais toutefois pas très importante :

- la superficie ;
- la part modale du bus.

Le graphique de distribution des résidus en fonction de la superficie montre effectivement une morphologie en cône s'évasant vers la droite, surtout du côté positif (Figure 17). Cette forme ressort par ailleurs nettement mieux par rapport au modèle général.

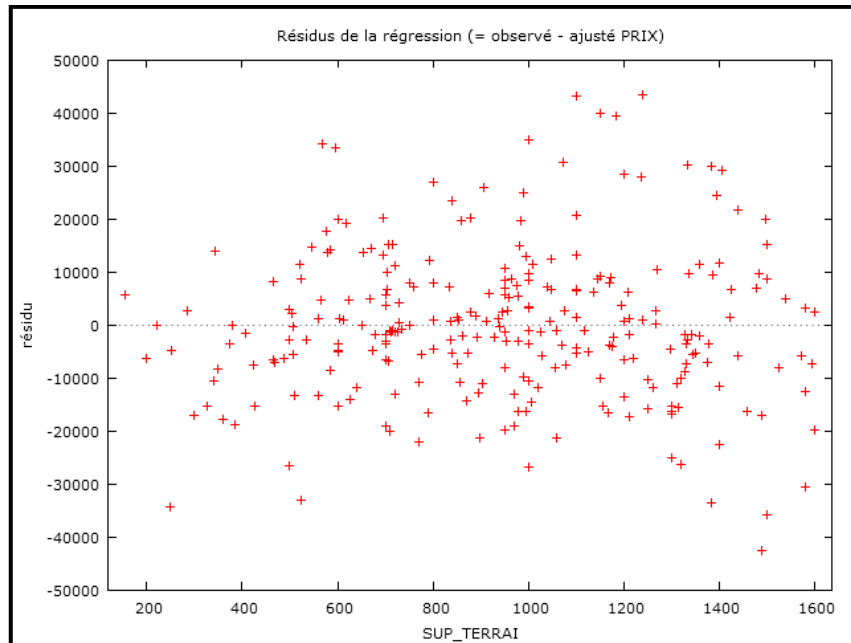


Figure 17 – Distribution des résidus en fonction de la superficie.

Nous devons dès lors construire un nouveau modèle en corrigeant cette hétéroscédasticité.

3.3.1.4. Correction de l'hétéroscédasticité

Nous effectuons donc une nouvelle régression de type MCG. Le modèle ainsi construit présente les paramètres suivants (Tableau 10) :

Moy. Var. dep	79025,92	Ec. typ. var. dep.	30125,80
Somme carré résidus	5,98e+10	Ec. typ. de régression	15497,15
R ²	0,899616	R ² ajusté	0,891554
F (20,249)	111,5743	p. critique (F)	5,6e-112
Log de vraisemblance	-574,1596	Critère d'Akaike	1190,319
Critère de Schwarz	1265,886	Hannan-Quinn	1220,664

Tableau 10 – Paramètres du modèle corrigeant l'hétéroscédasticité.

Les paramètres montrent une importante progression des performances du modèle. Le F montre une très grande robustesse, la p -value associée étant de plus très proche de 0. Le pouvoir explicatif est également bien meilleur, en progression de treize points par rapport au modèle MCO, à 0,892. Les critères de Schwarz et d'Akaike ainsi que l'indice Hannan-Quinn confirment la très nette amélioration du modèle.

L'analyse du graphique de la répartition des résidus en fonction de la superficie montre une distribution moins conique des occurrences, davantage groupées autour de la droite horizontale de valeur nulle (Figure 18). La correction de l'hétéroscédasticité a rendu la variable « part modale du bus » non-significative. C'est pourquoi nous l'avons retirée du modèle.

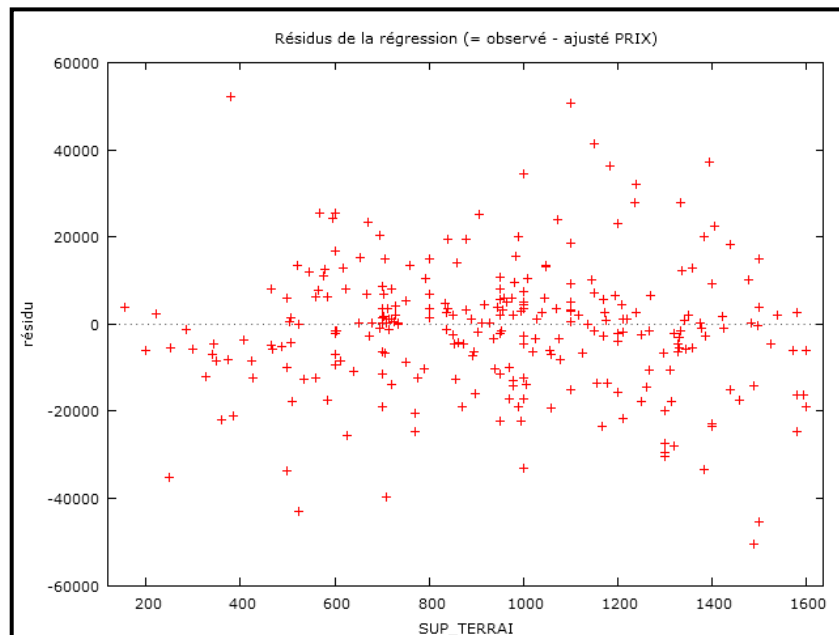


Figure 18 – Distribution des résidus en fonction de la superficie après correction de l'hétéroscédasticité.

Nous procédons à une nouvelle vérification de l'absence de colinéarité (annexe XV). Tel est le cas. Le temps d'accès à Liège conserve la plus haute valeur de VIF, avec 2,60.

3.3.1.5. Suppression des observations à résidus extrêmes et ajustement du modèle

Gretl identifie sept résidus supérieurs à 2,5 fois l'erreur standard. Contrairement aux précédents développements, il ne s'agit plus cette fois de résidus associés à des prix élevés pour une grande part. Nous procédons donc à une vérification de la localisation de ces résidus, comme conseillé par F. Des Rosiers (2001). Il apparaît que plusieurs biens ont des prix plus faibles par rapport aux valeurs prédites. Ils sont en effet situés dans un cadre calme, vert et agréable, mais d'autres facteurs viennent sans doute diminuer les prix. Un des terrains est par exemple en surplomb d'un petit chemin, nécessitant des travaux de terrassement. Un autre est situé au pied d'un viaduc de l'E42 près de Herve. Trois terrains ont des prix élevés : l'un d'entre eux est entouré de maisons luxueuses, entraînant de facto une augmentation de sa valeur. Un autre est entouré de constructions récentes dans une petite rue champêtre, tandis que le dernier, étant contre la ligne de chemin de fer Bruxelles – Liège, a sans doute une autre vocation que celle de recevoir une maison. Ce dernier terrain est une observation aberrante au sein du marché des terrains à bâtir.

Nous supprimons donc les observations liées à ces résidus extrêmes et réajustons notre modèle sur base de l'échantillon épuré. Les paramètres sont présentés ci-dessous (Tableau 11).

Moy. Var. dep	78555,32	Ec. typ. var. dep.	29772,31
Somme carré résidus	4,50e+10	Ec. typ. de régression	13615,68
R ²	0,922927	R ² ajusté	0,916901
F (19,243)	153,1510	p. critique (F)	1,8e-123
Log de vraisemblance	-543,8919	Critère d'Akaike	1127,784
Critère de Schwarz	1199,227	Hannan-Quinn	1156,495

Tableau 11 – Paramètres du modèle corrigeant l'hétéroscédasticité après suppression des résidus délinquants.

Sur base de ces paramètres, le modèle ainsi développé peut être qualifié de très bon. Ses performances prédictives et explicatives se sont accrues par rapport au modèle général. L'erreur type d'estimation calculée sur base des résidus s'élève à 13 137,79 euros, soit une erreur relative de 16,72 %. C'est une amélioration de 3,38 % par rapport au modèle général (20,10 %). Ces observations rejoignent celles de A. Goodman et T. Thibodeau (2003), à savoir qu'une segmentation augmente le pouvoir prédictif. Le R^2 ajusté s'élève à 0,917, soit une augmentation de dix points par rapport au modèle général. Le F de Fischer démontre une très grande robustesse statistique, avec une valeur de 153,15.

Malgré que la morphologie de la distribution des résidus ne prenne pas la forme d'une courbe de Gauss, le test du Chi carré indique que la distribution est normale, l'hypothèse nulle ne pouvant être rejetée par la p-value associée au test, cette dernière étant supérieure à la valeur critique de 0,05 (Figure 19). La majeure partie des occurrences est concentrée autour de la valeur nulle centrale.

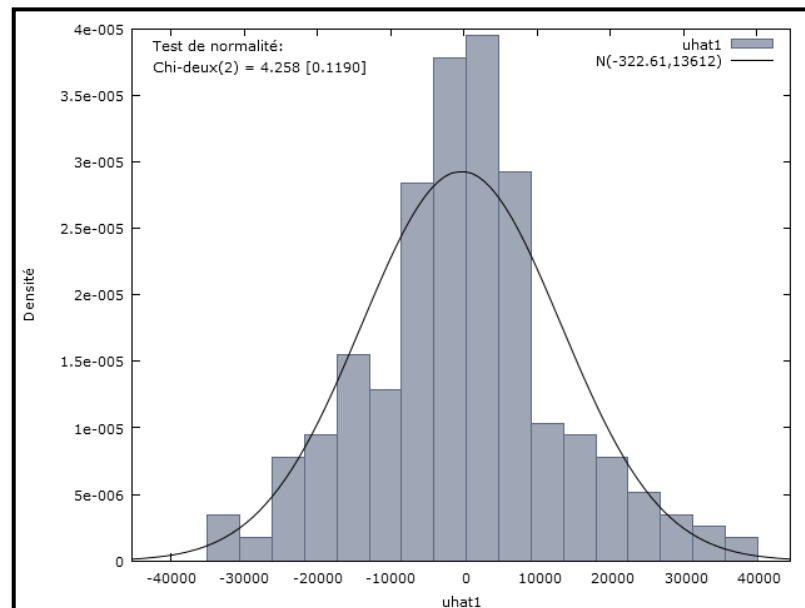


Figure 19 – Distribution des résidus après épuraton.

L'analyse du diagramme quantile-quantile des résidus apporte cependant une nuance au constat précédent (Figure 20). Ainsi, les résidus positifs s'éloignent quelque peu de la droite de Henry des résidus. L'ajustement du modèle n'est donc pas parfait pour les valeurs positives de résidus, mais étant donné le respect de la distribution normale mise en évidence par le test de Chi carré, nous pouvons considérer cet ajustement comme étant acceptable. Pour les résidus les plus extrêmes, nous retrouvons des comportements similaires à ceux mis en évidence lors du développement du modèle général : en immobilier, la modélisation n'est pas optimale pour ces valeurs extrêmes hautes de résidus (F. Des Rosiers, 2001). Ce mauvais ajustement pour les valeurs extrêmes hautes

se confirme via l'analyse visuelle du graphique de distribution des résidus en fonction du prix (Figure 21). Au-delà de 110 000 euros, les performances du modèle déclinent, pour pratiquement devenir nulles pour les biens ayant des prix supérieurs à 130 000 euros. Cette constatation est similaire à celles qui avaient été mises en évidence pour les modèles généraux.

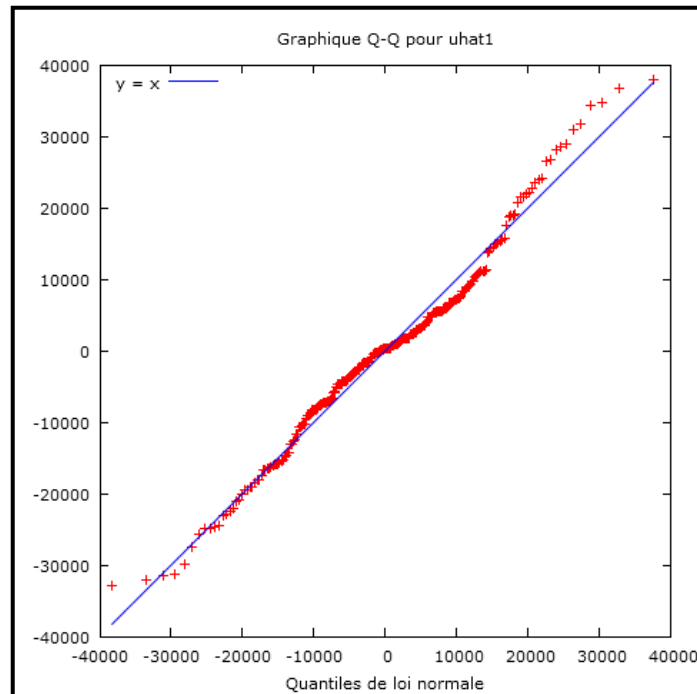


Figure 20 – Diagramme quantile-quantile des résidus.

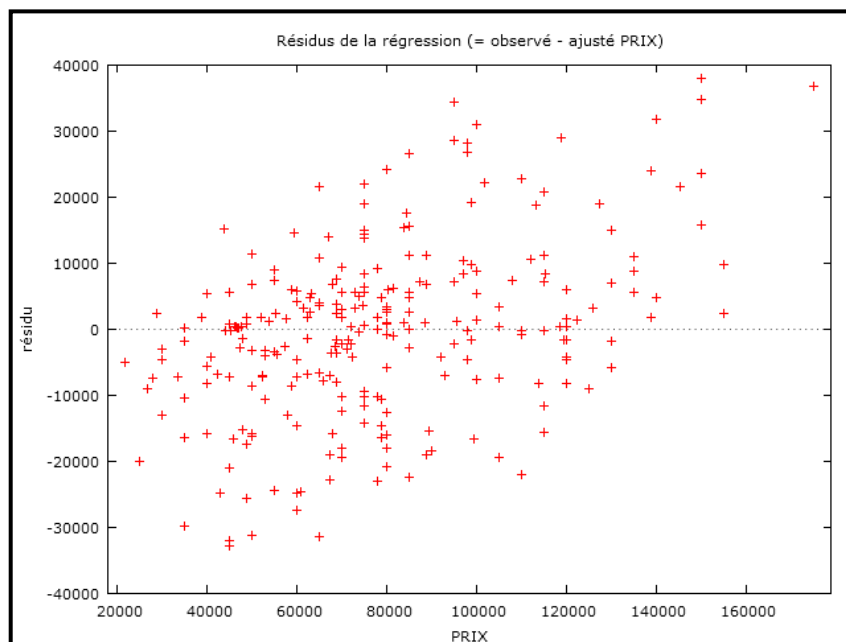


Figure 21 – Distribution des résidus en fonction du prix.

3.3.1.6. Analyse de l'autocorrélation spatiale des résidus

L'indice de Moran est proche de 0 (valeur exacte : 0,015194), et non significativement différent (p-value : 0,207000) de l'indice attendu (-00038). Ceci permet d'écarter la présence d'autocorrélation spatiale entre les résidus. Le diagramme de dispersion de Moran est consultable à l'annexe XVII. Notre modèle prend en compte suffisamment de variables pour représenter le marché des terrains à bâtir en périphérie de l'agglomération opérationnelle de Liège.

Au niveau local, nous voyons apparaître deux concentrations d'environnement haut (Figure 22). Pour rappel, de telles valeurs indiquent que le modèle sous-estime les prix en ces points. La première, dans l'ouest du bassin d'emploi, avait déjà été mise en évidence. Une explication, à savoir l'influence bruxelloise et les prix fonciers élevés de son bassin d'emploi, lui avait déjà été appliquée. La deuxième s'affirme au sud-ouest de Liège. Elle pourrait indiquer que les communes concernées – Neupré, Anthisnes, Nandrin – profitent d'un effet de perception positif, faisant de cette zone de périurbanisation liégeoise un secteur recherché pour l'achat de terrains à bâtir (Compagnie des notaires de Liège, 2012), et où les prix sont dès lors plus élevés que la prévision du modèle.

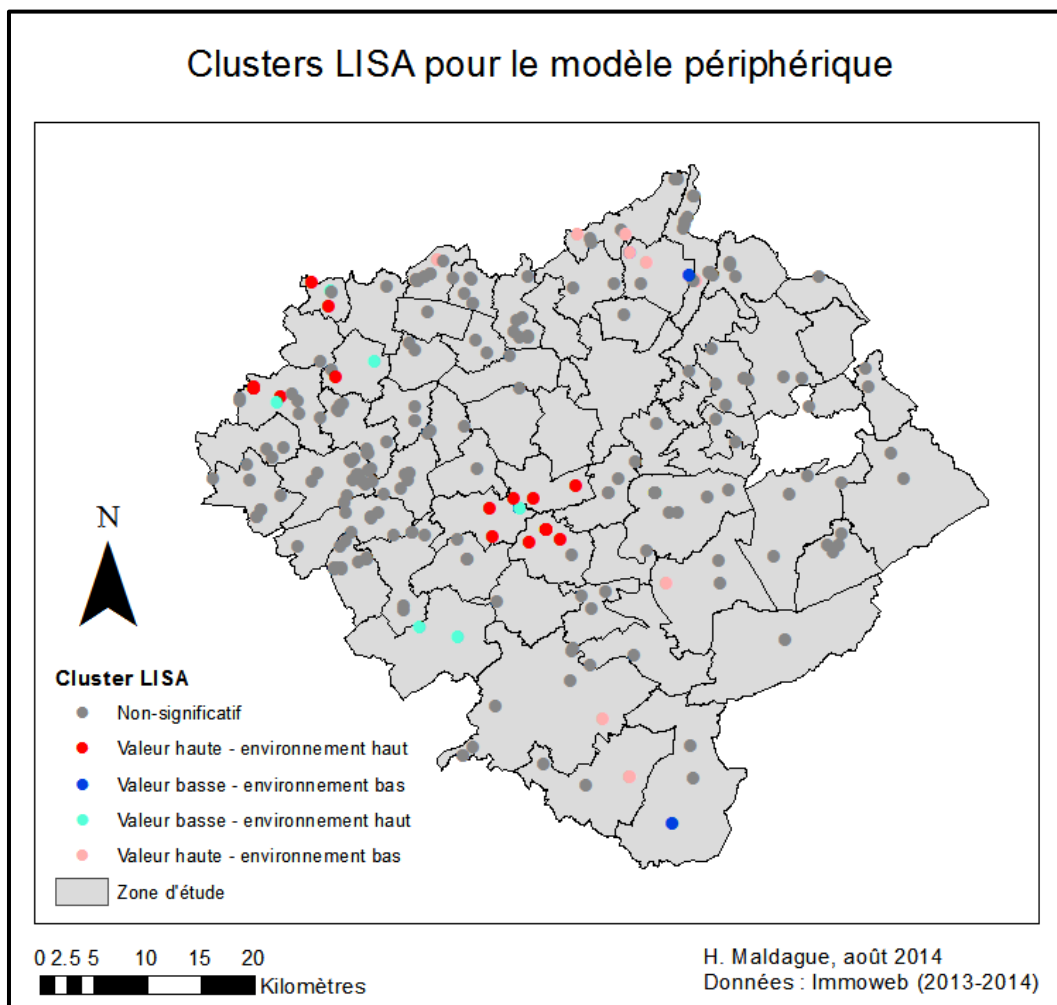


Figure 22 – Clusters LISA pour le modèle périphérique.

3.3.2. Présentation des résultats

Le modèle périphérique, à hétéroscédasticité corrigée et à résidus délinquants supprimés, est présenté ci-dessous (Tableau 12).

	<i>Coefficient</i>	<i>Erreur Std</i>	<i>t de Student</i>	<i>p. critique</i>	
Constante	44380,4	8186,24	5,4213	<0,00001	***
Appartenance à la commune de Spa	20548,8	6704,69	3,0648	0,00242	***
Superficie	52,4875	2,09271	25,0811	<0,00001	***
Localisation dans un cul-de-sac	5848,49	1997,05	2,9286	0,00373	***
Localisation dans un lotissement	-6600,15	1428,43	-4,6206	<0,00001	***
Bruit aérien > 60 dB	-9836,7	2758,35	-3,5662	0,00044	***
Revenus médians	1,55375	0,166623	9,3249	<0,00001	***
Cours d'eau à moins de 100 m	-4904,93	1611,61	-3,0435	0,00260	***
Distance à Tihange	0,8976	0,077538	11,5763	<0,00001	***
Commerce alimentaire à moins de 500 m	10304,6	3015,8	3,4169	0,00074	***
Distance aux arrêts de bus	4,78982	1,14642	4,1781	0,00004	***
Chemin de fer à moins de 700 m	-6037,77	2403,13	-2,5125	0,01264	**
Sortie d'autoroute à moins de 700 m	-12117,4	4903,07	-2,4714	0,01415	**
Temps d'accès à Liège	-13,3372	1,52401	-8,7514	<0,00001	***
Temps d'accès aux nodules commerciaux	-28,09	4,62194	-6,0775	<0,00001	***
Ecole secondaire à moins de 120 s	18621,2	5095,28	3,6546	0,00032	***
Etablissement supérieur à moins de 5 min	17537,4	3644,85	4,8115	<0,00001	***
Temps d'accès à Bruxelles	-7,67625	1,13609	-6,7567	<0,00001	***
Part modale des modes lents	-125154	28698,9	-4,3609	0,00002	***
Gare L à moins de 500 m	11835,7	4848,8	2,4409	0,01536	**
* <i>p</i> -value inférieure à 0,10 ** <i>p</i> -value inférieure à 0,05 *** <i>p</i> -value inférieure à 0,01					

Tableau 12 – Résultats du modèle périphérique liégeois.

Le tableau ci-dessous (Tableau 13) reprend les variables explicatives classées selon l'importance de leur coefficient standardisé, en valeur absolue.

Variable	B*	B
Superficie	0,58462	52,4875
Distance à Tihange	0,37520	0,8976
Temps d'accès à Liège	-0,25022	-13,3372
Revenu médian	0,21931	1,55375
Temps d'accès à Bruxelles	-0,20435	-7,67625
Temps d'accès aux nodules commerciaux	0,17671	-28,09
Part modale des modes lents	-0,13866	-125154
Ecole secondaire à moins de 120 s	0,13077	18621,2
Etablissement supérieur à moins de 5 min	0,10729	17537,4
Localisation dans un lotissement	0,10068	-6600,15
Commerces alimentaire à moins de 500 m	0,08978	10304,6
Appartenance à la commune de Spa	0,08463	20548,8
Chemin de fer à moins de 700 m	-0,07911	-6037,77
Gare L à moins de 500 m	0,07618	11835,7
Localisation dans un cul-de-sac	0,07477	5848,49
Cours d'eau à moins de 100 m	0,06801	-4904,93
Distance aux arrêts de bus	0,06144	4,78982
Sortie d'autoroute à moins de 700 m	0,06089	-12117,4
Bruit aérien > 60 dB	0,04521	-9836,7

Tableau 13 – Classement des variables explicatives par ordre d'importance.

A nouveau, la superficie reste la variable la plus explicative, suivie des variables d'accessibilité, du revenu et de la distance à la centrale nucléaire de Tihange. Les autres variables n'expliquent qu'une faible part des prix.

3.3.2.1. Variables structurelles

La superficie est, sans surprise, très fortement significative. Le prix du mètre carré, dans le cas présent (52,49 euros) est légèrement inférieur à celui du modèle général (54,16 euros), toutes choses égales par ailleurs.

3.3.2.2. Variables d'accessibilité

La proximité d'un commerce alimentaire est significative, contrairement au modèle général. C'est au seuil des 500 mètres que la plus-value s'exprime le mieux (10 304,60 euros). Ceci peut paraître important, mais la signification de la variable est à relativiser. En périphérie, ces commerces sont pour la plupart situés en cœur de village ou de ville, et souvent proches d'autres types de commerces et de services. Le fait que leur répartition soit moins dense qu'au sein de l'agglomération opérationnelle amène les ménages à considérer la proximité de ces commerces lors de l'achat d'un terrain.

Se trouver à moins de deux minutes en voiture d'une école secondaire et à moins de cinq minutes en voiture d'un établissement du supérieur entraîne une plus-value de 18 621,20 euros pour le premier cas et de 17 537,40 euros pour le deuxième cas. Un raisonnement similaire à celui des commerces alimentaires peut être appliqué ici : en périphérie, la moindre concentration des établissements scolaires amène sans doute les ménages à considérer leur proximité lors de l'achat d'un terrain. De plus, l'élimination de l'agglomération opérationnelle a peut-être fait ressortir une influence de la proximité du campus du Sart-Tilman. Ces résultats se rapprochent des études de C. J. Barnett (1985), R. Paris (2003) et A. Donzel (2008).

Nous pouvons raccrocher ces constatations aux observations faites par J. Strée et J.-M. Halleux (2013) via l'approche par contingence du marché des maisons au niveau de la commune de Wavre : l'éloignement des commerces et d'autres services (parmi lesquels les écoles primaires et secondaires) entraîne une moins-value d'environ 50 000 euros pour une maison de 330 000 euros.

La distance euclidienne aux arrêts de bus est toujours aussi significative. Le coefficient (3,65 euros par mètre d'éloignement) est inférieur à celui calculé par le modèle général (5,20 euros). Si nous reprenons l'argument des nuisances et de la densité, il est cohérent d'obtenir un coefficient réduit en périphérie, où la densité de population et les nuisances sont moins importantes. La part modale du bus reste non significative. Ceci semble confirmer la faible influence d'une bonne desserte en bus sur le prix des terrains.

Le temps d'accès à Liège est, comme dans les deux modèles généraux, très significatif. Le coefficient indique que chaque seconde d'éloignement entraîne une baisse de 13,34 euros. Ceci est comparable au coefficient calculé par le modèle général (-14,55 euros).

Il en va de même pour le temps d'accès aux nodules commerciaux. Chaque seconde d'éloignement entraîne une baisse de 28,09 euros, très similaire au coefficient du modèle général (-30,53 euros).

Le temps d'accès à Bruxelles est lui aussi toujours significatif, avec une baisse de 7,68 euros par seconde. Ce coefficient est plus prononcé que le paramètre calculé par le modèle général (-2,66 euros). Cette différence est relativement mal comprise.

La part modale des modes lents entraîne une moins-value qui équivaut à 1251,54 euros par pourcent supplémentaire. Nous pensons que la densité de population, utilisée dans le calcul de la part modale, est responsable de la valeur du coefficient observée.

La présence d'une gare desservie par train L à moins de 500 mètres engendre une plus-value de 11 835,70 euros. Les différentes variables binaires de proximité d'une gare, tout type de train confondu, ne sont pas significatives. Il en va de même pour les variables de proximité d'une gare IC/IR. Il est possible que l'éventuelle utilité procurée par les gares IC/IR soit masquée par une influence de l'environnement urbain dans lequel elles se trouvent pour la plupart (Huy, Waremme, Visé). A l'inverse, les gares L se localisent plutôt dans des villages ruraux, là où leur utilité peut s'exprimer dans le modèle hédonique. Ceci vient remettre en question la constatation faite dans le modèle général. La présence du train pourrait apparaître comme une utilité. Cependant, la part modale du train n'est pas significative, semblant montrer que la fréquence de desserte n'est pas réellement importante aux yeux des ménages.

3.3.2.3. Variables d'environnement physique

La localisation dans un lotissement exerce toujours une influence très significative sur les prix. Son coefficient est similaire à celui calculé pour le modèle général (-5028,83 euros). Outre le fait que nous puissions confirmer que l'appartenance à un lotissement entraîne une moins-value, nous pouvons de plus aborder le fait qu'il n'existe pas de réelles différences entre la périphérie et le bassin d'emploi pris dans sa totalité. Nous pouvons donc supposer que la situation est comparable pour les lotissements localisés à l'intérieur de l'agglomération opérationnelle, chose que nous n'avons pas pu vérifier étant donné le trop faible nombre de biens.

La localisation dans un cul-de-sac entraîne une plus-value significative. Le coefficient est comparable à celui du modèle général (-4738,63 euros). La situation entre la périphérie et le bassin d'emploi pris dans sa totalité est comparable.

Les biens soumis à un bruit de trafic aérien plus important que 60 dB sont toujours concernés par une moins-value significative. Le coefficient est un peu moins important que celui mis en évidence par le modèle général (-11 912,50 euros). Ceci provient du fait qu'une partie des biens proches de l'aéroport est située au sein de l'agglomération opérationnelle : ceux-ci ne sont donc pas repris au sein de l'échantillon. La moins-value s'exerce donc sur quelques biens plus éloignés de l'aéroport.

La proximité d'un cours d'eau semble toujours exercer une moins-value sur les terrains à bâtir situés à moins de 100 mètres de ceux-ci. Celle-ci se chiffre à -4904,93 euros, soit une valeur inférieure à celle déterminée pour le modèle général (-6105 euros). Ceci peut s'expliquer par une moindre présence d'industries le long des cours d'eau de la périphérie.

La distance euclidienne à la centrale de Tihange est également significative, comme dans le cas des deux modèles généraux. La plus-value (0,89 euro par mètre d'éloignement) est plus élevée que celle calculée par le modèle général (0,57 euro). La variable a sans doute été débarrassée d'autres influences inhérentes à l'agglomération opérationnelle.

Le modèle périphérique rend la variable de proximité des lignes de chemin de fer significative au seuil des 700 mètres, en engendrant une moins-value de 6037,77 euros. En périphérie, milieu en général plus calme, les nuisances engendrées par le trafic ferroviaire ressortent mieux en l'absence d'autres sources de bruit. De plus, une moins grande densité de constructions diminue l'effet barrière, et le bruit du trafic ferroviaire se répand plus loin. Ainsi, le bruit généré par les trains sera plus dérangeant en périphérie qu'en agglomération où il se fond davantage dans le bruit de la ville, élément qui semble être mis en évidence lorsque le bassin d'emploi est considéré dans sa totalité.

La proximité des routes de plus de 10 000 ou de 20 000 véhicules par jour n'est pas significative, ceci sans doute par manque de biens concernés, la plupart d'entre eux étant situés au sein de l'agglomération. Nous avons réintégré la variable de proximité des sorties d'autoroute. Le seuil des 700 mètres s'est révélé être le plus significatif, il concerne six biens. Ils sont affectés par une moins-value de 12 117,40 euros. Les nuisances engendrées par le trafic routier semblent donc prendre le dessus sur l'utilité procurée par la présence d'une sortie d'autoroute.

3.3.2.4. Variables de valorisation sociale

La variable binaire d'appartenance à Spa est à nouveau significative. Son coefficient est cependant plus faible par rapport au modèle général (20 548,80 euros contre 26 109,50), cet abaissement étant sans doute lié à des réajustements statistiques. L'explication donnée pour le modèle général, à savoir une offre réduite et une demande soutenue liée à un sentiment d'appartenance à la commune, est dès lors toujours valable.

Le revenu médian est à nouveau très significatif. Le coefficient (1,55 euros) est un peu plus élevé que celui du modèle général (1,12 euros). La contribution du revenu semble donc plus importante en périphérie.

3.4. Conclusion sur les modèles liégeois

L'ensemble des analyses effectuées semble montrer qu'il existe, pour certaines caractéristiques, deux marchés distincts : l'agglomération opérationnelle industrielle de 1971 et la périphérie. C'est notamment le cas pour les commerces alimentaires et les établissements scolaires, les sorties d'autoroute mais aussi le temps d'accès en voiture à Bruxelles. A l'inverse, d'autres caractéristiques sont pratiquement insensibles à la zone d'étude considérée, bassin d'emploi en entier ou périphérie. Néanmoins, nous nous limiterons à une hypothèse. Nous n'avons pas pu développer un modèle robuste pour l'agglomération opérationnelle afin de confirmer ces propos.

Il est ainsi possible de mener une comparaison entre certaines variables présentant une certaine robustesse : il s'agit de la superficie, du revenu, la localisation dans un cul-de-sac ou dans un lotissement, certaines variables de temps d'accès et de distance ainsi que les nuisances du trafic aérien et de la circulation automobile intense. A l'inverse, l'impact de certaines variables d'accessibilité (distance aux arrêts de bus, présence d'une gare) et de nombreuses variables d'environnement physique est plus difficile à discuter, compte tenu de l'influence d'autres facteurs difficilement distinguables. C'est notamment le cas de certaines nuisances visuelles et sonores qui peuvent être diminuées par des barrières, mais aussi de l'effet de milieux urbains denses, anciens ou industriels qui peuvent masquer l'utilité ou l'influence propre de l'objet sur lequel sont définies une ou plusieurs variables.

