

---

## Mémoire

**Auteur :** Delrez, Ludovic

**Promoteur(s) :** Halleux, Jean-Marie

**Faculté :** Faculté des Sciences

**Diplôme :** Master en sciences géographiques, orientation générale, à finalité spécialisée en urbanisme et développement territorial

**Année académique :** 2021-2022

**URI/URL :** <http://hdl.handle.net/2268.2/15203>

---

*Avertissement à l'attention des usagers :*

*Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.*

*Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.*

---



FACULTÉ DES SCIENCES  
DÉPARTEMENT DE GÉOGRAPHIE

*APPLICATION D'UN MODÈLE HÉDONIQUE SUR LE COMPLEXE  
RÉSIDENTIEL URBAIN DE LIÈGE  
COMPARAISON DES DONNÉES DU SPF FINANCES ET D'IMMOWEB*

Mémoire présenté par : **Ludovic DELREZ**

pour l'obtention du titre de

**Master en sciences géographiques, orientation générale, à finalité spécialisée en  
urbanisme et développement territorial**

Année académique :  
Date de défense :

**2021-2022  
Septembre 2022**

Président de jury :  
Promoteur :  
Jury de lecture :

Pr. **Serge SCHMITZ**  
Pr. **Jean-Marie HALLEUX**  
Pr. **Aurelia HUBERT**  
Dr. **Guénaël DEVILLET**



FACULTÉ DES SCIENCES  
DÉPARTEMENT DE GÉOGRAPHIE

***APPLICATION D'UN MODÈLE HÉDONIQUE SUR LE COMPLEXE  
RÉSIDENTIEL URBAIN DE LIÈGE  
COMPARAISON DES DONNÉES DU SPF FINANCES ET D'IMMOWEB***

Mémoire présenté par : **Ludovic DELREZ**

pour l'obtention du titre de

**Master en sciences géographiques, orientation générale, à finalité spécialisée en  
urbanisme et développement territorial**

Année académique :

**2021-2022**

Date de défense :

**Septembre 2022**

Président de jury :

**Pr. Serge SCHMITZ**

Promoteur :

**Pr. Jean-Marie HALLEUX**

Jury de lecture :

**Pr. Aurelia HUBERT**

**Dr. Guénaël DEVILLET**

*“I invoke the first law of geography: everything is related to everything else, but near things are more related than distant things” (Tobler, 2004).*

**CONVENTION D'UTILISATION DES DONNÉES DE L'AGDP DANS LE CADRE DES  
RECHERCHES AU SEIN DE L'UNIVERSITÉ DE LIÈGE**

Dans le cadre de la gestion professionnelle des données du cadastre, l'accès à ces données requiert la signature de cette convention d'usage avant tout accès à celles-ci.

Je soussigné, Dobry Ludovic (Nom Prénom),  
étudiant (Fonction), m'engage

- à n'utiliser les données que dans le cadre précis de ma recherche, de ma thèse ou de mon travail de fin d'étude ;
- à utiliser les données dans le respect des principes de finalité et de proportionnalité ;
- à ne pas transmettre les données à des tiers ;
- à ne pas télécharger ou copier les données en dehors de mon ordinateur personnel ;
- à accompagner toute reproduction numérique ou imprimée du texte-copyright suivant :  
"© (date) Administration Générale de la Documentation patrimoniale".
- à effacer définitivement les données au terme de leur utilisation.

L'utilisateur est responsable du respect de cette convention qui engage son université, son unité de recherche, ainsi que lui-même.

Fait le 08-12-2021 à Verviers

Signature de l'utilisateur



## Remerciements

Je souhaite tout d'abord remercier l'ensemble de mes professeurs et leurs assistants qui ont d'une certaine manière permis l'aboutissement de ce mémoire. J'ai pu à travers leurs enseignements mobiliser une grande partie de mon apprentissage. Je remercie Aurélia Hubert et Guénaël Devillet pour leurs relectures. Plus spécifiquement, j'aimerais apporter une attention particulière à Jean-Marie Halleux, mon promoteur, pour sa relecture et ses explications au long de la réalisation de ce document.

Je souhaite remercier Hubert Maldague pour ses conseils sur les aspects statistiques de la méthode hédonique. En ce qui concerne l'acquisition des données et les détails sur celles-ci, je remercie Sébastien Hendrickx, sans qui de nombreuses variables n'auraient pas pu être considérées.

Je remercie aussi mes amis dont Alice Boodts pour ses corrections et ses conseils, ainsi qu'Antoine Bosly pour le cours d'introduction au Python et son aide général.

Un grand merci également à ma famille et ma belle-famille qui m'ont accompagné dans ma rédaction. Je dédie notamment une pensée particulière à mes parents qui au cours de mon éducation m'ont toujours appris la persévérance, merci papa et merci maman. Je souhaite également remercier Jean-François Montenair et Nicole Dehard pour leurs corrections orthographiques respectives. Finalement, un merci singulier à Marie Montenair qui m'accompagne depuis toutes ces années.

## **Résumé**

Ce document est une actualisation d'un mémoire réalisé en 2012, celui de Perinne Dethier. L'approche hédonique est utilisée pour comparer deux jeux de données. Le premier set de données fait suite à une récolte de données Immoweb par Grignet en 2011. La seconde est une base de données cadastrales du SPF Finances qui reprend l'ensemble des transactions et modifications au sein des biens immobiliers (et fonciers). La modélisation hédonique est utilisée pour la comparaison des bases de données. Les comparaisons entre les deux jeux de données sont faites sur leurs coefficients de détermination ( $R^2$ ) et sur leurs erreurs relatives sur le prix.

La recherche s'oriente sur le marché immobilier des maisons d'habitation appartenant au complexe résidentiel urbain de Liège. Un grand nombre de variables relatif aux biens eux-mêmes mais également à leur environnement sont employés. Les variables propres à l'environnement des biens font intervenir un ensemble de thématiques diverses tel que le paysage, la mobilité, les aspects socio-économiques ou encore la proximité des commerces. La variable du prix étant à expliquer par une série de composante complexe.

## ***Abstract***

*This paper is an update of a dissertation done in 2012, that of Perinne Dethier. The hedonic approach is used to compare two datasets. The first dataset follows an Immoweb data harvest by Grignet in 2011. The second one is a cadastral database of the FPS Finance that includes all transactions and modifications within real estate (and land). Hedonic modeling is used to compare the databases. Comparisons between the two datasets are made on their coefficients of determination ( $R^2$ ) and on their relative errors on the price.*

*The research focuses on the real estate market of residential houses belonging to the urban residential complex of Liege. A large number of variables relating to the properties themselves but also to their environment are employed. The variables specific to the environment of the properties involve a set of diverse themes such as the landscape, mobility, socio-economic aspects or the proximity of shops. The price variable is explained by a series of complex components.*



# Table des matières

<b>INTRODUCTION</b> .....	<b>14</b>
CONTEXTE ET STRUCTURE DE LA RECHERCHE.....	14
QUESTION DE RECHERCHE ET HYPOTHÈSES.....	15
ZONE D'ÉTUDE : LE COMPLEXE RÉSIDENTIEL URBAIN DE LIÈGE.....	15
<b>CHAPITRE 1    ETAT DE L'ART</b> .....	<b>19</b>
1.1    LA DÉFINITION D'UN BIEN IMMOBILIER.....	19
1.1.1 <i>L'utilité d'un bien, les apports de Lancaster (1966)</i> .....	19
1.1.2 <i>Les biens immobiliers, des biens combinatoires</i> .....	19
1.1.3 <i>Les caractéristiques des biens immobiliers</i> .....	20
1.2    QU'EST-CE QUE L'APPROCHE HÉDONIQUE ?.....	21
1.2.1 <i>L'approche hédonique, une combinaison de variables</i> .....	21
1.2.2 <i>Modèle hédonique du prix des logements</i> .....	22
1.2.2.1 <i>Corrélation entre variables</i> .....	22
1.2.2.2 <i>Modèle de régression linéaire simple</i> .....	24
1.2.2.3 <i>Modèle de régression linéaire multiple</i> .....	25
1.2.2.4 <i>La transformation des modèles économétriques</i> .....	25
1.2.3 <i>Les paramètres d'évaluation du modèle</i> .....	26
1.2.3.1 <i>L'erreur type d'estimation (SEE)</i> .....	26
1.2.3.2 <i>Le coefficient de détermination (<math>R^2</math>)</i> .....	28
1.2.4 <i>Les problèmes économétriques</i> .....	28
1.2.4.1 <i>Multicollinéarité</i> .....	29
1.2.4.2 <i>La dépendance spatiale</i> .....	30
1.2.4.3 <i>L'hétérogénéité spatiale</i> .....	31
1.3    LE MARCHÉ DU LOGEMENT.....	33
1.3.1 <i>L'approche hédonique, un équilibre entre l'offre et la demande</i> .....	33
1.3.1.1 <i>Considération d'un marché implicite</i> .....	33
1.3.1.2 <i>Fonctionnement du point de vue du consommateur et du producteur</i> .....	33
1.3.2 <i>Le marché de l'immobilier, un marché inefficace</i> .....	34
1.3.3 <i>Segmentation du marché</i> .....	35
<b>CHAPITRE 2    MÉTHODOLOGIE</b> .....	<b>36</b>
2.1    DÉFINITION DES OBJECTIFS DE LA MODÉLISATION ET APPROCHE ANALYTIQUE.....	37
2.1.1 <i>Les modèles hédoniques finaux de Grignet et de Dethier</i> .....	37
2.2    CHOIX ET DESCRIPTION DU SECTEUR D'ANALYSE ET NATURE DE L'ÉCHANTILLON.....	38

2.2.1	<i>Le type de segmentation : les maisons d'habitation</i>	38
2.2.2	<i>La représentation cartographique : le complexe résidentiel urbain de Liège</i>	39
2.3	COLLECTE DE L'INFORMATION ET DÉFINITION DES VARIABLES	39
2.3.1	<i>L'origine des données et leurs descriptions</i>	40
2.3.1.1	<i>Les données cadastrales du SPF Finances</i>	40
2.3.1.2	<i>Les données externes à la BD cadastrale du SPF Finances</i>	41
2.4	LA DÉFINITION OPÉRATIONNELLE DES VARIABLES RETENUES	42
2.4.1	<i>Comparaison des données du SPF Finances et des données Immoweb</i>	43
2.4.2	<i>Les variables d'accessibilité</i>	45
2.4.2.1	<i>La distance-temps en voiture aux centres névralgiques de Liège</i>	46
2.4.2.2	<i>L'accessibilité potentielle aux centres d'emplois</i>	47
2.4.2.3	<i>Les cartographies d'accessibilité de la CPDT</i>	49
2.4.2.4	<i>L'accessibilité potentielle aux gares et aux arrêts de bus</i>	49
2.4.2.5	<i>Proximité d'une gare ou d'un arrêt de bus</i>	50
2.4.3	<i>Les variables de valorisation sociale</i>	51
2.4.3.1	<i>Le revenu médian</i>	51
2.4.3.2	<i>Le taux de chômage</i>	51
2.4.3.3	<i>La part des diplômés du supérieur</i>	51
2.4.4	<i>Les variables de qualité physique</i>	52
2.4.4.1	<i>Densité de population</i>	52
2.4.4.2	<i>Part du bâti récent dans les constructions</i>	53
2.4.4.3	<i>Proximité aux fonctions essentielles</i>	53
2.4.4.4	<i>Accessibilité potentielle aux nodules commerciaux</i>	54
2.4.4.5	<i>Influence des parcs urbains</i>	54
2.4.4.6	<i>Proximité aux trames verte et bleue</i>	54
2.4.4.7	<i>Nuisances visuelles liées aux sites d'extractions, aux industries et aux SAR</i>	55
2.4.4.8	<i>Nuisances sonores liées aux transports et aux industries</i>	55
2.4.4.9	<i>Indice de diversité de Shannon</i>	56
2.4.4.10	<i>Indice d'estimation du paysage I de Schmitz-Kiemstedt</i>	57
2.5	RÉSUMÉ DES VARIABLES ET DES ÉCARTS AUX VARIABLES DE DETHIER	58
<b>CHAPITRE 3</b>	<b>ANALYSE PAR L'APPROCHE HÉDONIQUE</b>	<b>66</b>
3.1	DESCRIPTION ET ANALYSE DE LA BASE DE DONNÉES	66
3.1.1	<i>Répartition spatiale des biens immobiliers</i>	66
3.1.2	<i>L'application des statistiques descriptives</i>	67

3.1.2.1	<i>Distribution et analyse descriptive du prix</i> .....	68
3.1.2.2	<i>Analyse descriptive et transformation des autres variables</i> .....	71
3.1.3	<i>Distribution spatiale du prix des biens immobiliers</i> .....	74
3.1.3.1	<i>Autocorrélation spatiale globale du prix</i> .....	74
3.1.3.2	<i>Autocorrélation spatiale locale du prix</i> .....	77
3.1.4	<i>Echantillonnage pour la validation-croisée</i> .....	79
3.2	ANALYSE DES CORRÉLATIONS.....	79
3.2.1	<i>Analyse des corrélations pour la base de données du SPF Finances</i> .....	79
3.2.2	<i>Analyse des corrélations pour la base de données Immoweb</i> .....	81
3.3	ANALYSE DE RÉGRESSION.....	83
3.3.1	<i>Modèle A, inspiration du modèle de Dethier</i> .....	83
3.3.2	<i>Modèle B, exploration en considérant l'ensemble des variables</i> .....	84
3.3.3	<i>Modèle C, affinage du modèle B</i> .....	85
3.4	ANALYSE DES RÉSIDUS.....	86
3.4.1	<i>Distribution des résidus obtenus par régression sur le prix</i> .....	86
3.4.2	<i>Vérification de l'autocorrélation spatiale des résidus</i> .....	88
3.4.3	<i>L'exclusion des résidus marginaux</i> .....	90
3.5	MISE AU POINT DU MODÈLE FINAL.....	90
3.5.1	<i>Modèle final (Modèle D)</i> .....	90
3.5.2	<i>Validation du modèle final</i> .....	92
3.5.3	<i>Les valeurs de prix des variables indépendantes</i> .....	93
	<b>CONCLUSION</b> .....	<b>96</b>
	<b>ANNEXES</b> .....	<b>103</b>

## Liste des cartes

<i>Carte 1. Le complexe résidentiel urbain de Liège, définition de Van der Haegen en 1991</i> .....	17
<i>Carte 2. Répartition des maisons d'habitation par commune en 2011 pour les deux jeux de données</i> .....	66
<i>Carte 3. Comparaison de l'accessibilité potentielle à l'emploi du bien immobilier en 2011</i> .....	74
<i>Carte 4. Comparaison de la répartition du prix des biens immobiliers</i> .....	75
<i>Carte 5. Autocorrélation spatiale locale (LISA) sur le prix de vente en 2011</i> .....	78
<i>Carte 6. Autocorrélation spatiale locale (LISA) sur les résidus de la régression</i> .....	89

## Liste des figures

<i>Figure 1. Structure type d'un complexe résidentiel urbain</i> .....	16
<i>Figure 2. Variables caractéristiques d'un bien immobilier</i> .....	20
<i>Figure 3. Nuage de points indiquant le type de relation entre deux variables</i> .....	22
<i>Figure 4. Exemple de régression non-linéaire</i> .....	26
<i>Figure 5. Estimation par les moindres carrés ordinaires (MCO)</i> .....	27
<i>Figure 6. Représentation graphique de l'autocorrélation spatiale</i> .....	31
<i>Figure 7. Illustration du paradoxe de Simpson</i> .....	32
<i>Figure 8. Construction du prix implicite pour une caractéristique Z1</i> .....	34
<i>Figure 9. Cadre méthodologique à la réalisation d'un modèle hédonique</i> .....	36
<i>Figure 10. Histogramme et diagramme Q-Q pour la distribution du prix du SPF Finances (brutes)</i> .....	69
<i>Figure 11. Histogramme et diagramme Q-Q pour la distribution du prix du SPF Finances (95 %)</i> .....	69
<i>Figure 12. Exemple de l'effet de la transformation logarithmique sur la distribution des superficies</i> .....	72
<i>Figure 13. Comparaison de la distribution de l'âge des biens immobiliers entre les deux sets de données</i> .....	73
<i>Figure 14. Comparaison de la distribution de l'accessibilité à l'emploi entre les deux sets de données</i> .....	73
<i>Figure 15. Comparaison des diagrammes de Moran (15 voisins) pour les distributions OSIRIS</i> .....	77
<i>Figure 16. Distribution des résidus obtenus par régression sur le prix</i> .....	87
<i>Figure 17. Diagramme quantile-quantile pour les résidus obtenus par régression sur le prix</i> .....	87
<i>Figure 18. Diagramme de dispersion entre le prix et les résidus du modèle de régression</i> .....	90

## Liste des tableaux

Tableau 1. Indicateur d'ajustements des modèles obtenus par Immoweb .....	38
Tableau 2. Comparaison des variables structurelles entre Immoweb et le SPF Finances .....	43
Tableau 3. Approximation de l'âge selon les considérations du cadastre (IGEAT, 2010).....	44
Tableau 4. Mode principal de déplacement (tous types de déplacements confondus) (Cornelis et al., 2012) .....	45
Tableau 5. Paramètres et facteurs correctifs considérés (Cornet et al., 2005) .....	49
Tableau 6. Classification du bruit se basant sur l'information du bruit de l'aéroport de Liège.....	56
Tableau 7. Résumé des variables retenues et les écarts aux variables de Dethier (2012) .....	59
Tableau 8. Analyse de voisinage par le ratio du plus proche voisin .....	67
Tableau 9. Statistique descriptive et indice de forme de la distribution du prix .....	70
Tableau 10. Résultat du test de normalité de Kolmogorov-Smirnov sur le prix (95 % de la distribution) .....	71
Tableau 11. Résultat de l'indice I de Moran selon le choix de la matrice de poids pour les données du SPF Finances .....	76
Tableau 12. Résultat de l'indice I de Moran selon le choix de la matrice de poids pour les données Immoweb ..	76
Tableau 13. Résumé des indicateurs d'ajustements du modèle A .....	83
Tableau 14. Résumé des indicateurs d'ajustements du modèle B .....	85
Tableau 15. Résumé des indicateurs d'ajustements du modèle C .....	86
Tableau 16. Résumé des indicateurs d'ajustements des modèles D .....	91
Tableau 17. Résumé des indicateurs d'ajustements des modèles D pour la validation-croisée.....	92
Tableau 18. Coefficient de détermination et son écart entre les modèles D2 et D3 .....	93

## Liste des annexes

Annexe 1. Description des données contenues dans la BD cadastrales du SPF Finances.....	103
Annexe 2. Source des données utilisées pour la construction des variables .....	105
Annexe 3. Script python utilisé pour le calcul des variables .....	108
Annexe 4. Pondération et légende relative à la couche d'occupation CLC de 2012 .....	111
Annexe 5. Représentations graphiques de la distribution du prix.....	112
Annexe 6. Statistiques descriptives de l'ensemble des variables .....	113
Annexe 7. Matrice de corrélation de Spearman pour les variables quantitatives.....	124
Annexe 8. Matrice de corrélation bisériale .....	128
Annexe 9. Modélisation hédonique du prix.....	134
Annexe 10. Résultat du VIF .....	146
Annexe 11. Coefficient et erreur des modèles finaux.....	149
Annexe 12. Les critères de choix entre les modèles hédoniques .....	151

## Acronymes

ACNAW : Autorité de contrôle des nuisances sonores aéroportuaires en Région wallonne

ACP : Analyse en Composante Principale

AGDP : Administration Générale de la Documentation Patrimoniale

AIC : Critère d'information d'Akaike (*Akaike information criterion*)

ANOVA : Analysis of variance

API : Interface de programmation d'application (*application programming interface*)

BD : Base de données

BIC : Critère d'information Bayésien (*Bayesian information criterion*)

CC : Code de construction

CPDT : Conférence Permanente du Développement Territorial

CRU : Complexe résidentiel urbain

EEA : Agence européenne pour l'Environnement (*European Environment Agency*)

IGN : Institut Géographique National

INS : Institut National de la Statistique

LISA : *Local Indicator for Spatial Association*

MCO : Moindre Carré Ordinaire

MNT : Modèle numérique de terrain

OSM : *Open Street Map*

OSIRIS : *Organisation of statistical information for research by institutions and scientists*

SAED : Sites d'activité économique désaffectés

SAER : Sites d'activité économique à réhabiliter

SAR : Sites à réaménager

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

SIG : Système d'information géographique

SPF : Service Public Fédéral

SPW : Service Public de Wallonie

SSE : Somme des carrés des erreurs (*Sum Squared Error*)

SST : Somme des carrés totaux (*Sum Squares Total*)

SSR : Somme des carrés de la régression (*Sum Squares Regression*)

SRWT : Société Régionale Wallonne du Transport

VIF : Facteur d'inflation de la variance (*Variance Inflation Factor*)

ZRMA : Zone Résidentielle des Migrants Alternants

# INTRODUCTION

## Contexte et structure de la recherche

La recherche présentée dans ce mémoire est une actualisation du travail de Perrine Dethier réalisé en 2012. Son document s'intitule « Etude comparative entre les préférences révélées et déclarées. Les maisons d'habitation en région liégeoise ». Elle-même avait poursuivi le mémoire de Yannick Grignet réalisé en 2011, qui s'appelle « Approche hédonique du marché immobilier sur base d'annonces internet : les maisons d'habitation au sein de la région urbaine de Liège ». Ainsi, le présent document est fortement inspiré sur la structure des travaux de mes prédécesseurs. Pour se distinguer de ces dits travaux, les données récupérées d'Immoweb par Grignet (2011) sont comparées aux données de l'Administration Générale de la Documentation Patrimoniale (AGDP | SPF Finances). L'objectif est de produire un modèle hédonique sur les deux jeux de données et de vérifier la fiabilité des données du SPF Finances au terme de cette étude. Dethier (2012) réussi à produire des résultats significatifs, par l'approche hédonique, au cours de sa propre recherche en utilisant les données Immoweb. En suivant la même procédure, il est espéré d'arriver à des conclusions similaires par l'utilisation des données du cadastre. Les données du SPF Finances ont été modifiées par une communauté de chercheurs (OSIRIS) pour pouvoir être plus aisément appliquées à un contexte scientifique. Les données sont obtenues par l'intermédiaire de Sébastien Hendrickx, chercheur au Lepur<sup>1</sup>. Le point 2.3.1.1 inventaire l'information contenue dans la base de données (BD) cadastrales de l'AGDP.

La recherche est partagée en trois grandes parties inégales, ainsi qu'une introduction et une conclusion. L'introduction contextualise la recherche, propose une question de recherche ainsi que des hypothèses, puis définit la zone d'étude. Au chapitre 1, l'état de l'art est une synthèse de la littérature existante qui expose la théorie relative à l'approche hédonique. Le chapitre 2 est la méthodologie, ce point est particulièrement conséquent car il reprend un détail des données et la construction des variables employées dans la méthode hédonique. Le chapitre 3 analyse les bases de données en termes de statistiques descriptives et de corrélations avant la

---

<sup>1</sup> Ces données sont soumises à des droits d'utilisation. Une convention signée interdit notamment de conserver ces données au terme de cette étude ou encore de les partager à des tiers. Cette convention se situe à la page 4 de ce document.

construction des différents modèles. Finalement, la conclusion synthétise les résultats et propose des perspectives pour approfondir la recherche.

## **Question de recherche et hypothèses**

Le présent document vise à tester la BD cadastrales de l'AGDP, en comparant les modèles hédoniques de prix obtenus par l'information échantillonnée sur Immoweb. *In fine*, une question sous-jacente peut être posée :

*La qualité de la base de données cadastrale de l'AGDP est-elle suffisante pour permettre des modélisations hédoniques relatives aux prix des logements ?*

En d'autres mots, l'objectif du document est de mettre en exergue si les données du SPF Finances permettent une modélisation au moins équivalente à la modélisation de Dethier (2012) sur la base de l'information Immoweb en termes de paramètres explicatifs.

De cette question l'hypothèse générale de la recherche, le nombre de biens immobiliers supérieurs de la BD cadastrale permettrait une amélioration des estimations de prix, par l'approche hédonique, en comparaison de l'information récoltée par l'intermédiaire d'Immoweb. Des sous-hypothèses peuvent être émises selon les variables introduites dans le modèle. La superficie des logements est proportionnelle au prix ; un prix supérieur est attendu pour les biens sans étage ou encore le prix au mètre carré est dégressif avec la distance au centre de Liège. Finalement, l'effet de chaque variable sur le prix pourra être isolé. Certaines variables relatives à l'environnement du bien peuvent également avoir un effet tantôt positif, tantôt négatif sur le prix. Ainsi, les nuisances sonores devraient influencer négativement le prix, à l'inverse d'un paysage de qualité qui devrait augmenter la valeur du bien.

## **Zone d'étude : le complexe résidentiel urbain de Liège**

Il semble opportun de redéfinir le complexe résidentiel urbain (CRU), utilisé pour cette recherche, en effet, les définitions peuvent varier. Le CRU se compose de deux parties distinctes : la région urbaine et la Zone Résidentielle des Migrants Alternants (ZRMA).

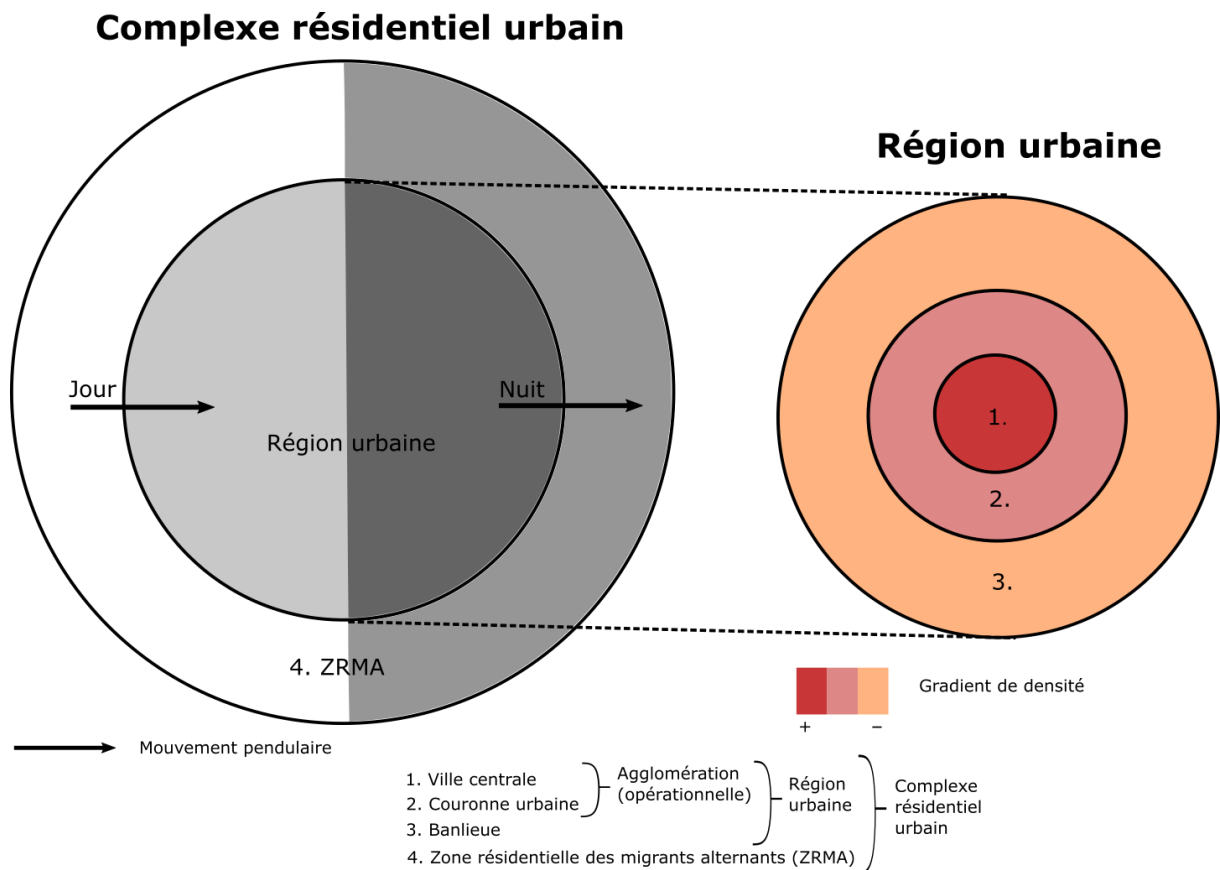
La **région urbaine** est définie par trois composantes principales. Celles-ci ont des caractéristiques propres mais sont interconnectées :

- La **ville centrale** est la commune où se situe la ville-centre, c'est-à-dire, le noyau urbain et les anciens quartiers à construction très denses.



- La **couronne urbaine** entoure la ville centrale, ce sont les quartiers qui possèdent toujours un bâti continu mais qui est moins dense. Ces quartiers sont également plus récents, ils datent majoritairement du XX<sup>e</sup> siècle.
- La **banlieue** borde l'agglomération (opérationnelle), ce sont les faubourgs de la ville. L'habitat y est moins dense et discontinu. La banlieue est le résultat d'un développement croissant de la ville.

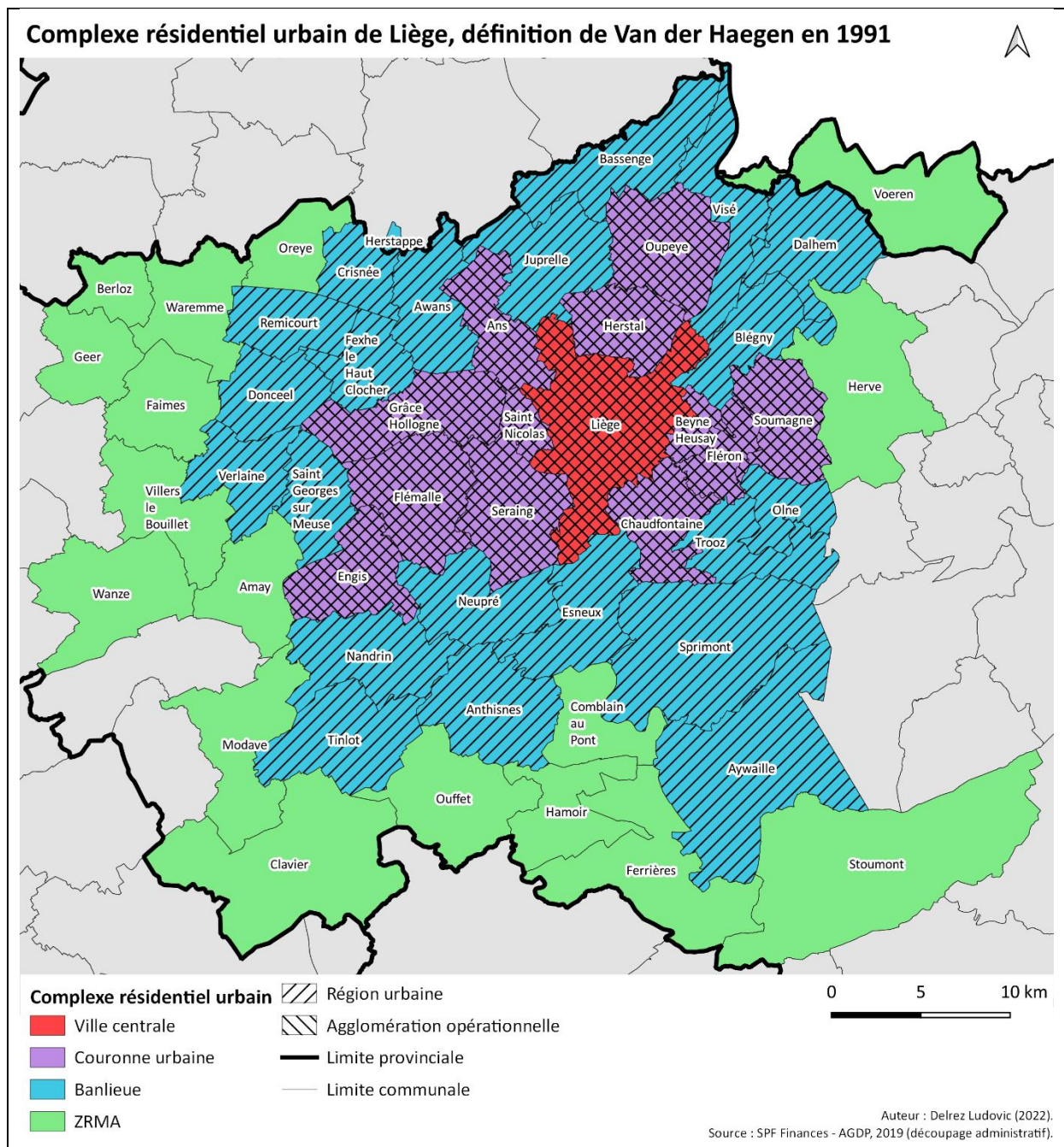
L'**agglomération opérationnelle** est l'association de la ville centrale et de la couronne urbaine. Ainsi, elle considère les communes avec un tissu urbain dense et continu (Figure 1) (Van Hecke *et al.*, 2009). La **ZRMA** est constituée des lieux de destination dont la part de navetteurs est importante. D'une certaine manière, ces lieux sont comme attirés par la région urbaine, car elle est une grande génératrice d'emploi. Elle concentre une série d'activités et de services (Van Hecke *et al.*, 2009). Ainsi, le **complexe résidentiel urbain** est la zone rassemblant l'agglomération urbaine et les lieux qu'elle polarise. En d'autres mots, cette zone coïncide avec l'aire d'influence de la ville (Figure 1).