Quoiqu'il en soit, il existe bel et bien un élément perturbateur au sein de l'agglomération opérationnelle, élément que nous avons attribué à un effet de perception et de valorisation. Le modèle développé sur la périphérie de cette agglomération semble bien plus homogène et surtout plus explicatif et plus robuste statistiquement.

4. Modélisation du prix des terrains à bâtir dans le bassin d'emploi de Charleroi

Dans le cadre de ce mémoire, nous souhaitons effectuer la comparaison entre deux modèles hédoniques développés sur des bassins d'emploi, et ce afin de tester si l'application d'un modèle hédonique peut également être menée sur une zone d'étude différente du bassin d'emploi de Liège. Les mémoires de Y. Grignet (2011), C. Delbar (2012) et P. Dethier (2012) ont montré qu'il était possible de modéliser le prix des maisons unifamiliales dans deux zones d'études différentes, à savoir la région urbaine de Liège et la région urbaine de Bruxelles. Une de nos hypothèses, à savoir qu'il est possible d'appliquer l'approche hédonique sur les terrains à bâtir de différentes zones d'étude, est vérifiée dans le présent chapitre.

Nous développons un modèle visant à établir la fonction hédonique liant le prix des terrains à bâtir et leurs caractéristiques dans le bassin d'emploi de Charleroi tel que défini par l'IWEPS. La liste des communes concernées est disponible à l'annexe II. Le bassin d'emploi carolorégien est moins étendu que son homologue liégeois. Il est plus fortement polarisé par Bruxelles (CPDT, 2011) : au nord, sa limite se confond avec la frontière de la province du Hainaut. A l'opposé, son extension est bien plus importante au sud. Celle-ci s'étend à cheval sur l'est de la province du Hainaut et l'ouest de la province de Namur. Cette plus faible superficie du bassin d'emploi est en partie responsable du plus faible nombre de biens collectés, en plus d'une offre assez limitée (Chambre des Notaires de la province du Hainaut, 2010). En effet, seulement 293 biens ont pu être correctement localisés.

4.1. Sélection des observations

Nous avons procédé à une segmentation identique à celle utilisée pour le développement des modèles liégeois. Nous avons pris en compte des biens localisés dans les zones d'habitat et d'habitat à caractère rural du Plan de Secteur et retiré les biens présentant une superficie supérieure à 1600 m². L'échantillon de départ, à l'instar de celui du bassin d'emploi de Liège, était également constitué de terrains à bâtir et de gisements fonciers (Tableau 14).

	Nombre	Moyenne	Ecart-Type	Minimum	Maximum
Superficie	293	1062,433	618,4013	40,000	3434,000

Tableau 14 – Statistiques descriptives appliquées aux superficies de l'ensemble de l'échantillon carolorégien.

Le choix d'une limite supérieure de superficie identique à celle appliquée au bassin d'emploi de Liège s'explique d'une part par une volonté de comparaison identique avec la situation liégeoise, mais aussi en raison de l'apparition prépondérante des gisements de terrains au-delà de la limite des 1600 m². L'existence des deux marchés a également été mise en évidence sur base des résidus de la superficie lors du développement d'un modèle portant sur les superficies comprises entre 0 et 2400 m² (Figure 23).

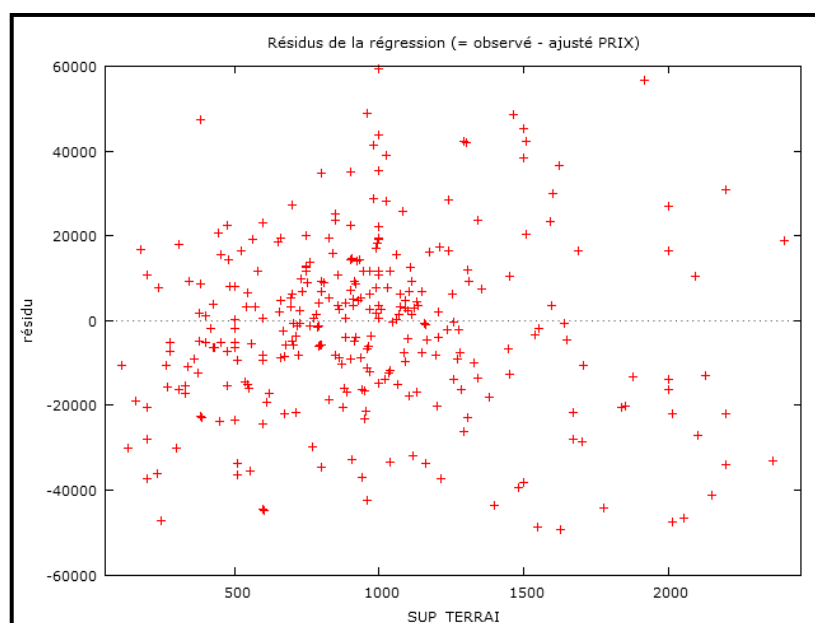


Figure 23 – Résidus en fonction de la superficie issus du modèle 0 – 2400 m².

4.2. Modélisation des prix pour l'ensemble du bassin d'emploi de Charleroi en 2013-2014

4.2.1. Développement du modèle

4.2.1.1. Statistiques descriptives

Sur base des statistiques descriptives, nous mettons en évidence un terrain à bâtir présentant une superficie de 40 m². Cette surface nous paraissant trop petite, nous choisissons de retirer l'observation concernée. De même, à l'instar de Liège, nous retirons tous les biens présentant des prix supérieurs à 175 000 euros. Il s'agit d'observations extrêmes, détachées de la distribution continue des valeurs de prix. Cette opération permet de plus d'obtenir une distribution normale des valeurs de la variable dépendante. La figure 24 illustre la répartition des 223 biens retenus au sein de la zone d'étude. Une plus forte concentration s'opère au sein de l'agglomération de Charleroi, tandis que le sud du bassin d'emploi ne compte que quelques observations. Cet échantillon présente les caractéristiques suivantes :

- prix compris entre 10 000 et 170 000 euros ;
- superficies comprises entre 113 et 1600 m² ;
- biens situés dans les zones d'habitat et d'habitat à caractère rural, et donc juridiquement urbanisables.

Variables	Minimum	Maximum	Moyenne	Ecart-type
Prix (euros)	10000	170000	72274,76	33567,47
Superficie (m ²)	113	1600	846,91	339,30

Tableau 15 – Statistiques descriptives pour le prix et la superficie.

A l'instar des modèles développés pour le bassin d'emploi de Liège, il est impératif de contrôler la distribution des valeurs de la variable dépendante. Celles-ci sont distribuées de manière proche de la loi normale (Figure 25).

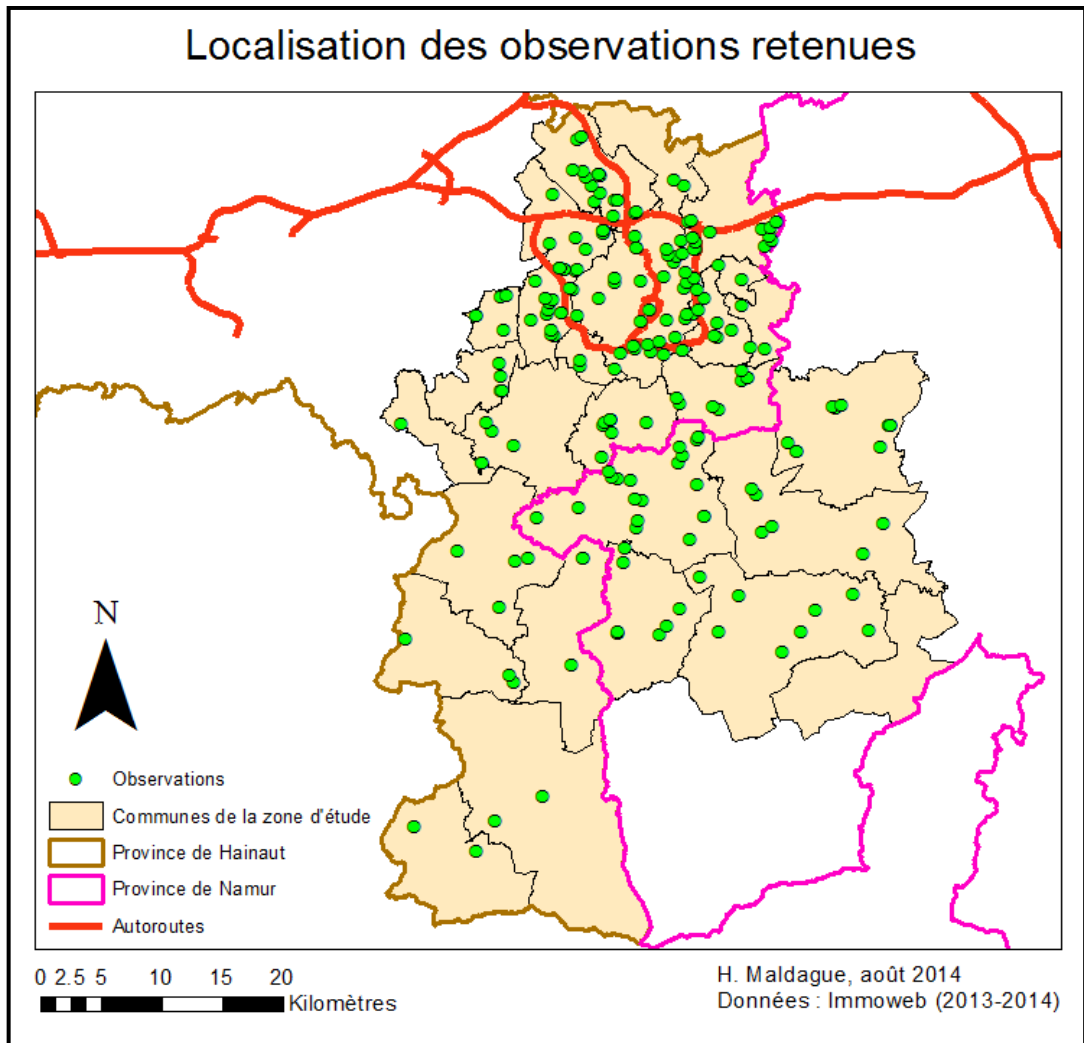


Figure 24 – Répartition des observations retenues au sein de la zone d'étude

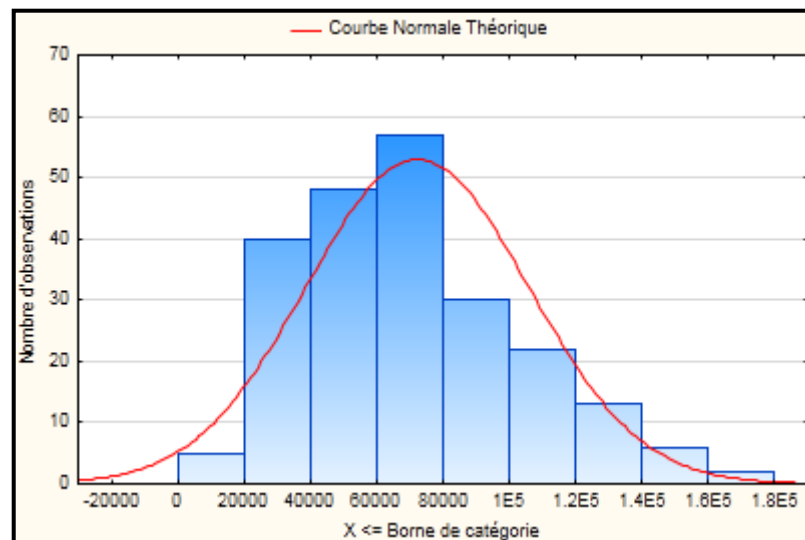


Figure 25 – Distribution des valeurs de prix.

Les variables structurelles quantitatives, la superficie et la pente, présentent des histogrammes continus (Figures 26 et 27).

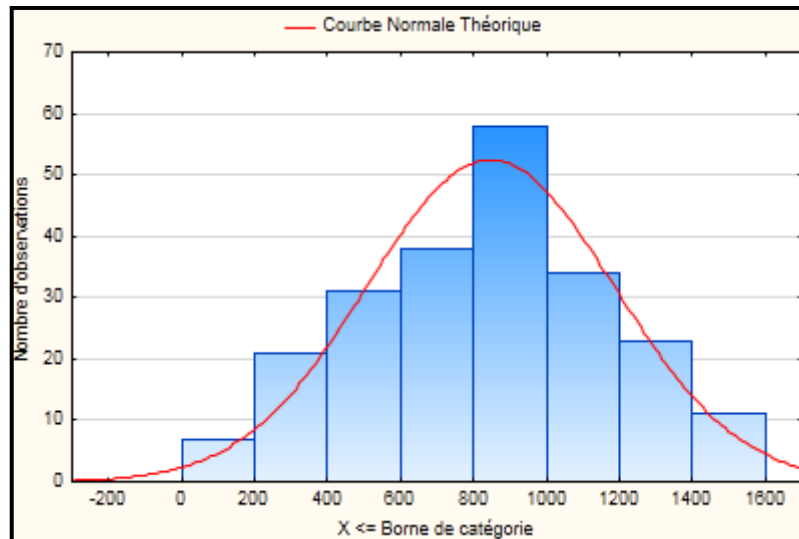


Figure 26 – Distribution des valeurs de superficie.

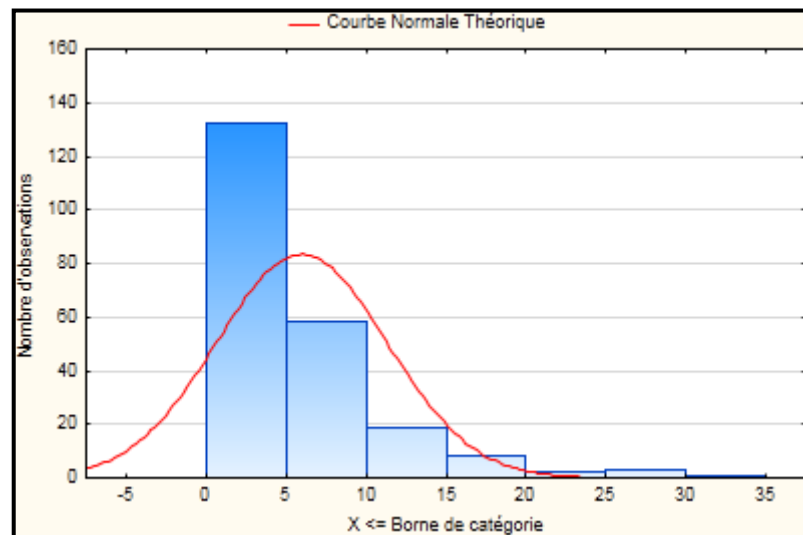


Figure 27 – Distribution des valeurs de pente.

4.2.1.2. Analyse des corrélations entre variables explicatives

Afin d'éviter les problèmes de colinéarité, nous appliquons une analyse de corrélation à nos variables explicatives. Le même seuil critique que celui utilisé à Liège est choisi, soit 0,75 en valeur absolue. La liste des couples de variables posant problème est exposée à l'annexe XVIII. Pour rappel, ne sont pas reprises les corrélations excessives entre temps d'accès et distance euclidienne à un même objet, les corrélations entre variables quantitatives et leurs dérivées binaires et les corrélations entre les variables binaires dérivées d'une même variable quantitative.

Un nombre important de couples de variables fortement corrélées a été trouvé. Ainsi, 136 combinaisons sur 17 424 présentent une corrélation supérieure à 0,75 en valeur absolue, soit environ 0,78 %. Ceci peut paraître minime, mais ces corrélations réduisent une nouvelle fois le nombre de variables qui peuvent être intégrées dans la régression. Nous retenons notamment le temps d'accès à Charleroi, celui-ci étant considéré comme le pôle central du bassin d'emploi. Peu de variables quantitatives d'accès ou de distance seront utilisées compte tenu des nombreuses

corrélations excessives qu'elles entretiennent avec une ou plusieurs autres variables. Leurs dérivées binaires permettent de solutionner ce problème.

4.2.1.3. Modélisation par MCO

A l'instar des développements effectués pour les modèles liégeois, de nombreux essais ont été effectués afin d'identifier les variables les plus explicatives et les plus pertinentes. Le revenu a constitué une pierre d'achoppement dans le développement du modèle carolorégien. En effet, le revenu exprimé à l'échelle des secteurs statistiques adopte, au fil des ajustements, une p-value très légèrement supérieure au seuil de non-significativité retenu, pour rappel 0,10. Afin de rester cohérent avec la méthodologie appliquée, nous devons donc la retirer de l'analyse. Nous avons cependant testé un modèle intégrant cette variable de revenu agrégée à l'échelle des anciennes communes. Celle-ci s'est révélée être significative, mais les paramètres du modèle final ainsi développé nous ont paru un peu faibles (R^2 ajusté : 0,756 ; F : 48,94). Le pouvoir prédictif du modèle laissait de plus à désirer, avec une erreur de 17 758,98 euros. Le non-alignement des résidus le long de la droite de Henry dénotait un ajustement imparfait du modèle. Nous avons pris la peine de vérifier nos données pour constater l'absence d'erreurs de notre part. Ceci effectué, nous avons pris la décision de retirer la variable de revenus de l'analyse, en concluant sur une explication non claire de la part de celle-ci. Les tests effectués sans cette variable permettent l'établissement d'un modèle plus robuste. C'est l'établissement de ce modèle qui est présenté ici.

Les paramètres du modèle sont repris ci-dessous (Tableau 16).

Moy. Var. dep	72274,76	Ec. typ. var. dep.	33567,47
Somme carré résidus	6,16e+10	Ec. typ. de régression	17334,28
R^2	0,753750	R^2 ajusté	0,733330
F (17,205)	36,91110	p. critique (F)	5,89e-63
Log de vraisemblance	-2483,618	Critère d'Akaike	5003,235
Critère de Schwarz	5064,565	Hannan-Quinn	5027,994

Tableau 16 – Paramètres du modèle retenu développé par MCO.

Le modèle développé sous régression MCO est relativement performant, avec une robustesse significative (F de 36,91 et p-value associée minimale) et un pouvoir explicatif assez élevé (R^2 ajusté de 0,733).

Le test de colinéarité (annexe XIX) indique qu'aucune des variables intégrées dans le modèle n'est concernée par ce problème. Le temps d'accès en voiture à Charleroi présente la valeur de VIF la plus élevée (2,65), toutefois loin des seuils d'avertissement.

Le test de Breusch-Pagan (annexe XX) montre que quelques variables sont concernées de manière significative – mais relativement réduite – par de l'hétéroscédasticité. Il s'agit de

- la superficie du terrain ;
- la présence d'un commerce alimentaire à moins de 700 mètres.

L'hétéroscédasticité se devine dans le graphique des résidus en fonction de la superficie par une vague morphologie évasée vers la droite de la distribution des résidus (Figure 28).

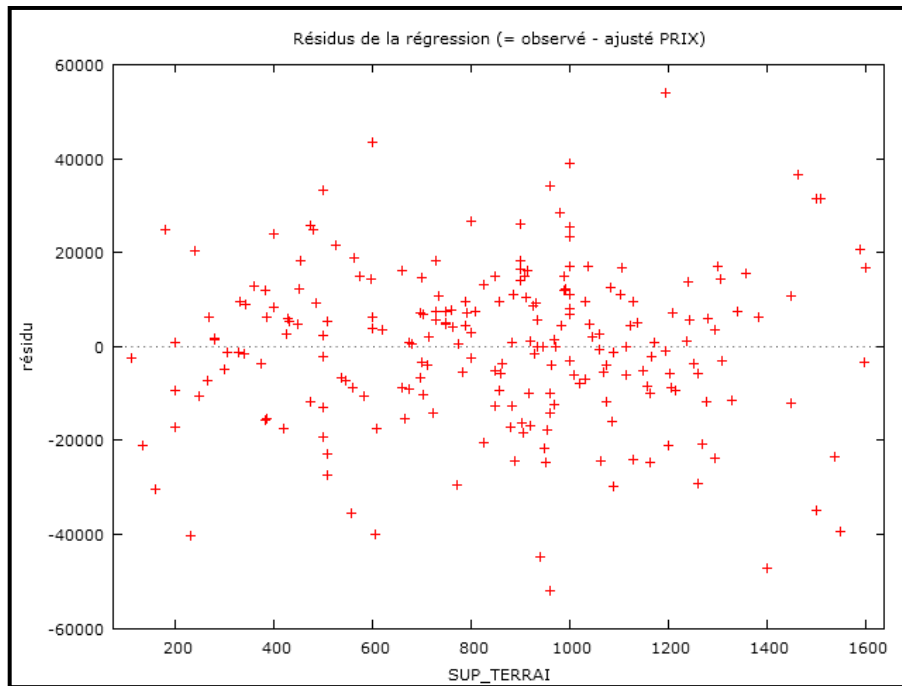


Figure 28 – Distribution des résidus en fonction de la superficie.

4.2.1.4. Correction de l'hétéroscédasticité

La première étape consistant à transformer les variables concernées, nous dérivons la variable de superficie en logarithme en base 10. Mais comme dans le cas des modèles liégeois, cette transformation n'apporte pas d'amélioration : les paramètres du modèle sont dégradés. De plus, nous souhaitons conserver la variable non transformée, étant donné qu'elle a été utilisée sous cette forme pour l'étude du bassin d'emploi de Liège. Il n'est pas possible de transformer la variable de proximité d'un commerce, étant donné son caractère binaire. Aussi, nous passons à l'étape suivante qui consiste à corriger l'hétéroscédasticité sur base d'un modèle développé selon une régression MCG. Les paramètres de ce modèle sont les suivants (Tableau 17) :

Moy. Var. dep	72274,76	Ec. typ. var. dep.	33567,47
Somme carré résidus	6,21e+10	Ec. typ. de régression	17411,60
R ²	0,815952	R ² ajusté	0,800690
F (17,205)	53,46129	p. critique (F)	1,16e-65
Log de vraisemblance	-463,6692	Critère d'Akaike	963,3385
Critère de Schwarz	1024,668	Hannan-Quinn	988,0966

Tableau 17 – Paramètres du modèle corrigeant hétéroscédasticité.

La correction de l'hétéroscédasticité apporte une amélioration des paramètres du modèle : celui-ci est donc amélioré. Le *F* de Fischer a gagné environ seize points par rapport au modèle MCO, indiquant une plus grande robustesse statistique. La performance explicative du modèle est également renforcée avec un *R² ajusté* de 0,801.

Les résidus en fonction de la superficie présentent une distribution légèrement plus resserrée autour de la droite horizontale de valeur nulle, indiquant que l'hétéroscédasticité a bel et bien été réduite (Figure 29).

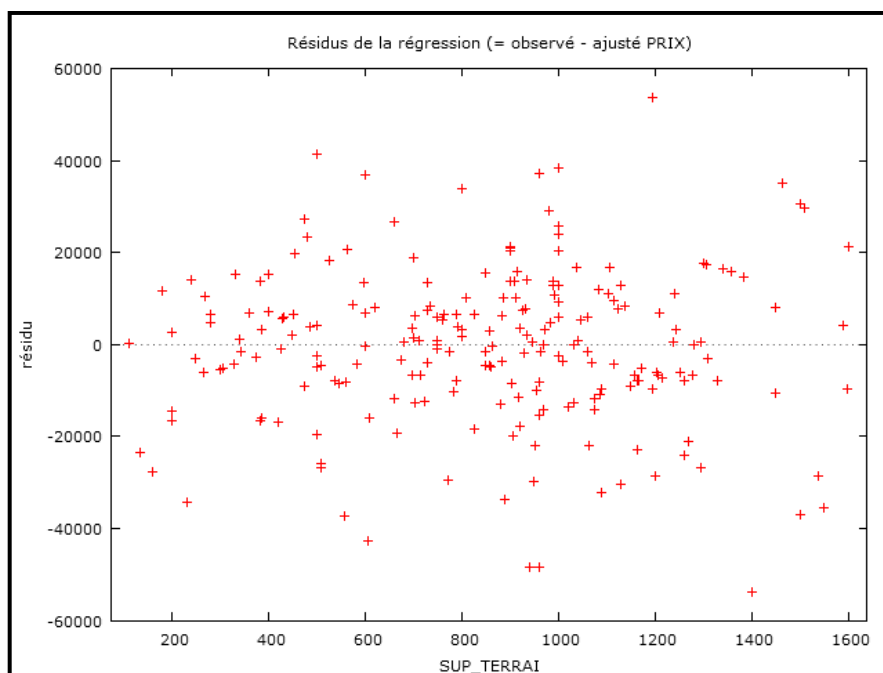


Figure 29 – Distribution des résidus en fonction de la superficie après correction de l'hétéroscédasticité.

Nous analysons à nouveau la colinéarité entre les variables. Ce problème reste absent de celles-ci, le VIF le plus élevé demeurant celui du temps d'accès en voiture à Charleroi, avec une valeur de 2,67.

4.2.1.5. Suppression des observations à résidus extrêmes et ajustement du modèle

Le logiciel Gretl identifie quatre résidus supérieurs à 2,5 fois l'erreur standard, en valeur absolue. L'un d'entre eux est associé à une observation à prix élevé, les trois autres sont des biens à prix plus bas. Le premier est sans doute un bien exceptionnel. Le deuxième est un bien à prix très faible par rapport à ce qui est attendu, probablement en raison de sa contiguïté avec la route nationale N5 sur l'entité de Walcourt. Le troisième est également un terrain à prix plus faible par rapport à la prévision ; ce terrain est situé dans une cité sociale. Le quatrième terrain a également un prix plus faible que prévu ; il est situé en fond de vallée, à proximité de la ligne SNCB 132 et est entouré de maisons à l'apparence délabrée.

Après suppression des observations présentant ces résidus délinquants, nous réajustons le modèle. Celui-ci présente les paramètres suivants (Tableau 18) :

Moy. Var. dep	72256,95	Ec. typ. var. dep.	33292,93
Somme carré résidus	5,26e+10	Ec. typ. de régression	16131,27
R ²	0,875443	R ² ajusté	0,865577
F (16,202)	88,73420	p. critique (F)	4,76e-82
Log de vraisemblance	-458,2953	Critère d'Akaike	950,5905
Critère de Schwarz	1008,205	Hannan-Quinn	973,8592

Tableau 18 – Paramètres du modèle corrigeant l'hétéroscédasticité après suppression des résidus délinquants.

Le pouvoir prédictif n'est pas optimal : l'erreur type d'estimation vaut 15 563,75 euros, soit une erreur relative de 21,54 %. Néanmoins, le modèle est très robuste statistiquement, avec un F qui s'est élevé d'un peu moins de 36 points (88,73). Le R^2 ajusté dénote également un modèle au plus grand pouvoir explicatif, avec une augmentation d'environ 6,5 points (0,866). Les critères d'Akaike

et de Schwarz ainsi que l'indice Hannan-Quinn sont en baisse, témoignant de l'amélioration des performances du modèle.

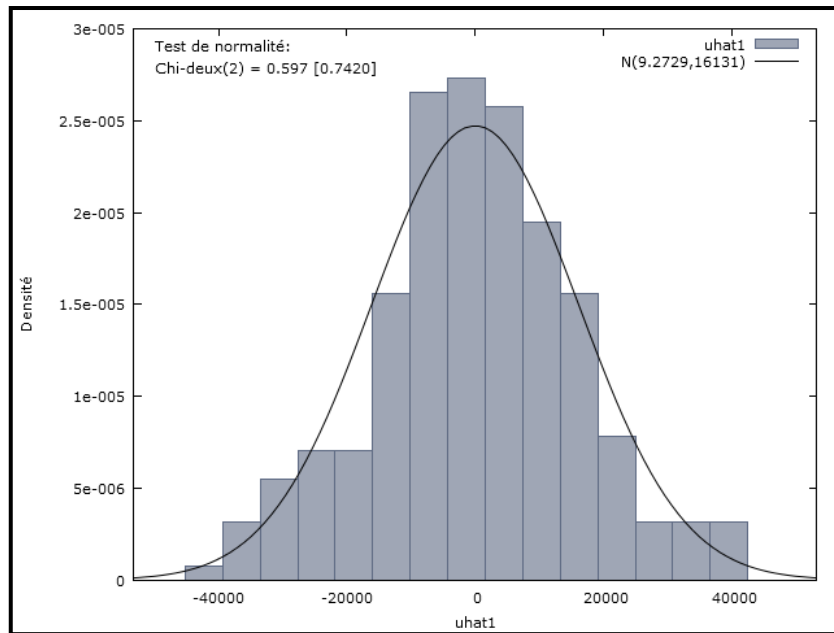


Figure 30 – Distribution des résidus après épuration.

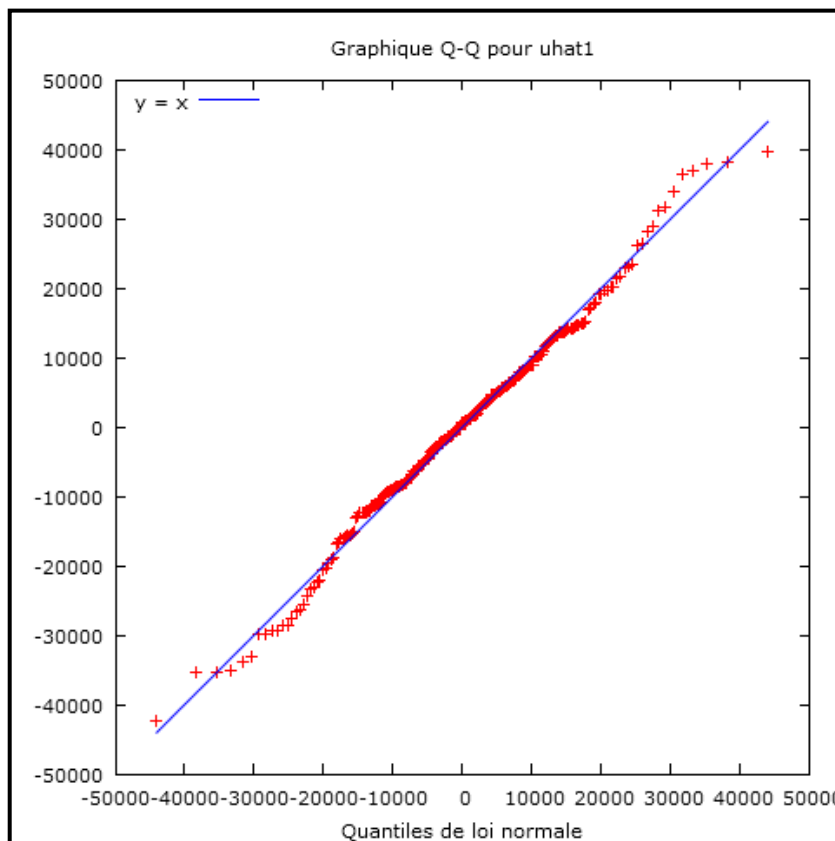


Figure 31 – Diagramme quantile-quantile des résidus.

Le test du Chi carré fourni par Gretl fait état d'une distribution normale des résidus (Figure 30). La p-value associée (0,74) est bien supérieure au seuil critique de 0,05, forçant l'acceptation de l'hypothèse nulle, pour rappel la normalité de la distribution des résidus. La suppression des

observations à résidus délinquants a permis de rendre normale une distribution des résidus qui ne l'était pas avant cette étape.

L'analyse du diagramme quantile-quantile des résidus montre que ceux-ci sont globalement alignés le long de la droite de Henry des résidus (Figure 31). Ils s'en écartent pour les valeurs les plus négatives et les plus positives. Le problème maintes fois mentionné de l'ajustement médiocre pour les valeurs hautes et les valeurs basses de résidus est à nouveau rencontré dans le cas présent.

Le graphique des résidus en fonction du prix met en évidence la diminution des performances du modèle pour les biens supérieurs à 100 000 euros (Figure 32). De plus, les résidus deviennent exclusivement positifs au-delà de 140 000 euros. Le problème de prévision pour les biens à prix élevés du marché foncier, mis en évidence pour le bassin d'emploi de Liège, semble se confirmer. Pour rappel, cette mauvaise prédiction pour les biens à prix élevés avait été mise en évidence pour l'immobilier (F. Des Rosiers, 2001). Le foncier semble dès lors également concerné.

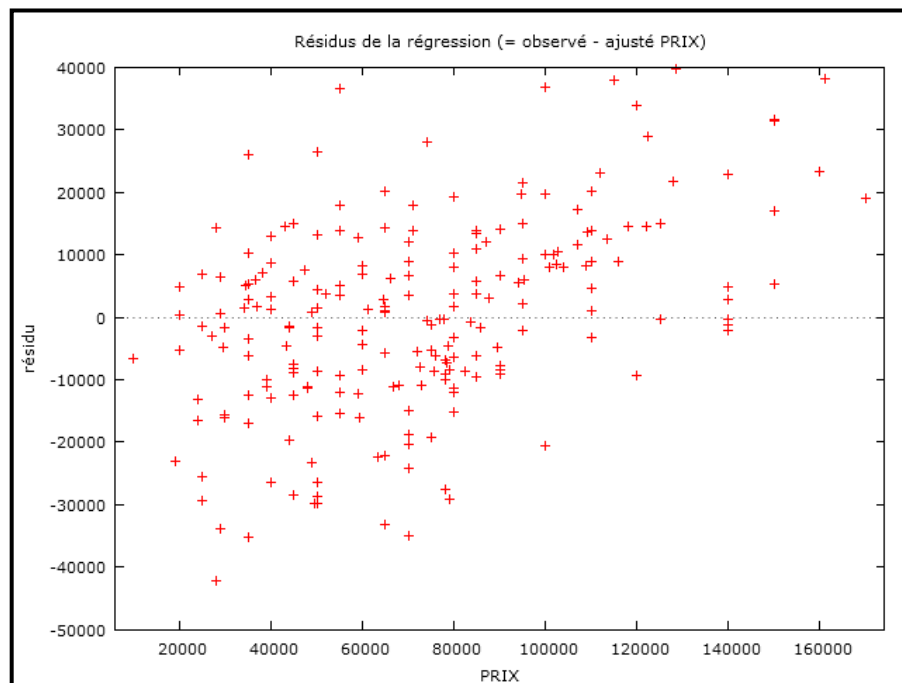


Figure 32 – Distribution des résidus en fonction du prix.

4.2.1.6. Analyse de l'autocorrélation spatiale

L'indice de Moran, proche de 0 (valeur exacte : -0,0169734), n'est pas significativement différent ($p=0,260000$) de la valeur attendue (-0,0046), ce qui permet d'exclure la présence d'autocorrélation spatiale entre les résidus du modèle. Le diagramme de dispersion de Moran est visible à l'annexe XXI. Le modèle est donc correctement ajusté et comporte suffisamment de variables pour rendre compte des caractéristiques du marché des terrains à bâtir du bassin d'emploi de Charleroi.

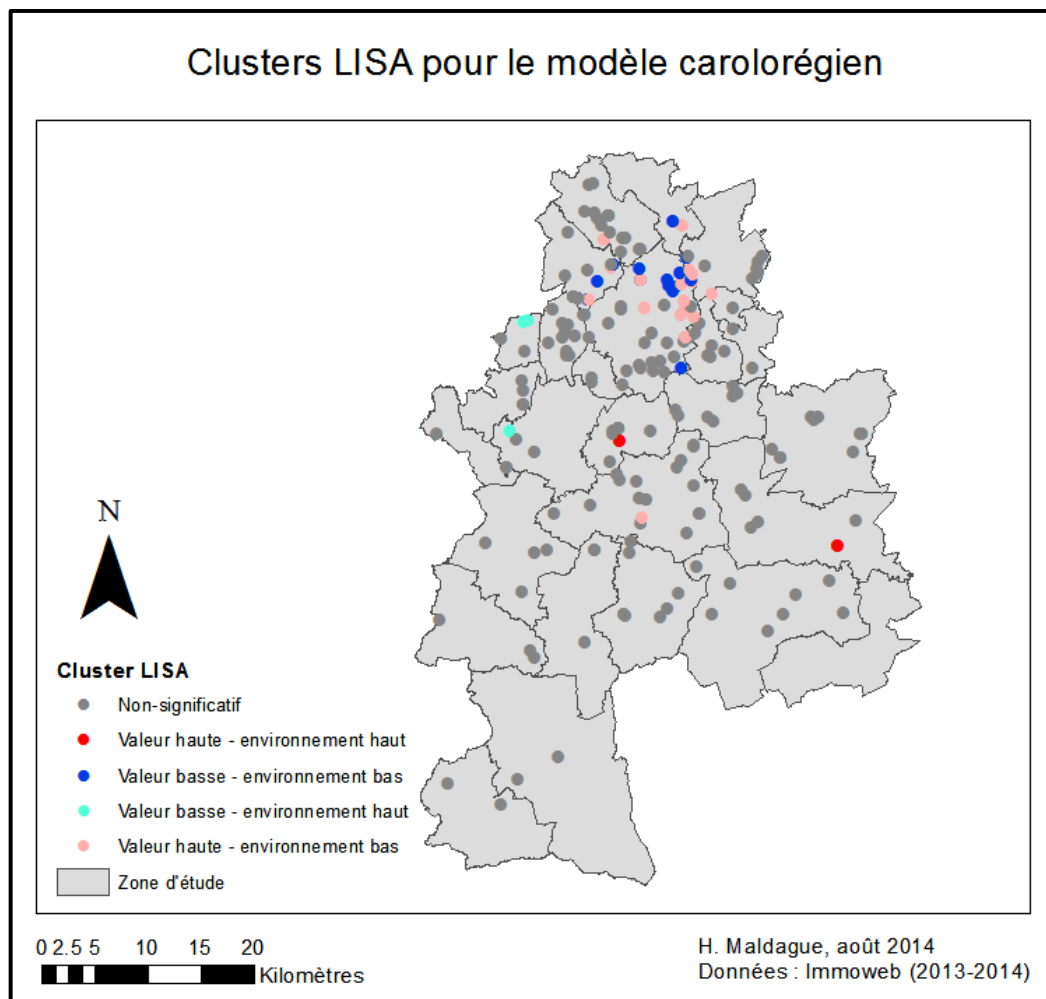


Figure 33 – Clusters LISA pour le modèle carolorégien.

Au niveau plus local, nous observons l'existence d'un cluster d'autocorrélation négative dans l'est de la commune de Charleroi, avec une série de valeurs hautes dans un environnement bas (Figure 33). Ces résidus sont côtoyés par des valeurs basses d'autocorrélation positive. Pour ces zones, la prédiction du modèle est donc imparfaite. Ceci n'est pas sans rappeler le phénomène de valorisation et de perception au sein de l'agglomération liégeoise. La proximité de l'aéroport de Charleroi-Bruxelles-Sud et des entreprises et industries de l'aéropôle est sans doute responsable de ce phénomène.

4.2.2. Présentation des résultats

Les résultats du modèle corrigeant l'hétéroscédasticité et débarrassé des observations à résidus délinquants sont présentés ci-dessous (Tableau 19).

Une nouvelle fois, peu de variables sont significatives au seuil 0,10 utilisé pour la p-value. Elles sont au nombre de seize. Afin de mener à bien la comparaison, nous avons, comme pour le modèle liégeois, recalculé les coefficients standardisés, qui sont classés par ordre d'importance ci-dessous, en valeur absolue (Tableau 20).

	<i>Coefficient</i>	<i>Erreur Std</i>	<i>t de Student</i>	<i>p. critique</i>	
Constante	11521	5582,49	2,0638	0,04032	**
Part de bâti dans un rayon de 200 m	-108,76	35,7999	-3,0380	0,00270	***
Superficie	57,8152	3,51719	16,4379	<0,00001	***
Localisation dans un lotissement	9450,16	2454,64	3,8499	0,00016	***
Présence d'un bâtiment sur le terrain	-18412,7	4921,36	-3,7414	0,00024	***
Bruit aérien > 60 dB	-44380,7	10945,7	-4,0546	0,00007	***
Ecole secondaire à moins de 60 secondes	26395,7	5067,98	5,2083	<0,00001	***
Temps d'accès aux nodules commerciaux	22,3032	6,11305	3,6485	0,00034	***
Ecole primaire/second. à moins de 30 s	15884,4	6254,65	2,5396	0,01185	**
Temps d'accès voiture à Charleroi (s)	-24,5707	2,19312	-11,2036	<0,00001	***
SAED à moins de 100 m	-32802	8883,54	-3,6925	0,00029	***
Ligne HT à moins de 500 m	-4655,26	2364,26	-1,9690	0,05032	*
Chemin de fer à moins de 500 m	-7564,38	2961,12	-2,5546	0,01137	**
Commerce alimentaire à moins de 700 m	17939,6	2510,56	7,1457	<0,00001	***
Distance aux arrêts de bus	-14,637	1,80481	-8,1100	<0,00001	***
Part modale du train	326911	24027,1	13,6059	<0,00001	***
Part de maisons à quatre façades	376,204	51,4837	7,3073	<0,00001	***
*p-value inférieure à 0,10 **p-value inférieure à 0,05 ***p-value inférieure à 0,01					

Tableau 19 – Résultats du modèle carolorégien.

Variable	B*	B
Superficie	0,58930	57,8157
Temps d'accès à Charleroi	-0,49269	-24,5707
Part modale du train	0,37329	326911
Part de maisons à quatre façades	0,29817	376,204
Commerce alimentaire à moins de 700 m	0,23421	17939,6
Ecole secondaire à moins de 60 s	0,16588	26395,7
SAED à moins de 100 m	-0,16120	-32802
Bruit aérien > 60 dB	-0,15530	-44380,7
Temps d'accès aux nodules commerciaux	0,14567	22,3032
Présence d'un bâtiment sur le terrain	-0,13560	-18412,7
Ecole primaire/second. à moins de 30 s	0,12444	15884,4
Localisation dans un lotissement	0,12338	9450,16
Distance aux arrêts de bus	-0,12236	-14,637
Part de bâti dans un rayon de 200 m	-0,09780	-108,76
Chemin de fer à moins de 500 m	-0,08440	-7564,38
Ligne HT à moins de 500 m	-0,06283	-4655,26

Tableau 20 – Classement des variables explicatives par ordre d'importance.

La variable apportant la plus grande part d'explication au prix est la superficie. A sa suite viennent des variables d'accessibilité. Les différentes variables d'environnement physique terminent le classement : ces variables expliquent plus faiblement les prix que la superficie et la plupart des variables d'accessibilité. Apparemment, les ménages, lors d'un achat, sont premièrement attentifs à la superficie et à l'accessibilité du terrain avant d'envisager l'environnement physique et les nuisances, à moins que celles-ci soient fortement dérangeantes (SAED et bruit aérien dans ce cas). Cet ordre est une nouvelle fois assez similaire à celui mis en évidence par Y. Grignet (2011).

4.2.2.1. Variables structurelles

La variable structurelle qu'est la superficie est la première caractéristique explicative du prix des terrains. Le mètre carré de terrain supplémentaire représente 57,82 euros, toutes choses étant

égales par ailleurs. Pour rappel, il existe un phénomène de marginalité décroissante des prix en fonction de la superficie. A l’instar du modèle général liégeois, nous avons cherché à mettre en évidence ce phénomène au sein du bassin d’emploi de Charleroi. Nous avons scindé notre échantillon selon les quartiles et avons développé des modèles pour chacun d’entre eux. Les développements s’effectuant sur des échantillons de faible effectif, il convient de rester prudent vis-à-vis des résultats. Néanmoins, comme illustré ci-dessous, la marginalité décroissante du prix en fonction de la superficie peut être approchée (Tableau 21).

Intervalle de superficie	coefficient	p-value
0 – 582	74,42	<0,00001
583 – 883	73,86	0,00611
884 - 1063	127,91	0,03970
1064 - 1600	41,65	0,00850

Tableau 21 – Illustration de la marginalité décroissante du prix selon la superficie.

Nous pouvons voir que, la troisième catégorie mise à part pour laquelle le coefficient est mal compris, la marginalité décroissante apparaît, bien qu’étant moins nette qu’à Liège ; nous attribuons une partie de l’explication aux faibles effectifs utilisés. Le prix du mètre carré, toutes choses égales par ailleurs, passe de 74,42 euros à 41,65 euros. Comme à Liège, la p-value augmente avec la superficie, sauf pour la troisième catégorie.

La présence d’un bâtiment entraîne une moins-value de 18 412,70 euros. Ceci est dû à la présence de petites constructions de type garage ou abri de jardin, appelées sans doute à être démolies par les futurs propriétaires du terrain. Cette démolition entraînant un coup supplémentaire, les acquéreurs potentiels se détournent de ce type de biens qui voient dès lors leur prix diminuer. Un tel raisonnement devrait s’appliquer aux modèles liégeois, mais ce n’est pas le cas. La moins-value est sans doute ici une particularité de notre échantillon. Il convient de rester prudent vis-à-vis de cette variable.

Le relief (pente et orientation au sud) n’a pas d’influence sur les prix. La variable de pente présente cependant, avant son retrait, une p-value légèrement supérieure au seuil critique (valeur exacte : 0,123), avec indication d’une moins-value pour des valeurs grandissantes. Il semble donc y avoir une influence, mais nous ne pouvons conclure à sa significativité.

4.2.2.2. Variables d’accessibilité

Se trouver à moins de 60 secondes en voiture d’une école secondaire engendre une plus-value de 26 395,70 euros. L’utilité procurée par une école secondaire est de plus renforcée par le fait que ces écoles sont situées à Charleroi et dans les petites villes équipées de l’Entre-Sambre-et-Meuse. Une plus-value est également engendrée par la présence à moins de 30 secondes en voiture d’une école maternelle et/ou primaire, celle-ci étant plus faible et moins fortement significative. La plus-value engendrée est de 15 884,40 euros. Les annonces Immoweb font régulièrement mention de la proximité d’une école, indiquant que cette proximité semble recherchée. Ces constatations sont cohérentes avec les résultats de plusieurs études consultées (C. J. Barnett, 1985 ; L. Tyrvaïnen et A. Miettinen, 2000 ; A. Donzel *et al*, 2008). Les autres seuils ne sont pas significatifs, hormis la présence d’une école secondaire à moins de deux minutes en voiture. Dans ce cas-ci, c’est une moins-value qui est mise en évidence ; il est supposé que le milieu urbain dense et bruyant que constitue l’agglomération de Charleroi en est responsable.

La présence d'un commerce alimentaire à moins de 700 mètres est responsable d'une plus-value de 17 939,60 euros. Cela justifie le fait que la proximité des commerces est régulièrement mentionnée dans les annonces Immoweb. La proximité d'un commerce s'est de plus révélée positive sur le prix dans deux études consultées (X. Gao et Y. Asami, 2007 ; Y. Liu *et al.*, 2010). Dans l'Entre-Sambre-et-Meuse, ces commerces se localisent quasi exclusivement au sein des petites villes. L'influence procurée par cette variable est à rapprocher de celle engendrée par la présence des écoles, à savoir l'influence des petits pôles pourvoyeurs de services de première nécessité. Les autres seuils ne se sont pas révélés être significatifs.

Nous pouvons à nouveau rapprocher les observations citées de l'étude de J. Strée et J.-M. Halleux (2013) qui met en évidence une moins-value d'environ 50 000 euros pour une maison localisée loin des commerces et services (écoles primaire et écoles secondaires) par rapport à la même maison proche de ces implantations et valant alors 330 000 euros. Ceci est assez proche de nos résultats.

De manière assez surprenante, le temps d'accès aux nodules commerciaux entraîne une plus-value : à chaque seconde d'éloignement, le prix augmente de 22,30 euros, soit 1338,19 euros par minute. Une étude parmi les travaux consultés est arrivée à une constatation similaire en utilisant la distance. L'auteur concluait à une influence non claire (C. J. Barnett, 1985). Dans le cas présent et après étude de la répartition des nodules, nous pouvons supposer que les nuisances liées à un environnement dense et au trafic des grands axes prennent le dessus sur l'utilité procurée par la proximité d'un nodule commercial.

Le temps d'accès à Charleroi est une variable très fortement significative. Son comportement est cohérent avec la plupart des études consultées. Pour chaque seconde d'éloignement, le prix diminue de 24,57 euros. Cela reflète à nouveau le rôle des villes en tant que fournisseur d'emplois et de services divers. Ainsi, le bien le plus éloigné, à 3677 secondes (1h01), coûte 90 346,46 euros moins cher, toutes choses égales par ailleurs, qu'un bien identique situé sur la Place Charles II, pour rappel choisie comme centre du bassin d'emploi.

Une distance croissante à un arrêt de bus entraîne une moins-value. Pour chaque mètre d'éloignement, le prix diminue de 14,64 euros. Pour le bien le plus éloigné, à 2566 mètres, ceci représente une moins-value de 37 558,54 euros. Il semble dès lors que les ménages perçoivent l'utilité procurée par la présence d'un arrêt de bus à proximité de leur future acquisition. Ces observations sont cohérentes avec plusieurs études consultées, aussi bien pour les terrains à bâtir (X. Gao et Y. Asami, 2007 ; Y. Liu *et al.*, 2010) que les maisons (S. Zheng et M. E. Kahn, 2008 ; R. Cervero et C. D. Kang, 2011; I.-H. Lin *et al.*, 2013). Néanmoins, la part modale du bus n'apporte pas d'influence significative. Les ménages semblent indifférents à la fréquence de passage des bus et au nombre de lignes desservant l'arrêt le plus proche. Ceci est peut-être dû à un manque d'information à ce niveau : il y a conscience de la proximité d'un arrêt, mais sa desserte est mal connue.

La part modale du train entraîne une influence significative. La plus-value paraît extrêmement importante, mais ceci est dû au domaine d'existence mathématique de la variable, compris entre 0 et 1. Exprimé autrement, un gain de 1 % de part modale entraîne une plus-value de 3269,11 euros. Entre le bien le moins accessible (3 %) et le bien le plus accessible (20 %), nous observons une différence de 55 574,87 euros. L'hypothèse que nous pouvons émettre est que les

ménages considèrent donc les possibilités offertes par le transport ferroviaire, en tenant compte des fréquences. Le train apparaît donc comme un élément utile, ceci pouvant s'expliquer par les liaisons rapides entre le bassin d'emploi de Charleroi et la métropole bruxelloise, relativement proche. En effet, les déplacements wallons en train se font surtout vers Bruxelles, ils sont perçus comme une bonne alternative à la voiture (CPDT, n.d.). Le nord du bassin d'emploi de Charleroi est de plus fort polarisé par Bruxelles (CPDT, 2008). La proximité d'une gare ne semble pas avoir d'impact, quelle que soit sa desserte. La plupart des biens sont situés à proximité de gares à faible desserte (Obaix-Buzet, vallée de l'Eau d'Heure), tandis que d'autres sont localisés dans des quartiers industriels ou à proximité de gares entourées d'un important faisceau de voies (Luttre, Charleroi).

La proximité d'une école du supérieur n'a pas d'influence. Elles sont peu nombreuses dans le bassin d'emploi de Charleroi et sont situées au sein de la commune éponyme ou à Fleurus. L'éventuelle utilité procurée se mêle à l'effet de l'environnement urbain dense dans lequel elles se trouvent. L'absence d'un établissement connu, comme une université, explique peut-être l'absence d'influence.

La proximité d'une sortie d'autoroute n'apporte pas d'influence significative sur les prix, sans doute en raison d'un enchevêtrement entre l'utilité procurée et les nuisances engendrées par le trafic. A ce titre, la proximité d'un axe routier ayant un trafic de plus de 10 000 ou 20 000 véhicules par jour ne semble pas engendrer d'influence sur les prix. Cependant, le seuil des 200 mètres autour d'une route de plus de 20 000 véhicules par jour réagit au cours de la modélisation, avec une p-value à peine supérieure au seuil retenu (valeur exacte : 0,108). Une influence semble exister, mais au vu du seuil choisi, nous ne pouvons conclure à sa significativité. La plupart des biens concernés se situant en milieu urbain, il est possible que l'influence soit parasitée par d'autres nuisances. Une autre explication est à rechercher dans l'effet de barrière sonore et visuelle engendré par les éléments du voisinage des terrains concernés, le bruit étant ainsi diminué, et l'influence sur le prix réduite.

La proximité d'un RAVeL n'engendre aucune influence sur les prix. Ceci concorde avec l'absence d'influence significative de la part modale des modes lents. Les déplacements à pieds et à vélo restant marginaux en Wallonie, il est logique de penser que les ménages ne prennent pas en compte la présence d'infrastructures proposées à destination des modes lents.

4.2.2.3. Variables d'environnement physique

Une part de bâti croissante dans un rayon de 200 mètres entraîne une moins-value sur le prix des terrains, à raison d'une baisse d'environ 108,76 euros par pourcent de bâti supplémentaire. Concomitamment, une part de maisons à quatre façades croissante au sein des secteurs statistiques entraîne une plus-value : chaque pourcent supplémentaire entraîne une croissance du prix de 376,20 euros. Ceci illustre clairement le goût des ménages pour, d'une part, un environnement aéré et peu densément bâti et pour, d'autre part, la maison quatre façades comme habitation (J.-M. Halleux, 2012). Ce phénomène est ici mis en évidence au travers de notre modèle. A l'inverse, la densité de population n'est pas significative. Nous avançons le fait que cette variable ne représente pas parfaitement la densité de construction. Ce que vendeurs et acheteurs perçoivent en premier lieu, c'est la densité des constructions et non la densité réelle de population. Dans les faits, il existe une corrélation modérée entre la proportion de maisons à

quatre façades et la densité de population, indiquant que la variable de proportion de maisons à quatre façades reprend une part de l'explication que pourrait procurer la densité de population.

La diversité n'est pas significative, sans doute en raison des goûts variés des ménages et de la difficulté de jugement vis-à-vis de ce concept. La part de bâti à 1000 mètres n'a pas d'influence non plus. Ceci contredit l'étude de J. Geoghegan *et al.* (1997).

L'appartenance à un lotissement engendre une plus-value de 9450,16 euros. Ceci semble indiquer que les lotissements sont très recherchés dans le bassin d'emploi de Charleroi. Le fait qu'un certain nombre de démarches administratives puisse être évité peut expliquer cette plus-value (Construire facile, 2012). D'autre part, la demande en terrain semblant être supérieure à l'offre, l'ouverture d'un lotissement peut paraître comme une opportunité pour les candidats acheteurs (Chambre des Notaires de la province du Hainaut, 2010). Nous pouvons rapprocher ce constat du fait que la création d'une parcelle en lotissement avec ouverture de voirie coûte environ 15 000 euros plus cher qu'une parcelle hors-lotissement dans le contexte français (CPDT, 2012).

Les nuisances sonores du trafic aérien ont un impact très fort sur les prix. Les biens subissant un bruit de plus de 60 dB connaissent une moins-value de 44 380,7 euros. Celle-ci peut paraître importante, mais elle peut être expliquée par la fréquence des décollages et des atterrissages à l'aéroport de Charleroi, au nombre de 125 par jour (Brussels South Charleroi Airport, 2014). Les vols ayant lieu entre 6h30 et 23h30, ceci représente une moyenne de 7,35 décollage/atterrissage par heure. Le seuil de 55 dB était à peine significatif, c'est pourquoi nous ne l'avons pas considéré.

La présence d'un site d'activité économique désaffecté à moins de 100 mètres entraîne une moins-value importante, à raison de 32 802,00 euros. Cet effet n'est plus visible de manière significative au seuil de 200 mètres, confirmant le fait que l'influence négative d'une friche ne s'exprime qu'à proximité immédiate du site en question (P. Dethier, 2012). La nuisance générée étant uniquement visuelle, celle-ci est rapidement arrêtée par les barrières visuelles de l'environnement urbain.

La présence d'une ligne à haute tension à moins de 500 mètres semble porter préjudice au prix des terrains, avec une moins-value de 4655,26 euros. La p-value relativement élevée souligne toutefois le fait que l'influence n'est pas très importante. Cette influence ne s'est par ailleurs pas exprimée aux autres seuils, malgré un nombre suffisamment important de biens (26 à moins de 100 mètres, 35 à moins de 200 mètres).

La présence d'une ligne de chemin de fer à moins de 500 mètres entraîne une moins-value de 7564,38 euros. Les autres seuils ne sont pas significatifs. De plus, la p-value relativement élevée montre, à l'instar du cas des lignes à haute tension, que le dérangement provoqué par le trafic ferroviaire n'est pas très important. Ceci peut s'expliquer par le fait que les 37 biens concernés se situent à proximité de lignes au trafic acceptable. De plus, il peut exister des barrières naturelles ou artificielles qui réduisent la portée de la vue et du bruit du trafic ferroviaire.

La localisation dans un cul-de-sac n'apporte pas de modification significative des prix. Les biens en question sont sans doute affectés par d'autres éléments dus à leur localisation : certains d'entre eux sont situés en milieu urbain dense, d'autres sont proches de voies ferrées ou de grands axes routiers.

La proximité d'une forêt ou d'un espace vert urbain n'engendre pas d'influence, sans doute en raison de leur abondance. En effet, 91 % des biens sont situés à moins de 500 mètres d'une forêt ou d'un espace vert.

La proximité des industries ne semble pas avoir d'effet sur les prix. Ceci peut, à l'instar de la situation liégeoise, être expliqué par le fait que des barrières visuelles et sonores limitent les nuisances. De plus, tous les types d'industries ne sont sans doute pas perçus de la même manière.

La proximité d'une carrière n'engendre pas d'influence, sans doute en raison d'un trop faible nombre de biens concernés (trois à moins de 200 mètres, huit à moins de 500 mètres).

Haut lieu du tourisme wallon, les lacs de l'Eau d'Heure ne semblent pas engendrer d'influence sur les prix des terrains. Ceci est en contradiction avec deux études consultées (S.-T. Anderson et S.-E. West, 2006 ; H. A. Sander et R. G. Haight, 2012). Le fait qu'il s'agisse d'une zone à l'écart des services fournis par Beaumont et Philippeville et à l'accès lent et malaisé peut expliquer cette absence d'influence. D'une manière proche, la proximité d'un cours d'eau n'engendre pas d'influence, au même titre que la localisation en zone inondable. Cette dernière variable concerne cependant un faible nombre de biens, à savoir cinq.

4.2.2.4. Variables de valorisation sociale

Comme nous l'avons expliqué auparavant, l'influence du revenu est difficile à objectiver. L'intégration de cette variable dans les modèles développés entraîne une dégradation des paramètres, d'où son retrait. Nous ne tirerons pas de conclusion quant à l'impact qu'elle peut avoir sur les prix. Dans les faits, nous émettons l'hypothèse qu'une partie de l'influence du revenu est reprise au travers de la variable de part de maisons à quatre façades. Il existe en effet une corrélation modérée entre ces variables.

Le taux de chômage et le taux de diplômés du supérieur n'entraînent pas d'influence sur les prix, sans doute parce que les données sont trop anciennes pour refléter la situation actuelle.

4.3. Conclusion sur le modèle carolorégien

Les résultats de ce modèle tendent à confirmer les conclusions retirées de l'établissement des modèles liégeois. Certaines variables ont une influence correctement cernée par le modèle carolorégien : il s'agit de la superficie et d'une bonne partie des variables d'accessibilité. A l'inverse, l'impact de l'environnement physique est mal identifié : de nombreuses variables ne sont pas significatives, alors que nous nous attendions au départ à voir apparaître une influence marquée sur la formation des prix. Une nouvelle fois, des particularités locales, telles que les barrières visuelles et sonores, viennent sans doute contrecarrer l'influence de certaines variables.

5. Comparaison entre les modèles liégeois et le modèle carolorégien

Dans ce chapitre, nous réalisons une comparaison entre le modèle développé sur les terrains à bâtir du bassin d'emploi de Charleroi et les modèles – général et périphérique – générés sur base des terrains du bassin d'emploi de Liège. Nous montrons, au cours de cette comparaison, que nous pouvons rassembler les variables en quatre groupes selon les possibilités de comparaison. Une majorité de déterminants entre dans la catégorie des critères pour lesquels la comparaison est malaisée, faute à différents paramètres que nous énoncerons ici.

Variables	Charleroi		Liège général		Liège périphérique	
	B	p-value	B	p-value	B	p-value
Superficie	57,82	<0,00001	54,16	<0,00001	52,48	<0,00001
Temps d'accès au pôle central ³²	-24,57	<0,00001	-14,55	<0,00001	-13,34	<0,00001
Part modale du train (%)	3269,11	<0,00001	/	>0,10	/	>0,10
Localisation dans un lotissement	9450,16	0,00016	-5028,83	0,00045	-6600,15	<0,00001
Bruit aérien supérieur à 60 dB	-44380,7	0,00007	-11912,5	0,00188	-9836,7	0,00044
Ligne à haute tension à 500 m	-4655,26	0,05032	-3909,92	0,01954	/	>0,10
Ecole secondaire à 60 s	26395,7	<0,00001	/	>0,10		
Ecole secondaire à 120 s			/	>0,10	18621,2	0,00032
Chemin de fer à moins de 500 m	-7564,38	0,01137	/	>0,10		
Chemin de fer à moins de 700 m			/	>0,10	-6037,77	0,01264
Commerce alim. à moins de 500 m			/	>0,10	10304,6	0,00074
Commerce alim. à moins de 700 m	17939,6	<0,00001	/	>0,10		
Présence d'un bâtiment	-18412,7	0,00024	/	>0,10	/	>0,10
Temps d'accès aux nodules	22,30	0,00034	-30,53	<0,00001	-28,09	<0,00001
Ecole maternelle/primaire à 30 s	15884,4	0,01185	/	>0,10	/	>0,10
Distance aux arrêts de bus	-14,64	<0,00001	5,20	0,01094	4,79	0,00004
Gare L à moins de 500 m	/	>0,10	/	>0,10	11835,7	0,01536
Part modale des modes lents	/	>0,10	/	>0,10	-1251,54	0,00002
Localisation dans un cul-de-sac	/	>0,10	4738,63	0,03779	5848,49	0,00373
Part de bâti à moins de 200 m	-108,76	0,00270	/	>0,10	/	>0,10
Part de maisons à quatre façades	376,204	<0,00001	/	>0,10	/	>0,10
Cours d'eau à moins de 100 m	/	>0,10	-6105,1	0,00305	-4904,93	0,00260
Route (>20000 veh/j) à 200 m	/	>0,10	-19604,1	0,02199	/	>0,10
Accès autor. à moins de 700 m	/	>0,10	/	>0,10	-12117,4	0,01415
SAED à moins de 100 m	-32802	0,00029	/	>0,10	/	>0,10
Revenu médian par sect. stat.	/	>0,10	1,12	<0,00001	1,55	<0,00001

Tableau 22 – Comparaison entre le modèle carolorégien et les deux modèles liégeois.

Le tableau ci-dessus (Tableau 22) reprend les coefficients des trois modèles développés ainsi que la p-value associée à chacun d'entre eux. Le premier regroupement contient les variables aux résultats robustes et cohérents avec la littérature, pour lesquelles une comparaison est possible, sans restriction. La deuxième catégorie reprend les déterminants pour lesquels une comparaison peut être réalisée, mais avec prudence. La troisième catégorie regroupe des variables qui se sont

³² Place Charles II pour Charleroi, Place Saint-Lambert pour Liège.

révélées significatives pour différents seuils de distance, qu'il est possible de considérer pour une comparaison. La quatrième catégorie reprend les variables pour lesquelles la comparaison est malaisée.

5.1. Variables comparables

Les variables sont au nombre de deux : la superficie et le temps d'accès au pôle central du bassin, à savoir la place Charles II pour le bassin d'emploi de Charleroi, la place Saint-Lambert pour le bassin d'emploi de Liège.