Auteur : Delrez Ludovic (2022).  
Sur la base de Van der Haegen, Van Hecke et Juchtmans (1996).

**Figure 1. Structure type d'un complexe résidentiel urbain**

Le choix de cette définition est relatif à deux raisons ; premièrement, elle permet des comparaisons avec les travaux de Dethier (2012) et Grignet (2011). Deuxièmement, le CRU prend en compte les communes polarisées par la région urbaine liégeoise. Ainsi, la Carte 1 correspond à la zone étudiée, c'est-à-dire au CRU liégeois comme il a été défini en 1991 par Van der Haegen, Van Hecke et Juchtmans (1996).



**Carte 1. Le complexe résidentiel urbain de Liège, définition de Van der Haegen en 1991**

L'agglomération opérationnelle est composée de la ville centrale – commune où se situe la ville-centre, Liège (Carte 1, en rouge) – et de la couronne urbaine (Carte 1, en mauve). Les communes englobées sont Liège, Ans, Beyne-Heusay, Chaudfontaine, Engis, Flémalle, Fléron, Grâce-Hollogne, Herstal, Oupeye, Saint-Nicolas, Seraing et Soumagne.

La région urbaine est constituée de l'agglomération opérationnelle et de la banlieue (Carte 1, en bleu). Les communes comprises dans la banlieue sont Anthisnes, Awans, Aywaille, Bassenge, Blégny, Crisnée, Dalhem, Donceel, Esneux, Fexhe-le-Haut-Clocher, Herstappe, Juprelle, Nandrin, Neupré, Olne, Remicourt, Saint-Georges-Sur-Meuse, Sprimont, Tinlot, Trooz, Verlaine et Visé.

Finalement, le CRU considère l'ajout de la ZRMA (Carte 1, en vert) à la région urbaine. Les navetteurs proviennent des communes d'Amay, de Berloz, de Clavier, de Comblain-au-Pont, de Faimés, de Ferrières, de Geer, de Hamoir, de Herve, de Modave, d'Ouffet, d'Oreye, de Stoumont, de Villers-le-Bouillet, des Fourons (Voeren), de Wanze et enfin de Waremme.

Les régions urbaines de Belgique ont été redéfinies à plusieurs reprises, la dernière actualisation datant de 2017, par Vanderstraeten et Van Hecke (2019). Néanmoins, Cette définition n'est pas utilisée. Ce choix est délibéré pour permettre des comparaisons avec le travail de Dethier (2012). De plus, les changements relatifs au CRU de Liège de 2017 sont mineurs et concernent l'ajout de trois communes périphériques (Manhay, Lierneux et Marchin) au CRU liégeois de 1991.

# Chapitre 1      ETAT DE L'ART

## 1.1 La définition d'un bien immobilier

### 1.1.1 L'utilité d'un bien, les apports de Lancaster (1966)

En économie, Lancaster (1966) remet en cause les théories traditionnelles de l'utilité<sup>2</sup> d'un bien au départ de trois hypothèses :

- 1) Un bien seul n'est pas à l'origine de l'utilité, ce sont les caractéristiques du bien qui donnent lieu à une utilité.
- 2) Un bien possède généralement plusieurs caractéristiques et celles-ci peuvent être partagées entre différents biens.
- 3) Un bien combiné peut posséder des caractéristiques différentes des biens considérés séparément.

Ainsi, Lancaster (1996) mit en avant l'hétérogénéité des biens, dont les biens immobiliers ne font pas exception.

### 1.1.2 Les biens immobiliers, des biens combinatoires

En plus d'être des biens hétérogènes, les biens immobiliers sont des biens combinatoires, car ils sont constitués d'une superficie foncière et d'une structure bâtie (Halleux, 2019). Cette définition fait écho à la définition légale d'un bien immobilier. Selon l'Art. 3.47 du Code civil : « *Sont immeubles par leur nature, les fonds de terre et les divers volumes les composant, déterminés en trois dimensions.* » (Justice belge, 2020). Les fonds de terre faisant référence au foncier et les volumes à la structure bâtie.

---

<sup>2</sup> L'**utilité (individuelle)** peut être définie comme la satisfaction résultante d'un individu à posséder une certaine quantité de biens, indépendamment de la valeur marchande de ceux-ci. Tandis que l'**utilité collective** est la somme des utilités individuelles d'une collectivité, qui fait référence à un bien-être global. Par exemple, lors de la mise en place d'un aéroport, l'utilité collective liée à l'activité d'un aéroport est supérieure aux utilités individuelles des résidents impactés par la présence de cet aéroport. D'un point de vue économique, l'utilité collective et l'utilité individuelle permettent de trancher en faveur d'un des deux partis. En l'occurrence, ici il semble que le bien-être général qu'apporte l'aéroport soit plus important que le bien-être général de l'ensemble des résidents. Finalement, nous pouvons définir l'**utilité marginale** comme la satisfaction obtenue par l'ajout d'une unité supplémentaire d'un bien. En général, l'utilité marginale tend à décroître avec l'ajout d'une unité supplémentaire (Bezbakh et Gherardi, 2000).

### 1.1.3 Les caractéristiques des biens immobiliers

Comme Lancaster (1966) le signifiait, un bien est défini par un ensemble de caractéristiques. Ces caractéristiques sont résumées sur la Figure 2. Les variables explicatives du prix des biens immobiliers sont dissociables en deux grandes catégories ; les variables de localisation (ou variables extrinsèques) et les variables intrinsèques (ou structurelles).

Les **variables de localisation** sont relatives à l'accessibilité et à la qualité environnementale. Les variables propres à l'environnement peuvent également être partagées en deux groupes : les variables de qualité physique (impact du paysage, microclimat, etc.) et les variables de valorisation sociale (niveau de richesse, effet de la criminalité, etc.).

Les **variables intrinsèques** sont des variables propres aux biens fonciers ou aux biens immobiliers. Pour les biens purement fonciers, les variables intrinsèques vont être relatives à la superficie des terrains, ce qui permet de ne pas considérer les données liées à la construction. Tandis que pour un bien immobilier, les variables intrinsèques concernent les éléments liés à la construction – en d'autres mots au nombre de pièces, au nombre de salles de bains ou encore à la présence (ou l'absence) d'un garage (Halleux, 2019).

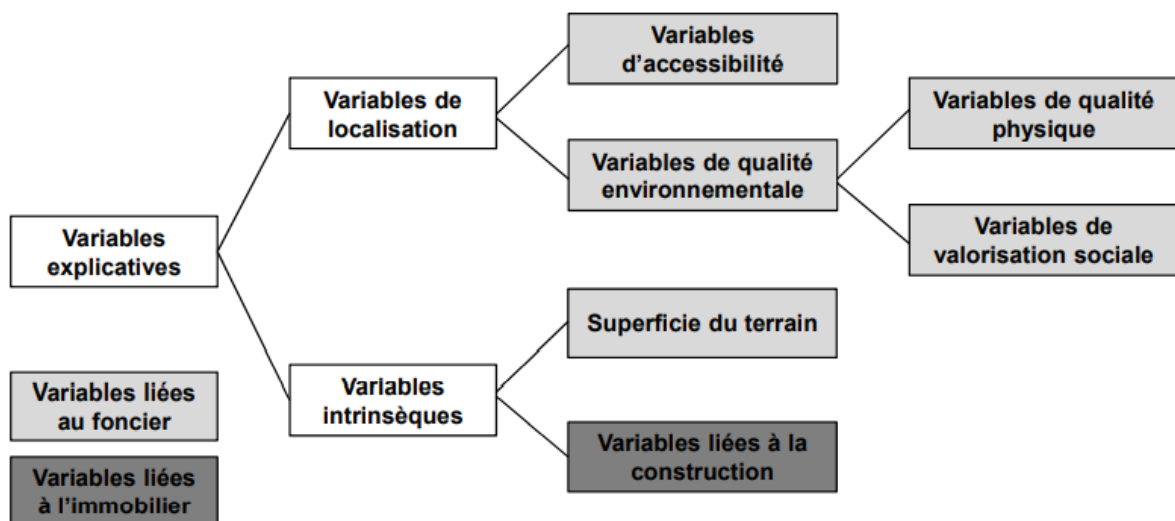


Figure 2. Variables caractéristiques d'un bien immobilier  
(Halleux, 2019)

## 1.2 Qu'est-ce que l'approche hédonique ?

### 1.2.1 L'approche hédonique, une combinaison de variables

Bien que l'approche hédonique soit employée dans un grand nombre de domaines, la détermination du prix sur les marchés fonciers et immobiliers y tient une part majeure (Anselin & Lozano-Gracia, 2009). Aujourd'hui, l'Université de Liège y a largement contribué, avec la réalisation de mémoires sur les maisons d'habitation (Grignet, 2011 ; Dethier, 2012), sur les appartements (Lorquet, 2016) ou encore sur le marché foncier (Maldague, 2014 ; Naets, 2015).

Le prix du marché est déterminé par l'intermédiaire de multiples variables. Celles-ci sont souvent catégorisées comme au point 1.1.3. Ainsi, le modèle hédonique permet de rapprocher les différentes caractéristiques d'un bien à un prix, à une valeur. Le prix correspond à la quantité monétaire qu'un acteur doit céder pour l'obtention d'une unité d'un bien donné, tandis que la valeur correspond à l'utilité retirée de ce bien (Bezbakh & Gherardi, 2000). Le prix est une valeur fixée à un instant donné pour l'échange, là où la valeur fluctue constamment (Halleux, 2019). Il est convenu que le prix hédonique soit, en réalité, la valeur marchande du bien. Plus précisément, cette valeur correspond au prix de transaction le plus probable (Des Rosiers, 2001).

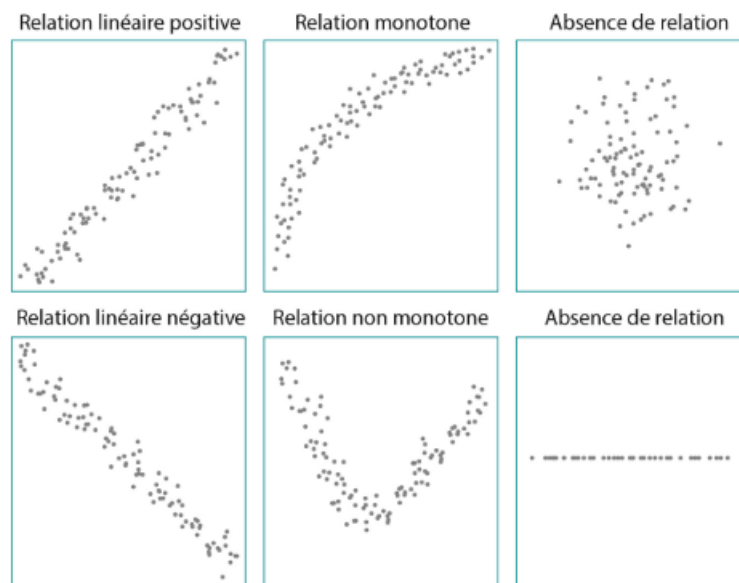
Pour rendre compte des variables récurrentes dans les analyses hédoniques, il semble opportun de montrer quelques-unes de leurs utilisations. En théorie, les composantes intrinsèques peuvent déjà expliquer 60 % à 70 % du prix d'un bien. Tandis que les composantes extrinsèques introduites dans le modèle voient leurs parts explicatives s'accroître – jusqu'à 50 % – au fur et à mesure qu'une centralité prend de l'importance (Des Rosiers *et al.*, 2008).

Les modèles hédoniques du prix de l'immobilier sont couramment employés pour évaluer la valeur imposable d'un logement et pour évaluer la capitalisation des externalités non-marchandes. Le modèle se base sur un large éventail de caractéristiques propres aux biens immobiliers, mais également sur des variables de localisation qui influencent le prix du bien. Par exemple, un bien immobilier présentant un attrait paysager particulier devrait présenter une valeur supérieure (Anselin & Lozano-Gracia, 2009). À l'inverse, des externalités négatives – dues à la proximité d'un aéroport, d'une autoroute (Palmquist, 1992) ou encore d'une centrale nucléaire (Maldague, 2014) – peuvent induire une baisse des prix.

## 1.2.2 Modèle hédonique du prix des logements

### 1.2.2.1 Corrélacion entre variables

Avant d'expliquer ce qu'est un modèle hédonique au sens mathématique, il apparaît opportun de spécifier ce qu'est une régression simple. Une régression linéaire (simple) permet d'expliquer la relation entre la variable à expliquer (y) (aussi appelée variable dépendante/endogène) et la variable explicative (x) (aussi appelée variable indépendante/exogène) à travers une droite dite de régression. L'analyste utilise en général une représentation graphique (tel qu'un nuage de points) et calcule un coefficient de corrélation pour déterminer l'intensité de la relation. La forme de la relation peut être sinusoïdale, exponentielle ou encore logarithmique. Le sens est dit monotone si la relation est strictement positive ou négative. Finalement, l'intensité de la relation peut également être commentée selon que les observations sont dispersées ou rapprochées les unes des autres dans un plan (Figure 3) (Feuillet *et al.*, 2019).



**Figure 3. Nuage de points indiquant le type de relation entre deux variables**  
(Feuillet *et al.*, 2019)

Le coefficient de corrélation de Bravais-Pearson ( $r$ ) se calcule selon l'Équation 1. Celui-ci varie de -1 à 1. Lorsque  $r = 0$ , il n'y a pas de corrélation entre les deux variables. À l'inverse, plus le coefficient se rapproche de 1 (ou -1) et plus la corrélation est forte (Feuillet *et al.*, 2019).

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Avec,

$\sigma_{xy}$  : covariance entre x et y ;

$\sigma_x$  : variance de x ;

$\sigma_y$  : variance de y ;

$\bar{x}$  : moyenne des x ;

$\bar{y}$  : moyenne des y.

**Équation 1. Coefficient de corrélation de Bravais-Pearson (r)**

Le coefficient de corrélation de Pearson ne doit pas être utilisé si la distribution est anormale. Ainsi, le coefficient de Spearman ( $\rho$ ) (Équation 2) est souvent employé pour les analyses non-paramétriques. Celui-ci évalue une valeur corrélative à partir de la position des  $x_i$  et  $y_i$ , plutôt que leurs valeurs propres.

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = 1 - \frac{6 \sum_{i=1}^n (R(x_i) - R(y_i))^2}{n(n^2 - 1)}$$

Avec,

$R(x_i)$  : rang de  $x_i$  ;

$R(y_i)$  : rang de  $y_i$  ;

$\bar{x}$  : moyenne des x ;

$\bar{y}$  : moyenne des y ;

$n$  : nombre d'observations.

**Équation 2. Coefficient de corrélation par rang de Spearman ( $\rho$ )**

Finalement, la relation entre les variables dichotomiques et les variables quantitatives est évaluée à travers l'utilisation d'un coefficient de corrélation à part entière, le coefficient de corrélation de point bisérial ( $r_{pb}$ ) (Équation 3).



$$r_{pb} = \frac{(m_1 - m_0)p_0p_1}{s_y}$$

Où,  $p_0 = \frac{n_0}{n}$  et  $p_1 = \frac{n_1}{n}$

Avec,

$n$  : nombre d'observations ;

$n_1$  : nombre d'observations présentant une valeur de 1 ;

$n_0$  : nombre d'observations présentant une valeur de 0 ;

$s_y$  : écart-type de  $y$  ;

$m_1$  : moyenne de  $y$  lorsque  $x = 0$  ;

$m_0$  : moyenne de  $y$  lorsque  $x = 1$ .

### Équation 3. Coefficient de corrélation biserial de point ( $r_{pb}$ )

Les trois coefficients de corrélation présentés sont interprétés comme le coefficient de Bravais-Pearson. Ils varient de -1 à 1, avec une absence de corrélation en 0.

#### 1.2.2.2 Modèle de régression linéaire simple

Dans le cas d'une relation linéaire simple, une droite permet la modélisation de la relation. La pente de la droite ( $\beta_1$ ) correspond ainsi à la variation de  $y$  ( $\Delta y$ ) sur la variation de  $x$  ( $\Delta x$ ). L'ordonnée à l'origine ( $\beta_0$ ) est une constante. Une erreur ( $\varepsilon_i$ ), appelée résidu, subsiste étant donné qu'un modèle n'est jamais parfait (Équation 4) (Feuillet *et al.* 2019). En effet, il subsiste toujours une différence entre la valeur observée de  $y$  et la valeur attendue de  $y$ , simplement car les variables ne rendent que partiellement compte de la réalité (Jawlik, 2016).

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$\beta_1 = \frac{\Delta y}{\Delta x}$$

Avec,

$y_i$  : la variable dépendante ;

$x_i$  : la variable indépendante ;

$\beta_0$  : l'ordonnée à l'origine (constante) ;

$\beta_1$  : le coefficient de régression associé à la variable indépendante ;

$\varepsilon_i$  : le résidu.