La superficie, en tant que première variable explicative du prix, est très fortement significative dans les trois cas. Le prix du mètre carré est légèrement plus élevé dans le bassin d'emploi de Charleroi. Ceci peut s'expliquer par une offre moins abondante qu'à Liège (Chambre des Notaires de la province du Hainaut, 2010 ; CPDT, 2012), pour une demande se maintenant depuis plusieurs années à un niveau élevé. La figure 34 présente la carte du taux de disponibilité foncière pour 2008, réalisée par la CPDT (2012). Ce dernier correspond au rapport entre le nombre d'hectares non artificialisés et le nombre d'hectares disponibles dans les zones d'habitat du Plan de Secteur. Nous pouvons constater que le nord du bassin d'emploi de Charleroi présente des valeurs faibles de cet indice. Or, la plupart des terrains de notre échantillon se situent dans cette zone. Le bassin d'emploi de Liège présente des valeurs un peu plus élevées de l'indice. Ceci confirme les propos de la Chambre des Notaires du Hainaut concernant l'offre se réduisant. Il est dès lors compréhensible que la valeur d'un mètre carré de terrain supplémentaire, toutes choses étant égales par ailleurs, soit légèrement plus élevée dans le bassin de Charleroi (57,82 euros) que dans le bassin de Liège (54,16 euros).

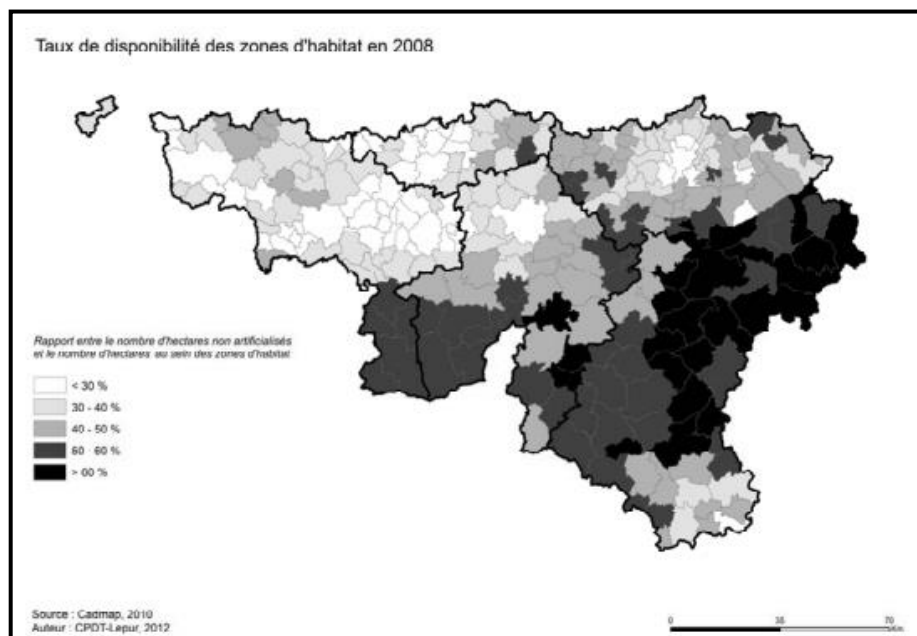


Figure 34 – Taux de disponibilité foncière en 2008 (Source : CPDT, 2012).

Le temps d'accès en voiture à Liège d'une part, à Charleroi d'autre part, est très significatif. Les coefficients, respectivement -24,57 et -14,55 euros par seconde supplémentaire, font tous les deux état d'une moins-value. Celle-ci est environ une fois et demie plus importante à Charleroi qu'à Liège. Le fait qu'il n'existe pas, au sein du bassin d'emploi de Charleroi, de pôle secondaire

peut expliquer cette diminution plus rapide. Au sein du bassin d'emploi de Liège, de petits pôles secondaires comme Verviers, Huy et Visé apportent peut-être un rôle modérateur dans cette décroissance. Il est également possible que la moins-value soit accentuée par l'impact du temps d'accès à Bruxelles qui n'a pu être étudié en raison d'une corrélation excessive.

5.2. Variables comparables avec prudence

La part modale du train engendre une plus-value dans le bassin d'emploi de Charleroi, au contraire de celui de Liège. Ceci avait auparavant été expliqué par une distance plus courte entre Charleroi et Bruxelles qu'entre Liège et Bruxelles. Considérant que le train est une alternative intéressante à la voiture pour l'accès à Bruxelles, la plus-value carolorégienne est compréhensible. A l'inverse, ce statut d'alternative ne semble pas perçu comme tel en province de Liège (Direction Mobilité, 2010 ; cité par P. Dethier, 2012).

La localisation au sein d'un lotissement entraîne une influence opposée sur le prix entre les deux bassins d'emplois. Les raisons avaient déjà été avancées dans les précédents chapitres. Si nous considérons que la création d'une parcelle en lotissement avec ouverture de voirie nécessite un coût moyen de 15 000 euros (CDPT, 2012) et que le promoteur foncier va répercuter ce coût sur le prix qu'il proposera à la vente du terrain, la plus-value de 9450,16 euros observée dans le bassin d'emploi de Charleroi est cohérente. A l'inverse, la moins-value déterminée dans le bassin d'emploi de Liège peut paraître surprenante, mais elle ressort à la fois dans le modèle global et le modèle périphérique. Il semble que l'offre en terrain plus importante à Liège permette aux demandeurs de trouver plus facilement un terrain hors lotissement, ce qui leur autorise une plus grande liberté vis-à-vis des choix de construction et d'aménagement. Dès lors, les terrains en lotissements sont caractérisés par une certaine surabondance (Compagnie des Notaires de Liège, 2013), celle-ci entraînant un abaissement du prix de ces biens, toutes choses égales par ailleurs, et ce conformément aux lois de l'offre et de la demande.

L'influence du bruit du trafic aérien à raison de plus de 60 dB semble plus importante à Charleroi qu'à Liège. Ceci peut s'expliquer par la différence de trafic aérien entre les deux aéroports. Ainsi, en 2012, l'aéroport de Charleroi a connu 44 414 décollages et atterrissages, soit une moyenne de 121,68 mouvements par jour. Pour la même année, l'aéroport de Liège a connu 24 674 décollages et atterrissages, soit une moyenne de 67,60 mouvements par jour. De plus, le trafic était en progression à Charleroi entre 2011 et 2012, au contraire de Liège (Statbel, 2014).

Dans les deux bassins d'emplois, la présence d'une ligne à haute tension à moins de 500 mètres engendre une moins-value (-3909,92 euros à Liège, -4655,26 euros à Charleroi). L'influence semble plus forte dans le bassin d'emploi de Charleroi que dans celui de Liège. Néanmoins, la non-significativité de cette variable dans le modèle périphérique incite à la prudence. Il est cependant possible que le nombre de biens concernés soit trop faible pour rendre compte de l'éventuelle moins-value.

5.3. Variables comparables par seuils différents

La présence d'une école secondaire à moins de 60 secondes engendre une importante et très significative plus-value à Charleroi (26 395,70 euros). Dans le modèle périphérique liégeois, cette dernière s'exprime également, mais à moins de 120 secondes d'une école secondaire, et sa valeur

est légèrement moins élevée (18 621,20 euros). Le faible nombre de biens concernés empêche sans doute la plus-value de s'exprimer via le seuil des 60 secondes.

La proximité d'une ligne de chemin de fer entraîne une moins-value au seuil de 500 mètres pour le bassin de Charleroi. Cela n'est pas le cas pour le bassin liégeois considéré dans sa totalité. Cependant, cette moins-value apparaît bel et bien dans le modèle périphérique ; c'est au seuil des 700 mètres qu'elle s'exprime le mieux, avec une valeur de -6037,77 euros, très similaire à celle de Charleroi (-7564,38 euros). L'explication avancée était que les nuisances du chemin de fer avaient un impact plus fort en-dehors de l'agglomération liégeoise, alors qu'au sein de celle-ci, elles se diluent dans l'environnement bruyant de la ville. Ceci est cohérent avec les travaux de Y. Liu *et al.* (2010).

La présence d'un commerce alimentaire à moins de 700 mètres entraîne une plus-value égale à 17 939,60 euros dans le bassin d'emploi de Charleroi. Le modèle périphérique liégeois met cette plus-value en évidence pour le seuil des 500 mètres (10 304,60 euros). A Charleroi comme à Liège, la proximité d'un commerce alimentaire semble influencer positivement le prix des terrains à bâtir concernés. Néanmoins, il convient de rester prudent pour cette plus-value comme pour celle procurée par la présence des écoles secondaires. En effet, ces influences ne s'expriment pas au sein du modèle général liégeois.

5.4. Variables difficilement comparables

L'ensemble des variables entrant dans cette catégorie ne peut permettre une comparaison entre les différents modèles. Ceci pour les raisons suivantes :

- les échantillons n'étant pas très fournis, il est relativement difficile de mettre en évidence l'influence de l'environnement, notamment des nuisances s'exerçant sur de faibles distances. Peu de biens sont en effet concernés par ces nuisances ;
- les contextes comparés sont très différents compte tenu notamment de l'existence des barrières visuelles et sonores qui tendent à limiter ou faire disparaître l'influence de certains déterminants ;
- enfin, les variables d'environnement physique interviennent de manière réduite dans l'explication des prix des terrains à bâtir. L'étude des coefficients standardisés pour le modèle carolorégien et les deux modèles liégeois a clairement mis en évidence ce phénomène. Cela rejoint les constatations faites par Y. Grignet (2011) lors de l'étude du marché des maisons unifamiliales dans la région urbaine de Liège.

La présence d'un bâtiment sur le terrain est seulement significative à Charleroi. Il est difficile de conclure que cette même variable n'a pas d'influence sur les prix à Liège. Le critère ainsi créé reflète avant tout les caractéristiques de l'échantillon plutôt que celles du marché dans son ensemble.

Le temps d'accès aux nodules commerciaux entraîne des comportements opposés des prix. Ainsi, dans le bassin d'emploi de Liège, la moins-value de 30,54 euros portée au prix des terrains pour chaque seconde supplémentaire est cohérente avec ce qui est attendu. A l'inverse, l'impact de cette même variable dans le bassin d'emploi de Charleroi n'est pas clair. Cette variable est probablement polluée par d'autres facteurs tels que le milieu urbain dense et les nuisances qui en découlent.

La présence d'une école maternelle ou primaire à moins de 30 secondes en voiture entraîne une plus-value dans le bassin d'emploi de Charleroi. Leur présence semble procurer une utilité, ce qui n'est mis en évidence dans le bassin d'emploi de Liège. Après avoir étudié la localisation des biens concernés dans le bassin d'emploi de Liège, nous avons pu mettre en évidence qu'un bon nombre d'entre eux est situé à proximité d'industries. Les nuisances viennent sans doute masquer l'utilité procurée par la présence d'une école maternelle ou primaire.

La distance à un arrêt de bus entraîne également des influences opposées selon le bassin d'emploi. A Charleroi, la moins-value de 14,64 euros par mètre supplémentaire est cohérente avec ce qui peut être attendu : les ménages prennent en compte l'utilité procurée par la proximité d'un arrêt. La plus-value – 5,20 euros par mètre supplémentaire – observée à Liège est sans doute le résultat de nuisances. Le réseau liégeois est en effet développé le long des grands axes routiers, là où le réseau carolorégien propose, quant à lui, une desserte intéressante y compris sur des axes à trafic moins dense. Nous avons vérifié la localisation des arrêts de bus par rapport aux routes nationales et aux autoroutes. En considérant de manière arbitraire un buffer de 500 mètres autour de ce réseau, nous constatons une proportion plus importante d'arrêts localisés dans ce buffer pour le bassin d'emploi de Liège (62,04 % contre 54,73 % à Charleroi). D'autre part, les arrêts dans leur ensemble sont situés dans des milieux plus denses à Liège qu'à Charleroi. En faisant la moyenne des densités desservies par les arrêts de bus, nous obtenons une moyenne de 1944,82 habitants/km² pour le bassin d'emploi de Liège et de 1678,36 habitants/km² pour le bassin d'emploi de Charleroi. Nous pouvons ajouter que la part modale du bus ne semble avoir aucune influence sur le prix des terrains à bâtir. Outre les différences de comportement des prix constatées entre Liège et Charleroi que nous imputons en partie à d'autres influences, nous pouvons avancer que le bus ne semble pas présenter une réelle alternative de mobilité susceptible d'impacter de manière conséquente le prix des terrains à bâtir.

L'impact de la proximité d'une gare est difficilement comparable en raison des différences de contexte. De plus, l'existence d'une plus-value pour le modèle périphérique liégeois apporte une nuance quant à la supposée absence d'influence de la présence du train. Une pollution de l'impact de la variable par d'autres éléments est possible.

L'influence de la part modale des modes lents n'est pas comparable en raison de la pollution de celle-ci par d'autres facteurs.

Il est compréhensible que les ménages recherchent des rues calmes, calme qu'ils trouvent notamment dans les voiries en cul-de-sac du bassin d'emploi de Liège (plus-values de 4738,63 euros pour le modèle général, de 5848,49 euros pour le modèle périphérique). Il est probable que la plus-value puisse exister dans le bassin d'emploi de Charleroi. Celle-ci se serait sans doute exprimée si les biens concernés de l'échantillon carolorégien ne se situaient pas pour une grande part à proximité de grands axes routiers et en milieu urbain dense.

Une part de bâti croissante dans un rayon de 200 mètres entraîne une moins-value dans le bassin d'emploi de Charleroi (-108,76 euros par pourcent). Dans celui de Liège, ce n'est pas le cas. Il est cependant délicat de tirer des conclusions sur base de ces différences. Il est constaté un impact positif d'une part croissante de maisons à quatre façades (376,20 euros par pourcent). Le fait que le revenu, variable modérément corrélée aux deux premières, ait un comportement inexplicable sur le prix des terrains à bâtir du bassin d'emploi de Charleroi peut apporter un éclairage.

L'influence du revenu s'exprime en partie au travers de la part de bâti et la part de maisons à quatre façades.

La proximité d'un cours d'eau est significative pour le seuil des 100 mètres dans le bassin d'emploi de Liège (moins-values de 6105,10 euros pour le modèle général, de 4904,93 euros pour le modèle périphérique). Aucun impact n'est mis en évidence pour le bassin de Charleroi. Nous pensons à une pollution de l'influence des cours d'eau par d'autres éléments environnementaux, comme nous l'avons exposé lors de l'analyse du modèle général liégeois. Nous préférons dès lors ne pas tirer de conclusion sur ce point dans la comparaison entre les deux bassins d'emploi.

La présence d'une route au trafic de plus de 20 000 véhicules par jour à moins de 200 mètres entraîne une moins-value de 19 604,10 euros dans le bassin d'emploi de Liège. Cette moins-value ne se retrouve pas à Charleroi. Néanmoins, comme cela a été expliqué auparavant, la variable pour Charleroi a été éliminée de justesse (p-value de 0,108). Il existerait donc bien un effet, mais celui-ci ne peut pas être considéré comme significatif. L'explication provient probablement du fait que les axes concernés sont moins chargés à Charleroi qu'à Liège. En effet, certains axes liégeois voient passer plus de 50 000 véhicules par jour, là où la plupart des biens carolorégiens concernés sont situés à proximité d'axes voyant passer 20 000 – 30 000 véhicules par jour. L'effet des barrières visuelles et sonores ne peut être négligé. Ainsi, la comparaison est rendue malaisée par deux contextes très différents. Il en va de même pour la proximité des accès autoroutiers.

La présence d'un site d'activité économique désaffecté à moins de 100 mètres entraîne une importante moins-value de 32 802,00 euros à Charleroi. Le même phénomène n'est cependant pas observé à Liège. Nous expliquons cela par l'absence de barrières visuelles pour les biens concernés dans le bassin de Charleroi : ils sont en effet situés dans des lieux peu urbanisés, avec une vue directe sur les friches en question, notamment à Couillet, Châtelineau et Luttre.

Pour des raisons précédemment évoquées, nous n'avons pas considéré le revenu médian agrégé au niveau des anciennes communes au modèle de Charleroi, le rôle de cette variable n'étant pas clair et les paramètres du modèle étant dégradés. A titre d'indication, ce modèle montrait une plus-value de 1,87 euro par euro de revenus supplémentaire. Ceci est proche de la contribution du revenu par secteurs statistiques dans le bassin d'emploi de Liège, où le coefficient s'élève à 1,12 euro pour le modèle général, à 1,55 euro pour le modèle périphérique.

5.5. Conclusion

La comparaison entre les modèles des deux bassins d'emplois est malaisée, notamment pour les variables d'environnement. Nous avons pu mettre en évidence, pour plusieurs d'entre elles, des phénomènes de pollution par d'autres éléments étrangers à ces variables. Il existe des particularités ultra-locales, comme par exemple les barrières visuelles et sonores, qui ne peuvent être aisément modélisées. Ces particularités sont supposées être en partie responsables des différences observées entre les bassins d'emploi de Liège et de Charleroi. Ce phénomène est d'autant plus important que le nombre de biens concernés par les différentes variables d'environnement est parfois faible. L'appel aux résultats du modèle périphérique liégeois a néanmoins permis de relativiser l'apparente non-significativité de certains impacts environnementaux.

6. Modélisation du prix des terrains à bâtir de la fin des années nonante

L'occasion nous a été offerte de développer une comparaison historique entre la situation des années nonante et la situation actuelle. Les données historiques couvrant une grande part des arrondissements de Liège et de Waremme, nous avons, outre le modèle historique, tenté de développer un modèle actuel sur base des observations situées au sein de cette même zone d'étude. La liste des communes en faisant partie est disponible à l'annexe III. Les résultats de ce modèle actuel sont toutefois décevants : très peu de variables sont significatives, et la distribution non-normale des résidus montre que l'ajustement n'est pas correct. Nous utiliserons toutefois ces résultats, combinés à ceux des modèles général et périphérique liégeois pour établir la comparaison temporelle.

6.1. Sélection des observations

L'échantillon a été délimité sur base des critères similaires à ceux employés pour le développement des modèles actuels :

- superficies de 100 à 1600 m² ;
- prix inférieurs à 46 543 euros de 1997. En considérant une multiplication du prix moyen des terrains sur la zone d'étude de 3,76 et la limite de 175 000 euros employée pour les modèles actuels, nous obtenons ce seuil. La multiplication du prix moyen a été calculée sur base des données concernant le prix du mètre carré de terrain fournies par le SPF Economie ;
- élimination des biens situés en-dehors des zones d'habitat et d'habitat à caractère rural (situation arrêté pour l'année 1999).

6.2. Modélisation du prix des terrains à bâtir de la fin des années nonante

6.2.1. Développement du modèle

Compte tenu des données à notre disposition, nous avons premièrement choisi de développer un modèle avec les transactions de l'année 1998. Nous avons cherché à minimiser au maximum les erreurs dues à la localisation imparfaite liée à l'absence d'adresses complètes. Nous avons testé plusieurs développements en utilisant différents seuils de longueur de rue³³. Cependant, en diminuant le seuil, nous nous retrouvons rapidement confronté à un trop faible nombre de données. Nous avons dès lors choisi d'intégrer les données de 1996 à l'échantillon.

L'échantillon contient ainsi 565 biens. Nous avons une nouvelle fois testé plusieurs développements selon la longueur des rues, en choisissant les seuils de 150, 300 et 500 mètres :

- nous n'avons pas retenu le modèle développé avec les rues de moins de 150 mètres, en raison d'un mauvais ajustement de celui-ci, mis en évidence dans le diagramme quantile-quantile des résidus. Ceux-ci ne sont absolument pas alignés le long de la droite de Henry ;

³³ Nous avons pris soin, lors de la préparation des données, de noter la longueur de la rue. Voir point 2.2.5.

- le modèle développé sur base des rues de moins de 500 mètres présente des paramètres robustes et un bon pouvoir explicatif. Néanmoins, le R^2 ajusté n'est pas meilleur que celui du troisième modèle, développé avec les rues de moins de 300 mètres et comptant dès lors moins d'observations. En relativisant le nombre de biens, nous constatons que le troisième modèle est le plus satisfaisant, c'est pourquoi nous l'avons retenu. Nous détaillons brièvement son développement et nous attardons sur les résultats dans ce chapitre.

6.2.1.1. Statistiques descriptives

Lors du développement des statistiques descriptives, nous nous sommes retrouvés confrontés à une distribution non-normale des valeurs de prix (Figure 35). Etant donné qu'aucune transformation ne semblait pouvoir rendre cette distribution normale, nous avons effectué un écrémage de l'échantillon pour les valeurs de prix inférieures à 10 000 euros. Nous avons retiré aléatoirement 21 biens de l'intervalle 0 – 5000 euros et 26 biens de l'intervalle 5000 – 10000 euros. En procédant de la sorte, l'histogramme des valeurs de prix est plus proche d'une distribution normale (Figure 36).

Les figures 37 et 38 montrent que les valeurs de superficie et de pente du terrain forment un histogramme continu. Il n'y a pas de valeurs extrêmes à retirer. L'échantillon ainsi constitué compte 266 biens. La figure 39 montre la répartition des biens au sein de la zone d'étude. Une concentration plus importante s'observe au sein de l'agglomération de Liège.

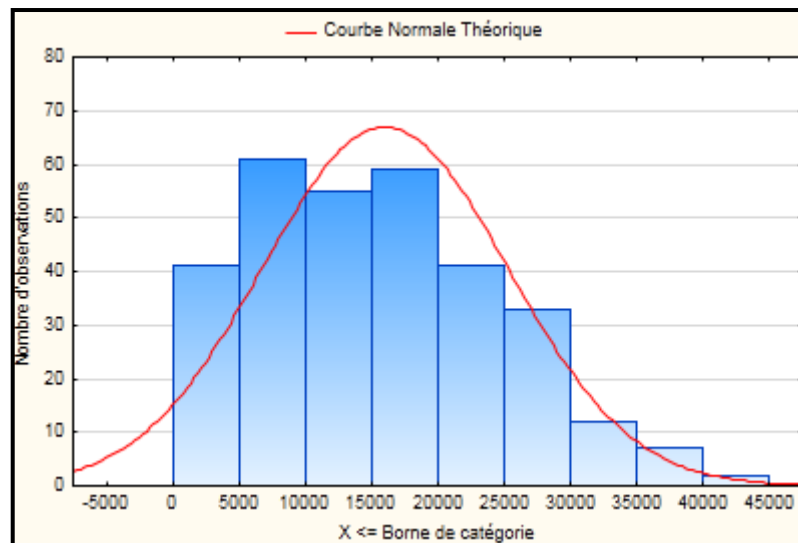


Figure 35 – Distribution des valeurs de prix avant écrémage.

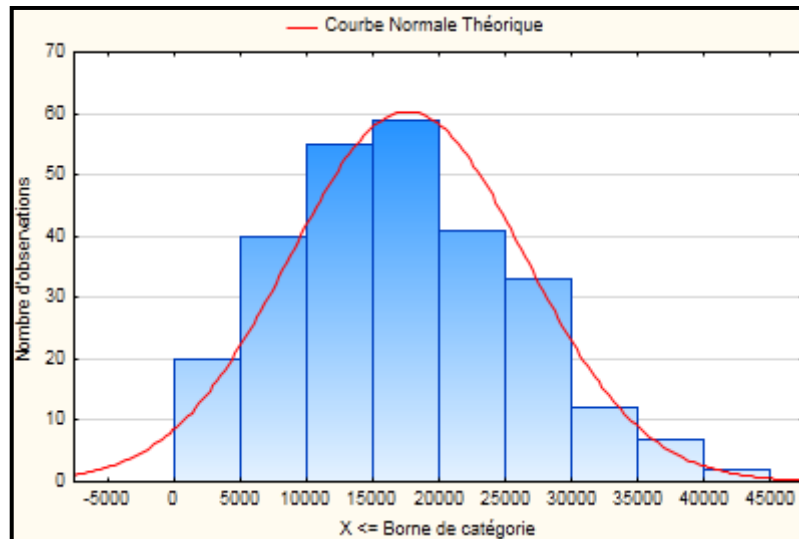


Figure 36 – Distribution des valeurs de prix après écrémage.

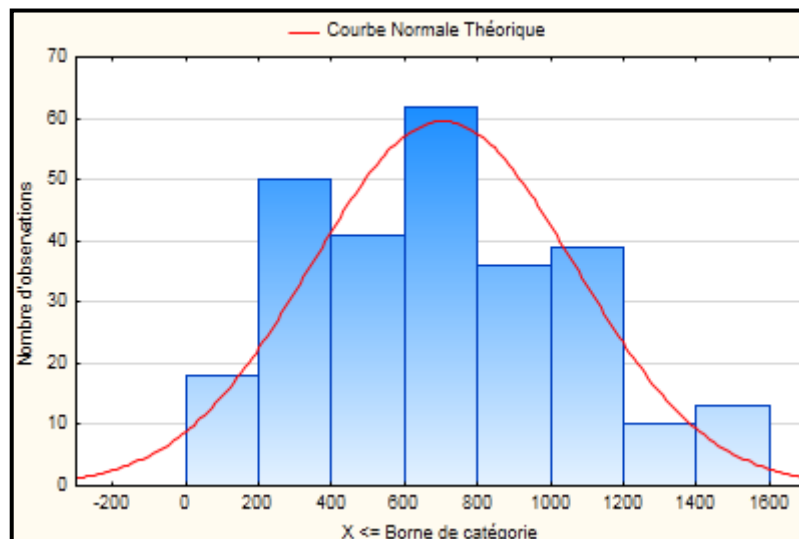


Figure 37 – Distribution des valeurs de superficie.

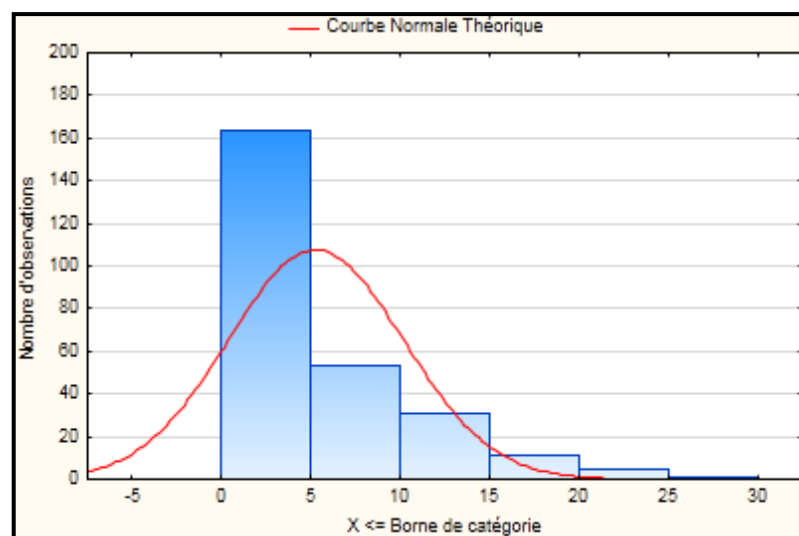


Figure 38 – Distribution des valeurs de pente.

Variables	Minimum	Maximum	Moyenne	Ecart-type
Prix (euros)	1041,15	42141,90	17626,17	8890,45
Superficie (m ²)	105,0	1532,0	704,92	360,66

Tableau 23 – Statistiques descriptives pour le prix et la superficie.

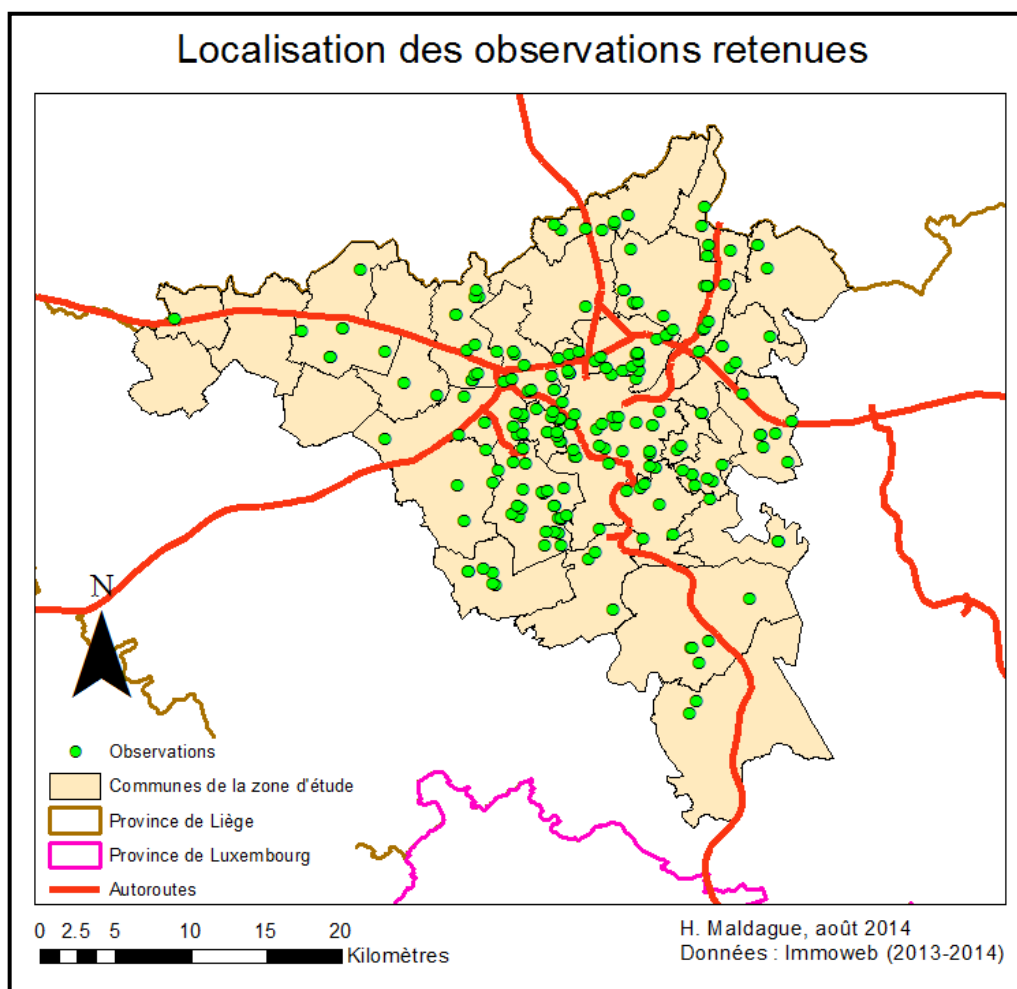


Figure 39 – Répartition des biens retenus au sein de la zone d'étude.

6.2.1.2. Analyse des corrélations entre variables explicatives

Parmi les variables concernées par des problèmes de corrélation excessive, nous retrouvons des couples similaires à ceux mis en évidence lors du développement des modèles liégeois. La liste des couples concernés est disponible à l'annexe XXII. Pour rappel, le seuil critique retenu est 0,75 en valeur absolue. Une nouvelle fois, nous choisissons de considérer la distance à Tihange, le temps d'accès à Liège et le temps d'accès à Bruxelles en raison de l'importance de ces localisations.

6.2.1.3. Modélisation par MCO

Le modèle développé par MCO montre une certaine robustesse statistique accompagnée d'un pouvoir explicatif à améliorer (Tableau 24). Le test de colinéarité (annexe XXIII) écarte tout problème de ce type étant donné l'éloignement de la valeur la plus élevée de VIF (temps d'accès à Liège : 1,56) des seuils critiques. Le test de Breusch-Pagan (annexe XXIV) indique cependant une présence d'hétéroscédasticité. Une variable est concernée : la superficie.

Moy. Var. dep	17392,36	Ec. typ. var. dep.	8658,976
Somme carré résidus	6,97e+09	Ec. typ. de régression	5238,844
R ²	0,649147	R ² ajusté	0,633952
F (11,254)	42,72265	p. critique (F)	1,53e-51
Log de vraisemblance	-2649,284	Critère d'Akaike	5322,568
Critère de Schwarz	5365,570	Hannan-Quinn	5339,843

Tableau 24 – Paramètres du modèle retenu développé par MCO.

6.2.1.4. Correction de l'hétéroscédasticité

Le modèle à hétéroscédasticité corrigée présente les paramètres suivants (Tableau 25) :

Moy. Var. dep	17392,36	Ec. typ. var. dep.	8658,976
Somme carré résidus	7,85e+09	Ec. typ. de régression	5547,494
R ²	0,694307	R ² ajusté	0,682319
F (10,249)	57,91706	p. critique (F)	6,63e-60
Log de vraisemblance	-582,7801	Critère d'Akaike	1187,560
Critère de Schwarz	1126,979	Hannan-Quinn	1203,396

Tableau 25 – Paramètres du modèle corrigeant l'hétéroscédasticité.

L'amélioration des performances du modèle est notable, avec un F plus élevé et un R^2 ajusté également en augmentation. Une nouvelle analyse de la colinéarité (annexe XXIII) écarte à nouveau ce problème.

6.2.1.5. Suppression des observations à résidus extrêmes et ajustement du modèle

Le logiciel Gretl repère six résidus supérieurs à 2,5 fois l'erreur standard. Deux d'entre eux présentent des prix élevés, les quatre autres sont moins chers. Ce sont des terrains mal situés et pour lesquels les nuisances de l'environnement font diminuer leur prix.

Nous supprimons les six observations ayant des résidus extrêmes et ajustons notre modèle. Ses paramètres sont présentés ci-dessous (Tableau 26).

Moy. Var. dep	17227,19	Ec. typ. var. dep.	8461,906
Somme carré résidus	6,47e+09	Ec. typ. de régression	5097,536
R ²	0,723317	R ² ajusté	0,712205
F (10,249)	65,09457	p. critique (F)	8,25e-71
Log de vraisemblance	-553,8909	Critère d'Akaike	1129,782
Critère de Schwarz	1168,949	Hannan-Quinn	1145,528

Tableau 26 – Paramètres du modèle corrigeant l'hétéroscédasticité après suppression des résidus délinquants.

Le modèle est assez bon, présentant une robustesse statistique certaine et un pouvoir explicatif satisfaisant. Celui-ci ne peut sans doute être considérablement amélioré compte tenu des localisations approximatives des biens au centre des rues concernées. Le pouvoir prédictif est assez médiocre avec une erreur relative de 29,07 %.

L'analyse des résidus montre une distribution normale de ceux-ci, ceci étant confirmé par le test du Chi carré (Figure 40). Le graphique quantile-quantile (Figure 41) confirme la distribution et montre, globalement, un alignement des résidus proche de la droite de Henry. L'ajustement du modèle n'est toutefois plus parfait pour les résidus élevés en valeur absolue, pour des raisons déjà maintes fois évoquées.

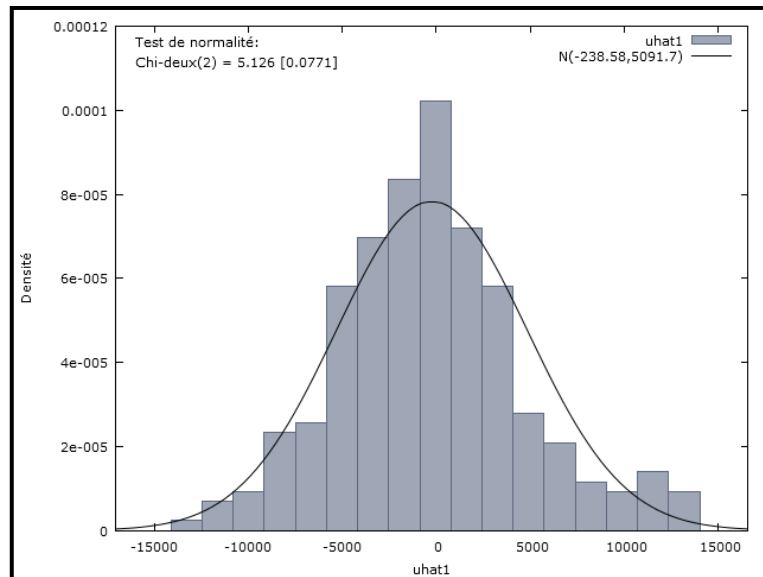


Figure 40 – Distribution des résidus après épuration.

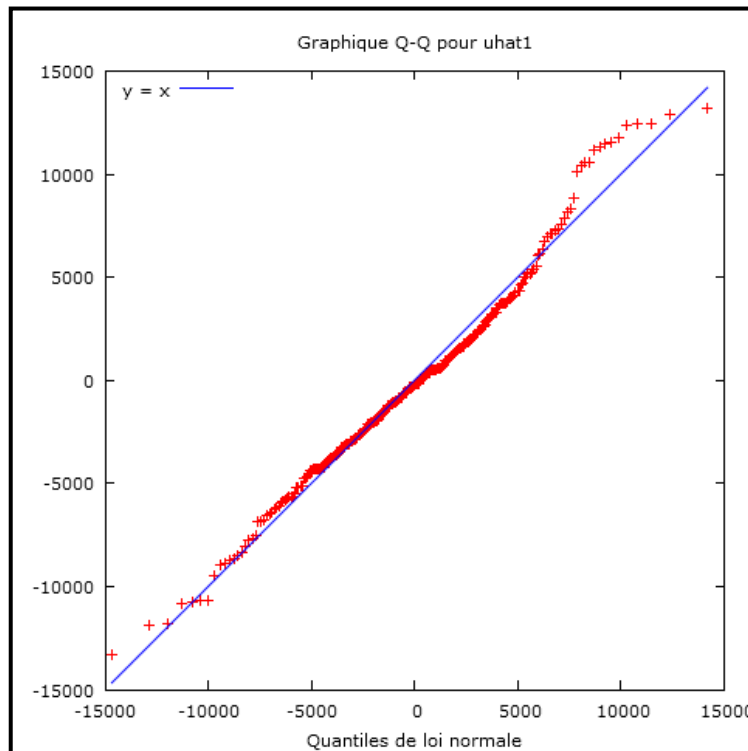


Figure 41 – Diagramme quantile-quantile des résidus.

6.2.1.6. Analyse de l'autocorrélation spatiale des résidus

L'autocorrélation spatiale des résidus est absente du modèle. L'indice de Moran (-0,0048) n'est pas statistiquement différent (p-value : 0,491) de l'indice attendu (-0,0039). Le diagramme de dispersion est disponible à l'annexe XXV. L'analyse de l'autocorrélation locale ne montre pas de concentrations étendues de résidus importants, si ce n'est dans le nord-ouest de l'agglomération. Cette concentration peut être rattachée à l'effet de perception et de valorisation rencontré pour les modèles actuels (Figure 42).

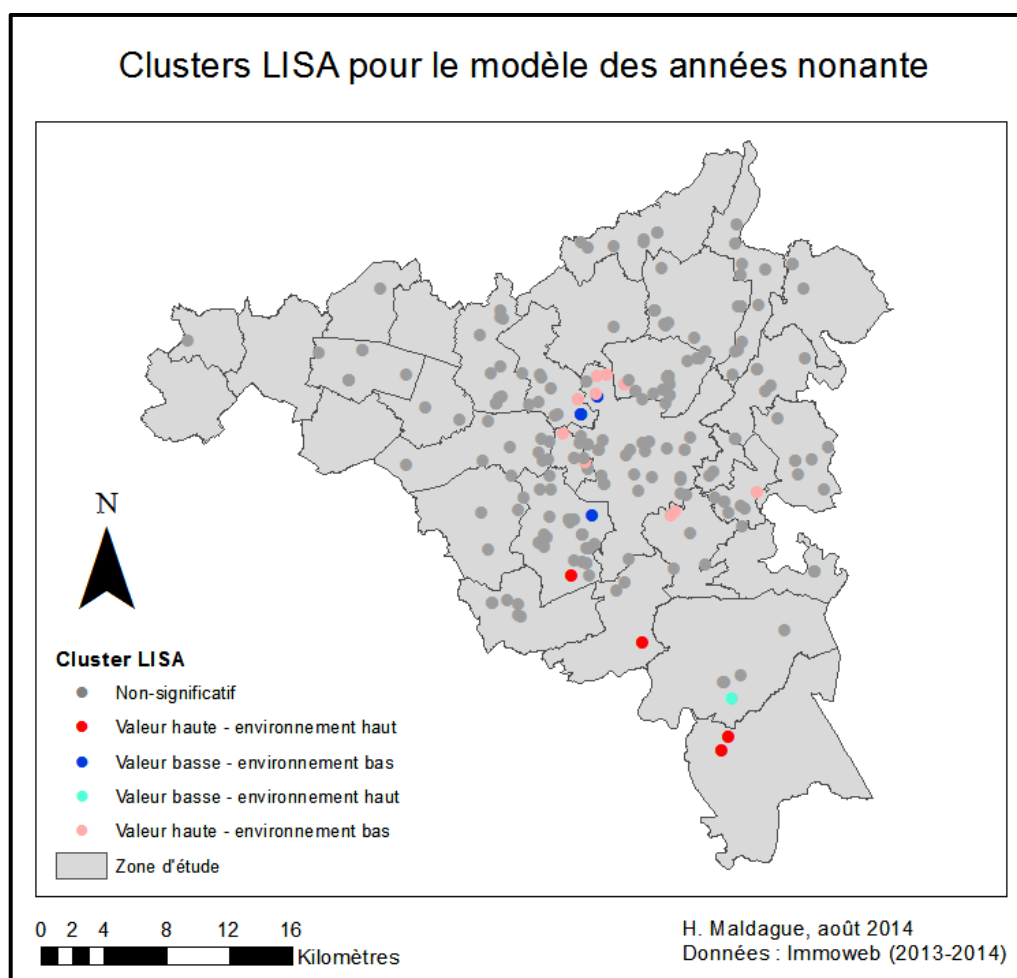


Figure 42 – Clusters LISA pour le modèle des années nonante.

6.2.2. Présentation des résultats

Outre les résultats du modèle (Tableau 27), nous présentons également les coefficients standardisés recalculés manuellement (Tableau 28).

	<i>Coefficient</i>	<i>Erreur Std</i>	<i>t de Student</i>	<i>p. critique</i>	
Constante	-1968,83	2747,15	-0,7167	0,47424	
Superficie	19,3306	0,966769	19,9951	<0,00001	***
Pente au niveau du terrain (%)	-298,163	60,8481	-4,9001	<0,00001	***
Temps d'accès voiture à Liège (s)	-11,6948	1,34375	-8,7031	<0,00001	***
Part modale du bus	95947	18377,8	5,2208	<0,00001	***
Sortie d'autoroute à moins de 500 m	-4595,21	1366,73	-3,3622	0,00089	***
Cours d'eau à moins de 100 m	3254,91	1531,15	2,1258	0,03450	**
Distance à Tihange	0,156804	0,0460966	3,4016	0,00078	***
Densité de la population	-0,333122	0,173625	-1,9186	0,05617	*
Industrie à moins de 200 m	-3174,07	935,79	-3,3919	0,00081	***
Revenus médians ³⁴	0,315373	0,0600977	5,2477	<0,00001	***
*p-value inférieure à 0,10 **p-value inférieure à 0,05 ***p-value inférieure à 0,01					

Tableau 27 – Résultats du modèle liégeois de la fin des années nonante.

³⁴ Corrigés de l'inflation : ramenés en euros constants 2013.

Variable	B*	B
Superficie	0,82604	19,3306
Temps d'accès voiture à Liège	-0,38973	-11,6948
Part modale du bus	0,24718	95947
Pente au niveau du terrain	-0,19379	-298,163
Revenus médians	0,18338	0,315373
Cours d'eau à moins de 100 m	0,11540	3254,91
Distance à Tihange	0,11412	0,156804
Sortie d'autoroute à moins de 500 m	-0,10861	-4595,21
Industrie à moins de 200 m	-0,09040	-3174,07
Densité de population	-0,08233	-0,333122

Tableau 28 – Classement des variables explicatives par ordre d'importance.

Nous avons également corrigé les coefficients non-standardisés de l'inflation (infl.) d'une part et de l'augmentation du prix des terrains (terr.) d'autre part, afin de pouvoir mener la comparaison avec les modèles actuels par la suite (Tableau 29). Les données d'inflation et celles relatives aux prix moyens des terrains proviennent de l'Institut National des Statistiques.

Variable	Euros 1997	Euros 2013 (infl.)	Euros 2013 (terr.)
Superficie	19,33	26,75	72,62
Temps d'accès voiture à Liège	-11,69	-16,19	-43,94
Part modale du bus	95947	132810	360760,72 ³⁵
Pente au niveau du terrain	-298,163	-412,72	-1121,09
Revenus médians	0,32	0,44	1,18
Cours d'eau à moins de 100 m	3254,91	4505,45	12227,39
Distance à Tihange	0,16	0,22	0,59
Sortie d'autoroute à moins de 500 m	-4595,21	-6360,69	-17262,37
Industrie à moins de 200 m	-3174,07	-4393,55	-11923,71
Densité de population	-0,33	-0,46	-1,25

Tableau 29 – Correction des valeurs des coefficients de régression.

Nous retrouvons ici l'ordre général rencontré dans chaque modélisation de ce travail. La superficie est la variable qui contribue le plus à l'explication du prix. Elle est suivie par les déterminants d'accessibilité, la pente du terrain et les revenus. Les variables d'environnement physique expliquent plus faiblement les prix.

Dans la présentation des résultats ci-dessous, nous exprimerons les plus-values et les moins-values engendrées par les différentes variables en euros corrigés de l'augmentation du prix des terrains.

6.2.2.1. Variables structurelles

La superficie reste la première variable explicative en termes d'importance. Corrigée de l'augmentation des prix des terrains, nous obtenons une valeur de 76,40 euros par mètre carré. A l'instar des modèles actuels, nous cherchons à mettre en évidence la marginalité décroissante des prix en fonction de la superficie en scindant notre échantillon sur base des quartiles. Les résultats sont présentés ci-dessous (Tableau 30). Les coefficients sont corrigés de l'augmentation du prix des terrains.

³⁵ Le coefficient est ici exprimé pour une augmentation de 100 % de part modale. Le prix augmente de 3607,6 euros par pourcent de part modale supplémentaire.

Intervalle de superficie	coefficient	p-value
0 – 395	44,55	0,0014
396 – 690	131,44	<0,00001
690 - 994	69,49	0,00581
995 - 1600	65,33	0,00001

Tableau 30 – Illustration de la marginalité décroissante des prix selon la superficie.

Nous constatons que la marginalité décroissante n'est pas mise en évidence dans ce cas-ci. Plutôt que de conclure à son absence, nous préférons ne pas tirer de conclusions, les effectifs utilisés étant une nouvelle fois réduits.

Une pente grandissante entraîne une moins-value sur les prix (-1121,09 euros par pourcent supplémentaire). Cette diminution du prix avec la pente est compréhensible dans la mesure où un terrain pentu nécessitera des travaux de terrassement. Il est donc moins recherché, entraînant une diminution de son prix.

L'orientation au sud n'apporte aucune plus-value. Nous avons déjà évoqué les défauts de cette variable auparavant, à savoir que les effets bénéfiques ou néfastes de l'orientation sont liés à une pente de plus en plus forte.

6.2.2.2. Variables d'accessibilité

Le temps d'accès en voiture à Liège reste une variable fortement explicative. Le coefficient, une fois corrigé de l'augmentation du prix des terrains, est environ trois fois plus important que celui mis en évidence pour les modèles actuels, reposant sur des zones d'étude différentes. Aussi, le bien le plus éloigné de la Place Saint-Lambert en terme de temps, à 25 minutes et 30 secondes, est 69 864,60 euros moins cher que le même bien sur la place même.

La part modale du bus est significative : un terrain bien desservi connaît une plus-value de 3607,60 euros par pourcent supplémentaire. Entre le bien le moins desservi et celui le mieux desservi, il existe une différence de 46 898,80 euros. Le fait que bon nombre de biens sont situés dans l'agglomération liégeoise, caractérisée par une bonne desserte des bus, amènerait les ménages à considérer la disponibilité en transports en commun pour se rendre au centre-ville par exemple. La proximité d'un arrêt de bus n'engendre par contre aucune influence. A l'instar des propos tenus dans les modélisations actuelles, les nuisances liées à un environnement bâti dense ne sont pas à négliger : elles peuvent masquer l'utilité procurée par la proximité d'un arrêt de bus.

La proximité d'un RAVeL n'apporte aucun impact sur les prix. Il en va de même pour la part modale des modes lents. Les ménages intéressés par un terrain ne semblent pas considérer ces deux paramètres lors de l'achat d'un terrain. En effet, les déplacements à pied et à vélo étaient, comme ils le sont encore aujourd'hui, relativement minoritaires par rapport à d'autres modes de transport en Wallonie.

Ni la proximité des gares, ni la part modale du train n'engendrent d'impact sur les prix. Sans doute, à l'image de la situation actuelle, le train n'offrirait-il pas de réelle alternative à la voiture notamment (Direction Mobilité, 2010).

6.2.2.3. Variables d'environnement physique

La proximité de l'eau engendre une plus-value de 12 227,39 euros si le cours d'eau en question est situé à moins de 100 mètres. Nombre de biens concernés sont situés en milieu urbain où la présence du cours d'eau apparaît comme une plus-value. Ils sont pour la plupart situés à l'écart des industries.

Sans surprise et bien que située en-dehors de la zone d'étude, la centrale nucléaire de Tihange apparaissait déjà comme un élément menaçant à la fin des années nonante. Si nous éliminons l'effet de l'augmentation des prix, nous obtenons un coefficient de 0,59 euro par mètre supplémentaire. Ainsi, le bien le plus éloigné, à 39,724 km, est plus cher de 23 437,16 euros corrigés de l'augmentation du prix des terrains, par rapport au même bien qui serait situé sur le site de la centrale, toutes choses égales par ailleurs.

La proximité d'une sortie d'autoroute entraîne une moins-value, sans doute en raison des nuisances générées par le trafic de l'autoroute et des axes menant à ces sorties. A noter que la proximité des routes supportant un trafic de plus de 10 000 ou 20 000 véhicules par jour n'est pas significative, quel que soit le seuil utilisé. Le rôle des barrières visuelles ne serait pas à négliger dans ce cas.

La densité de population est à peine significative. Par habitant/km² supplémentaire, le prix diminue de 1,25 euro, toutes choses égales par ailleurs. Dans l'ensemble, cette variable n'a pas un rôle très important dans l'explication des prix. Le fait que la densité soit appréciée différemment selon les localisations et les ménages (P. Dethier, 2012) pourrait expliquer ce constat.

La présence d'une zone industrielle à moins de 200 mètres entraîne une moins-value équivalente à 11 923,71 euros. Ceci est essentiellement dû au fait qu'un grand nombre de biens concernés ont une vue relativement dégagée sur ces sites industriels, notamment du côté de Seraing, Ans, Rocourt, de la Basse Meuse et de Grâce-Hollogne. Les barrières visuelles et sonores interviennent moins dans le cas présent. L'influence négative de la proximité des industries s'illustre donc plus clairement dans ce cas-ci.

Ni la part de bâti, ni la diversité de l'environnement n'engendre de plus ou de moins-value. Ceci contredit l'étude de J. Geoghegan *et al.* (1997). Pour la diversité, nous pouvons néanmoins reprendre les propos de Y. Grignet (2011), à savoir que la diversité est un concept difficilement jugeable.

La proximité d'une carrière n'engendre pas de moins-value sur les prix. Nous pouvons supposer que le faible nombre de biens concernés n'est pas suffisant pour rendre compte de cette éventuelle moins-value.

La présence d'une forêt ou d'un espace vert ne semble pas influencer les prix. Nous pouvons reprendre les mêmes arguments que ceux exposés pour les modélisations actuelles : la forêt peut être perçue de manière favorable (récréation) ou défavorable (danger en cas de tempête, barrière visuelle). De plus, leur relative abondance pourrait laisser les ménages indifférents.

La localisation en zone inondable n'entraîne pas de moins-value, sans doute en raison de la non-connaissance des risques et des étendues des éventuelles inondations pouvant survenir.

6.2.2.4. Variables de valorisation sociale

Le revenu est significatif. Corrigé de l'augmentation du prix des terrains, il s'élève à 1,18 euros. Cette valeur est cohérente et proche de celles mises en évidence par les modèles actuels liégeois.

Ni le taux de chômage, ni le taux de diplômés du supérieur n'a d'impact significatif sur les prix. Ceci pourrait indiquer qu'il n'y avait pas de convergence dans la perception de ces deux paramètres de la part des ménages intéressés par l'achat d'un terrain. Ils semblent davantage tenir compte du revenu. Ajoutons toutefois qu'il existe une corrélation modérée entre le taux de chômage et le revenu. De manière implicite, le revenu reprend donc une partie de l'impact du taux de chômage.

6.3. Conclusion

Malgré l'absence d'adresse complète dans nos données, l'utilisation des terrains à bâtir localisés dans les rues les plus courtes, ici moins de 300 mètres, permet une approche relativement correcte du comportement des prix des biens en fonction de leur accessibilité et du niveau socio-économique du quartier dans lequel ils se trouvent. A l'inverse, l'influence des différentes variables de l'environnement est plus ou moins bien cernée selon les cas. La densité de population et la proximité des industries et des cours d'eau entraînent des comportements de prix conformes à la littérature. A l'inverse, d'autres variables comme la part du bâti, la proximité des carrières, des lignes de chemin de fer et des routes à trafic important ne sont pas significatives, et ce alors que les différentes études consultées montrent l'impact de ces éléments sur les prix. Ceci indique une nouvelle fois que l'environnement est difficilement cerné au travers de nos modèles. Dans le chapitre suivant, nous tenterons d'établir une comparaison entre certains résultats du modèle des années nonante et les résultats des modélisations actuelles.

7. Comparaison temporelle liégeoise

Dans ce chapitre, nous menons la comparaison entre le modèle de la fin des années nonante, le modèle de comparaison et les modèles général et périphérique. Bien que ne reposant pas sur la même aire d'étude, nous avons choisi d'utiliser ces deux derniers développements afin de pallier les manquements de la modélisation des prix actuels sur la même zone d'étude que celle du modèle historique.

Le tableau ci-dessous (Tableau 31) reprend les variables comparées pour le modèle historique, le modèle de comparaison, le modèle général et le modèle périphérie liégeois. Nous avons regroupé les variables en deux catégories : celles pour lesquelles la comparaison est aisée car elles sont cohérentes avec la littérature et celles pour lesquelles la comparaison est plus délicate.

Variables	96-98 ³⁶	Comp. 13-14	Général 13-14	Périph. 13-14
Superficie	72,62 ***	66,58 ***	54,16 ***	52,48 ***
Distance à Tihange	0,59 ***		0,57 ***	0,90 ***
Revenus médians	1,18 ***	1,69 ***	1,12 ***	1,55 ***
Pente au niveau du terrain	-1121,09 ***		NS ³⁷	NS
Temps d'accès à Liège	-43,94 ***		-14,55 ***	-13,34 ***
Temps d'accès à Bruxelles	NS		-2,66 **	-7,68 ***
Part modale du bus	3607,61 ***		NS	NS
Sortie d'autoroute à moins de 500 m	-17262,37 ***		NS	NS
Sortie d'autoroute à moins de 700 m	NS		NS	-12117,4 **
Cours d'eau à moins de 100 m	12227,39 **		-6105,1 ***	-4904,93 ***
Cours d'eau à moins de 200 m		-11927,3 ***		
Densité de population	-1,25 *		NS	NS
Industrie à moins de 200 m	-11923,71 ***		NS	NS

Tableau 31 – Comparaison temporelle entre les modèles liégeois.

Le prix du mètre carré de terrain, considéré seul, semble avoir légèrement diminué au cours du temps. Pour la même zone d'étude, le coefficient est passé de 72,62 euros à 66,58 euros. Bien que la différence ne soit pas très importante, il est relativement difficile de trouver une cause à cette diminution. Nous pouvons émettre l'hypothèse que la valeur du mètre carré a diminué par rapport aux attributs de localisation.

Le coefficient de la distance à Tihange pour le modèle de la fin des années nonante est similaire à celui du modèle général actuel, et légèrement inférieur à celui du modèle périphérique. L'influence de la centrale nucléaire ne semble pas avoir évolué au cours du temps.

Fin des années nonante comme actuellement, le revenu médian, en tant que variable de valorisation sociale, exerce un impact relativement similaire sur les prix. Pour la même zone d'étude, la contribution d'un euro supplémentaire de revenus a légèrement augmenté, avec un coefficient passant de 1,18 à 1,69 euros. Ceci tend à montrer que la différence de prix entre un secteur à bas revenus et un secteur à hauts revenus s'est accentuée en un peu plus de quinze ans. Ce phénomène peut avoir plusieurs composantes. Nonobstant l'évolution du prix moyen des

³⁶ Coefficients exprimés en euros corrigés de l'augmentation du prix des terrains.

³⁷ NS : Non significatif.

terrains, un terrain dans un quartier à hauts revenus est plus cher aujourd'hui et/ou un terrain dans un quartier à bas revenus est moins cher.

Il est relativement difficile de tirer des conclusions quant à l'évolution de l'impact de la pente sur les prix des terrains. Nous ne pouvons qu'émettre une hypothèse, qu'il s'agirait de vérifier : une situation en pente semble moins problématique aujourd'hui qu'à la fin des années nonante. Les ménages seraient moins dérangés par la pente du terrain, d'autant plus que cette localisation peut offrir une vue agréable sur les zones situées en contrebas.

L'évolution du temps d'accès au centre-ville est également difficilement comparable, dans la mesure où notre modèle de comparaison n'a pas pu cerner l'impact de cette variable sur les prix. Néanmoins, il semble que la décroissance des prix engendrée par l'éloignement du centre-ville de Liège se soit amenuisée avec le temps. Ceci est peut-être le reflet d'une délocalisation d'activités en périphérie, mais aussi l'entrée de Liège au sein de logiques territoriales de niveau supérieur, polarisées en premier lieu par Bruxelles et, dans une moindre mesure, par Maastricht. Le temps d'accès à Bruxelles ne s'est d'ailleurs pas montré significatif pour la fin des années nonante, illustrant une attraction encore relativement faible de la capitale fédérale sur la métropole liégeoise.

Pour la fin des années nonante, la contribution de l'accessibilité en bus est cohérente avec ce qui peut être attendu sur base de la littérature. Nous avons toutefois avancé un effet de la répartition des observations. De plus, l'absence d'un modèle de comparaison nous empêche de confronter cette donnée avec la situation actuelle.

L'influence de la proximité d'une autoroute semble demeurer négative à travers le temps. Si nous considérons de plus les moins-values engendrées par la proximité des routes à fort trafic et mises en évidence dans le modèle général liégeois, nous constatons que le trafic routier important est demeuré une nuisance sonore dérangeante entre la fin des années nonante et la période actuelle. Chiffrer l'évolution des moins-values engendrées est par contre impossible.

Il est difficile de conclure à un changement d'impact des cours d'eau sur le prix des terrains. Nous pensons davantage à un effet de localisation des observations. Comme nous l'avons dit précédemment, la majorité des terrains concernés dans les années nonante sont éloignés d'éléments potentiellement perturbants comme les industries. Ces cours d'eau, en couronne urbaine, offrent un cadre verdoyant qui peut sans doute être considéré comme une aménité. A l'opposé, les biens concernés de la situation actuelle sont proches des industries. La comparaison temporelle de l'influence des cours d'eau n'est donc pas possible.

Il semble y avoir eu une disparition de l'influence de la densité de population sur les prix des terrains en une quinzaine d'années. Ceci pourrait indiquer qu'aujourd'hui, la densité ne soit plus perçue de la même manière, ce qui rejoint le constat de P. Dethier (2012). Néanmoins, il s'agit de rester une nouvelle fois prudent vis-à-vis de ces propos.

La comparaison est également délicate pour l'évolution de l'influence des industries proches. Nous avons exposé auparavant le problème des variables visuelles et sonores qui atténuent les nuisances. Ce même problème équivaut pour la proximité des lignes de chemin de fer et des carrières pour lesquelles aucune conclusion ne peut être tirée.

Le train ne semblait pas constituer une alternative valable à la voiture à la fin des années nonante. La plupart des modèles liégeois n'ont pas montré de plus-values engendrées par la proximité d'une gare, ni de la part modale du train. La seule exception provient du modèle périphérique liégeois qui présentait la proximité d'une gare L comme génératrice d'une plus-value. Il est ainsi possible que l'accès au train commence, depuis peu, à être envisagé comme un atout.

Par contre, les prix demeurent insensibles à la part modale des modes lents et à la proximité d'un RAVeL. L'argument d'un faible attrait pour le vélo et la marche explique probablement cette absence d'influence.

Ce chapitre a mis en lumière la difficulté de comparaison entre la situation de la fin des années nonante et la situation actuelle. Le fait que nous n'ayons pu développer un modèle correct de comparaison y est pour beaucoup. Nous pouvons ajouter à cela le fait qu'une nouvelle fois, l'influence des différentes composantes de l'environnement physique n'a pu être systématiquement mise en évidence. Parmi nos variables, certaines ont été polluées par des influences externes. Ainsi, la comparaison entre les variables représentatives de l'environnement physique est rendue très difficile.

Conclusion générale et perspectives

Cette première application de l'approche hédonique aux terrains à bâtir en Belgique – à notre connaissance – ouvre de réelles perspectives quant aux possibilités qu'offre ce champ d'étude, en regard de la littérature consultée sur ce domaine. A l'instar des mémoires concernant le marché des maisons unifamiliales à Liège et à Bruxelles, nous avons pu montrer qu'il était tout à fait possible d'envisager une approche similaire pour les terrains à bâtir et de la généraliser à différentes zones d'études. Nous avons pu, à notre manière, contribuer au développement de l'étude des marchés fonciers et immobiliers belges via l'approche hédonique.

Aussi bien pour le bassin d'emploi de Liège que pour le bassin d'emploi de Charleroi, nous avons ainsi pu construire des modèles robustes statistiquement et au pouvoir explicatif important. Nous sommes parvenus, grâce aux variables utilisées et à la méthodologie employée, à expliquer 80 à 90 % de la variation du prix des terrains à bâtir. Seule la performance prédictive se révèle non-optimale avec une erreur relative tournant autour de 20 %. Nous avons attribué une partie de cette erreur à l'inefficacité des marchés fonciers et immobiliers (A. W. Evans, 1995 ; cité par Y. Grignet, 2011). Lors de l'analyse de la contribution des différentes variables à la formation du prix, nous avons approché l'ordre mis en évidence par Y. Grignet (2011) pour le marché des maisons unifamiliales de la région urbaine de Liège : la superficie est la variable la plus explicative du prix, suivie par les caractéristiques d'accessibilité, le revenu et certains déterminants d'environnement physique responsables d'importantes nuisances sonores (bruit aérien, proximité d'une ligne de chemin de fer) ou perçus comme une menace (distance à la centrale nucléaire de Tihange). L'impact des autres variables d'environnement physique est plus discutable : certaines d'entre elles reprennent d'autres influences externes ou sont atténuées par des particularités locales parfois difficilement modélisables (notamment les barrières visuelles et sonores). Nous avons également pu mettre en évidence la marginalité décroissante du prix en fonction de la superficie selon la méthode utilisée par Y. Grignet (2011), C. Delbar (2012) et P. Dethier (2012).

Malgré le fait que nous ne disposions pas de données possédant une adresse complète, nous avons pu construire un modèle satisfaisant pour représenter la situation de la fin des années nonante. La robustesse statistique et le pouvoir explicatif sont certes moins élevés que ceux des modèles actuels, mais toutefois acceptables. Seule la performance prédictive peut être considérée comme moyenne. A l'instar des modèles actuels, la superficie contribue le plus à la formation des prix, suivie par les variables d'accessibilité et le revenu et, enfin, par les déterminants de l'environnement physique. L'influence de ces derniers peut n'être pas toujours correctement cernée.

Pour l'ensemble de nos modèles, nous avons pu tirer profit de l'analyse de l'autocorrélation spatiale locale pour identifier d'éventuelles variables manquantes. Celles-ci sont à intégrer dans les développements pour corriger la dépendance entre les résidus, comme ce fut le cas lors du développement du modèle général liégeois. Nous avons également montré que la diminution de la performance prédictive des modèles pour les biens immobiliers de haut standing s'applique également au marché des terrains à bâtir.