### Équation 4. Modèle de régression linéaire simple

### 1.2.2.3 *Modèle de régression linéaire multiple*

Le modèle de régression linéaire simple ne considère qu'une seule variable explicative  $x$ , tandis que le modèle de régression multiple considère plusieurs variables explicatives ( $x_i$ ). Ainsi, la représentation est multidimensionnelle, il y a autant de dimensions que de variables introduites dans le modèle (Jawlik, 2016). La fonction hédonique du prix est équivalente à une régression multiple (Équation 5) (Des Rosiers, 2001).

$$y_i = \beta_0 + \sum_k^p \beta_k x_{ik} + \varepsilon_i$$
$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon_i$$

Dans le cas d'une analyse hédonique, les paramètres peuvent être décrits comme suit :

- $y_i$  : le prix de l'immobilier (variable dépendante) ;
- $x_{ik}$  : la caractéristique  $k$  du bien  $i$  (variable indépendante) ;
- $\beta_0$  : l'ordonnée à l'origine (constante) ;
- $\beta_k$  : le coefficient de régression associé à la caractéristique  $k$  ;
- $p$  : le nombre de variables explicatives ;
- $\varepsilon_i$  : les résidus.

#### **Équation 5. Modèle de régression linéaire multiple**

### 1.2.2.4 *La transformation des modèles économétriques*

Il faut également remarquer que le modèle n'est pas toujours linéaire (Rosen, 1974). Des traitements sur les variables sont à envisager au cas par cas pour leur rendre une certaine linéarité, même si la forme linéaire d'une variable reste celle à privilégier pour faciliter l'interprétation des résultats, dans certains cas, une transformation de la variable s'impose. Plusieurs modèles non-linéaires peuvent être rencontrés en fonction de la transformation envisagée. Les transformations peuvent s'appliquer aux variables indépendantes ( $x$ ) mais également à la variable dépendante ( $y$ ) (Terra, 2005 ; Jawlik, 2016).


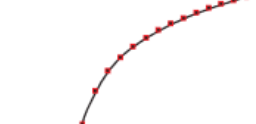
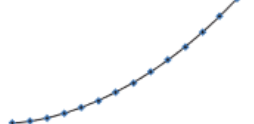
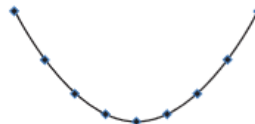
Exponential	Logarithmic	Power	Polynomial
$y = a * b^x$	$y = a + (b * \log(x))$	$y = a * x^b$	$y = a_i x^i + a_{i-1} x^{i-1} + \dots + a_1 x + a_0$
			

Figure 4. Exemple de régression non-linéaire  
(Jawlik, 2016)

Les transformations suggérées sont par exemple la transformation logarithmique, la transformation réciproque (consiste à prendre l'inverse de la variable,  $1/X$ ), la transformation exponentielle (généralement se traduit par la mise au carré de la variable,  $X^2$ ), la transformation quadratique (par l'utilisation d'un polynôme du deuxième ou troisième degré), la transformation « ratio » (par exemple, en considérant le prix au mètre carré), etc. (Figure 4) (Des Rosiers *et al.*, 2008 ; Jawlik, 2016).

### 1.2.3 Les paramètres d'évaluation du modèle

#### 1.2.3.1 L'erreur type d'estimation (SEE)

La somme des carrés totaux (SST) correspond à la somme des écarts quadratiques par rapport à la moyenne, c'est une mesure de l'ensemble de la variation de  $y$  autour de sa moyenne. La somme des carrés des résidus (ou des erreurs) (SSE) est la somme des écarts quadratiques par rapport à la droite de régression. SSE est une composante de SST, l'indice mesure la variation de  $y$  non expliquée, en d'autres mots l'erreur du modèle. Ainsi, au plus le SSE est petit, meilleur est l'ajustement à la droite de régression. La somme des carrés de la régression (SSR) est la seconde composante de SST. Le SSR mesure la variation de  $y$  expliquée par le modèle (Jawlik, 2016).

$$SST = \sum (y_i - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y_i - \hat{y})^2$$

$$SST = SSR + SSE$$

$$SSR = SST - SSE$$

Avec,

$y_i$  : valeur observée

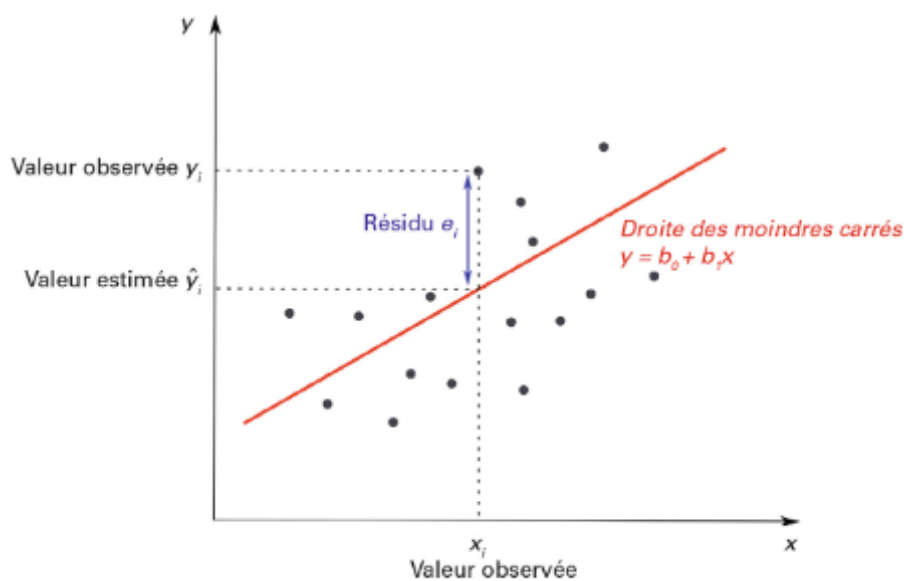
$\hat{y}$  : valeur attendue

$\bar{y}$  : valeur moyenne

Pour mesurer la qualité de l'ajustement, le SSR n'est pas à considérer seul, car un grand SST induit inévitablement un grand SSR. Ainsi, en considérant la division des deux indices, le coefficient de détermination ( $R^2$ ) est trouvé.

$$R^2 = \frac{SSR}{SST}$$

Le meilleur ajustement correspond à la droite qui minimise la somme des écarts quadratiques, c'est-à-dire la droite obtenue par moindres carrés ordinaires (MCO), celle qui minimise SSE (ou encore qui maximise  $R^2$  et SSR). En géographie, un intérêt particulier est porté sur les résidus du modèle ( $\varepsilon_i$ ) (et sur leurs cartographies) qui renseignent sur des spécificités locales, souvent laissées pour compte au profit d'une tendance globale. Pour cela, la droite des moindres carrés permet de minimiser la somme totale du carré des résidus, tel que cette droite s'ajuste au mieux aux nuages de points (Figure 5). La valeur résiduelle ( $\varepsilon_i$ ) équivaut à la valeur ajustée de  $y$  (appelée  $\hat{y}_i$ ) dont est soustrait la valeur observée de  $y$  (Jawlik, 2016 ; Feuillet *et al.*, 2019).



**Figure 5. Estimation par les moindres carrés ordinaires (MCO)**  
(Feuillet *et al.*, 2019)

### 1.2.3.2 Le coefficient de détermination ( $R^2$ )

Le choix entre plusieurs modèles est déterminé par plusieurs paramètres relatifs à la régression. Un des premiers validateurs regardé est le coefficient de détermination ( $R^2$ ). Ce coefficient renseigne sur la part expliquée du modèle, plus précisément sur la part de la variation totale de  $y$  expliquée. Le calcul du  $R^2$  découle du  $R$  multiple (celui-ci est semblable au coefficient de corrélation sur son interprétation). Le  $R^2$  varie entre 0 et 1. Plus le coefficient est élevé, meilleure est l'adéquation du modèle de régression (Équation 6) (Jawlik, 2016).

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

#### Équation 6. Coefficient de détermination $R^2$

Un inconvénient du  $R^2$  est que, lorsque le nombre de variables explicatives ( $x$ ) augmente, le  $R^2$  augmente également. Étant donné que cette augmentation ne rend pas nécessairement compte d'un ajustement de qualité, le  $R^2$  ajusté (ou  $\bar{R}^2$ ) est utilisé. Le  $R^2$  ajusté est une correction du  $R^2$  qui considère le nombre de variables  $x$  introduites dans le modèle (Équation 7) (Jawlik, 2016).

$$\bar{R}^2 = R^2 - (1 - R^2) \frac{p}{n - p - 1}$$

Avec,

$\bar{R}^2$  : coefficient de détermination ajusté ;

$R^2$  : coefficient de détermination ;

$n$  : nombre d'observations ;

$p$  : nombre de variables prédictives  $x$ .

#### Équation 7. Coefficient de détermination ajusté ( $\bar{R}^2$ )

## 1.2.4 Les problèmes économétriques

Avant d'entamer une validation du modèle par des tests statistiques, la robustesse est évaluée à travers une série de paramètres tels que la colinéarité, l'hétéroscédasticité ou encore l'autocorrélation.

#### 1.2.4.1 Multicollinéarité

Premièrement, les variables explicatives (x) doivent en principe être indépendantes les unes des autres pour éviter les problèmes de multicollinéarité<sup>3</sup>. Dans la pratique, les relations existantes entre les variables sont inévitables, il convient dès lors de limiter celles-ci. Il est admis qu'une corrélation supérieure à 80 % entre deux variables indépendantes (x) est à exclure. De plus, les variables binaires sont moins sensibles à la redondance étant donné qu'elles ciblent une caractéristique unique (Des Rosiers, 2001). Un indice de multicollinéarité est calculé par le logiciel GeoDa. Il doit être inférieur à 20 ou 30. De plus, la valeur VIF (*Variance Inflation Factor*) informe sur la redondance des variables explicatives. L'indice est calculé par le logiciel ArcGis. Il est conseillé que le VIF soit inférieur à 3, voire à 2 (Feuillet *et al.*, 2019). Dans la pratique, un VIF de 5 à 10 est admis, tout dépend de la précision attendue du modèle (Esri, 2018). Pour cette recherche, un VIF inférieur à 7,5 sera considéré.

$$VIF(x_j) = \frac{1}{1 - R_j^2}$$

Où  $j = 2, \dots, j$

Avec,  $R_j^2$  : coefficient de détermination  $R^2$  pour les variables considérées excepté  $x_j$ .

#### Équation 8. Facteur d'inflation de la variance (VIF)

De manière générale, la procédure est itérative et permet de percevoir l'effet des variables sur le modèle. Elle consiste soit à supprimer une à une les variables présentant une multicollinéarité trop importante (*forward stepwise*) ou, soit à l'inverse, d'introduire les variables une à une dans le modèle (*backward stepwise*). Le même principe d'exclusion est appliqué aux valeurs présentant une part explicative trop faible. Dans certains cas, les variables indépendantes (x) fortement corrélées entre elles peuvent être rassemblées en une nouvelle variable unique (Des Rosiers, 2001 ; Feuillet *et al.*, 2019). L'analyse déductive (*forward stepwise*) est utilisée pour générer un modèle au départ d'un modèle prédéfini par Dethier (2012). Sinon, la méthode utilisée est davantage inductive (*backward stepwise*). De

---

<sup>3</sup>La multicollinéarité est définie comme des interdépendances qui peuvent exister entre deux (ou plusieurs) variables indépendantes x. Dans la mesure où ce phénomène n'est pas considéré, l'effet respectif des corrélations entre les multiples variables indépendantes x et la variable dépendante y ne peut clairement être isolé (Jawlick, 2016).

plus, un test t de Student est effectué sur chaque variable, il indique la significativité de la variable dans le modèle. Finalement, un test de Fisher détermine la significativité globale du modèle. Plus précisément, il évalue si les variables indépendantes (x) parviennent à expliquer la variable dépendante (y), le prix. Les tests sont évalués à un niveau de significativité de 0,05. (Des Rosiers, 2001).

#### 1.2.4.2 La dépendance spatiale

La **dépendance spatiale** peut être définie comme la relation de proximité qui peut exister entre une observation et ces observations voisines, leurs degrés de ressemblance (Feuillet *et al.*, 2019). Par exemple, Anselin et Lozano-Gracia (2009) mettent en avant qu'une dépendance spatiale forte peut être observée lorsque les caractéristiques des logements proches sont davantage semblables à celles des logements éloignés.

La définition statistique de la dépendance spatiale est l'autocorrélation spatiale. L'**autocorrélation spatiale** correspond à la corrélation d'une variable avec elle-même. Comme le souligne Le Gallo (2002), en présence d'autocorrélation spatiale, l'inférence statistique par MCO n'est pas permise. Naturellement, l'autocorrélation spatiale doit être mesurée et testée, l'indice I de Moran offre une mesure globale de l'autocorrélation spatiale (Équation 9) (Feuillet *et al.*, 2019).

$$I = \frac{n \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_i (x_i - \bar{x})^2}$$

Où  $i \neq j$  et  $W = \sum_i \sum_j w_{ij}$

Avec,

$i = i^{\text{ème}}$  observation ;

$n =$  nombre d'observations ;

$W =$  somme de tous les poids spatiaux ;

$w_{ij} =$  poids spatial entre l'individu  $i$  et  $j$  ;

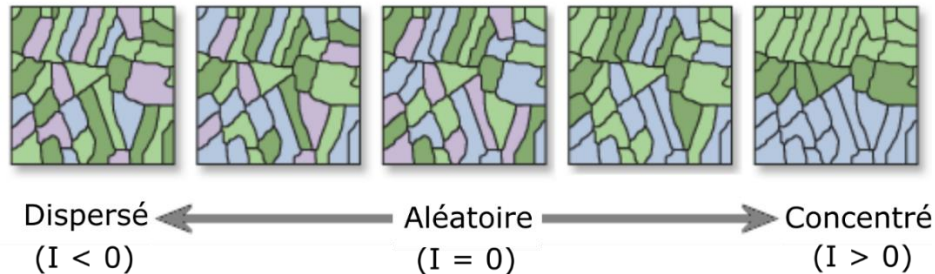
$\bar{x} =$  moyenne des valeurs ;

$x_i =$  valeur de l'observation  $i$  ;

$x_j =$  valeur de  $x$  dans le voisinage de  $i$ .

**Équation 9. L'indice d'autocorrélation spatiale global, I de Moran**

L'indice I de Moran est généralement standardisé entre -1 et 1. En -1, la distribution spatiale est dispersée, en 0 la distribution spatiale est aléatoire et en 1 la distribution spatiale est concentrée (Figure 6).



**Figure 6. Représentation graphique de l'autocorrélation spatiale**  
(Esri, s.d., modifié)

Un **indice d'autocorrélation spatiale locale** est déterminé à partir de la décomposition de l'indice I de Moran. Il s'appelle l'indicateur **LISA** (*Local Indicator for Spatial Association*) (Équation 10) (Feuillet *et al.*, 2019).

$$I_i = \frac{n(x_i - \bar{x}) \sum_j w_{ij}(x_j - \bar{x})}{\sum_j (x_j - \bar{x})^2}$$

**Équation 10. L'indice d'autocorrélation spatiale locale, LISA**

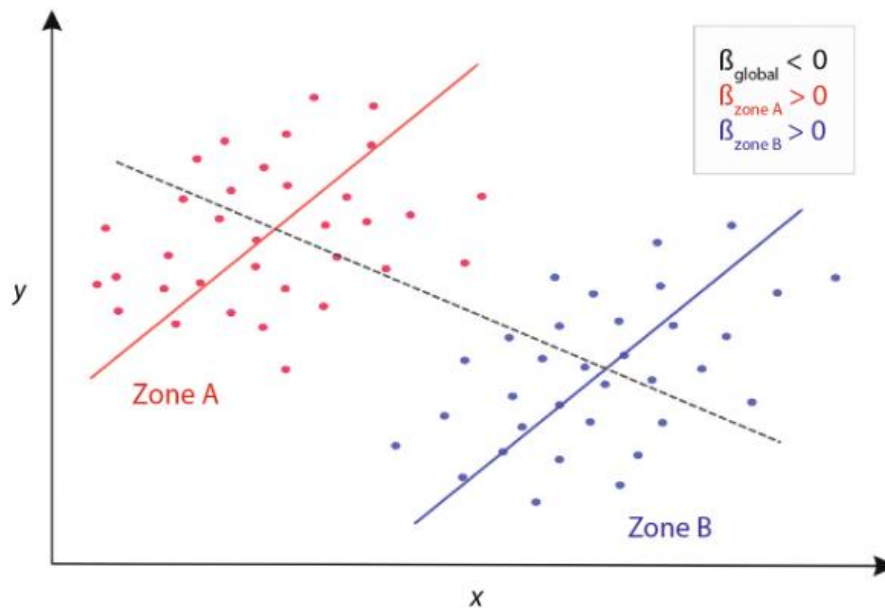
L'indice permet de connaître où des regroupements de valeurs similaires (ou dissemblables) sont observées sur le territoire. Par ailleurs, des regroupements locaux homogènes peuvent exister même si l'indice d'autocorrélation global indique une absence d'autocorrélation. Une méthode pour appréhender plus aisément les variations locales d'autocorrélations est le diagramme de Moran (Feuillet *et al.*, 2019).

#### 1.2.4.3 L'hétérogénéité spatiale

**L'hétérogénéité spatiale** indique si une instabilité spatiale est observée ou non. Elle correspond à une variabilité des paramètres statistiques (moyenne et variance) de certaines caractéristiques spatialisées, dépendant du lieu et de l'échelle d'analyse. Cette instabilité est représentée statistiquement par l'hétéroscédasticité des résidus. Il faut que la variance des



résidus soit constante pour avoir une homoscedasticité<sup>4</sup>. En d'autres mots, l'espace influence les paramètres statistiques de certaines variables, ce qui influence les corrélations et les covariances du modèle de régression (Des Rosiers, 2001 ; Feuillet *et al.*, 2019).



**Figure 7. Illustration du paradoxe de Simpson**  
(Feuillet et al., 2019)

Pour illustrer ce principe, la Figure 7 montre le paradoxe de Simpson. Lors d'une analyse de régression linéaire globale, il y a une non-prise en compte des phénomènes locaux, par un lissage des valeurs au profit de la moyenne. La relation statistique qui existe globalement est négative. Or, en considérant deux sous-groupes (zone A et zone B), les relations deviennent positives. Ainsi, en considérant uniquement la tendance globale, les singularités locales ne sont plus considérées. Une solution utilisée pour rendre compte des lois générales et des phénomènes locaux sont les régressions géographiquement pondérées et les régressions multiniveaux par une analyse en composante principale (ACP) (Feuillet *et al.*, 2019). Dans le cas de l'approche hédonique, les biens luxueux risquent d'induire un étalement de la distribution, due à leurs dispersions de valeurs souvent plus marquées. Ainsi, le modèle risque de surestimer la valeur des biens standards (Des Rosiers *et al.*, 2008). Dans GeoDa, les tests employés pour détecter l'hétéroscedasticité des résidus sont ceux de Breusch-Pagan et de

<sup>4</sup> Le terme s'oppose ainsi à la notion d'hétéroscedasticité.

Koender-Bassett (basé sur un test de Chi-2). Le test de Jarque-Bera vérifie également la normalité des résidus. Les tests sont évalués à un niveau de significativité de 0,05.

## 1.3 Le marché du logement

### 1.3.1 L'approche hédonique, un équilibre entre l'offre et la demande

#### 1.3.1.1 Considération d'un marché implicite

Un bien  $z$  peut être décrit en  $n$  caractéristiques. Cela se traduit par un vecteur  $z = (z_1, z_2, \dots, z_n)$ , où  $z_i$  est la quantité de la  $i^{\text{e}}$  caractéristique du bien considéré. La diversité des biens par leurs caractéristiques suppose un large spectre de produits différenciés disponibles sur le marché (Rosen, 1974). Dans un sens, sur le marché immobilier, cela se vérifie car tout bien immobilier est au moins différent en termes de localisation. En d'autres mots, les biens immobiliers sont hétérogènes (Halleux, 2019). Finalement, un prix implicite (ou contribution marginale) du bien  $p(z) = p(z_1, z_2, \dots, z_n)$  découle de l'ensemble de ces caractéristiques (Rosen, 1974).

#### 1.3.1.2 Fonctionnement du point de vue du consommateur et du producteur

Rosen (1974) généralise le prix du marché  $p(z) = p(z_1, z_2, \dots, z_n)$  à un équilibre du marché, constitué d'un grand nombre de consommateurs et d'un grand nombre de producteurs. Par ailleurs, l'hypothèse d'équilibre du marché suppose que la concurrence soit parfaite.

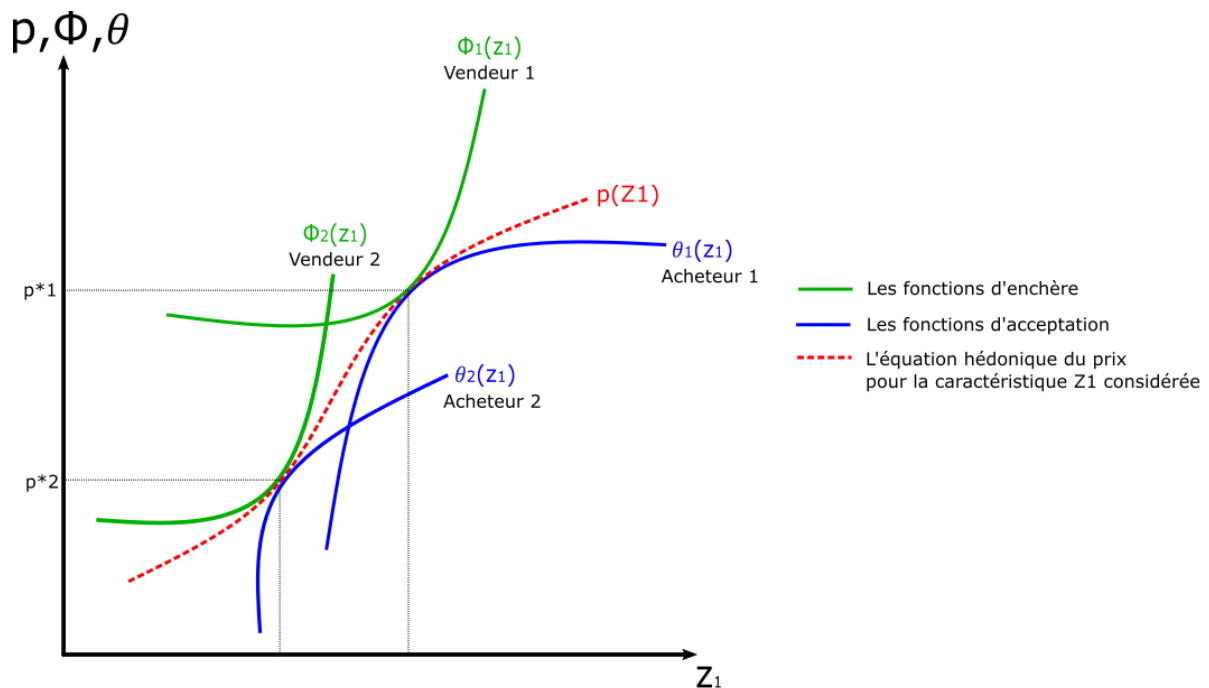
Pour les consommateurs, la fonction d'enchère  $\theta$  exprime la propension à payer d'un individu pour un bien complexe  $z$ . Elle peut être résumée par l'équation  $\theta = \theta(z_1, z_2, \dots, z_n; u, y, \alpha)$ .

La fonction fait intervenir la préférence d'achat  $\alpha$  et l'utilité  $u$ , qui est proportionnelle aux revenus  $y$  de l'acheteur. Au contraire, pour les producteurs, la fonction d'acceptation  $\Phi$  donne le prix minimum auquel les vendeurs sont prêts à proposer leur bien  $z$ . L'équation s'écrit :  $\Phi = \Phi(z_1, z_2, \dots, z_n; \pi, M, \beta)$ , elle fait intervenir les paramètres de production  $\beta$  et les bénéfices  $\pi$ , qui sont fonction du niveau de production  $M$ .

Lorsque toutes ces variables sont considérées ensemble afin de considérer le prix d'un bien (Figure 8), il est considéré que l'individu veut maximiser son utilité ( $u$ ), tout en minimisant le prix ( $p$ ) du bien sur le marché. Il cherche un optimum ( $p^*$ ) pour acquérir l'ensemble des caractéristiques du bien qu'il souhaite. Dès lors, la fonction d'enchère (ou de valeur)  $\theta(z)$  est supposée strictement concave. L'optimum peut être trouvé par le point de tangence entre la surface de prix  $p(z)$  et la surface d'enchère  $\theta(z)$ . À l'inverse, le producteur va chercher à

minimiser ses coûts (coûts des matériaux, coûts de la main d'œuvre, etc.), tout en maximisant ses bénéfices ( $\pi$ ). Ainsi, et parallèlement au traitement de la demande, la fonction d'acceptation  $\Phi(z)$  est supposée strictement convexe (Rosen, 1974).

Finalement, l'équation hédonique du prix  $p(z)$  détermine l'équilibre du marché qui correspond à l'addition de toutes les interactions de l'offre et la demande d'un bien donné (Anselin & Lozano-Gracia, 2009 ; Des Rosiers *et al.*, 2008).



Auteur : Delrez Ludovic (2022).  
 Au départ de Des Rosiers, Thériault et Dubé (2008),  
 sur la base des travaux de Rosen (1974).

Figure 8. Construction du prix implicite pour une caractéristique  $Z_1$

### 1.3.2 Le marché de l'immobilier, un marché inefficace

Même si les travaux de Rosen (1974) considèrent le marché comme étant à l'équilibre pur et parfait, le marché de l'immobilier ne l'est pas et ce pour plusieurs raisons :

- Le marché immobilier n'est pas atomistique, il n'y a pas un grand nombre de vendeurs et d'acheteurs.
- Il n'existe pas de prix unique pour un bien donné, les biens immobiliers sont au moins différents sur leurs localisations (hétérogénéité des biens). Le prix n'est pas fixé par des barèmes pour chaque type de caractéristiques du bien. Par exemple, une piscine va augmenter la valeur du bien de « x » milliers d'euros.

- Il y a un manque de transparence de l'information entre l'acheteur et le revendeur. De manière générale, le vendeur a une connaissance plus approfondie du bien, cette dissymétrie d'information est à son avantage.
- La mobilité des acteurs est restreinte à une zone géographique donnée généralement dépendante du lieu d'emploi (due à une immobilité physique des biens).

L'ensemble de ces éléments rend le marché immobilier relativement inefficace (Evans, 1995 ; Des Rosiers, 2001 ; Halleux, 2019). L'erreur moyenne des modèles de prix est similaire à l'erreur des évaluateurs expérimentés des biens ; c'est-à-dire 10 % du prix. Cette marge d'erreur est tout de même assez importante et rend compte d'un marché efficient à seulement 90 %. Ainsi, le prix d'un bien immobilier n'est que partiellement expliqué par le marché. Ce prix dépend des négociations entre le vendeur et l'acheteur, de la période durant laquelle ce bien reste sur le marché ou encore du comportement des agents immobiliers (Evans, 1995). Pour ces raisons, le modèle soumis est acceptable si l'erreur moyenne sur le prix est proche de 10 %. Des Rosiers (2001) admet une marge de 15 %, même si une marge de 10 % est préférable.

### **1.3.3 Segmentation du marché**

Sur Immoweb, le marché de l'immobilier est segmenté en douze catégories. Il y a une différenciation selon le type d'activités (logement, garage, bureau, commerce, industrie ou autres) qui nécessite des besoins et donc une utilité différente du bien. Il y a également une distinction entre le type de biens : maisons ou appartements. Immoweb distingue le marché locatif (à louer), du marché acquisitif (à vendre), ainsi que le marché ancien ou neuf avec des indications sur l'âge ou la rénovation des biens. Une catégorie spécifique pour le marché foncier (terrain) est également disponible (Immoweb, 2022). Dethier (2012) a basé son analyse sur 548 biens immobiliers échantillonnés sur Immoweb entre novembre 2010 et juin 2011 par Grignet. L'ensemble des observations est repris sur le marché acquisitif des maisons unifamiliales. Les données ont également été croisées avec les informations du cadastre pour obtenir l'âge des biens.

Des Rosiers, Thériault et Dubé (2008) soulignent l'importance du choix d'un marché homogène. Par exemple, considérer l'habitat d'un standing élevé et l'habitat bas de gamme dans un marché unique ne rend pas compte de la réalité de l'ensemble des ménages. Il y a une nécessité de segmenter les biens immobiliers par gamme de prix comparables pour rendre compte d'un marché homogène.

## Chapitre 2 Méthodologie

Des Rosiers (2001) propose une méthodologie pour la mise en œuvre de l'approche hédonique des prix. Dethier (2012) utilise le même cadre d'analyse. Celui-ci a l'avantage d'être détaillé, ce qui permet de réduire les erreurs méthodologiques. La Figure 9 reprend la méthode construite en dix points.

1. Définition des objectifs de la modélisation et approche analytique
2. Choix et description du secteur d'analyse et nature de l'échantillon
  - Le type de segmentation
  - La représentation cartographique
3. Collecte de l'information et définition des variables
  - Le traitement de la base de données
  - La sélection d'un sous-échantillon pour valider ultérieurement le modèle
  - La définition opérationnelle des variables retenues
4. Description et analyse de la base de données
  - L'application des statistiques descriptives
  - La transformation mathématique de variables
  - L'application des statistiques descriptives après corrections et épuration
5. Analyse de corrélation
  - Le test de corrélation simple
  - Le test de fiabilité
6. Analyse de régression
  - La procédure de régression standard
  - La réduction du nombre de variables
  - La transformation mathématique des variables indépendantes
  - La transformation mathématique de la variable dépendante
  - La procédure de régression par étape
7. Analyse des résidus
  - L'identification des résidus délinquants
  - La représentation graphique des résidus
8. Mise au point du modèle final
9. Validation du modèle final à l'aide du sous-échantillon retenu à l'étape 3
10. Production d'indications de la valeur

**Figure 9. Cadre méthodologique à la réalisation d'un modèle hédonique  
(Des Rosiers, 2001)**

## 2.1 Définition des objectifs de la modélisation et approche analytique

Comme discuté précédemment, l'objectif de cette recherche est de tester les données de l'AGDP à travers une modélisation des prix (par l'approche hédonique). Pour cela, les données Immoweb sont utilisées comme point de comparaison. Sur certains aspects, les données de l'AGDP sont peu fiables pour une utilisation scientifique. Cette problématique est développée au point 2.3.1.1. Globalement, l'information utilisée manque d'exhaustivité, car les données sont mises à jour uniquement lors de demandes de permis ou de la vente du bien. Or, plusieurs informations contenues dans la BD cadastrales ne nécessitent pas de demandes de permis comme l'installation d'une salle de bains ou d'un chauffage central. Ces biais pourraient aboutir à un modèle avec des parts explicatives ( $R^2$ ) trop faibles. En parallèle, l'ensemble des traitements proposés est appliqué à la BD Immoweb. Ils seront souvent moins détaillés car d'autres travaux ont déjà établi des modèles hédonique au départ de ces données (Grignet, 2011 ; Dethier, 2012). De plus, la comparaison des deux jeux de données n'est pas si évidente, il est essayé de limiter les écarts entre ceux-ci. Par exemple, l'ensemble des variables sont également calculées pour les données Immoweb. Le point de comparaison principal est le coefficient de détermination ( $R^2$ ). Il est utilisé pour informer sur la performance du modèle (tout en respectant l'ajustement de celui-ci). Un second indicateur employé est l'erreur moyenne du modèle sur le prix. Pour rappel, l'erreur moyenne du modèle doit approcher les 10 % pour être acceptable (Evans, 1990).

Enfin, une validation-croisée permet de vérifier la robustesse du modèle. La méthode consiste à une sélection d'un sous-échantillon de contrôle. Celui-ci n'est pas employé dans l'analyse de régression. Il permet *in fine* de valider le modèle en testant le modèle sur cet échantillon. Dans le cas où les paramètres du modèle sont également validés pour ce modèle, alors la régression est considérée comme robuste (Des Rosiers, 2001).

### 2.1.1 Les modèles hédoniques finaux de Grignet et de Dethier

Dethier (2012) améliora le modèle de Grignet (2011), avec un coefficient de détermination ( $R^2$ ) supérieur, notamment en considérant l'hétéroscédasticité des résidus à son modèle. Par ailleurs, considérer cet aspect a grandement contribué à réduire l'erreur relative du modèle (Tableau 1).

**Tableau 1. Indicateur d'ajustements des modèles obtenus par Immoweb**

	Grignet (2011)	Dethier (2012)
R <sup>2</sup>	0,8000	0,8800
R <sup>2</sup> ajusté	0,7900	0,8747
Prix moyen	188 802 €	183 801 €
Erreur type	37 480 €	23 642 €
Erreur relative	19,85 €	12,86 %
Test F	F : 69,3500 Prob(>F), (30, 517) : 0,0000*	F : 168,9200 Prob(>F), (20, 507) : 0,0000*

En termes de résultats, il est attendu que le modèle généré soit au moins aussi performant que ceux de Grignet (2011) et Dethier (2012). Le même segment de marché – les maisons unifamiliales – est ainsi sollicité et le modèle aspire à une prévision avec une erreur moyenne proche de 10 % (Evans, 1995).

## **2.2 Choix et description du secteur d'analyse et nature de l'échantillon**

La BD du SPF Finances possède un champ relatif à la date de l'acte de vente. Celui-ci varie temporellement de 2007 à 2015. Cette information pourrait être utilisée pour une analyse temporelle. Par exemple, l'analyse des données de 2007 et 2015 pourrait mettre en évidence des tendances générales en termes de prix pour les biens considérés. Néanmoins, ceci ne sera pas exploré dans le cadre de cette recherche.