La comparaison interbassin montre qu'il existe différentes catégories de variables : certaines facilement comparables, d'autres qui le sont avec prudence ou au travers de seuils de distance différents et, enfin, des variables pour lesquelles la comparaison est délicate en raison de la pollution de leur influence par d'autres paramètres ou de l'altération de cette influence par des barrières visuelles et sonores. Ces variables sont essentiellement représentatives de l'environnement physique. Ainsi, nous devons admettre que l'impact de ses différents aspects a été imparfaitement mis en évidence, et ce malgré l'ajout de nouvelles variables et l'étude des terrains à bâtir, candidats idéals pour une meilleure mise en évidence des influences environnementales (V. R. Sharma, 2013).

Ce constat équivaut également pour la comparaison temporelle : celle-ci est rendue malaisée par les constats effectués au paragraphe précédent. Dans ce cas-ci, le problème est aggravé par la localisation approximative et par la localisation par défaut des terrains à bâtir au milieu de la rue. De plus, nous ne sommes pas parvenus à développer un modèle intéressant sur la même zone d'analyse pour l'étude de la situation actuelle. Celui-ci aurait sans doute permis d'apporter davantage de solidité à notre comparaison.

Nos travaux pourraient être améliorés en utilisant un échantillon plus fourni. Nous pensons notamment aux variables d'environnement qui semblent nécessiter davantage de biens concernés pour réellement mettre en évidence leur influence. La taille de nos échantillons a souffert de la crise économique qui a engendré un important ralentissement du marché ces dernières années, notamment en Province de Liège, où des délais de plus en plus longs entre la mise en vente et la transaction sont observés (Compagnie de Notaires de Liège, 2014). Aussi, le renouement avec la croissance économique pourrait permettre un redémarrage du marché, et ainsi offrir de meilleures possibilités d'étude des marchés des terrains à bâtir. Nous invitons également à envisager l'application de l'approche hédonique à la région de Bruxelles, où le marché semble nettement plus dynamique, avec des délais de vente moins importants compte tenu de la forte demande caractérisant l'hinterland de la capitale fédérale.

Bien que l'obtention de telles informations semble compliquée de prime abord, nous pensons que le développement de modèles similaires sur base des données notariales pourrait contribuer à l'amélioration et/ou à la vérification des constatations mises en évidence dans le présent travail. Ceci permettrait notamment de supprimer l'imprécision issue du prix de l'annonce, proche mais supposé toutefois légèrement différent du prix réel de la transaction. En effet, la négociation amène régulièrement à un réajustement du prix lors de l'achat.

Certaines variables pourraient être améliorées et/ou testées. Nous pensons que l'utilisation de la forme du terrain amènerait une amélioration du modèle. Nous ne possédons pas assez d'observations pour étudier correctement cette dernière variable. Nous avons en outre souhaité tester l'effet du champ de vision, mais nous nous sommes heurté à différents problèmes conceptuels et temporels, nous empêchant d'investiguer en détail cette variable. Nous pensons néanmoins qu'elle pourrait améliorer le modèle. Le critère de proximité des forêts pourrait être affiné en les considérant selon leur type (mixte, feuillus, résineux). La fréquence de passage des trains et des bus pourrait également être utilisée. De la même manière, le développement de modèles sur base d'intervalles plus réduits de prix contribuerait sans doute à l'optimisation des pouvoirs explicatif et prédictif des régressions. Enfin, parmi les nuisances de l'environnement,

celles engendrées par les éoliennes et les sites SEVESO pourraient également faire l'objet de variables explicatives.

Comme nous l'avons précisé précédemment, nous pensons que de nouvelles opportunités se présenteront lorsque le marché des terrains à bâtir, victime d'un climat socio-économique morose, sortira de la léthargie dans laquelle il est actuellement plongé. L'application de l'approche hédonique au marché des terrains à bâtir reste dès lors un champ d'investigation prometteur.

Bibliographie

Littérature

- AHLFELDT, G. & MAENNIG, W. 2008. Impact of sports arenas on land values: evidence from Berlin. *Ann Reg Sci*, **44**, pp. 205-227.
- ANDERSON, S.-T. & WEST, S.-E. 2006. Open space, residential property values, and spatial context. *Regional Science and Urban Economics*, **36**, pp. 773-789.
- ANDERSSON, D. E., SHYR, O. F., FU, J. 2010. Does high-speed rail accessibility influence residential property prices? Hedonic estimate from southern Taiwan. *Journal of Transport Geography*, **18**, pp. 166-174.
- ANDOH, K. & OHTA, M. 1997. A hedonic analysis of land prices in Yamanashi Prefecture, Japan. *Review of Urban and Regional Development Studies*, **9**, pp. 146-158.
- BAE, C.-H. C., JUN. M.-J., PARK, H. 2003. The impact of Seoul's subway Line 5 on residential property values. *Transport Policy*, **10**, pp. 85-94.
- BARNETT, C. J. 1985. An Application of the Hedonic Price Model to the Perth Residential Land Market. *Economic Record*, **61**(1), pp. 476-481.
- BAUDRY, M. & MASLIANSKAÏA-PAUTREL, M. 2011. Une remise en cause de la méthode des prix hédoniques en présence d'une segmentation du marché. *Revue économique*, **62**, pp. 567-578.
- BORCARD, D. n.d. *Transformation de données : normalisation, stabilisation des variances*, Université de Montréal. 4 p.
- BOWES, D. R. & IHLANFELDT, K. R. 2001. Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*, **50**, pp. 1-25.
- CERVERO, R. & KANG, C. D. 2011. Bus rapid transit impacts on land use and land values in Seoul, Korea. *Transport Policy*, **18**, pp. 102-116.
- CHAMBRE DES NOTAIRES DE LA PROVINCE DU HAINAUT 2010. *Le marché immobilier dans la province du Hainaut en 2009 – l'arrondissement de Charleroi*. Site de la Fédération Royale du Notariat belge (http://www.notaire.be/admin/files/assets/subsites/8/documenten/1297248240_rapport-02-2010-charleroi.pdf), page consultée le 22 juillet 2014.
- CHO, S.-H., POUDYAL, N. C., ROBERTS, R. K. 2008. Spatial analysis of the amenity value of green open space. *Ecological Economics*, **66**, pp. 403-416.
- CHOUVERT J. & TRAVERS, M. 2010. La capitalisation immobilière des espaces verts dans la ville d'Angers. *Revue économique*, **61**, pp. 821-836.

COLWELL, P. F. & MUNNEKE, H. J. 1997. The structure of Urban Land Prices. *Journal of Urban Economics*, **41**, pp. 321-336.

COMBY, J. 2003. La formation de la valeur sur les six marchés fonciers. *Etudes foncières*, **101**, pp. 18-23.

COMBY, J. 2013. *Vocabulaire foncier*. Site de J. Comby, expert en marchés fonciers et immobiliers (http://www.comby-foncier.com/vocabulaire_foncier.pdf), page consultée le 27 avril 2013.

COMPAGNIE DES NOTAIRES DE LIÈGE 2012. *Analyse du marché immobilier – année 2011 – province de Liège*. Site de la Fédération Royale du Notariat belge (http://www.notaire.be/admin/files/assets/subsites/8/documenten/1329299694_analyse-du-marche-immobilier-2011-liege-def.pdf), page consultée le 18 juillet 2014.

COMPAGNIE DES NOTAIRES DE LIÈGE 2013. *Analyse du marché immobilier – année 2012 – province de Liège*. Site de la Fédération Royale du Notariat belge (http://www.notaire.be/admin/files/assets/subsites/8/documenten/1360145275_liege-marche-immo-2012-compilation-def.pdf), page consultée le 18 juillet 2014.

COMPAGNIE DES NOTAIRES DE LIÈGE 2014. *Analyse du marché immobilier – année 2013 – province de Liège*. Site de la Fédération Royale du Notariat belge (http://www.notaire.be/admin/files/assets/subsites/8/documenten/1392206266_analyse-marche-immo-province-liege-2014-def.pdf), page consultée le 18 juillet 2014.

COMPAGNIE DES NOTAIRES DE LIÈGE 2014b. *Analyse du marché immobilier – année 2013 – province de Luxembourg*. Site de la Fédération Royale du Notariat belge (http://www.notaire.be/admin/files/assets/subsites/8/documenten/1391789613_province-du-luxembourg-marche-immobilier-analyse-generale-et-prix.pdf), page consultée le 19 juillet 2014.

CONFÉRENCE PERMANENTE DU DÉVELOPPEMENT TERRITORIAL 2008. *Atlas des dynamiques territoriales*. Site de la CPDT (http://cpdt.wallonie.be/sites/default/files/atlas_des_dynamiques_territoriales/walbassins-navette-dom.pdf), page consultée le 23 juillet 2014.

CONFÉRENCE PERMANENTE DU DÉVELOPPEMENT TERRITORIAL 2011. *Rapport final de la subvention 2010-2011 – Annexe : Diagnostic territorial de la Wallonie, approche spatiale – Structure fonctionnelle du territoire wallon : hiérarchie urbaine et aires d'influence*, 51 p.

CONFÉRENCE PERMANENTE DU DÉVELOPPEMENT TERRITORIAL 2012. *Contribution au rapport final – subvention 2011-2012 – volume annexe – Recherche I.5 – Production de l'habitat et enjeux territoriaux*, 72 p.

CONFÉRENCE PERMANENTE DU DÉVELOPPEMENT TERRITORIAL 2014. *Contribution au rapport intermédiaire – subvention 2013-2014 – volume annexe – Recherche I.5 – Production de l'habitat et enjeux territoriaux – Partie 3 : Les modes de gestion : préparation de l'évaluation des mesures et de la rencontre avec le panel d'experts*, 38 p.

CONFÉRENCE PERMANENTE DU DÉVELOPPEMENT TERRITORIAL n.d. *Le quartier de gare, un atout à valoriser ?* Site de la CPDT (http://www.cpd.be/old/Data/formation/com_mg.pdf), page consultée le 22 juillet 2014.

CONSTRUIRE FACILE 2012. *Construire sa maison en lotissement ou hors lotissement ?* Site de conseils en construction immobilière (<http://www.construirefacile.com/terrain/construire-sa-maison-en-lotissement-ou-hors-lotissement>), page consultée le 15 juillet 2014.

CORNET, Y., DAXHELET, D., HALLEUX, J.-M., KLINKENBERG A.-C., LAMBOTTE J.-M. 2005. *Cartographie de l'accessibilité par les alternatives à la voiture*, dans F. Witlox & V. Van Acker (Eds.), *Les Journées Géographiques Belges, Mobilité, Société et Environnement en cartes*. Zelzate : DCL Print & Sign, pp. 355-364.

DEKKERS, J. E. C. & WILLEMIJN VAN DER STRAATEN, J. 2009. Monetary valuation of aircraft noise: a hedonic analysis around Amsterdam Airport. *Ecological Economics*, **68**, pp. 2850-2858.

DELBAR, C. 2012. *Approche hédonique du marché immobilier – Les maisons unifamiliales de la région urbaine de Bruxelles*. Mémoire présenté pour l'obtention du diplôme de Master en Sciences Géographiques, Orientation Géomatique et Géométrie à l'Université de Liège, inédit, 89 p.

DE POLIGNAC, B. & MONCEAU, J.-P. 2002. *Expertise immobilière. Guide pratique*. Deuxième Edition, Paris : Eyrolles, 490 p.

DES ROSIERS, F. 2001. *La modélisation statistique en analyse et évaluation immobilières : guide méthodologique*. Gestion Urbaine et Immobilière, Université de Laval.

DETHIER, P. 2012. *Etude comparative entre les préférences révélées et les préférences déclarées – Les maisons d'habitation en région liégeoise*. Mémoire présenté pour l'obtention du diplôme de Master en Sciences Géographiques, Orientation Développement Territorial et Géomatique à l'Université de Liège, inédit, 144 p.

DE VOR, F. & DE GROOT, H. L. F. 2011. The impact of industrial sites on residential property values: a hedonic pricing analysis from the Netherlands. *Regional Studies*, **45**(5), pp. 609-623.

DEYMIER, G. 2003. *Capitalisation immobilière des gains d'accessibilité : Etat de l'art et étude de cas sur l'agglomération lyonnaise*. Thèse réalisée en vue de l'obtention de Docteur ès Sciences Economiques, mention Economie des Transports à l'Université Lumière Lyon 2.

DIRECTION MOBILITÉ 2010. *Diagnostic des déplacements domicile – travail 2008, rapport final*. Service Public Fédéral Mobilité et Transports, 46 p.

DODGSON, J. S. & TOPHAM, N. 1990. Valuing residential properties with the hedonic method: A comparison with the results of professional valuation. *Housing Studies*, **5**(3), pp. 209-213.

DONNAY, J.-P. 2011. *Analyse spatiale – ch. 5 : entités zonales*. Cours de troisième année de bachelier en Sciences Géographiques, Orientation Générale, de l'Université de Liège, inédit.

- DONZEL, A., FRANÇOIS, D., GENIAUX, G., NAPOLEONE, C. 2008. *Les déterminants socio-économiques des marchés fonciers. Rapport final*. Territoires Méditerranéens, 19, 120 p. Lien : <http://prodinra.inra.fr/record/38980>
- DUBÉ, J. 2010. *Vers une approche spatio-temporelle pour la modélisation des prix hédoniques : une application à la région de Québec, 1986 – 2004*. Thèse présentée à la Faculté des études supérieures de l'Université de Laval, Laval, Québec.
- EBERTZ, A. 2013. The Capitalization of Public Services and Amenities into Land Prices – Empirical Evidence from German Communities. *International Journal of Urban and Regional Research*, **37**(6), pp. 2116-2128.
- EVANS, A. W. 1995. The property market: Ninety per cent efficient?. *Urban Studies*, **32**(1), pp. 5-29.
- FAUSOLD, C. J. & LILIEHOLM, R. J. 1999. The Economic Value of Open Space: a review and synthesis. *Environmental Management*, **23**(2), pp. 307-320.
- FUMIHIRO, Y., HIDEAKI, O., KOTA, A. 2011. Nuclear Power-Related Facilities and Neighboring Land Price: A Case Study on the Mutsu-Ogawara Region, Japan. *Risk Analysis*, **31**(12), pp.1969-1994.
- GAO, X. & ASAMI, Y. 2007. Effect of urban landscapes on land prices in two Japanese cities. *Landscape and Urban Planning*, **81**(1), pp. 155-166.
- GEOGHEGAN, J., WAINGER, L. A., BOCKSTAEEL, N. E. 1997. Spatial Landscape indices in a hedonic framework: an ecological economics analysis using GIS. *Ecological Economics*, **23**(1), pp. 251-264.
- GOFFETTE-NAGOT, F., REGINSTER, I., THOMAS, I. 2011. Spatial Analysis of Residential Land Prices in Belgium: Accessibility, Linguistic Border, and Environmental Amenities. *Regional Studies*, **45**(9), pp. 1253-1268.
- GRIGNET, Y. 2011. *Approche hédonique du marché immobilier sur base d'annonces internet : les maisons d'habitations au sein de la région urbaine de Liège*. Mémoire présenté pour l'obtention du diplôme de Master en Sciences Géographiques, Orientation Géomatique et Géométrie à l'Université de Liège, inédit, 70 p.
- HALLEUX, J.-M. 2005. *Structuration spatiale des marchés fonciers et production de l'urbanisation morphologique : application à la Belgique et à ses nouveaux espaces résidentiels*. Thèse de Doctorat en Sciences Géographiques, Université de Liège, inédit, 308 p.
- HALLEUX, J.-M. 2005b. Le rôle des promotions foncières et immobilières dans la production des périphéries : application à la Belgique et à ses nouveaux espaces résidentiels. *Revue géographique de l'Est*, **45**(3-4), pp. 161-173.
- HALLEUX, J.-M. 2009. Modélisation hédonique des prix immobiliers : quelles opportunités pour la collectivité ? *Revue trimestrielle pour le géomètre-expert*, **62**, pp. 42-46.
- HALLEUX, J.-M. 2012. *Développement et Aménagement des Territoires*. Cours de première année de Master en Géographie Générale et Géomatique - Géométrie de l'Université de Liège, inédit.

- HALLEUX, J.-M. 2012b. *Marchés fonciers et immobiliers*. Cours de deuxième année de Master en Géographie générale et Géomatique – Géométrie de l'Université de Liège, inédit.
- HALLEUX, J.-M. 2014. *Discussion à propos de la segmentation entre le marché des terrains à bâtir et celui des gisements fonciers*. Réunion de suivi du mémoire.
- IHLANFELDT, K. R. 2006. The effect of land use regulation on housing and land prices. *Journal of Urban Economics*, **61**, pp. 420-435.
- JEHAIN, J.-B. 2014. *Entretien téléphonique à propos des hauts prix observés au sein de la commune de Spa*.
- KIM, K. S., PARK, S. J., KWEON, Y.-J. 2007. Highway traffic noise effects on land price in an urban area. *Transportation Research Part D*, **12**, pp. 275-280.
- LAKE, I. R., LOVETT, A. A., BATEMAN, I. J., LANGFORD, I. H. 1998. Modeling environmental influences on property prices in an urban environment. *Computer, Environment and Urban Systems*, **32**(2), pp. 121-136.
- LANCASTER, K. J. 1966. A new approach of consumer theory. *Journal of Political Economy*, **74**, pp. 132-157.
- LANCELOT, R. & LESNOFF, M. 2005. *Sélection de modèles avec l'AIC et critères d'information dérivés*. CIRAD, Montpellier.
- LEGGETT, C. G. & BOCKSTAEL, N. E. 1999. Evidence of the Effects of Water Quality on Residential Land Prices. *Journal of Environmental Economics and Management*, **39**, pp. 121-144.
- LIN, I.-H., WU, C., DE SOUSA, C. 2013. *Examining the economic impact of park facilities on neighboring residential property values*. *Applied Geography*, **45**, pp. 322-331.
- LIU, Y., ZHENG, B., TURKSTRA, J., HUANG, L. 2010. A hedonic model comparison for residential land value analysis. *International Journal of Applied Earth Observation and Geoinformation*, **12**, pp. 181-193.
- LUTTIK, J. 2000. The value of trees, water and open space as reflected by house prices in the Netherlands. *Landscape and Urban Planning*, **48**, pp. 161-167.
- MATSUI, K. & FUKUSHIGE, M. 2012. Land prices and landscape preservation restriction in a metropolitan area: the case of Kyoto City. *Review of urban and regional development studies*, **24**(1-2), pp. 17-34.
- MEDDA, F. 2012. Land value capture finance for transport: a review. *Journal of Transport Geography*, **25**, pp. 154-161.
- PACE, R. K., BARRY, R., CLAPP, J.M., RODRIQUEZ, M. 1998. Spatiotemporal autoregressive models of neighborhood effects. *The Journal of Real Estate Finance and Economics*, **17**, pp. 15-33.

- PACE, R. K., BARRY, R., SIRMANS, C. F. 1998b. Spatial statistics and real estate. *The Journal of Real Estate Finance and Economics*, **17**(1), pp. 5-13.
- PARIS, R. 2003. Essai de mesure des effets externes engendrés par une opération d'aménagement à travers une fonction hédonique des prix fonciers. *Revue d'Économie Régionale & Urbaine*, pp. 671-689.
- RENARD, V. 1975. L'allocation des sols urbains : modèles et réalités. *Revue économique*, **26**(1), pp. 91-110.
- ROSEN, S. 1974. Hedonic prices and implicit Markets: Product differentiation in pure competition. *The Journal of Political Economy*, **82**(1), pp. 34-55.
- SANDER, H. A. & HAIGHT, R. G. 2012. Estimating the economic value of cultural ecosystem services in an urbanizing area using hedonic pricing. *Journal of Environmental Management*, **113**, pp. 194-205.
- SHARMA, V. R. 2013. Forest proximity and residential land values. *Journal of Forest Economics*, **19**, pp. 78-86.
- SHONKWILER, J. S. & REYNOLDS, J. E. 1986. A note on the use of hedonic price models in the analysis of land prices at the Urban Fringe. *Land Economics*, **62**(1), pp. 58-63.
- SOWAER n.d. *Riverains – Dans quelle zone habitez-vous ?*. Site de la Société Wallonne des Aéroports (<http://www.sowaer.be/sowaer.php?cat=5&sscat=42>), page consultée pour la dernière fois le 15 juillet 2014.
- SRIKHUM, P. 2012. *Statistiques spatiales et études immobilières*. Thèse pour l'obtention du titre de Docteur en Sciences de Gestion, Université Paris-Dauphine.
- STRÉE, J. & HALLEUX, J.-M. 2013. *Production de l'habitat et enjeux territoriaux. Partie 2 : les modes d'habiter*. R.I.5 : Contribution au rapport final de la subvention 2012-2013, CPDT. Rapport de recherche.
- TAVIERS, M., BONNET, E., CHEVÉ, M. APPÉRE, G. 2009. Perception des risques industriels dans une zone estuarienne : une analyse hédonique spatiale. *Economie et Prévision*, **4-5**(190-191), pp. 135-158.
- THISSE, J.-F. & THOMAS, I. 2007. Bruxelles et Wallonie, une lecture de la vie économique urbaine. *Reflets et Perspectives de la Vie Economique*, **46**, pp. 75-93.
- TYRVAINEN, L. & MIETTINEN, A. 2000. Property prices and urban forest amenities. *Journal of Environmental Economics and Management*, **39**, pp. 205-233.
- VALENTE, J., WU, S., GELFAND, A., SIRMANS, C.-F. 2005. Apartment rent prediction using spatial modeling. *Journal of real estate research*, **27**(1), pp. 105-136.
- VAN DER HAEGEN, H. & PATTYN, M. 1979. *Les régions urbaines belges*, *Bulletin de Statistiques*. Institut National des Statistiques, Ministère des Affaires Economiques.

VANDERMOTTEN, C., ISTAZ, D., MARISSAL, P. 1996. *Marchés fonciers et immobiliers acquisitifs et locatifs : 1994-1995*. Jambes : DGATLP, 39 p.

VAN HECKE, E., HALLEUX, J.-M., DECROLY, J.-M., MÉRENNE-SCHOUMAKER, B. (2009). *Noyaux d'habitat et régions urbaines dans une Belgique urbanisée*. Bruxelles : SPF Economie, 201 p.

WALTERT, F. & SCHLÄPFER, F. 2010. Landscape amenities and local development: a review of migration, regional economic and hedonic pricing studies. *Ecological Economics*, **70**, pp. 141-152.

WILLIAMS, R. (2011). *Heteroscedasticity*. University of Notre Dame.

ZHENG, S. & KAHN, M. E. 2008. Land and residential property markets in a booming economy: New evidence from Beijing. *Journal of Urban Economics*, **63**, pp. 743-757.

Statistiques

BRUSSELS SOUTH CHARLEROI AIRPORT (2014). *Statistiques et données clés*. Site de l'aéroport e Charleroi-Bruxelles-Sud (<http://www.charleroi-airport.com/presse/statistiques-et-donnees-cles/index.html>), page consultée le 22 juillet 2014.

SPF – MOBILITÉ-TRANSPORT (2011). *Recensement général de la circulation 2009*. n°52.

STATBEL (2014). *Transport aérien*. Site de l'Institut National des Statistiques (http://statbel.fgov.be/fr/statistiques/chiffres/circulation_et_transport/transport/aerien/), page consultée le 23 juillet 2014.

Table des annexes

I. Liste des communes du bassin d'emploi de Liège	I
II. Liste des communes du bassin d'emploi de Charleroi	I
III. Liste des communes de la zone d'étude de la fin des années nonante	I
IV. Code Java pour la collecte des données sur Immoweb	II
V. Codes d'erreur de géolocalisation du batch géocodeur	VII
VI. Liste des variables créées pour la situation actuelle	VIII
VII. Liste des variables créées pour la situation de la fin des années nonante	XI
VIII. Script Python pour la fonction <i>Tabulate Area</i>	XIII
IX. Interprétation du diagramme de dispersion de Moran	XIV
X. Liste des couples de variables concernés par une colinéarité excessive pour le modèle général liégeois	XIV
XI. Tests de colinéarité du modèle général liégeois.....	XV
1. Modèle MCO.....	XV
2. Modèle corrigeant l'hétéroscédasticité	XV
XII. Test de Breusch-Pagan du modèle général liégeois.....	XVI
XIII. Diagramme de dispersion de Moran pour le modèle actuel général liégeois.....	XVII
XIV. Liste des couples de variables concernés par une colinéarité excessive pour le modèle périphérique liégeois	XVII
XV. Tests de colinéarité du modèle périphérique liégeois.....	XVIII
1. Modèle MCO.....	XVIII
2. Modèle corrigeant l'hétéroscédasticité	XIX
XVI. Test de Breusch-Pagan du modèle périphérique liégeois.....	XIX
XVII. Diagramme de dispersion de Moran pour le modèle périphérique liégeois	XX
XVIII. Liste des couples de variables concernés par une colinéarité excessive pour le modèle carolorégien	XX
XIX. Tests de colinéarité du modèle carolorégien	XXIII
1. Modèle MCO.....	XXIII
2. Modèle corrigeant l'hétéroscédasticité	XXIII
XX. Test de Breusch-Pagan du modèle carolorégien	XXIV
XXI. Diagramme de dispersion de Moran pour le modèle carolorégien.....	XXIV

XXII. Liste des couples de variables concernés par une colinéarité excessive pour le modèle liégeois des années nonante.....	XXV
XXIII. Test de colinéarité du modèle liégeois des années nonante	XXV
1. Modèle MCO.....	XXV
2. Modèle corrigeant l'hétéroscédasticité	XXVI
XXIV. Test de Breusch-Pagan du modèle liégeois des années nonante.....	XXVI
XXV. Diagramme de dispersion de Moran pour le modèle liégeois des années nonante	XXVII

I. Liste des communes du bassin d'emploi de Liège

Amay	Dalhem	Héron	Remicourt
Ans	Dison	Herstal	Saint-Nicolas
Anthisnes	Donceel	Huy	Seraing
Aubel	Durbuy	Jalhay	Soumagne
Awans	Engis	Juprelle	Spa
Aywaille	Erezée	Liège	Sprimont
Bassenge	Esneux	Limbourg	Stoumont
Berloz	Faimes	Manhay	Theux
Beyne-Heusay	Ferrières	Marchin	Thimister-Clermont
Blegny	Fexhe-le-Haut-Clocher	Modave	Tinlot
Braives	Flémalle	Nandrin	Trooz
Burdinne	Fléron	Neupré	Verlaine
Chaufontaine	Geer	Olne	Villers-le-Bouillet
Clavier	Grâce-Hollogne	Oreye	Visé
Comblain-au-Pont	Hamoir	Ouffet	Wanze
Crisnée	Herve	Oupeye	Waremme

II. Liste des communes du bassin d'emploi de Charleroi

Aiseau	Courcelles	Les Bons Villers	Montigny-le-Tilleul
Anderlues	Doische	Lobbes	Philippeville
Beaumont	Farciennes	Gerpennes	Pont-à-Celles
Cerfontaine	Fleurus	Ham-sur-Heure-Nal.	Sivry-Rance
Charleroi	Florennes	Merbes-le-Château	Thuin
Chatelet	Fontaine-l'Evêque	Mettet	Walcourt
Chimay	Froidchapelle	Momignies	

III. Liste des communes de la zone d'étude de la fin des années nonante

Ans	Crisnée	Herstal	Seraing
Awans	Dalhem	Juprelle	Soumagne
Aywaille	Donceel	Liège	Sprimont
Bassenge	Esneux	Neupré	Trooz
Berloz	Fexhe-le-Haut-Clocher	Oreye	Visé
Beyne-Heusay	Flémalle Fléron	Oupeye	Waremme
Blegny	Geer	Remicourt	
Chaufontaine	Grâce-Hollogne	Saint-Nicolas	

IV. Code Java pour la collecte des données sur Immoweb

```
import java.util.*;
import java.util.regex.Pattern;
import java.lang.*;
import java.lang.String;
import java.text.Normalizer;

public class Immoweb {
    public static void main(String[] args) throws Exception {

        System.out.print("Coller les URL séparées par une virgule");

        // open up standard input
        BufferedReader br = new BufferedReader(new
InputStreamReader(System.in));

        String liste_adresse = null;

        try {
            liste_adresse = br.readLine();
        } catch (IOException ioe) {
            System.out.println("IO error");
            System.exit(1);
        }

        String temp,find,write;
        String[] adresse;

        int nombre_dadresse;

        adresse = liste_adresse.split(",");
        nombre_dadresse= adresse.length;
        System.out.println("nombre d'adresses "+nombre_dadresse);

        //création fichier csv
        FileWriter caract = new FileWriter("immoweb.csv");
        BufferedWriter out_caract = new BufferedWriter(caract);
        out_caract.write("url;id;PRIX;PRIX AU METRE
CARRE;AUTORISATION LOTISSEMENT;PERMIS URBANISME;PERMIS BATIR;CITATION
INFRA URBA;AFFECTATION URBANISTIQUE LA PLUS RECENTE;LARGUEUR DU
TERRAIN;PROFONDEUR DU TERRAIN(m);SUP_TERRAIN;FRONT_RUE;RACCORDEMENT
POSSIBLE;RACCORDEMENT EGOUT;ACCORDEMENT EAU,GAZ,ELEC;ORIENTATION
FACADE;TYPE_CONSTRUCTION;ENVIRONNEMENT\n");

        for(int i=0;i<adresse.length;i++){

            String
prix_carre="",permis_batir="",citation="",affect_urba="",raccordement_ok=
"",egout="",eau_gaz_elec="",permis_urb="",ID="",prix="",lotissement="",la
rgeur_terrain="",profondeur_terrain="",front_rue="",sup_terrain="",orient
ation="",type_constr="",environnement="";

            boolean
flag_prix_carre=false,flag_ID=false,flag_prix=false,flag_sup_terrain=fals
e;
```



```

        System.out.println(adresse[i]);

        URL url = new URL(adresse[i]);
        //lit l'annonce i
        BufferedReader in = new BufferedReader(new
InputStreamReader(url.openStream(),"UTF8"));

        //supprime la dernière partie de l'URL (après le &)
        String URL = url.toString();
        String[] tokens = URL.split("&");
        URL=tokens[0];
        //System.out.println(URL);

        //récupère l'ID
        tokens = URL.split("=");
        ID = tokens[tokens.length-1];
        if(ID.matches("[0-9]*")) flag_ID=true;
        //System.out.println("ID "+ID);

        //chaque ligne du html
        String inputLine=in.readLine();

        //création fichier html
        FileWriter HTML = new FileWriter(ID+".html");
        BufferedWriter out_HTML = new BufferedWriter(HTML);

        //System.out.println(URL);
        int j=0;

        while ( inputLine != null){

            inputLine=inputLine+"\n";
            //si la ligne est non vide
            if(!Pattern.matches("\\s+", inputLine)){

                //numéro de ligne
                j++;
                //écrit la ligne dans le fichier HTML
                out_HTML.write(inputLine);

                //prix
                //si input line contient chiffre,
                point, espace vide
                if(inputLine.matches("[\\d\\s\\.]+")){
                    //remplacer espace vide par
                    vide
                    temp=
                    inputLine.replaceAll("\\s+","");
                    //remplacer point par vide
                    temp=temp.replaceAll("\\.", "");
                    //si contient au moins 5
                    chiffre i.e. minimum 10000 euros
                    if(temp.matches("[\\d]{5,}")){
                        prix=temp;
                        flag_prix=true;
                    }
                }
            }
        }

```

```

//repertorier prix au metre carre
if(flag_prix_carre==true){
    temp=
inputLine.replaceAll("\\s+", "");
    prix_carre=temp;
    flag_prix_carre=false;
}
//trouver prix au metre carre
find="<li>Prix au m² :";
if ((inputLine.indexOf(find) != -1)){
    flag_prix_carre=true;
}

//autorisation de lotissement

find="Autorisation de lotissement : Oui";
if ((inputLine.indexOf(find) != -1)){
    lotissement="oui";
}
find="Autorisation de lotissement : Non";
if ((inputLine.indexOf(find) != -1)){
    lotissement="non";
}

//permis d'urbanisme

find="Permis d'urbanisme obtenu : Oui";
if ((inputLine.indexOf(find) != -1)){
    permis_urb="oui";
}
find="Permis d'urbanisme obtenu : Non";
if ((inputLine.indexOf(find) != -1)){
    permis_urb="non";
}

//permis de batir
find="permis de bâtir obtenu";
if ((inputLine.indexOf(find) != -1)){
    permis_batir="oui";
}

// Citation pour infrastructure urbanistique
find="Citation pour infraction urbanistique
: Oui";

if ((inputLine.indexOf(find) != -1)){
    citation="oui";
}
find="Citation pour infraction urbanistique
: Non";

if ((inputLine.indexOf(find) != -1)){
    citation="non";
}

// Affectation urbanistique la plus récente

find="Affectation urbanistique la plus
récente sur base des dénominations utilisées dans le registre des plans
:";

if (inputLine.indexOf(find) != -1){
    temp =
inputLine.replaceAll("<p>", "");
    temp=temp.replaceAll("</p>", "");
}