En effet, dans l'objectif de comparaison avec les données Immoweb, une première sélection est effectuée pour les ventes de 2011. La date choisie se calque sur la période d'échantillonnage de Grignet pour l'année académique 2010 et 2011. Ainsi, un léger décalage est reconnu entre les données sélectionnées et les données de Grignet (2011).

### **2.2.1 Le type de segmentation : les maisons d'habitation**

La base de données cadastrales permet de segmenter le marché en plusieurs catégories. Un premier champ différencie les « maisons d'habitation » et les « immeubles à appartements ». Cette recherche portera exclusivement sur les « maisons d'habitation », toujours dans l'objectif comparatif avec Dethier (2012). Un second champ, plus descriptif, permet un affinage de la segmentation en « maison de cité », « villa », « bungalow » ou encore « maison bel-étage ». A priori, cette sous-typologie ne sera pas appliquée à cette recherche. Néanmoins,

une variable binaire multi-catégorielle – afin de rendre compte des différences de prix pour chaque sous-type de maison – est envisageable.

### **2.2.2 La représentation cartographique : le complexe résidentiel urbain de Liège**

L'échelle d'analyse concerne le complexe résidentiel urbain de Liège (Carte 1). Plus la zone d'étude est étendue, plus il y a un risque d'avoir plusieurs sous-marchés. En conséquence, il est recommandé de multiplier les variables explicatives. Une autre solution peut résider dans un découpage de la zone en plusieurs sous-secteurs. Le découpage en segmentation doit être effectué par des unités de voisinage semblable (Des Rosiers, 2001). Or, d'une certaine manière, les sous-ensembles constituant le CRU pourraient être utilisés à cet effet. Ces ensembles étant définis par des phases successives de développement de la ville.

### **2.3 Collecte de l'information et définition des variables**

Pour établir des points de comparaisons avec le modèle hédonique de Dethier (2012), il est recommandé de minimiser les écarts méthodologiques à celui-ci. Il y a une volonté de conserver des variables semblables à celles utilisées par Dethier pour sa modélisation. Cependant, une reproductibilité complète n'est pas possible. Tout d'abord, il y a un manque de détails sur les sources des données utilisées dans la réalisation des variables extrinsèques de Grignet et Dethier. Ensuite, cette seconde réalisation s'effectue à 10 ans d'écart de la première et les données utilisées ne sont plus nécessairement disponibles. Cette contrainte sur les données induit inévitablement des différences sur les variables produites. De même, les variables intrinsèques aux biens sont distinctes dû à la nature différente des données cadastrales et d'Immoweb. Ainsi, les écarts méthodologiques étant inévitables. L'ensemble des variables produites est également appliqué au set de données de Grignet (2011). Ceci permet de formuler un modèle hédonique au départ des mêmes variables pour les deux jeux de données.

Le choix des données introduites dans le modèle est arbitré par différentes études de cas et fondements théoriques, ainsi que par le modèle de Dethier (2012) qui constitue la base du modèle souhaité.

Par ailleurs, une certaine vigilance est de mise car l'approche hédonique ne fait qu'analyser les interactions entre les variables du modèle à travers des corrélations. En aucun cas, à elle seule, elle ne peut valider le lien de causalité entre une ou plusieurs variables. Néanmoins, si aucune variable importante n'a été omise et que l'ensemble des variables introduites est



fondée de manière cohérente, la méthode peut tout à fait expliquer une grande partie du phénomène observé (Des Rosiers, 2001).

### **2.3.1 L'origine des données et leurs descriptions**

#### *2.3.1.1 Les données cadastrales du SPF Finances*

Les données relatives aux biens immobiliers proviennent de l'inventaire des données de la documentation patrimoniale de l'AGDP. Ces données ont initialement un objectif fiscal, relatif à la fixation du revenu cadastral. L'utilisation de ces données à des fins scientifiques doit respecter quelques contraintes (IGEAT, 2010).

Premièrement, le niveau le plus fin mis à disposition est généralement l'échelle communale. Deuxièmement, l'agrégation importante des données ne permet pas toujours des croisements avec d'autres données existantes. Troisièmement, les données cadastrales ne sont pas mises à jour de manière systématique. Cela rend certaines variables peu fiables, comme les champs relatifs à la description des biens (IGEAT, 2010).

Les données de la documentation patrimoniale sont riches mais elles manquent à certains égards de structuration et de fiabilité pour des analyses scientifiques. Le projet OSIRIS – mené par une équipe de recherche (ULB, VUL et KUL) entre 2007 et 2010 – a apporté un regard nouveau sur ces données à travers un rapport de synthèse détaillant les métadonnées des BD Loco et CadNet. De plus, une série de recommandations pour l'utilisation des données a été proposée à des fins scientifiques (Hendrickx *et al.*, 2016 ; IGEAT, 2010).

La donnée à disposition est une couche vectorielle constituée de points (128 868 entités spatialisées) pour l'ensemble de la province de Liège. Les champs attributaires peuvent être séparés en catégories de variables. Celles-ci sont détaillées en Annexe 1. L'essentiel de l'information est également détaillé dans les points suivants.

##### **2.3.1.1.1 Les variables d'identification**

Plusieurs champs correspondent à des clefs d'identification unique ; soit des biens immobiliers, soit de la parcelle cadastrale ou encore de l'acte de vente. Pour plus de précision, le champ d'identification des parcelles s'appelle le Capakey. Ce code informe notamment sur le nombre de modifications des entités cadastrées et permet de les localiser sur le plan parcellaire. De plus, les coordonnées cartographiques des points en XY (en Lambert Belge 72) permettent de les localiser précisément (Annexe 1) (IGEAT, 2010).

#### 2.3.1.1.2 Les variables propres aux parcelles cadastrales

D'après R. Brunet (1993), le cadastre est une source d'information publique – à l'échelle fédérale, il est mis en place par l'AGDP (SPF Finances) – qui informe sur l'appropriation et la division des terres. Chaque division cadastrale reprend un ensemble de parcelles cadastrales répertoriées sur un plan cadastral. Un code propre à chaque parcelle cadastrale permet son identification. Les parcelles peuvent être associées à des constructions (IGEAT, 2010). Les éléments du domaine public ne sont pas numérotés sur le plan cadastral. Cela correspond aux rues, aux places et aux chemins, aux rivières (les cours d'eau et les canaux), à certaines surfaces non bâties comme les gares et les chemins de fer, etc. (Annexe 1) (Hendrickx *et al.*, 2016 ; IGEAT, 2010).

#### 2.3.1.1.3 Les variables du code de construction et de l'acte de vente

Les variables relatives au code de construction (CC) sont des variables internes aux biens immobiliers. Ces variables sont à l'origine utilisées pour calculer le revenu cadastral. Or, les modifications notifiées au cadastre ne sont pas exhaustives. Cela est d'autant plus vrai que les biens immobiliers sont âgés. Le cadastre met à jour le revenu cadastral lors de la vente du bien ou lors d'une demande de permis. Ainsi, l'installation d'un chauffage central, la mise en place d'une salle de bains ou encore la création d'une nouvelle chambre ne sont pas toujours comptabilisés dans la base de données (Annexe 1) (IGEAT, 2010). D'autres éléments, comme la date de la dernière modification, la date de construction du bien, la description du bien ou encore le nombre de pièces seront approfondis au point 2.4.1.

#### 2.3.1.2 Les données externes à la BD cadastrale du SPF Finances

La source des données nécessaires à la construction des variables de localisation sont reprises en Annexe 2. La plupart des informations sont disponibles publiquement (SPW, 2008 ; SPF, 2008 ; SPF, 2011 ; SPF, 2013 ; SPF, 2014 ; SPW, 2018 ; SPF, 2018 ; EEA, 2020) et certaines informations font l'objet d'une demande d'acquisition (SPW, 2015 ; SPW, 2020). La récupération des données s'effectue à une date postérieure à Dethier (2012). Les anciennes données ne sont pas toujours disponibles, des alternatives ont dès lors été envisagées. Lorsqu'une donnée est indisponible, la recherche d'une donnée semblable parfois anachronique est suggérée. Une série d'informations a été obtenue par l'intermédiaire des bureaux d'études rattachés à l'Université de Liège, dont le Lepur (Centre de Recherche en Sciences de la Ville, du Territoire et du Milieu rural) et le SEGEFA (Service d'Etude en

Géographie Economique Fondamentale et Appliquée). Tout comme les données cadastrales, elles ont été partagées par Sébastien Hendrickx. Les données de Grignet (2011) ont été récupérées par l'intermédiaire de Perinne Dethier, elle-même.

Il est intéressant – dans le second objectif de perfectionnement du modèle – d'une part, de favoriser les données avec une résolution spatiale la plus fine possible et, d'autre part, de considérer certaines variables supplémentaires non considérées dans le modèle de Dethier (2012).

## **2.4 La définition opérationnelle des variables retenues**

Les traitements réalisés sur les variables ont été effectués par les logiciels ArcGIS et QGIS. Certains post-traitements concernant la mise en forme des données ont été réalisés sur Excel. Pour chaque nouvelle variable générée, un attribut aux biens immobiliers a été ajouté via la fonction QGIS « Joindre les attributs par localisation » ou via la fonction « *Extract value to point* » d' ArcGIS.

De nombreuses variables implémentées dans le modèle sont relatives à la proximité géographique. Selon Torre (2011) la proximité géographique est un concept neutre par essence, car ce sont les individus, selon leur perception, qui attribuent une dimension négative ou positive, ainsi qu'une utilité à cet état de proximité. Ainsi, selon le point de vue des individus, la proximité d'une rivière, d'une forêt, d'un site industriel ou encore d'un paysage bucolique induit tantôt un sentiment positif, tantôt négatif, c'est ce sentiment qui par extension affecte le prix des biens immobiliers. Par exemple, lors de l'achat d'un bien, il est supposé que la proximité d'une forêt soit un élément perçu positivement, au contraire de la proximité d'un site industrielle. Ainsi, deux proximités géographiques peuvent se distinguer : la proximité recherchée et la proximité subie (Torre, 2011). Des distances euclidiennes ont été calculées pour certaines variables. Ces distances continues sont transformées en classe de distances pour simuler une proximité. La proximité est évaluée à des distances seuils de 100, 300, 500, 700 ou 1 500 mètres. De la même manière, Dethier (2012) utilisa des zones tampons entre 100 et 700 mètres. Maldague (2014) et Lorquet (2016) étendirent légèrement les distances seuils de 1 500 jusqu'à 2 000 mètres.

Pour la modélisation des distances, des distances-temps voitures ont été modélisées au départ d'une couche raster obtenue par le SEGEFA. Celle-ci est constituée du réseau routier rastérisé pour toute la Belgique. Ainsi, les étapes de rastérisation du réseau routier et de reclassification des vitesses ne sont pas nécessaires. La couche possède une résolution de 100x100 m, ce qui

paraît raisonnable étant donné notre échelle d'analyse. Néanmoins, la donnée est certainement anachronique, due à l'utilisation d'un réseau routier actuel. Les déplacements piétons ont été définis par des distances euclidiennes en approximation. Les distances à vélo ou en transport en commun n'ont pas été considérées car leurs modélisations nécessiteraient l'utilisation d'une API (Google, OSM ou encore ViaMichelin), or ceci sort du cadre de nos compétences. Un plugin QGIS, ORS Tool, permet tout de même de modéliser ce type de déplacement. Néanmoins, son utilisation est très limitée au regard de l'étendue du set de données employés.

#### 2.4.1 Comparaison des données du SPF Finances et des données Immoweb

La BD cadastrale ne dispose pas d'information sur les équipements de luxes tels que la présence d'une piscine ou d'un terrain de tennis. De même, certaines caractéristiques d'équipements extérieurs comme la présence d'une terrasse ou d'un jardin ne sont pas disponibles. L'information relative à certaines caractéristiques intérieures des biens est également indisponible, comme la présence d'une véranda, d'une buanderie ou encore le type d'équipement de la cuisine. Par conséquent, ces paramètres seront ignorés, même si ceux-ci sont fréquents dans les analyses hédoniques. Finalement, l'état de rénovation du bien n'est pas connu, l'information la plus proche disponible est l'année de modification déclarée du bien (Tableau 2).

**Tableau 2. Comparaison des variables structurelles entre Immoweb et le SPF Finances**

<b>Immoweb</b>	<b>SPF Finances</b>
<b>Variable dépendante y :</b>	
Prix de mise en vente par le vendeur (hors frais notariaux et TVA)	Prix lors de l'acte de vente (hors frais notariaux et TVA)
<b>Variables explicatives intrinsèques :</b>	
Etat de rénovation du bien	Année de modification du bien
Superficie du terrain	Superficie de la parcelle cadastrale
Superficie habitable	Superficie du bâti * Nombre de pièces
Nombre de façades	Configuration du bien (mitoyen, semi-mitoyen ou 4 façades)
Nombre d'étages	Nombre d'étages
Nombre de salles de bains	Nombre de salles de bains
Nombre de salles de douches	
Nombre de chambres	Nombre de logements
Equipement de la cuisine	Indisponible

Présence d'une cave	Construction à partir d'un champ attributaire : - Maison sans cave habitable - Maison avec cave habitable
Présence d'un grenier	Présence d'un étage mansardé
Présence d'une terrasse	Indisponible
Présence d'un jardin	Indisponible
Présence d'une véranda	Indisponible
Présence d'une piscine	Indisponible
Présence d'une buanderie	Indisponible

Quelques variables sont rendues binaires afin de déterminer leurs éventuelles influences relatives dans l'explication du prix. Ces variables concernent le nombre de façades, le nombre d'étages ou encore le nombre de chambres. Quelques éléments peuvent être mis en avant comme le fait qu'une salle de bains est un bien attendu dans toutes les maisons. Ainsi, la présence d'une seule salle de bains ne devrait pas influencer le prix du bien, à l'inverse de son absence ou de sa multiplication. Des champs attributaires sont utilisés pour mettre en évidence la présence d'une cave (habitable ou non) et des sous-typologies de biens (bungalow, villa, maison de cité, etc.).

L'âge du bien, l'âge de rénovation ou la superficie habitable ont été déterminés au départ des variables intrinsèques aux biens immobiliers. Pour l'année de construction, les biens datant d'avant 1930 sont regroupés et sont codifiés différemment (Tableau 3) (IGEAT, 2010).

**Tableau 3. Approximation de l'âge selon les considérations du cadastre (IGEAT, 2010)**

Code du cadastre	Tranche d'âge considérée	Approximation pour l'étude
0001	Avant 1850	1850
0002	1850–1874	1874
0003	1875–1899	1899
0004	1990–1918	1918
0005	1919–1930	1930

Il est choisi de prendre la borne supérieure pour le calcul de l'âge du bien (1930, 1918, 1899, 1874 et 1850). Cette contrainte devra être considérée dans l'interprétation de cette variable, les âges étant approximés avant 1930. L'âge de rénovation correspond à l'âge du bien s'il n'a pas été rénové. Dans le cas d'une modification, le bien est considéré comme rajeuni en prenant en compte l'année de modification et non plus l'année de construction. Un indice binaire de la modification du bien est également ajouté pour les éventuels problèmes de

redondance liés aux variables temporelles. Par ailleurs, la même problématique pour les biens les plus âgés est à relativiser (Équation 11).

$$\text{Age du bien} = \text{Année de référence (2011)} - \text{Année de construction}$$

$$\text{Age de rénovation} = \text{Année de référence (2011)} - \text{Année de modification}$$

**Équation 11. Calcul de l'âge du bien et de l'âge de rénovation**

La superficie habitable n'est pas disponible dans les données du cadastre. Cette information semble essentielle pour expliquer le prix des logements. La variable est approximée en multipliant la surface bâtie par le nombre d'étages (Équation 12).

$$\text{Surface habitable} = \text{superficie bâtie} * \text{nombre d'étages}$$

**Équation 12. Calcul de la surface habitable**

**2.4.2 Les variables d'accessibilité**

Afin de contextualiser rapidement les modes de transports principaux sur Liège, quelques chiffres relatifs à l'enquête BELDAM de 2010 sont à considérer. Les parts modales pour l'ensemble des habitants de la région urbaine de Liège sont de 66 % d'utilisateurs de la voiture, 19 % de marcheurs, 13 % d'usagers des transports publics et 1 % d'utilisateurs du vélo. Ces chiffres au niveau belge sont semblables avec 65 % d'utilisateurs de la voiture, 16 % de marcheurs, 9 % d'usagers des transports publics et 8 % d'utilisateurs du vélo. Dans le transport public, le TEC liégeois a une part non-négligeable d'usagers. En revanche, la part de cyclistes fait grandement défaut en comparaison du niveau fédéral (Tableau 4) (Cornelis *et al.*, 2012).