```

```

tokens = temp.split(": ");
temp=tokens[1];
temp=temp.replaceAll(" ", "_");
temp=temp.replaceAll("\\s+", "");
");
affect_urba= temp;
}

//largeur terrain
find="largeur du terrain à rue: ";
if (inputLine.indexOf(find) != -1){
temp =
inputLine.replaceAll("<p>","");
terrain à rue: ");
tokens = temp.split("largeur du
temp=tokens[1];
tokens=temp.split(" ");
temp=tokens[0];

largeur_terrain= temp;
}
//profondeur terrain
find="profondeur du terrain: ";
if (inputLine.indexOf(find) != -1){
temp =
inputLine.replaceAll("<p>","");
terrain: ");
tokens = temp.split("profondeur du
temp=tokens[1];
tokens=temp.split(" ");
temp=tokens[0];

profondeur_terrain= temp;
}
//terrain à front de rue
find="Terrain à front de rue";
if ((inputLine.indexOf(find) != -1)){
front_rue="oui";
}
//raccordement possible
find="Raccordement possible";
if ((inputLine.indexOf(find) != -1)){
raccordement_ok="oui";
}
//raccordement egout
find="Raccordement à l'égout";
if ((inputLine.indexOf(find) != -1)){
egout="oui";
}
//eau gaz electricite
find="Gaz/Eau/Electricité";
if ((inputLine.indexOf(find) != -1)){
eau_gaz_elec="oui";
}
//type de construction
find="type de construction : ";
if (inputLine.indexOf(find) != -1){
temp =
inputLine.replaceAll("<p>","");
temp=temp.replaceAll("</p>","");
tokens = temp.split(": ");
temp=tokens[1];

```

```

temp=temp.replaceAll(" ", "");
temp=temp.replaceAll("/n", "");
temp=temp.replaceAll("\\s+",
");
type_constr= temp;
}
// Orientation de la facade rue

find="Orientation de la façade rue :";
if (inputLine.indexOf(find) != -1){
temp =
inputLine.replaceAll("<p>", "");
temp=temp.replaceAll("</p>", "");
tokens = temp.split(": ");
temp=tokens[1];
temp=temp.replaceAll("\\s+",
");
orientation= temp;
}

//superficie terrain
find="<li>Terrain de";
if (inputLine.indexOf(find) != -1){
temp =
inputLine.replaceAll("<li>", "");
temp =
inputLine.replaceAll("</li>", "");
//System.out.println(inputLine);
tokens = temp.split(" ");
temp=tokens[2];
sup_terrain=
temp.replaceAll("\\s+", "");
flag_sup_terrain=true;
}

//Environnement
find="Résidentiel";
if (inputLine.indexOf(find) != -1){
environnement="Résidentiel";
}
else {
find="Campagne";
if (inputLine.indexOf(find) !=
-1){
environnement="Campagne";
}
else {
find="Urbain";
if
(inputLine.indexOf(find) != -1){
environnement="Urbain";
}
}
}
}

inputLine=in.readLine();
}

```

```

//          if(cuisine!=""){

        write=URL+";"+ID+";"+prix+";"+prix_carre+";"+lotissement+";"+permis
_urb+";"+permis_batir+";"+citation+";"+affect_urba+";"+largeur_terrain+";
"+profondeur_terrain+";"+sup_terrain+";"+front_rue+";"+raccordement_ok+";
"+egout+";"+eau_gaz_elec+";"+orientation+";"+type_constr+";"+environnemen
t+"\n";

        out_caract.write(write);
/*          System.out.println("lotissement");
          System.out.println(lotissement);
          System.out.println("largeur terrain");
          System.out.println(largeur_terrain);
          System.out.println("profondeur terrain");
          System.out.println(profondeur_terrain);
          System.out.println("front rue");
          System.out.println(front_rue);
          */
          System.out.println("TYPE=");
          System.out.println(type_constr);
          System.out.println("orientation="+orientation);
//      }

        out_HTML.flush();

        out_HTML.close();
        in.close();
    }

        out_caract.flush();
        out_caract.close();

System.out.println("\nfichier immoweb.csv créé et annonces récupérées au
format HTML");

    }
}

```

V. Codes d'erreur de géolocalisation du batch géocodeur

Numéro de code	Signification
8 ou 9	La localisation exacte a été trouvée sur base de l'algorithme.
5	La localisation exacte n'a pu être trouvée : la commune et la rue ont bien été repérées, mais pas le numéro. Le bien a été localisé par défaut au milieu de la rue.
3	La localisation exacte n'a pu être trouvée : la commune a été repérée, mais la rue n'a pas été reconnue, ni le numéro. Le bien a été localisé par défaut au milieu de la commune.

VI. Liste des variables créées pour la situation actuelle

Nom	Description	Source	Type
SUP_TERRAI	Superficie	Immoweb	Q
pente_lieg/char	Pente au niveau du terrain	SPW-DGO3	Q
orientS	Orientation au sud	SPW-DGO3	B
eau_gaz_elec	Mention d'un raccordement	Immoweb	B
egout	Mention d'un raccordement	Immoweb	B
LARGEUR (Liè.)	Largeur du terrain à front de rue	Immoweb	Q
RATIOLP (Liè.)	Ratio largeur - profondeur	Immoweb	Q
ETROIT (Liè.)	Terrain présentant un ratio L/P inférieur à 0,30	Immoweb	B
ETROIT + (Liè.)	Terrain présentant un ratio L/P inférieur à 0,15	Immoweb	B
LARGE (Liè.)	Terrain présentant un ratio L/P supérieur à 1	Immoweb	B
L/S (Liè.)	Largeur à front de rue divisée par la superficie	Immoweb	Q
BATIMENT	Présence d'un bâtiment sur le terrain	Immoweb	B
dealimenta	Distance aux commerces alimentaires	SEGEFA	Q
dealimenta100	Commerce alimentaire à moins de 100 m	SEGEFA	B
dealimenta500	Commerce alimentaire à moins de 500 m	SEGEFA	B
dealimenta700	Commerce alimentaire à moins de 700 m	SEGEFA	B
debus	Distance aux arrêts de bus	TEC	Q
debus100	Arrêt de bus à moins de 100 m	TEC	B
debus500	Arrêt de bus à moins de 500 m	TEC	B
debus700	Arrêt de bus à moins de 700 m	TEC	B
del2014	Distance aux gares L	J. Charlier	Q
del100	Gare L à moins de 100 m	J. Charlier	B
del500	Gare L à moins de 500 m	J. Charlier	B
del700	Gare L à moins de 700 m	J. Charlier	B
t_l2014	Temps d'accès minimal à une gare L	J. Charlier	Q
degares	Distance aux gares	J. Charlier	Q
degares100	Gare à moins de 100 m	J. Charlier	B
degares500	Gare à moins de 500 m	J. Charlier	B
degares700	Gare à moins de 700 m	J. Charlier	B
degares1000	Gare à moins de 1000 m	J. Charlier	B
degares1500	Gare à moins de 1500 m	J. Charlier	B
t_gares	Temps d'accès minimal à une gare	J. Charlier	Q
deicir2014	Distance aux gares IC/IR	J. Charlier	Q
deicir500	Gare IC/IR à moins de 500 m	J. Charlier	B
deicir1000	Gare IC/IR à moins de 1000 m	J. Charlier	B
deicir1500	Gare IC/IR à moins de 1500 m	J. Charlier	B
t_icir2014	Temps d'accès minimal à une gare IC/IR	J. Charlier	Q
desortieauto	Distance aux sorties d'autoroute	LEPUR	Q
auto100	Sortie d'autoroute à moins de 100 m	LEPUR	B
auto500	Sortie d'autoroute à moins de 500 m	LEPUR	B
auto700	Sortie d'autoroute à moins de 700 m	LEPUR	B
auto1000	Sortie d'autoroute à moins de 1000 m	LEPUR	B
auto1500	Sortie d'autoroute à moins de 1500 m	LEPUR	B
t_sort_auto	Temps d'accès minimal à une sortie d'autoroute	LEPUR	Q
t_ans (Liè.)	Temps d'accès à Ans en secondes	Personnelle	Q
t_herstal (Liè.)	Temps d'accès à Herstal	Personnelle	Q
t_huy (Liè.)	Temps d'accès à Huy	Personnelle	Q

t_liege (Liè.)	Temps d'accès à Liège	Personnelle	Q
t_sart_til (Liè.)	Temps d'accès au Sart-Tilman	Personnelle	Q
t_seraing (Liè.)	Temps d'accès à Seraing	Personnelle	Q
t_verviers (Liè.)	Temps d'accès à Verviers	Personnelle	Q
t_namur	Temps d'accès à Namur	Personnelle	Q
t_bruelles	Temps d'accès à Bruxelles	Personnelle	Q
t_charlero (Cha.)	Temps d'accès à Charleroi	Personnelle	Q
t_fleurus (Cha.)	Temps d'accès à Fleurus	Personnelle	Q
t_gosselie (Cha.)	Temps d'accès à Gosselies	Personnelle	Q
t_aeropole (Cha.)	Temps d'accès à l'aéroport	Personnelle	Q
t_gilly (Cha.)	Temps d'accès à Gilly	Personnelle	Q
t_chatelet (Cha.)	Temps d'accès à Châtelet	Personnelle	Q
t_matern_prim	Temps d'accès min. à une école maternelle/primaire	LEPUR	Q
matprim30	Ecole maternelle ou primaire à moins de 30 s	LEPUR	B
matprim60	Ecole maternelle ou primaire à moins de 60 s	LEPUR	B
t_secondai	Temps d'accès minimal à une école secondaire	LEPUR	Q
second60	Ecole secondaire à moins de 60 s	LEPUR	B
second120	Ecole secondaire à moins de 120 s	LEPUR	B
t_superieu	Temps d'accès à un établissement scolaire supérieur	LEPUR	Q
super60	Etablissement scol. supérieur à moins de 60 s	LEPUR	B
super300	Etablissement scol. supérieur à moins de 300 s	LEPUR	B
t_nodules	Temps d'accès minimal à un nodule commercial	SEGEFA	Q
ltrainresi	Part modale du train au lieu de résidence	LEPUR	Q
lmlresid	Part modale des modes lents au lieu de résidence	LEPUR	Q
lbusresid	Part modale des bus au lieu de résidence	LEPUR	Q
SHA200	Indice de Shannon dans un rayon de 200 m	C. L. C. ³⁸	Q
PURBA200	Part d'artificialisation dans un rayon de 200 m	C. L. C.	Q
PBATI200	Part de bâti dans un rayon de 200 m	C. L. C.	Q
SHA1000	Indice de Shannon dans un rayon de 1000 m	C. L. C.	Q
PURB1000	Part d'artificialisation dans un rayon de 1000 m	C. L. C.	Q
PBATI1000	Part de bâti dans un rayon de 1000 m	C. L. C.	Q
densite09	Densité de population en 2009 (secteurs stat.)	INS	Q
PC4facades	Part de maisons isolées en 2001 (secteurs stat.)	INS	Q
LOTISSEMEN	Localisation dans un lotissement	Immoweb	B
CUL_DE_SAC	Localisation dans un cul-de-sac	Google Maps	B
LDEN60	Bruit aérien de plus de 60 dB	SOWAER	B
LDEN55	Bruit aérien de plus de 55 dB	SOWAER	B
de_eau	Distance au réseau hydrographique principal	ANB ³⁹	Q
de_eau100	Réseau hydrographique principal à moins de 100 m	ANB	B
de_eau200	Réseau hydrographique principal à moins de 200 m	ANB	B
de_eau500	Réseau hydrographique principal à moins de 500 m	ANB	B
de_eau_pre	Distance au réseau hydrographique complet	LEPUR	Q
de_eau_pre100	Réseau hydrographique complet à moins de 100 m	LEPUR	B
de_eau_pre200	Réseau hydrographique complet à moins de 200 m	LEPUR	B
de_eau_pre500	Réseau hydrographique complet à moins de 500 m	LEPUR	B
de_lacs (Cha.)	Distance aux lacs de l'Eau d'Heure	Personnelle	Q
delacs2000 (Cha.)	Lacs de l'Eau d'Heure à moins de 2000 m	Personnelle	B
delacs5000 (Cha.)	Lacs de l'Eau d'Heure à moins de 5000 m	Personnelle	B

³⁸ Corine Land Cover

³⁹ Atlas Numérique de Belgique

de_ravel_2	Distance au réseau RAVeL	SPW-DGO1	Q
de_ravel100	RAVeL à moins de 100 m	SPW-DGO1	B
de_ravel500	RAVeL à moins de 500 m	SPW-DGO1	B
de_tihange (Liè.)	Distance à la centrale nucléaire de Tihange	Personnelle	Q
de_tihange2000	Centrale nucléaire de Tihange à moins de 2000 m	Personnelle	B
de_tihange5000	Centrale nucléaire de Tihange à moins de 5000 m	Personnelle	B
de_tihange10000	Centrale nucléaire de Tihange à moins de 10000 m	Personnelle	B
decdf2014	Distance aux lignes de chemin de fer	CPDT/J. Charlier	Q
decdf100	Ligne de chemin de fer à moins de 100 m	CPDT/J. Charlier	B
decdf500	Ligne de chemin de fer à moins de 500 m	CPDT/J. Charlier	B
decdf700	Ligne de chemin de fer à moins de 700 m	CPDT/J. Charlier	B
deextracti	Distance aux carrières	C. L. C.	Q
deextracti100	Carrière à moins de 100 m	C. L. C.	B
deextracti200	Carrière à moins de 200 m	C. L. C.	B
deextracti500	Carrière à moins de 500 m	C. L. C.	B
deindustri	Distance aux industries	COSW, C. L. C.	Q
deindustri100	Industrie à moins de 100 m	COSW, C. L. C.	B
deindustri200	Industrie à moins de 200 m	COSW, C. L. C.	B
deindustri500	Industrie à moins de 500 m	COSW, C. L. C.	B
deindustri1000	Industrie à moins de 1000 m	COSW, C. L. C.	B
delht	Distance aux lignes à haute tension	LEPUR-CPDT	Q
delht100	Ligne à haute tension à moins de 100 m	LEPUR-CPDT	B
delht200	Ligne à haute tension à moins de 200 m	LEPUR-CPDT	B
delht500	Ligne à haute tension à moins de 500 m	LEPUR-CPDT	B
deroute10k	Distance aux routes (trafic > 10000 véhicules/jour)	SPF M-T ⁴⁰	Q
deroute10k100	Route (> 10000 véhicules) à moins de 100 m	SPF M-T	B
deroute10k200	Route (> 10000 véhicules) à moins de 200 m	SPF M-T	B
deroute10k500	Route (> 10000 véhicules) à moins de 500 m	SPF M-T	B
deroute20k	Distance aux routes (trafic > 20000 véhicules/jour)	SPF M-T	Q
deroute20k100	Route (> 20000 véhicules) à moins de 100 m	SPF M-T	B
deroute20k200	Route (> 20000 véhicules) à moins de 200 m	SPF M-T	B
deroute20k500	Route (> 20000 véhicules) à moins de 500 m	SPF M-T	B
desaed	Distance aux SAED	SPW-DGO4	Q
desaed100	SAED à moins de 100 m	SPW-DGO4	B
desaed200	SAED à moins de 200 m	SPW-DGO4	B
desaed500	SAED à moins de 500 m	SPW-DGO4	B
foretvert2	Distance aux forêts et espaces verts	COSW, C. L. C.	Q
foretvert100	Forêt ou espace vert à moins de 100 m	COSW, C. L. C.	B
foretvert500	Forêt ou espace vert à moins de 500 m	COSW, C. L. C.	B
zoneinond	Aléa d'inondation	SPW-DGO3	B
zoneinond3	Aléa d'inondation de niveau fort	SPW-DGO3	B
chomage2010	Taux de chômage en 2010 (communes)	IWEPS	Q
diplomes01	Taux de diplômés du supérieur en 2001 (communes)	INS	Q
tauxIPP	Taux de la taxe communale (IPP)	LEPUR	Q
chomage2001	Taux de chômage en 2001 (secteurs statistiques)	INS	Q
revmed2011	Revenu médian en 2011 (secteurs statistiques)	INS	Q

⁴⁰ Mobilité - Transport

VII. Liste des variables créées pour la situation de la fin des années nonante

Nom	Description	Source	Type
superficie	Superficie	Immoweb	Q
pente_lieg	Pente au niveau du terrain	SPW-DGO3	Q
orientS	Orientation au sud	SPW-DGO3	B
debus	Distance aux arrêts de bus	TEC	Q
debus100	Arrêt de bus à moins de 100 m	TEC	B
debus500	Arrêt de bus à moins de 500 m	TEC	B
debus700	Arrêt de bus à moins de 700 m	TEC	B
degares	Distance aux gares	J. Charlier	Q
degares100	Gare à moins de 100 m	J. Charlier	B
degares500	Gare à moins de 500 m	J. Charlier	B
degares700	Gare à moins de 700 m	J. Charlier	B
degares1000	Gare à moins de 1000 m	J. Charlier	B
degares1500	Gare à moins de 1500 m	J. Charlier	B
t90_gares	Temps d'accès minimal à une gare	J. Charlier	Q
desortieauto	Distance aux sorties d'autoroute	LEPUR	Q
auto100	Sortie d'autoroute à moins de 100 m	LEPUR	B
auto500	Sortie d'autoroute à moins de 500 m	LEPUR	B
auto700	Sortie d'autoroute à moins de 700 m	LEPUR	B
auto1000	Sortie d'autoroute à moins de 1000 m	LEPUR	B
auto1500	Sortie d'autoroute à moins de 1500 m	LEPUR	B
t90_sort_auto	Temps d'accès minimal à une sortie d'autoroute	LEPUR	Q
t90_ans	Temps d'accès à Ans en secondes	Personnelle	Q
t90_herstal	Temps d'accès à Herstal	Personnelle	Q
t90_huy	Temps d'accès à Huy	Personnelle	Q
t90_liege	Temps d'accès à Liège	Personnelle	Q
t90_tilman	Temps d'accès au Sart-Tilman	Personnelle	Q
t90_seraing	Temps d'accès à Seraing	Personnelle	Q
t90_verviers	Temps d'accès à Verviers	Personnelle	Q
t90_namur	Temps d'accès à Namur	Personnelle	Q
t90_bruzelles	Temps d'accès à Bruxelles	Personnelle	Q
cliptrares	Part modale du train au lieu de résidence	LEPUR	Q
clipmlres	Part modale des modes lents au lieu de résidence	LEPUR	Q
clipbusre	Part modale des bus au lieu de résidence	LEPUR	Q
SHA200	Indice de Shannon dans un rayon de 200 m	C. L. C.	Q
PURB200	Part d'artificialisation dans un rayon de 200 m	C. L. C.	Q
PBATI200	Part de bâti dans un rayon de 200 m	C. L. C.	Q
SHA1000	Indice de Shannon dans un rayon de 1000 m	C. L. C.	Q
PURB1000	Part d'artificialisation dans un rayon de 1000 m	C. L. C.	Q
PBATI1000	Part de bâti dans un rayon de 1000 m	C. L. C.	Q
densite01	Densité de population en 2001 (secteurs stat.)	INS	Q
PC4facades	Part de maisons isolées en 2001 (secteurs stat.)	INS	Q
de_eau	Distance au réseau hydrographique principal	ANB	Q
de_eau100	Réseau hydrographique principal à moins de 100 m	ANB	B
de_eau200	Réseau hydrographique principal à moins de 200 m	ANB	B
de_eau500	Réseau hydrographique principal à moins de 500 m	ANB	B
de_eau_pre	Distance au réseau hydrographique complet	LEPUR	Q

de_eau_pre100	Réseau hydrographique complet à moins de 100 m	LEPUR	B
de_eau_pre200	Réseau hydrographique complet à moins de 200 m	LEPUR	B
de_eau_pre500	Réseau hydrographique complet à moins de 500 m	LEPUR	B
de_ravel199	Distance au réseau RAVeL	SPW-DGO1	Q
de_ravel100	RAVeL à moins de 100 m	SPW-DGO1	B
de_ravel500	RAVeL à moins de 500 m	SPW-DGO1	B
de_tihange (Liè.)	Distance à la centrale nucléaire de Tihange	Personnelle	Q
de_tihange2000	Centrale nucléaire de Tihange à moins de 2000 m	Personnelle	B
de_tihange5000	Centrale nucléaire de Tihange à moins de 5000 m	Personnelle	B
de_tihange10000	Centrale nucléaire de Tihange à moins de 10000 m	Personnelle	B
decdf2000	Distance aux lignes de chemin de fer	CPDT J. Charlier	Q
decdf100	Ligne de chemin de fer à moins de 100 m	CPDT J. Charlier	B
decdf500	Ligne de chemin de fer à moins de 500 m	CPDT J. Charlier	B
decdf700	Ligne de chemin de fer à moins de 700 m	CPDT J. Charlier	B
deextr2000	Distance aux carrières	C. L. C.	Q
deextr100	Carrière à moins de 100 m	C. L. C.	B
deextr200	Carrière à moins de 200 m	C. L. C.	B
deextr500	Carrière à moins de 500 m	C. L. C.	B
deind	Distance aux industries	C. L. C.	Q
deind100	Industrie à moins de 100 m	C. L. C.	B
deind200	Industrie à moins de 200 m	C. L. C.	B
deind500	Industrie à moins de 500 m	C. L. C.	B
deind1000	Industrie à moins de 1000 m	C. L. C.	B
der10k	Distance aux routes (trafic > 10000 véhicules/jour)	SPF M-T	Q
der10k100	Route (> 10000 véhicules) à moins de 100 m	SPF M-T	B
der10k200	Route (> 10000 véhicules) à moins de 200 m	SPF M-T	B
der10k500	Route (> 10000 véhicules) à moins de 500 m	SPF M-T	B
der20k	Distance aux routes (trafic > 20000 véhicules/jour)	SPF M-T	Q
der20k100	Route (> 20000 véhicules) à moins de 100 m	SPF M-T	B
der20k200	Route (> 20000 véhicules) à moins de 200 m	SPF M-T	B
der20k500	Route (> 20000 véhicules) à moins de 500 m	SPF M-T	B
devert	Distance aux forêts et espaces verts	C. L. C.	Q
devert100	Forêt ou espace vert à moins de 100 m	C. L. C.	B
devert500	Forêt ou espace vert à moins de 500 m	C. L. C.	B
zoneinond	Aléa d'inondation	LEPUR	Q
zoneinond3	Aléa d'inondation de niveau fort	LEPUR	Q
chomage2001	Taux de chômage en 2001 (secteurs statistiques)	INS	Q
revmed97	Revenu médian en 1997 (secteurs statistiques)	INS	Q
chomage1997	Taux de chômage en 1997 (communes)	IWEPS	Q
diplomes01	Taux de diplômés du supérieur en 2001 (communes)	INS	Q

VIII. Script Python pour la fonction *Tabulate Area*

```
import sys, string, os, arcgisscripting

gp = arcgisscripting.create()
gp.CheckOutExtension("spatial")
gp.AddToolbox("C:\Program Files\ArcGIS\Desktop10.2\ArcToolbox\Toolboxes\Spatial Analyst
Tools.tbx")

i = 0

# changer dans la condition while le i en fct du nombre de fichiers shp à traiter
# changer les chemin d'accès aux fichiers
# les fichiers shp à traiter doivent se nommer BUFFER0.shp, BUFFER1.shp, BUFFER2.shp etc (à
paramétrer dans l'outil split layer by attributes ou renommer les fichiers par lot avec un outil
comme Antrenamer)

while (i<=1612):

    BUFFER = "D:\CARTOTHEQUE\\buffer\Liege901000oook\BUFFER"+str(i)+".shp"
    Landcover_LB72 = "D:\CARTOTHEQUE\Corine Land Cover\clip2000"
    Tab_BUFFER = "D:\CARTOTHEQUE\\tabulate3L901000\TAB_BUFFER"+str(i)
    cellSize = 90
    inField1 = "buff"
    inField2 = "VALUE"
    gp.TabulateArea_sa(BUFFER, inField1, Landcover_LB72, inField2, Tab_BUFFER, cellSize)

    i = i + 1
```

IX. Interprétation du diagramme de dispersion de Moran

N° de quadrant	Position	Auto-corrélation	Interprétation
1	Haute-droite	Positive	L'entité est affectée d'une valeur haute, et donc de résidu élevé, dans un environnement élevé (valeurs de voisinage élevées). Au niveau de l'entité, il existe une sous-estimation du modèle.
2	Basse-droite	Négative	L'entité est isolée et de valeur faible, et donc de résidu négatif, dans un environnement élevé (valeurs de voisinage élevées). Au niveau de l'entité, il existe une surestimation du modèle.
3	Basse-gauche	Positive	L'entité est affectée d'une valeur basse, et donc de résidu négatif, dans un environnement bas (valeurs de voisinage basses). Au niveau de l'entité, il existe une surestimation du modèle.
4	Haute-gauche	Négative	L'entité est isolée et de valeur élevée, et donc de résidu élevé, dans un environnement bas (valeurs de voisinage faibles). Au niveau de l'entité, il existe une sous-estimation du modèle.

X. Liste des couples de variables concernés par une colinéarité excessive pour le modèle général liégeois

La liste ci-dessous reprend les couples de variables présentant un coefficient de corrélation supérieur à 0,75. A noter que certaines corrélations n'y figurent pas :

- Les corrélations entre les variables quantitatives et leurs décompositions en variables binaires
- Les corrélations entre les variables binaires provenant de la décomposition d'une même variable quantitative
- Les corrélations entre les variables de temps d'accès et de distance euclidienne à un même objet.

Part de bâti dans un rayon de 200 m	0,97	Part d'artificialisation dans un rayon de 200 m
Part de bâti dans un rayon de 1000 m	0,94	Part d'artificialisation dans un rayon de 1000 m
Distance à Tihange	0,96	Temps d'accès voiture à Huy
Distance à Tihange	-0,76	Temps d'accès à Verviers
Distance à Tihange	0,84	Temps d'accès à Namur
Distance au chemin de fer	0,88	Distance aux gares
Distance au chemin de fer	0,85	Temps d'accès minimal à une gare
Gare L à moins de 500 m	1,00	Gare à moins de 500 m
Gare L à moins de 700 m	0,95	Gare à moins de 500 m
Gare L à moins de 500 m	0,76	Gare à moins de 700 m

Gare L à moins de 700 m	0,80	Gare à moins de 700 m
Gare à moins de 700 m	0,88	Distance au chemin de fer
Commerce alimentaire à moins de 100 m	1,00	Et. enseignement supérieur à moins de 100 m
Distance aux gares IC/IR	0,76	Temps d'accès minimal à un établissement sup.
Distance aux routes à trafic >10000 veh/jour	0,76	Distance aux routes à trafic >20000 veh/jour
Distance aux routes à trafic >10000 veh/jour	0,77	Temps d'accès minimal à un établissement sup.
Route à trafic >10000 veh/jour à moins de 500 m	0,87	Route à trafic >20000 veh/jour à moins de 500 m
Route à trafic >10000 veh/jour à moins de 200 m	0,76	Route à trafic >20000 veh/jour à moins de 200 m
Route à trafic >10000 veh/jour à moins de 100 m	0,81	Route à trafic >20000 veh/jour à moins de 200 m
Distance aux routes à trafic >20000 veh/jour	0,91	Distance aux sorties d'autoroute
Distance aux routes à trafic >20000 veh/jour	0,81	Temps d'accès à Ans
Distance aux routes à trafic >20000 veh/jour	0,80	Temps d'accès à Herstal
Distance aux routes à trafic >20000 veh/jour	0,84	Temps d'accès à Liège
Distance aux routes à trafic >20000 veh/jour	0,89	Temps d'accès minimal à une sortie d'autoroute
Distance aux routes à trafic >20000 veh/jour	0,91	Distance aux sorties d'autoroute
Temps d'accès à Namur	0,88	Temps d'accès à Huy
Distance aux sorties d'autoroute	0,77	Temps d'accès à Herstal
Distance aux sorties d'autoroute	0,77	Temps d'accès à Liège
Temps d'accès à Ans	0,86	Temps d'accès à Herstal
Temps d'accès à Ans	0,94	Temps d'accès à Liège
Temps d'accès à Ans	0,91	Temps d'accès à Seraing
Temps d'accès à Herstal	0,95	Temps d'accès à Liège
Temps d'accès à Herstal	0,83	Temps d'accès à Verviers
Temps d'accès à Huy	0,87	Temps d'accès à Namur
Temps d'accès à Liège	0,80	Temps d'accès au Sart-Tilman
Temps d'accès à Liège	0,87	Temps d'accès à Seraing
Temps d'accès à Seraing	0,89	Temps d'accès au Sart-Tilman
Temps d'accès minimal à un établissement sec.	0,77	Distance aux nœuds commerciaux
Temps d'accès à Namur	0,77	Temps d'accès à Bruxelles

XI. Tests de colinéarité du modèle général liégeois

1. Modèle MCO

Facteurs d'inflation de variance	
Valeur minimale possible = 1.0	
Valeurs > 10.0 peut indiquer un problème de colinéarité	
eau_gaz_elec	1.055
SPA	1.303
SUP_TERRAI	1.184
LOTISSEMEN	1.256
LDEN60	1.096
tauxIPP	1.380
revmed2011	1.459
de_tihange	2.284
debus	1.326
decdf2014700	1.420
degares700	1.297
deroute20k200	1.056
t_liege	2.561
t_nodules	1.804
t_bruxelles	1.973
Lbusresid	1.593
VIF(j) = 1/(1 - R(j)^2), où R(j) est un coefficient de corrélation multiple entre la variable j et les autres variables indépendantes	

2. Modèle corrigeant l'hétéroscédasticité

Facteurs d'inflation de variance

Valeur minimale possible = 1.0

Valeurs > 10.0 peut indiquer un problème de colinéarité

eau_gaz_elec	1.070
SPA	1.314
SUP_TERRAI	1.223
LOTISSEMEN	1.283
CUL_DE_SAC	1.234
LDEN60	1.139
tauxIPP	1.424
chomage2001	1.979
revmed2011	2.181
de_eau_pre100	1.236
de_tihange	2.601
debus	1.360
delht500	1.540
t_liege	3.527
t_nodules	1.781
t_bruzelles	2.234
deroute20k200	1.067
AGOP	2.201

$VIF(j) = 1/(1 - R(j)^2)$, où $R(j)$ est un coefficient de corrélation multiple entre la variable j et les autres variables indépendantes

XII. Test de Breusch-Pagan du modèle général liégeois

Test de Breusch-Pagan pour l'hétéroscédasticité

MCO, utilisant les observations 1-356

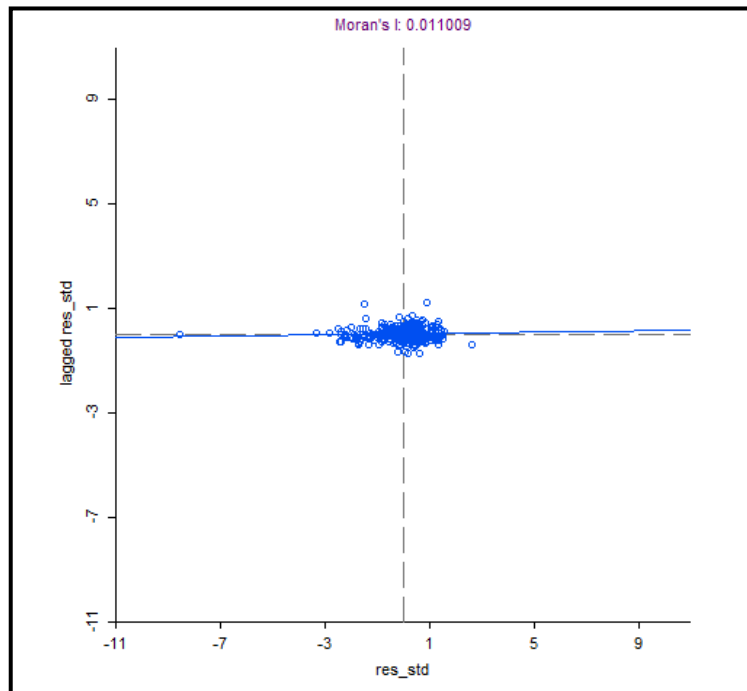
variable dépendante: à l'échelle $uhat^2$

	coefficient	erreur std.	t de Student	p. critique	
const	0.358829	1.73238	0.2071	0.8360	
eau_gaz_elec	0.594604	0.263788	2.254	0.0248	**
SPA	-0.442621	1.13389	-0.3904	0.6965	
SUP_TERRAI	0.000702838	0.000323684	2.171	0.0306	**
LOTISSEMEN	-0.609502	0.262689	-2.320	0.0209	**
LDEN60	-0.909027	0.663545	-1.370	0.1716	
tauxIPP	0.0852868	0.180406	0.4727	0.6367	
revmed2011	-3.48292e-05	2.87694e-05	-1.211	0.2269	
de_tihange	9.29967e-06	1.39949e-05	0.6645	0.5068	
debus	4.01151e-05	0.000348024	0.1153	0.9083	
decdf2014700	-0.229478	0.298960	-0.7676	0.4433	
degares700	1.64253	0.559144	2.938	0.0035	***
deroute20k200	-0.577761	0.775077	-0.7454	0.4565	
t_liege	-0.000519460	0.000273596	-1.899	0.0585	*
t_nodules	-0.000560313	0.000784411	-0.7143	0.4755	
t_bruzelles	3.30299e-05	0.000210360	0.1570	0.8753	
Lbusresid	22.0141	11.8058	1.865	0.0631	*

Somme des carrés expliquée = 218.07

Statistique de test: LM = 109.034990,
avec p. critique = $P(\text{Chi-deux}(16) > 109.034990) = 0.000000$

XIII. Diagramme de dispersion de Moran pour le modèle actuel général liégeois



XIV. Liste des couples de variables concernés par une colinéarité excessive pour le modèle périphérique liégeois

La liste ci-dessous reprend les couples de variables présentant un coefficient de corrélation supérieur à 0,75. A noter que certaines corrélations n'y figurent pas :

- Les corrélations entre les variables quantitatives et leurs décompositions en variables binaires
- Les corrélations entre les variables binaires provenant de la décomposition d'une même variable quantitative
- Les corrélations entre les variables de temps d'accès et de distance euclidienne à un même objet.