**Tableau 4. Mode principal de déplacement (tous types de déplacements confondus) (Cornelis *et al.*, 2012)**

	Région urbaine de Liège	Belgique
Voiture	66 %	65 %
Marche	19 %	16 %
Vélo	1 %	8 %
Transport public	13 %	9 %
Divers (taxis, motos, etc.)	1 %	2 %

#### 2.4.2.1 La distance-temps en voiture aux centres névralgiques de Liège

Les distances-temps sont implémentée en mode *raster*, afin de déterminer une valeur en tout point du territoire par propagation en dehors du réseau routier. L'ensemble des opérations est effectuée sur le logiciel ArcGIS. Les traitements poursuivis pour la suite de la procédure sont les suivants (Dethier, 2012 ; Devillet, 2021) :

- 1) Utilisation de la fonction « *cost distance* » au départ de la place Saint-Lambert et de Liège-Guillemins ;
- 2) Utilisation de la fonction « *raster calculator* » pour diviser chaque valeur de pixel par 600 000<sup>5</sup> afin de convertir l'unité de mesure en minute ;
- 3) Utilisation de la fonction « *extract value to point* » pour obtenir la distance de parcours en minute pour l'ensemble des biens immobiliers.

Comme Dethier (2012) le signalait déjà, l'utilisation de la fonction « *cost distance* » d'ArcGIS ne considère pas les sens interdits et parcourt l'ensemble du réseau (même les chemins non carrossables). La différence de l'utilisation de la fonction « *cost distance* » par rapport à la fonction « *euclidean distance* » est que la distance va être pondérée par une surface de coût, en l'occurrence cette surface est le réseau routier pour lequel une vitesse limite est attribuée selon la typologie de route rencontrée. À l'inverse, la distance euclidienne va simplement propager les distances dans toutes les directions sans considération de facteurs de pondération (Esri, S.D.).

Le choix de la place Saint-Lambert comme origine est à raisonner avec la notion d'hypercentre. En effet, le bi-pôle des places Saint-Lambert et Cathédrale constitue l'hypercentre de Liège. Les rues interconnectant les deux places sont particulièrement fréquentées. D'autres rues à forte fréquentation font le lien avec la place Cathédrale. Néanmoins, elles appartiennent à un tissu plus ancien et dense qui ne permet pas le développement de large surface commerciale (Devillelet *et al.*, 2014).

Depuis 2009, la gare de Liège-Guillemins est une gare TGV (TGV Station Liège-Guillemins, 2010). Ce choix de variable fait écho au caractère métropolisant d'une gare TGV. Même si Liège ne peut réellement être considéré comme une métropole la présence d'une gare TGV reste un atout certain. Si cette variable est trop corrélée à l'accessibilité à la place Saint-

---

<sup>5</sup> Conversion : de centiseconde à minute division par 6 000, puis considération de la résolution à 100 m (6 000\*100 = 600 000).

Lambert, une agrégation des deux variables est possible par addition des rasters. D'autres centralités auraient pu être considérées dans l'analyse. Par choix, il a été décidé de se limiter à ces deux lieux.

#### 2.4.2.2 L'accessibilité potentielle aux centres d'emplois

Rodrigue (2020) définit l'accessibilité comme une mesure de la capacité d'un lieu d'origine à être atteint par différents lieux de destinations. Cette logique est implémentable en raster, les pixels représentant l'accessibilité d'un lieu donné. La différence entre l'accessibilité et l'accessibilité potentielle est que la distance potentielle est pondérée par différents attributs. Ces attributs peuvent prendre diverses valeurs comme un nombre de places de stationnement, une superficie commerciale ou encore une fréquence de passages pour un arrêt de bus. L'Équation 13 reprend la formulation de l'accessibilité potentielle d'un lieu.

$$A(P) = \sum_i^n P_i + \sum_j^n P_j/d_{ij}$$

Où  $i \neq j$

Avec,  $A(P)$  = matrice d'accessibilité potentielle ;

$d_{ij}$  = distance entre les lieux  $i$  et  $j$  ;

$P_j$  = attributs de pondération du lieu  $j$  ;

$n$  = nombre de destinations.

#### Équation 13. Accessibilité potentielle

Pour cette variable, l'accessibilité potentielle est pondérée par le nombre d'emplois (Équation 13). Dethier (2012) avait à sa disposition une estimation du nombre d'emplois par secteur statistique du Lepur. A contrario, la donnée utilisée ici est la population active occupée selon le lieu de travail par secteur statistique. Elle provient du dernier recensement de population (Census, 2011). La temporalité des données correspond à notre période de référence, c'est-à-dire 2011. La zone d'emplois considérée coïncide aux secteurs statistiques compris dans le CRU de Liège auquel les centralités adjacentes de Verviers et de Huy ont été ajoutées (soit 2 030 secteurs statistiques). Par étape, Dethier (2012) suivit cette procédure :

- 1) Jointure de la statique population active occupée selon le lieu de travail aux secteurs statistiques spatialisés ;
- 2) Calcul de la superficie des secteurs statistiques en kilomètre carré ;
- 3) Calcul de la densité d'emploi par kilomètre carré ;



- 4) Détermination d'un seuil pour la sélection des densités d'emploi les plus élevées. Ce seuil correspond à la moyenne de la densité d'emploi ( $\bar{x}_{d\text{ emploi}}$ ) augmenté d'un écart type ( $sd_{d\text{ emploi}}$ ) pour la zone de référence ;

$$\begin{aligned} \text{Seuil}_{d\text{ emploi}} &= \bar{x}_{d\text{ emploi}} + sd_{d\text{ emploi}} \\ \text{Seuil}_{d\text{ emploi}} &= 492,26 + 598,54 = 1090,82 \approx 1000 \end{aligned}$$

- 5) Sélection des centroïdes<sup>6</sup> des secteurs statistiques avec une densité d'emploi supérieure ou égale à 1 000. Ainsi, le seuillage permet de mettre en évidence 325 centres d'emplois (destinations) importants pour la zone considérée (CRU de Liège, Huy et Verviers).

Ensuite, il a fallu déterminer les distances-temps voiture entre l'origine des biens et les destinations qui sont les centres d'emplois. A nouveau, un canevas type est poursuivi. Il se déroule selon les étapes suivantes :

- 6) Utilisation de la fonction « *split by attributes* » afin d'individualiser chaque centre d'emploi dans un *shapefile* unique ;
- 7) Utilisation de la fonction « *cost distance* » pour générer une image de coût pour chacun des centres individualisés ;
- 8) Division des surfaces de coût par 600 000 pour la conversion en minute ;
- 9) Création d'un raster constant reprenant le nombre d'emplois du centre d'emploi considéré ;
- 10) Division du raster constant par la distance-temps pour le centre d'emploi considéré ;
- 11) Utilisation de la fonction « *extract multi value to point* » sur chaque image pour l'associer aux biens immobiliers.

Ensuite, les données *shapefiles* ont été modifiées dans Excel. Les valeurs ont été sommées pour calculer l'accessibilité potentielle. Les étapes 7 et 10 ont nécessité une implémentation en python par l'utilisation d'une boucle au vu du grand nombre d'observations. Le script est repris à l'Annexe 3a et correspond à ce que Perinne Dethier avait elle-même réalisé.

---

<sup>6</sup>. Ce qui correspond au centre de gravité des secteurs statistiques concernés.

### 2.4.2.3 Les cartographies d'accessibilité de la CPDT

Les cartographies d'accessibilité de la CPDT déterminent l'accessibilité autre que la voiture selon le lieu de résidence et le lieu de travail. L'accessibilité a été considérée avec une série de paramètres et de facteurs correctifs pour estimer la part modale relative à chaque mode de transport sur l'ensemble du territoire wallon (Tableau 5) (Cornet *et al.*, 2005).

**Tableau 5. Paramètres et facteurs correctifs considérés (Cornet *et al.*, 2005)**

	Lieu de résidence	Lieu de travail
Bus	<ul style="list-style-type: none"> <li>- Fréquence cumulée des bus aux arrêts</li> <li>- Distance aux arrêts (maximum 500 m)</li> <li>- Dénivellation par rapport à l'arrêt (60 m)</li> <li>- Distance aux petites, moyennes et grandes villes</li> </ul>	<ul style="list-style-type: none"> <li>- Fréquence cumulée des bus aux arrêts</li> <li>- Distance aux arrêts (maximum 500 m)</li> <li>- Dénivellation par rapport à l'arrêt (60 m)</li> </ul>
Trains	<ul style="list-style-type: none"> <li>- Fréquence cumulée des trains aux gares</li> <li>- Distance aux gares (maximum 3 000 m)</li> <li>- Dénivellation par rapport à la gare (300 m)</li> <li>- Distance aux grandes villes</li> </ul>	<ul style="list-style-type: none"> <li>- Fréquence cumulée des trains aux gares</li> <li>- Distance aux gares (maximum 700 m)</li> <li>- Dénivellation par rapport à la gare (60 m)</li> <li>- Desserte en bus depuis les principales gares</li> </ul>
Modes Lents	<ul style="list-style-type: none"> <li>- Potentiel de population présente dans un rayon de 1 600 m</li> <li>- Potentiel de population dans un rayon de 8 km</li> <li>- Dénivellation (80 m)</li> </ul>	<ul style="list-style-type: none"> <li>- Potentiel de population présente dans un rayon de 1 600 m</li> <li>- Potentiel de population dans un rayon de 8 km</li> <li>- Dénivellation (80 m)</li> <li>- Distance aux moyennes et grandes villes</li> </ul>

Seules les cartes relatives aux lieux de résidence sont utilisées afin de déterminer l'accessibilité aux différents modes de transports selon la localisation des biens immobiliers. La donnée a également été agrégée dans une carte d'accessibilité synthétique qui rend compte de l'accessibilité générale du lieu de résidence. Les données utilisées pour générer ces cartes d'accessibilités sont relatives au recensement de population de 1991, ce qui peut paraître relativement ancien par rapport à notre année de référence (Cornet *et al.*, 2005).

### 2.4.2.4 L'accessibilité potentielle aux gares et aux arrêts de bus

Des couches *shapefiles* comportant les gares, les arrêts de bus ainsi que leurs fréquentations sont utilisées. Ces informations font suite à des récoltes de données auprès du SRWT par le Lepur (Bianchet *et al.*, 2018). L'accessibilité potentielle (Équation 13) pondérée par la fréquence peut être calculée. La réalisation est identique à la procédure expliquée au point 2.4.2.2.

Pour les gares, le nombre de montées à bord (en 2015) a été choisi comme champ de pondération. Les fermetures et éventuelles ouvertures de gares n'ont pas été contrôlées. Dans les faits, celles-ci devraient concerner majoritairement des gares avec de faibles fréquences de passages qui impacteront peu le calcul de la pondération. Pour les gares, le choix d'une distance modélisée par une surface de coût fait sens pour les ménages qui se déplacent en voiture vers la gare la plus proche pour se rendre sur leur lieu de travail (Équation 13).

A l'inverse, les déplacements vers un arrêt de bus se font majoritairement à pied. Les distances euclidiennes sont choisies. Le calcul de l'accessibilité potentielle pour les arrêts de bus est similaire à ce que Dethier (2012) introduisit dans son modèle. La pondération se base sur la fréquentation à l'arrêt concerné. Ce champ correspond à la somme des validations journalières par arrêt par jour pour l'ensemble du mois de septembre 2017 (Bianchet *et al.*, 2018).

#### 2.4.2.5 Proximité d'une gare ou d'un arrêt de bus

La proximité des gares est également évaluée, cette fois-ci à travers des distances euclidiennes. L'objectif est de mettre en avant les éventuelles nuisances ou au contraire les aménités pouvant être présentes dans les quartiers de gares. Certains auteurs sont parvenus à démontrer certains comportements du prix de l'immobilier. Par exemple, la proximité d'un quartier de gare (inférieur à 400 m) induit une baisse de la valeur des biens, tandis que la valeur augmente entre 900 m et 1 900 m. Cela démontre une non-linéarité de la fonction hédonique selon la distance de proximité considérée pour une gare (Lieske *et al.*, 2021).

À nouveau, le choix d'une distance de proximité propre à une gare est arbitraire. L'Atlas des gares de Wallonie propose une typologie des gares au départ d'une zone de 800 mètres, ce qui correspond à 10 minutes de déplacement à pied (Bianchet *et al.*, 2019). Dans la même logique, Lieske, Van den Nouwelant et Han (2021) proposent des distances par pas de 100 m jusqu'à 2 000 m. Pour la variable produite, quatre seuils de distance ont été générés. Le choix des distances est de 300, 500, 700 et 1 500 mètres pour les gares, tandis qu'elles sont de 100, 300, 500 et 700 mètres pour les arrêts de bus TEC. Les distances euclidiennes sont déterminées par l'outil « *euclidian distance* » au départ de l'ensemble des gares wallonnes (et des arrêts de bus de TEC) puis croisées avec la couche des biens immobiliers (fonction : *extract by location*).

### 2.4.3 Les variables de valorisation sociale

Les caractéristiques socio-économiques des consommateurs ne sont pas toujours implémentées dans les modèles hédoniques en pratique, car cela touche une faible proportion du marché (Terra, 2005). Certaines études, davantage orientées sur l'impact socio-économique, montrent l'effet de la criminalité sur le prix de l'immobilier (Delgado & Wences, 2019). De manière générale, les variables socio-économiques sélectionnées sont surtout des variables en lien avec des statistiques socio-économiques, à l'échelle la plus fine possible, telles que le taux de chômage, la part de population par tranche d'âge ou encore la part de population d'origine étrangère (Travers *et al.*, 2013).

#### 2.4.3.1 Le revenu médian

Dethier (2012) avait à sa disposition les données des revenus médians par secteur statistique entre 2002 et 2008. Elle considéra la moyenne des valeurs pour amoindrir l'effet de la crise économique de 2008 sur son jeu de données. La donnée du revenu médian de 2011 étant aujourd'hui disponible, il apparaît peu justifié de répliquer cette étape.

#### 2.4.3.2 Le taux de chômage

Cette variable est certainement redondante par rapport au revenu médian (2.4.3.1). Elle est tout de même introduite afin de multiplier les variables de valorisation sociale. Dans le cas d'une corrélation forte entre les deux variables, une des deux variables sera soustraite au modèle. Le calcul du taux de chômage est effectué pour l'ensemble des secteurs statistiques pour l'année 2011, selon l'Équation 14.

$$\text{Taux de chômage} = \frac{\text{Nombre de chômeurs}}{\text{Population active totale}} * 100$$

**Équation 14. Taux de chômage**

#### 2.4.3.3 La part des diplômés du supérieur

Une variable considérée par Dethier (2012) est le niveau d'éducation par secteur statistique. Son objectif était de mettre en évidence que la concentration de jeunes diplômés, s'installant dans un quartier, pouvait induire une gentrification des lieux. La relation n'était finalement pas significative. Pour essayer de rendre compte de ce processus, la part de diplômés de

l'enseignement supérieur est évaluée pour l'année 2011 à la place de l'année 2001. La formulation de la part des diplômés est calculée selon l'Équation 15.

$$\begin{aligned} & \textit{Part des diplômés du supérieur} \\ &= \frac{\textit{Nombre de diplômés du supérieur (agé de plus de 18 ans)}}{\textit{Population totale (agé de plus de 18 ans)}} * 100 \end{aligned}$$

**Équation 15. Part des diplômés du supérieur**

#### **2.4.4 Les variables de qualité physique**

Les espaces verts urbains et l'eau sont des aménités environnementales dans les villes. Ils contribuent à créer des îlots de fraîcheur dans les milieux denses. Ils permettent les rencontres. Ils engendrent des espaces naturels et écologiques en milieu urbain (Gibsons *et al.*, 2013). Il faut tout de même spécifier que, dans certains cas, les habitations adjacentes aux parcs peuvent être exposées à des désagréments, surtout lorsque ces parcs se détériorent, ce qui peut amener à une baisse du prix des logements. De plus, la multiplication de ces parcs en milieu urbain peut rendre l'impact sur le prix peu significatif, comme aux Pays-Bas (Halleux, 2005). Les variables liées aux espaces verts sont directement implémentables dans un SIG comme les distances euclidiennes (aux forêts, aux landes, aux vignes, etc.) ou des indices spécifiques (indice de Shannon) (Travers *et al.*, 2013). Plusieurs auteurs montrent l'importance de la proximité à l'eau (Van Dijk *et al.*, 2016 ; Luttik, 2020). Des phénomènes, à plus large échelle, parfois difficilement quantifiables en termes d'externalités, peuvent également être considérés par l'approche hédonique tel que la pollution atmosphérique (Maslianskaïa-Pautrel, 2009), les nuisances sonores dues aux avions (Winke, 2017 ; Rahmatian & Cockerill, 2006) ou encore les aménités liées au paysage (Luttik, 2020).

##### *2.4.4.1 Densité de population*

La densité de population est calculée par secteur statistique pour l'année 2011, selon l'Équation 16.

$$\textit{Densité de population} = \frac{\textit{Population totale par secteur statistique}}{\textit{Superficie du secteur statistique (en km}^2\text{)}}$$

**Équation 16. Densité de population**

Une seconde variable, *a priori* plus précise, est la population par pixel d'un kilomètre carré. Elle est évaluée sur l'ensemble du territoire belge. Un choix sera établi entre les deux indices en fonction d'un éventuel gain de part explicative d'une des deux variables. L'utilisation de la densité de population permet de considérer indirectement la densité bâtie autour de l'environnement du bien.