Part de bâti dans un rayon de 200 m	0,99	Part d'artificialisation dans un rayon de 200 m
Part de bâti dans un rayon de 1000 m	0,96	Part d'artificialisation dans un rayon de 1000 m
Taux de chômage (2010) par communes	0,86	Taux de dipl. du supérieur (2001) par commune
Distance à Tihange	-0,76	Temps d'accès à Verviers
Distance à Tihange	0,96	Temps d'accès à Huy
Distance à Tihange	0,84	Temps d'accès à Namur
Distance à Tihange	0,95	Temps d'accès min. à un commerce alimentaire
Temps d'accès minimal à une gare	0,87	Distance au chemin de fer
Gare IC/IR à moins de 500 m	1,00	Industrie à moins de 100 m
Gare à moins de 500 m	1,00	Gare L à moins de 500 m
Gare à moins de 700 m	0,81	Gare L à moins de 500 m
Gare à moins de 700 m	0,85	Gare L à moins de 700 m
Gare à moins de 500 m	0,95	Gare L à moins de 700 m
Distance aux gares	0,87	Distance au chemin de fer
Route à trafic >10000 veh/jour à moins de 500 m	0,92	Route à trafic >20000 veh/jour à moins de 500 m
Distance aux routes à trafic >20000 veh/jour	0,90	Distance aux sorties d'autoroute

Distance aux routes à trafic >20000 veh/jour	0,79	Temps d'accès à Ans
Distance aux routes à trafic >20000 veh/jour	0,79	Temps d'accès à Herstal
Distance aux routes à trafic >20000 veh/jour	0,84	Temps d'accès à Liège
Distance aux routes à trafic >20000 veh/jour	0,89	Temps d'accès min. à une sortie d'autoroute
Temps d'accès à Liège	0,75	Distance aux sorties d'autoroute
Temps d'accès à Herstal	0,75	Distance aux sorties d'autoroute
Temps d'accès à Ans	0,86	Temps d'accès à Herstal
Temps d'accès à Herstal	0,81	Temps d'accès à Verviers
Temps d'accès à Huy	0,88	Temps d'accès à Namur
Temps d'accès à Liège	0,77	Temps d'accès au Sart-Tilman
Temps d'accès à Liège	0,74	Temps d'accès min. à une sortie d'autoroute
Temps d'accès à Liège	0,88	Temps d'accès à Seraing
Temps d'accès à Seraing	0,85	Temps d'accès au Sart-Tilman
Temps d'accès à Namur	0,78	Temps d'accès à Bruxelles
Temps d'accès à Ans	0,94	Temps d'accès à Liège
Temps d'accès à Ans	0,92	Temps d'accès à Seraing
Temps d'accès à Herstal	0,95	Temps d'accès à Liège

XV. Tests de colinéarité du modèle périphérique liégeois

1. Modèle MCO

Facteurs d'inflation de variance	
valeur minimale possible = 1.0	
valeurs > 10.0 peut indiquer un problème de colinéarité	
SPA	1.461
SUP_TERRAI	1.323
BATIMENT	1.172
LDEN60	1.130
revmed2011	1.583
de_eau_pre100	1.178
de_tihange	2.349
debus	1.517
decdf2014700	1.650
de1500	1.544
t_liege	2.852
t_nodules	1.682
super300	1.185
t_bruzelles	2.284
lm1resid	1.866
Lbusresid	1.542
desaed100	1.411
second120	1.685
auto700	1.167
deicir500	1.574

VIF(j) = 1/(1 - R(j)^2), où R(j) est un coefficient de corrélation multiple entre la variable j et les autres variables indépendantes

2. Modèle corrigeant l'hétéroscédasticité

Facteurs d'inflation de variance

Valeur minimale possible = 1.0
Valeurs > 10.0 peut indiquer un problème de colinéarité

SPA	1.437
PBATI200	1.518
SUP_TERRAI	1.301
LOTISSEMEN	1.286
CUL_DE_SAC	1.326
LDEN60	1.148
revmed2011	1.686
de_eau_pre100	1.220
de_tihange	2.496
dealimenta500	1.249
debus	1.499
decdf2014700	1.582
auto700	1.176
t_liege	2.601
t_nodules	1.730
second120	1.592
super300	1.146
t_bruzelles	2.467
lmlresid	2.109
del500	1.389

$VIF(j) = 1/(1 - R(j)^2)$, où $R(j)$ est un coefficient de corrélation multiple entre la variable j et les autres variables indépendantes

XVI. Test de Breusch-Pagan du modèle périphérique liégeois

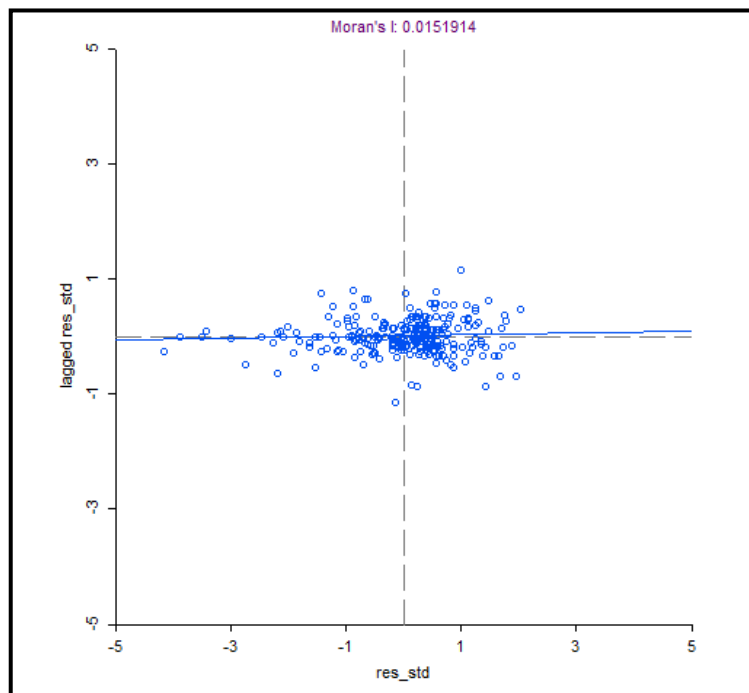
Test de Breusch-Pagan pour l'hétéroscédasticité
MCO, utilisant les observations 1-270
Variable dépendante: à l'échelle uhat^2

	coefficient	erreur std.	t de Student	p. critique	
const	-1.90777	1.54141	-1.238	0.2170	
SPA	-0.579314	0.997534	-0.5807	0.5619	
SUP_TERRAI	0.000816716	0.000343867	2.375	0.0183	**
BATIMENT	0.827541	0.800540	1.034	0.3023	
LDEN60	-0.661492	0.786043	-0.8415	0.4008	
revmed2011	-4.58348e-06	2.98069e-05	-0.1538	0.8779	
de_eau_pre100	0.130454	0.261808	0.4983	0.6187	
de_tihange	4.91486e-06	1.22952e-05	0.3997	0.6897	
debus	0.000118504	0.000325789	0.3637	0.7164	
decdf2014700	-0.0539728	0.324708	-0.1662	0.8681	
del500	-0.263415	0.655883	-0.4016	0.6883	
t_liege	-0.000485478	0.000303951	-1.597	0.1115	
t_nodules	0.000382533	0.000692219	0.5526	0.5810	
super300	-0.237200	0.604556	-0.3924	0.6951	
t_bruzelles	0.000202265	0.000191114	1.058	0.2909	
lmlresid	3.95456	3.98705	0.9919	0.3222	
lbusresid	53.7577	22.5948	2.379	0.0181	**
desaed100	-0.344951	0.980353	-0.3519	0.7252	
second120	0.0487714	0.564983	0.08632	0.9313	
auto700	-0.485343	0.730722	-0.6642	0.5072	
deicir500	-0.572734	2.05862	-0.2782	0.7811	

Somme des carrés expliquée = 83.3978

Statistique de test: LM = 41.698924,
avec p. critique = P(Chi-deux(20) > 41.698924) = 0.003026

XVII. Diagramme de dispersion de Moran pour le modèle périphérique liégeois



XVIII. Liste des couples de variables concernés par une colinéarité excessive pour le modèle carolorégien

La liste ci-dessous reprend les couples de variables présentant un coefficient de corrélation supérieur à 0,75. A noter que certaines corrélations n'y figurent pas :

- Les corrélations entre les variables quantitatives et leurs décompositions en variables binaires
- Les corrélations entre les variables binaires provenant de la décomposition d'une même variable quantitative
- Les corrélations entre les variables de temps d'accès et de distance euclidienne à un même objet.

Présence d'un raccord. à l'eau/gaz/électricité	0,81	Présence d'un raccordement à l'égout
Part de bâti dans un rayon de 200 m	0,93	Part d'artificialisation dans un rayon de 200 m
Part de bâti dans un rayon de 1000 m	0,93	Part d'artificialisation dans un rayon de 1000 m
Taux de chômage (2010) par communes	0,95	Taux de dipl. du supérieur(2001) par commune
Taux de chômage (2010) par communes	-0,79	Revenus médians agrégés par commune
Taux de dipl. du supérieur(2001) par commune	-0,81	Revenus médians agrégés par commune
Temps d'accès minimal à un ét. du supérieur	0,94	Temps d'accès min. à une sortie d'autoroute
Temps d'accès minimal à un ét. du supérieur	0,78	Temps d'accès minimal à une gare L
Temps d'accès minimal à un ét. du supérieur	0,87	Temps d'accès à Gosselies
Temps d'accès à Gilly	0,97	Temps d'accès minimal à un ét. du supérieur
Temps d'accès à Fleurus	0,91	Temps d'accès minimal à un ét. du supérieur
Temps d'accès à Châtelet	0,94	Temps d'accès minimal à un ét. du supérieur
Temps d'accès à Charleroi	0,96	Temps d'accès minimal à un ét. du supérieur
Temps d'accès à l'aéroport	0,91	Temps d'accès minimal à un ét. du supérieur
Temps d'accès minimal à un ét. du supérieur	0,95	Distance aux sorties d'autoroute
Temps d'accès minimal à un ét. du supérieur	0,91	Distance aux routes à trafic >20000 veh/jour

Temps d'accès minimal à un ét. du supérieur	0,83	Distance aux routes à trafic >10000 veh/jour
Temps d'accès minimal à un ét. du supérieur	0,80	Distance aux gares L
Temps d'accès minimal à un ét. du supérieur	0,75	Distance au chemin de fer
Temps d'accès minimal à un ét. du supérieur	0,82	Temps d'accès à Bruxelles
Temps d'accès minimal à un ét. du supérieur	0,82	Temps d'accès minimal à une gare L
Temps d'accès à Gosselies	0,95	Temps d'accès min. à une sortie d'autoroute
Temps d'accès à Gilly	0,93	Temps d'accès min. à une sortie d'autoroute
Temps d'accès min. à une sortie d'autoroute	0,90	Temps d'accès à Fleurus
Temps d'accès min. à une sortie d'autoroute	0,84	Temps d'accès à Châtelet
Temps d'accès min. à une sortie d'autoroute	0,97	Temps d'accès à Charleroi
Temps d'accès min. à une sortie d'autoroute	0,96	Temps d'accès à l'aéroport
Temps d'accès min. à une sortie d'autoroute	0,92	Distance aux routes à trafic >20000 veh/jour
Temps d'accès min. à une sortie d'autoroute	0,84	Distance aux routes à trafic >10000 veh/jour
Temps d'accès min. à une sortie d'autoroute	0,84	Distance aux gares L
Temps d'accès min. à une sortie d'autoroute	0,76	Distance aux gares
Temps d'accès min. à une sortie d'autoroute	0,78	Distance au chemin de fer
Temps d'accès min. à une sortie d'autoroute	0,93	Temps d'accès à Bruxelles
Temps d'accès minimal à une gare L	0,76	Temps d'accès à Gosselies
Temps d'accès minimal à une gare L	0,93	Temps d'accès minimal à une gare
Temps d'accès minimal à une gare L	0,82	Temps d'accès à Charleroi
Temps d'accès minimal à une gare L	0,76	Temps d'accès à l'aéroport
Temps d'accès minimal à une gare L	0,82	Distance aux sorties d'autoroute
Temps d'accès minimal à une gare L	0,88	Distance aux routes à trafic >20000 veh/jour
Temps d'accès minimal à une gare L	0,84	Distance aux routes à trafic >10000 veh/jour
Temps d'accès minimal à une gare L	0,80	Distance aux lignes à haute tension
Temps d'accès minimal à une gare L	0,78	Distance aux gares IC/IR
Temps d'accès minimal à une gare L	0,94	Distance aux gares
Temps d'accès minimal à une gare L	0,92	Distance au chemin de fer
Temps d'accès minimal à une gare L	0,76	Temps d'accès à Bruxelles
Temps d'accès minimal à une gare IC/IR	0,82	Temps d'accès minimal à une gare
Temps d'accès minimal à une gare IC/IR	0,80	Distance aux gares
Temps d'accès minimal à une gare IC/IR	0,78	Distance au chemin de fer
Temps d'accès à Gosselies	0,89	Temps d'accès à Gilly
Temps d'accès à Gosselies	0,95	Temps d'accès à Fleurus
Temps d'accès à Gosselies	0,77	Temps d'accès à Châtelet
Temps d'accès à Gosselies	0,91	Temps d'accès à Charleroi
Temps d'accès à Gosselies	0,99	Temps d'accès à l'aéroport
Temps d'accès à Gosselies	0,95	Distance aux sorties d'autoroute
Temps d'accès à Gosselies	0,87	Distance aux routes à trafic >20000 veh/jour
Temps d'accès à Gosselies	0,81	Distance aux routes à trafic >10000 veh/jour
Temps d'accès à Gosselies	0,77	Distance à une gare L
Temps d'accès à Gosselies	0,76	Temps d'accès à Namur
Temps d'accès à Gosselies	0,99	Temps d'accès à Bruxelles
Temps d'accès à Gilly	0,93	Temps d'accès à Fleurus
Temps d'accès à Gilly	0,96	Temps d'accès à Châtelet
Temps d'accès à Gilly	0,93	Temps d'accès à l'aéroport
Temps d'accès à Gilly	0,93	Distance aux sorties d'autoroute
Temps d'accès à Gilly	0,89	Distance aux routes à trafic >20000 veh/jour
Temps d'accès à Gilly	0,81	Distance aux routes à trafic >10000 veh/jour
Temps d'accès à Gilly	0,76	Distance aux gares L
Temps d'accès à Gilly	0,78	Temps d'accès à Namur
Temps d'accès à Gilly	0,84	Temps d'accès à Bruxelles
Temps d'accès aux gares	0,93	Temps d'accès minimal à une gare L
Temps d'accès aux gares	0,82	Temps d'accès minimal à une gare IC/IR
Temps d'accès aux gares	0,78	Distance aux routes à trafic >20000 veh/jour
Temps d'accès aux gares	0,82	Distance aux routes à trafic >10000 veh/jour
Temps d'accès aux gares	0,89	Distance aux gares L
Temps d'accès aux gares	0,87	Distance aux gares IC/IR
Temps d'accès aux gares	0,93	Distance au chemin de fer
Temps d'accès à Fleurus	0,86	Temps d'accès à Charleroi
Temps d'accès à Fleurus	0,96	Temps d'accès à l'aéroport

Temps d'accès à Fleurus	0,90	Distance aux sorties d'autoroute
Temps d'accès à Fleurus	0,85	Distance aux routes à trafic >20000 veh/jour
Temps d'accès à Fleurus	0,78	Distance aux routes à trafic >10000 veh/jour
Temps d'accès à Fleurus	0,90	Temps d'accès à Namur
Temps d'accès à Fleurus	0,90	Temps d'accès à Bruxelles
Temps d'accès à Fleurus	0,88	Temps d'accès à Châtelet
Temps d'accès à Châtelet	0,87	Temps d'accès à Charleroi
Temps d'accès à Châtelet	0,82	Temps d'accès à l'aéroport
Temps d'accès à Châtelet	0,84	Distance aux sorties d'autoroute
Temps d'accès à Châtelet	0,82	Distance aux routes à trafic >20000 veh/jour
Temps d'accès à Châtelet	0,80	Temps d'accès à Namur
Temps d'accès à Charleroi	0,93	Temps d'accès à l'aéroport
Temps d'accès à Charleroi	0,97	Distance aux sorties d'autoroute
Temps d'accès à Charleroi	0,91	Distance aux routes à trafic >20000 veh/jour
Temps d'accès à Charleroi	0,84	Distance aux routes à trafic >10000 veh/jour
Temps d'accès à Charleroi	0,84	Distance aux gares L
Temps d'accès à Charleroi	0,76	Distance aux gares
Temps d'accès à Charleroi	0,78	Distance au chemin de fer
Temps d'accès à Charleroi	0,86	Temps d'accès à Bruxelles
Temps d'accès à l'aéroport	0,95	Distance aux sorties d'autoroute
Temps d'accès à l'aéroport	0,88	Distance aux routes à trafic >20000 veh/jour
Temps d'accès à l'aéroport	0,81	Distance aux routes à trafic >10000 veh/jour
Temps d'accès à l'aéroport	0,77	Distance aux gares L
Temps d'accès à l'aéroport	0,79	Temps d'accès à Namur
Temps d'accès à l'aéroport	0,96	Temps d'accès à Bruxelles
Distance aux sorties d'autoroute	0,92	Distance aux routes à trafic >20000 veh/jour
Distance aux sorties d'autoroute	0,83	Distance aux routes à trafic >10000 veh/jour
Distance aux sorties d'autoroute	0,84	Distance aux gares L
Distance aux sorties d'autoroute	0,76	Distance aux gares
Distance aux sorties d'autoroute	0,78	Distance au chemin de fer
Temps d'accès à Bruxelles	0,92	Distance aux sorties d'autoroute
Distance aux routes à trafic >20000 veh/jour	0,91	Distance aux routes à trafic >10000 veh/jour
Distance aux routes à trafic >20000 veh/jour	0,81	Distance aux lignes à haute tension{
Distance aux routes à trafic >20000 veh/jour	0,88	Distance aux gares L
Distance aux routes à trafic >20000 veh/jour	0,78	Distance aux gares IC/IR
Distance aux routes à trafic >20000 veh/jour	0,81	Distance aux gares
Distance aux routes à trafic >20000 veh/jour	0,85	Distance au chemin de fer
Distance aux routes à trafic >20000 veh/jour	0,83	Temps d'accès à Bruxelles
Route à trafic >20000 veh/jour à moins de 500 m	0,81	Distance aux gares
Distance aux routes à trafic >10000 veh/jour	0,77	Distance aux lignes à haute tension
Distance aux routes à trafic >10000 veh/jour	0,84	Distance aux gares L
Distance aux routes à trafic >10000 veh/jour	0,79	Distance aux gares IC/IR
Distance aux routes à trafic >10000 veh/jour	0,85	Distance aux gares
Distance aux routes à trafic >10000 veh/jour	0,87	Distance au chemin de fer
Distance aux routes à trafic >10000 veh/jour	0,78	Temps d'accès à Bruxelles
Distance aux lignes à haute tension	0,79	Distance aux gares L
Distance aux lignes à haute tension	0,78	Distance aux gares
Distance aux lignes à haute tension	0,80	Distance au chemin de fer
Distance aux gares L	0,78	Distance aux gares IC/IR
Distance aux gares L	0,93	Distance aux gares
Distance aux gares L	0,92	Distance au chemin de fer
Distance aux gares L	0,76	Temps d'accès à Bruxelles
Gare L à moins de 500 m	0,84	Gare à moins de 500 m
Distance aux gares IC/IR	0,88	Distance aux gares
Distance aux gares IC/IR	0,86	Distance au chemin de fer
Distance aux gares	0,96	Distance au chemin de fer
Gare à moins de 500 m	0,84	Gare L à moins de 500 m
Distance aux lacs de l'Eau d'Heure	-0,76	Temps d'accès à Bruxelles

XIX. Tests de colinéarité du modèle carolorégien

1. Modèle MCO

```
Facteurs d'inflation de variance
valeur minimale possible = 1.0
valeurs > 10.0 peut indiquer un problème de colinéarité

    PBATI200    1.366
    SUP_TERRAI  1.167
    LOTISSEMEN  1.292
    BATIMENT    1.169
    t_charlero  2.647
    pente_char  1.241
    ltrainresi  1.420
    PC4facades  1.864
    second60    1.301
    matprim30   1.223
    desaed100   1.087
    delht500    1.296
    dealimenta700 1.486
    de_eau_pre200 1.209
    LDEN60     1.192
    debus      1.311
    t_nodules   2.324

VIF(j) = 1/(1 - R(j)^2), où R(j) est un coefficient de corrélation multiple
entre la variable j et les autres variables indépendantes
```

2. Modèle corrigeant l'hétéroscédasticité

```
Facteurs d'inflation de variance
valeur minimale possible = 1.0
valeurs > 10.0 peut indiquer un problème de colinéarité

    PBATI200    1.351
    SUP_TERRAI  1.153
    LOTISSEMEN  1.259
    BATIMENT    1.152
    LDEN60     1.193
    second60    1.299
    t_nodules   2.382
    matprim30   1.218
    t_charlero  2.668
    desaed100   1.169
    deravel100  1.123
    delht500    1.301
    decdf500    1.305
    dealimenta700 1.425
    debus      1.268
    ltrainresi  1.498
    PC4facades  1.896

VIF(j) = 1/(1 - R(j)^2), où R(j) est un coefficient de corrélation multiple
entre la variable j et les autres variables indépendantes
```

XX. Test de Breusch-Pagan du modèle carolorégien

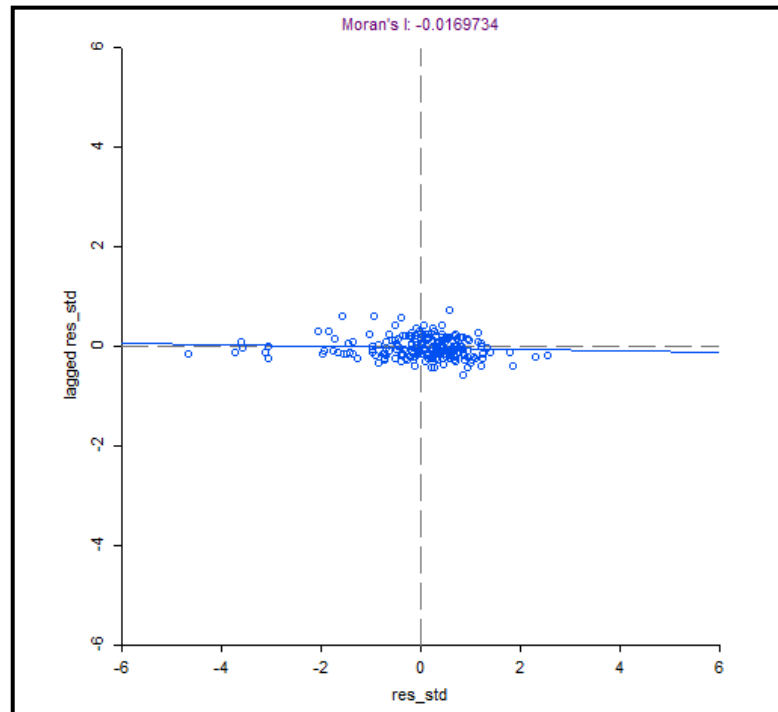
Test de Breusch-Pagan pour l'hétéroscédasticité
MCO, utilisant les observations 1-223
variable dépendante: à l'échelle uhat^2

	coefficient	erreur std.	t de Student	p. critique	
const	0.268417	0.712906	0.3765	0.7069	
PBATI200	0.00251738	0.00426826	0.5898	0.5560	
SUP_TERRAI	0.000886651	0.000347339	2.553	0.0114	**
LOTISSEMENT	-0.514664	0.285323	-1.804	0.0727	*
BATIMENT	0.192798	0.485089	0.3974	0.6915	
t_charlero	-0.000351235	0.000268143	-1.310	0.1917	
pente_char	0.00105076	0.0227839	0.04612	0.9633	
ltrainresi	0.340643	3.44383	0.09891	0.9213	
PC4facades	0.00694606	0.00564389	1.231	0.2198	
second60	-0.677825	0.599921	-1.130	0.2599	
matprim30	0.673345	0.466412	1.444	0.1504	
desaed100	-0.0966022	0.701420	-0.1377	0.8906	
delht500	-0.156139	0.277926	-0.5618	0.5749	
dealimenta700	0.676190	0.305950	2.210	0.0282	**
de_eau_pre200	-0.0387613	0.264534	-0.1465	0.8836	
LDEN60	-1.54806	1.03163	-1.501	0.1350	
debus	-0.000581092	0.000450726	-1.289	0.1988	
t_nodules	0.000218834	0.000765951	0.2857	0.7754	

Somme des carrés expliquée = 60.5231

Statistique de test: LM = 30.261567,
avec p. critique = $P(\text{Chi-deux}(17) > 30.261567) = 0.024519$

XXI. Diagramme de dispersion de Moran pour le modèle carolorégien



XXII. Liste des couples de variables concernés par une colinéarité excessive pour le modèle liégeois des années nonante

La liste ci-dessous reprend les couples de variables présentant un coefficient de corrélation supérieur à 0,75. A noter que certaines corrélations n'y figurent pas :

- Les corrélations entre les variables quantitatives et leurs décompositions en variables binaires
- Les corrélations entre les variables binaires provenant de la décomposition d'une même variable quantitative
- Les corrélations entre les variables de temps d'accès et de distance euclidienne à un même objet.

Part de bâti dans un rayon de 200 m	0,95	Part d'artificialisation dans un rayon de 200 m
Part de bâti dans un rayon de 1000 m	0,89	Part d'artificialisation dans un rayon de 1000 m
Distance à Tihange	-0,83	Temps d'accès à Verviers
Distance à Tihange	0,79	Temps d'accès à Huy
Distance à Tihange	0,76	Temps d'accès à Seraing
Distance au chemin de fer	0,78	Distance aux gares
Distance aux sorties d'autoroute	0,76	Distance aux routes à trafic >20000 veh/jour
Temps d'accès à Bruxelles	0,87	Temps d'accès à Ans
Temps d'accès à Bruxelles	0,88	Temps d'accès à Namur
Temps d'accès à Ans	0,81	Temps d'accès à Liège
Temps d'accès à Ans	0,84	Temps d'accès à Namur
Temps d'accès à Huy	0,86	Temps d'accès à Seraing
Temps d'accès à Huy	0,95	Temps d'accès à Namur
Temps d'accès à Seraing	0,77	Temps d'accès au Sart-Tilman
Temps d'accès à Seraing	0,81	Temps d'accès à Namur
Temps d'accès à Namur	0,88	Temps d'accès à Bruxelles
Taux de chômage (1997) par communes	0,98	Taux de diplômés supérieurs (2001) par sect. stat.
Distance aux routes à trafic >20000 veh/jour	0,80	Distance aux routes à trafic >10000 veh/jour
Route à trafic >20000 veh/jour à moins de 500 m	0,81	Route à trafic >10000 veh/jour à moins de 500 m

XXIII. Test de colinéarité du modèle liégeois des années nonante

1. Modèle MCO

Facteurs d'inflation de variance	
valeur minimale possible = 1.0	
valeurs > 10.0 peut indiquer un problème de colinéarité	
superficie	1.278
t90_liege	1.588
pente_lieg	1.077
cliplbusre	1.977
revmed97	1.312
de_eau_pre100	1.099
decdf100	1.042
der10k500	1.535
deextr500	1.222
deind200	1.287
auto500	1.245
VIF(j) = 1/(1 - R(j)^2), où R(j) est un coefficient de corrélation multiple entre la variable j et les autres variables indépendantes	

2. Modèle corrigeant l'hétéroscédasticité

```
Facteurs d'inflation de variance
Valeur minimale possible = 1.0
Valeurs > 10.0 peut indiquer un problème de colinéarité

de_eau_pre100    1.082
de_tihange      1.116
  auto500       1.131
  t90_liege     1.624
  pente_lieg    1.058
  cliplbusre    2.099
  revmed97      1.487
  superficie     1.310
  deind200      1.226
  densite01     1.860

VIF(j) = 1/(1 - R(j)^2), où R(j) est un coefficient de corrélation multiple
entre la variable j et les autres variables indépendantes
```

XXIV. Test de Breusch-Pagan du modèle liégeois des années nonante

```
Test de Breusch-Pagan pour l'hétéroscédasticité
MCO, utilisant les observations 1-266
variable dépendante: à l'échelle uhat^2

-----
                coefficient   erreur std.   t de Student   p. critique
-----
const           0.885023      0.820352      1.079          0.2817
superficie      0.000835482    0.000309156   2.702          0.0073   ***
t90_liege      -0.000487028   0.000439406  -1.108         0.2687
pente_lieg     0.0139349     0.0185892     0.7496         0.4542
cliplbusre     9.21864       6.04161       1.526          0.1283
revmed97      -3.46399e-05  3.19373e-05  -1.085         0.2791
de_eau_pre100  0.426230     0.340613     1.251          0.2120
decdf100       0.124305     0.674556     0.1843         0.8539
der10k500     -0.134130     0.267056    -0.5023         0.6159
deextr500      0.0935062     0.411429     0.2273         0.8204
deind200       0.149057     0.455083     0.3275         0.7435
auto500       -0.325767     0.490299    -0.6644         0.5070

Somme des carrés expliquée = 42.57

Statistique de test: LM = 21.285010,
avec p. critique = P(Chi-deux(11) > 21.285010) = 0.030536
```


XXV. Diagramme de dispersion de Moran pour le modèle liégeois des années nonante