#### 2.4.4.2 Part du bâti récent dans les constructions

Le calcul de la part de logements bâtis après 1981 permet de rendre compte des lieux dont le développement est relativement récent (de moins de 30 ans). Ainsi, les secteurs statistiques présentant une large part de constructions récentes, présenteront un parc de logements relativement jeune, avec des prix *a priori* plus élevés. Le calcul de cet indicateur se base sur les informations des logements par secteurs statistiques en 2011, selon l'Équation 17.

$$\text{Part de logements bâtis après 1981} = \frac{\text{Nombre de logements construits depuis 1981}}{\text{Nombre de logements construits existants}} * 100$$

**Équation 17. Part de logement bâtis après 1981**

#### 2.4.4.3 Proximité aux fonctions essentielles

L'information utilisée relative aux commerces provient du SEGEFA. Les données disponibles datent de 2019. Cet aspect est relativement contraignant pour cette analyse compte tenu de l'évolution rapide des commerces. Dethier (2012) choisit d'utiliser des espaces tampons (*buffer*) de 100, 500 et 700 mètres autour des biens immobiliers. Ensuite, elle croisa cette information avec une couche contenant uniquement les commerces alimentaires de type supérette < 400 m<sup>2</sup>, les plus grandes surfaces commerciales nécessitant bien souvent l'usage de la voiture.

Moreno (2020) met en évidence le besoin de proximité avec les six fonctions sociales (habiter, travailler, s'approvisionner, se soigner, s'éduquer, s'épanouir) au sein des villes. Il est ainsi à l'origine du concept de la ville du quart d'heure. Cette vision, bien que débattue, peut faire l'objet d'une nouvelle série de variables de proximité. Il donne également une limite à l'influence potentielle des fonctions environnantes avec le quart d'heure. Des vitesses moyennes de piétons sont proposées dans la littérature. Les vitesses varient de 1,00 m/s jusqu'à 1,50 m/s (due à de nombreux facteurs dont l'âge). La vitesse moyenne généralement retenue est de 1,47 m/s (Willis *et al.*, 2004). En considérant cette vitesse, en 15 min de

déplacement, 1 323 m sont parcourus. Étant donné l'imprécision de la méthode, une zone maximum de 1 500 m a été considérée. Dans cette vision, l'évaluation de proximité est étendue à d'autres fonctions qui semblent essentielles telles que les pharmacies, les boulangeries et les épiceries (supérette < 400 m<sup>2</sup>). Le choix de ces fonctions est purement personnel. L'information est tout de même rendue binaire pour exclure les éléments ayant *in fine* une part significative trop faible. Ainsi, la variable est déterminée pour des distances seuils de 300, 500, 700 et 1 500 mètres.

#### 2.4.4.4 *Accessibilité potentielle aux nodules commerciaux*

Le calcul de l'accessibilité potentielle est pondéré par la superficie commerciale des nodules (Équation 13). À nouveau, les données proviennent du SEGEFA et sont anachroniques par rapport à 2011. Une première étape à la construction de cette variable a été l'agrégation par nodule des superficies commerciales contenues dans chaque point de vente. Ensuite, une jointure a été réalisée par le champ d'identification des nodules commerciaux. Le reste de la procédure est identique au point (2.4.2.2), hormis l'utilisation des surfaces commerciales pour la pondération de l'accessibilité.

#### 2.4.4.5 *Influence des parcs urbains*

Pour les parcs urbains, la relation entre la taille et le rayon d'attraction est souvent mis en avant tel qu'est considéré : un rayon de 1000 m pour un parc > 30 ha, de 500 m pour un parc entre 10 à 30 ha, et de 250 m pour un parc < 10 ha (Halleux, 2005). Néanmoins, cette information n'est pas disponible pour toute l'échelle d'analyse. Seuls les parcs urbains de Liège sont identifiés sous forme de points. La taille des parcs n'est pas connue. Elle est approximée à l'aide des parcelles cadastrales. Les parcs sont ensuite classés selon leurs tailles.

#### 2.4.4.6 *Proximité aux trames verte et bleue*

En ce qui concerne la proximité des forêts, la méthode reste similaire à ce que Dethier (2012) a effectué. La couche d'occupation du sol de 2012 est utilisée pour extraire les forêts (mixtes, feuillues ou résineuses) à une résolution de 100 mètres (EEA, 2020) (fonction : « *Extract by attributes* »). Ensuite, des seuils de distances euclidiennes de 100, 300, 500 et 700 mètres sont définis pour classer la proximité relative des espaces verts par rapport aux habitations (fonction : « *Extract value to point* »). De plus, l'eau peut avoir un impact significatif sur le prix, par la proximité des grandes rivières, la vue sur les plans d'eau ainsi que de la distance aux sites de baignades (Van Dijk *et al.*, 2016 ; Luttik, 2000). Afin de rendre compte de

l'aspect qualitatif des cours d'eau, il apparaît opportun de considérer une nouvelle variable liée à l'eau. Les surfaces d'eau sont extraites de l'occupation du sol, puis à nouveau une distance euclidienne relative aux habitations y est associée.

#### 2.4.4.7 *Nuisances visuelles liées aux sites d'extractions, aux industries et aux SAR*

Plusieurs variables ayant un effet négatif sur les prix sont proposées par Dethier (2012). La procédure est semblable au point 2.4.4.6. A partir de la couche d'occupation du sol sont extraits, d'une part, les sites d'extractions miniers et, d'autres part, les sites industrielles. Ensuite, la catégorisation des distances euclidiennes par rapport à l'habitat concerne des distances euclidiennes de 100, 300, 500 et 700 mètres. Ceci a pour objectif de surtout mettre en avant l'effet visuel, souvent négatif, de ce type d'activité par rapport à l'habitat. D'autres aspects macros, comme la vue sur les affleurements rocheux pour les sites d'exploitations miniers, seront considérés par la variable Kiemstedt-Schmitz (au point 2.4.4.10).

Dethier (2012) prend également en considération les sites d'activités économiques désaffectés (SAED), ce qui permet de rendre compte de l'impact des friches industrielles sur l'environnement bâti. Aujourd'hui, les SAED sont repris dans les sites à réaménager (SAR) avec les sites d'activités économiques à réhabiliter (SAER) (SPW, 2022). La procédure suivie est identique à celle utilisée pour déterminer les nuisances visuelles des sites d'extractions et des sites industriels.

#### 2.4.4.8 *Nuisances sonores liées aux transports et aux industries*

Dethier (2012) considère un *buffer* de 100 mètres autour des axes routiers principaux pour rendre compte du bruit. Cette variable n'est pas retenue car plusieurs couches relatives aux bruits sont utilisées en complémentarité. Premièrement, la couche de bruit de 2006 relative aux réseaux routiers principaux et à la voie ferrée est utilisée pour l'ensemble de la zone d'étude sauf Liège. En effet, pour la commune de Liège, une couche de bruit de 2012 est disponible. Celle-ci informe également sur le niveau de bruit de certaines grandes industries. Cette dernière catégorie n'a pas été retenue car elle ne concernait que très peu d'habitations. Enfin, une donnée manquante est l'impact de l'aéroport. La donnée a été digitalisée sur la base de la représentation cartographique disponible en ligne<sup>7</sup>. Une grille reprenant les valeurs de décibels pour chaque zone permet d'attribuer un niveau de bruit pour chaque zone propre à

---

<sup>7</sup> La carte de départ est disponible ici : <https://eblg.aerovision.cloud/appmap>



l'aéroport de Liège (ACNAW, S.D.). L'indication du bruit pour l'aéroport est regroupée en quatre catégories de décibels, tandis que le SPW a défini six catégories de bruit pour les industries, voies ferrées et axes routiers principaux. Ainsi, l'information est harmonisée en quatre catégories (Tableau 6).

**Tableau 6. Classification du bruit se basant sur l'information du bruit de l'aéroport de Liège**

Classes du SPW	Nouvelles classes
< 55 db	Non considérée
De 55 à 59 db	De 55 à 59 db
De 60 à 64 db	De 60 à 64 db
De 65 à 69 db	De 65 à 69 db
De 70 à 74 db	> 70 db
> 75 db	

Finalement, les couches sont individualisées par type d'impact sonore à l'échelle du CRU de Liège.

#### 2.4.4.9 Indice de diversité de Shannon

L'hétérogénéité du paysage est mise en avant par l'Équation 18 qui porte sur l'indice de diversité de Shannon-Weaver (1949). Plus l'indice est élevé, plus la diversité du paysage est grande. En général, celui-ci varie de 1,5 à 3,5. Il dépasse rarement les 4,5. Il peut atteindre zéro dans le cas où une seule occupation du sol est rencontrée (Ortiz-Burgos, 2016).

$$H' = - \sum_{i=1}^n p_i \ln(p_i)$$

Avec,

$H'$  : l'indice de diversité ;

$p_i$  : proportion de l'élément  $i$  dans le paysage ;

$n$  : nombre de classes d'occupation du sol dans le paysage considéré.

#### Équation 18. Indice de diversité de Shannon-Weaver (1949)

Dethier (2012) proposait déjà l'indice en travaillant par *buffer* de 100 et 500 mètres autour de chaque bien immobilier. D'après Wu (2004), il est préférable de multiplier les échelles d'analyse du paysage car l'échelle spatiale affecte les indices du paysage. Selon l'analyse de Cavailhès *et al.* (2009), l'influence du paysage sur le prix est significative pour les vues directes, c'est-à-dire de 100 à 300 m. Au-delà, l'influence sur le prix semble être insignifiante.

Pour vérifier cette relation la variable est évaluée à 300 m et 500 m. La couche d'occupation a une résolution à 100 m ainsi, une zone tampon de cette envergure n'a pas été envisagée (IGN, 2014). Pour la réalisation du calcul, la procédure est décrite comme suit :

- 1) Individualisation des biens immobiliers via l'utilisation de la fonction « *split by attributes* » ;
- 2) Création d'espaces tampons de 300 et 500 m autour de chaque bien (fonction : « *buffer* ») ;
- 3) Calcul de l'aire de chaque classe d'occupation du sol par rapport aux zones tampons définies préalablement (fonction : « *tabulate area* ») ;
- 4) Fusion des fichiers individualisés (fonction : « *merge* ») pour permettre la jointure avec la couche des biens immobiliers. Avant la jointure, l'indice de Shannon est calculé dans Excel.

Etant donné le grand nombre d'opérations à répéter une boucle implémentée en python est utilisée pour les étapes 2 à 4 (Annexe 3b).

#### 2.4.4.10 Indice d'estimation du paysage *I* de Schmitz-Kiemstedt

Une méthode d'estimation du paysage, est proposée par Schmitz (2020). Cette méthode est inspirée d'une formule empirique déterminée par Kiemstedt (1967). Elle a l'avantage d'être applicable à l'ensemble de notre zone d'étude ; elle peut aisément être implantée dans un logiciel SIG, ce qui est particulièrement adapté pour la réalisation d'un modèle hédonique. La formulation de Schmitz (2020) permet de déterminer une valeur paysagère locale *I*, selon l'Équation 19.

$$I = E + DC = 12,5 \log_2(Alt_{max} - Alt_{min}) + \left( \sum_i^n A_i \right) \sqrt{N}$$

Avec  $i = 1, \dots, n$ .

Où, E : Energie du relief ;

D : Diversité ;

C : Complexité ;

A : Poids attribué par type d'occupation du sol ;

N : Nombre d'occupations du sol différent par maille.

**Équation 19. Indice d'estimation du paysage de Schmitz (basé sur Kiemstedt)**

La procédure suivie pour l'application de cette formule est établie par Bruckmann et Pepe (2021), l'ensemble des traitements est réalisé sur QGis. Une grille avec une maille carrée de

500 mètres de côté a été définie sur l'ensemble de l'aire d'étude. Une telle résolution spatiale semble être un bon compromis pour observer des variations paysagères à l'échelle de l'aire urbaine, tout en minimisant le temps de calcul. Les deux couches utilisées pour le calcul de l'indice sont une image du MNT de 2013-2014 (SPW,2015) et une couche vectorielle de l'occupation du sol en 2012 (IGN, 2014).

L'énergie du relief ( $E$ ) reflète la topographie du paysage. Plus la différence d'altitude est importante à l'intérieur de la maille, plus l'énergie du relief est grande. Pour calculer  $E$ , la fonction « Statistique de zone » est utilisée afin d'obtenir la différence d'altitude maximum et minimum pour chaque cellule de la grille sur la base du MNT. Puis, sur chaque pixel, via la « Calculatrice de champs », le logarithme en base 2 est appliqué. Le résultat est enfin multiplié par 12,5 (Bruckmann et Pepe, 2021).

La diversité ( $D$ ) est la somme des poids attribués par type occupation du sol ( $A$ ). Pour chaque attribut d'occupation du sol, une pondération arbitraire selon le type de sol est définie. Le choix de la pondération est repris en Annexe 4. Ensuite, la somme des poids pour chaque maille est calculée via la fonction « Joindre un attribut par localisation » (Bruckmann et Pepe, 2021).

La complexité ( $C$ ) est la diversité d'occupation du sol au sein de chaque maille. Pour la calculer, un comptage du type d'occupation du sol est effectué sur l'ensemble des mailles, la fonction utilisée est à nouveau « Joindre un attribut par localisation ». Ensuite, la racine carrée est appliquée par l'intermédiaire de la « Calculatrice de champs ». Finalement, toujours via la même calculatrice, l'indice  $I$  est calculé (Bruckmann et Pepe, 2021).

## **2.5 Résumé des variables et des écarts aux variables de Dethier**

Des écarts aux données utilisées par Dethier (2012) sont observés pour plusieurs raisons. La première est le manque de données contemporaines à la période de référence. Deuxièmement, les choix pris dans la réalisation de certains traitements sont adaptés aux données en possession.

Plusieurs variables supplémentaires auraient pu être imaginées, car elles auraient pu avoir un impact sur le prix. Par exemple, la notion de risque est peu explorée, la distance à une centrale nucléaire, à un site Seveso, à un sol pollué ou encore à une zone d'aléa sont autant de variables pouvant induire une diminution du prix. L'accessibilité en mode lent, excepté par la carte d'accessibilité du Lepur (Cornet *et al*, 2005) n'est que peu représentée dans le modèle. Finalement, certaines variables de proximité comme la distance aux hôpitaux, aux écoles ou

encore aux équipements sportifs n'ont pas été générées. L'ensemble des variables produites est rassemblée dans le Tableau 7. Une BD ACCESS assure la correspondance des valeurs à chaque bien immobilier. Le même champ d'identification, propre aux habitations, est utilisé pour effectuer la jointure entre les valeurs et les entités spatialisées.

**Tableau 7. Résumé des variables retenues et les écarts aux variables de Dethier (2012)**

Nom de la variable	Description de la variable	Type de données	Ecart à Dethier
PRIX	Prix lors de l'acte de vente (hors frais notariaux et TVA).	Quantitatif continu	Semblable
SURF_FONC	Superficie de la parcelle en mètre carré.	Quantitatif continu	Semblable
SURF_HAB	Superficie habitable en mètre carré.	Quantitatif continu	Semblable
NB_FACADE	Configuration des maisons (2, 3 ou 4 façades).	Quantitatif discret	Semblable
N2_FACADE	Maisons avec une configuration mitoyenne.	Binaire	Inédit
N3_FACADE	Maisons avec une configuration semi-mitoyenne.	Binaire	Semblable
N4_FACADE	Maisons avec une configuration quatre façades.	Binaire	Semblable
NB_ETAGE	Nombre d'étages (1, 2, 3 ou 4).	Quantitatif discret	Semblable
N0_ETAGE	Maisons avec un rez-de-chaussée uniquement.	Binaire	Semblable
N1_ETAGE	Maisons avec un seul étage.	Binaire	Inédit
N2_ETAGE	Maisons avec deux étages ou plus de deux étages.	Binaire	Semblable
MANSARDE	Présence d'un étage mansardé.	Binaire	Inédit
AGE_BIEN	Age du bien immobilier (estimation après 81 ans).	Quantitatif continu	Semblable
AGE_RENO	Age de rénovation du bien immobilier.	Quantitatif continu	Inédit
RENOVATION	Indique si le bien immobilier a été rénové.	Binaire	Variante
NB_GARAGE	Indique le nombre de garages du bien immobilier.	Quantitatif discret	Inédit
N0_GARAGE	Indique les biens immobiliers sans garage.	Binaire	Semblable
N1_GARAGE	Indique les biens immobiliers avec un garage.	Binaire	Semblable
N2_GARAGE	Indique les biens immobiliers avec deux ou plus de deux garages.	Binaire	Semblable
NB_SDB	Indique le nombre de salles de bains du bien immobilier.	Quantitatif discret	Semblable

N0_SDB	Indique les biens immobiliers sans salle de bains.	Binaire	Semblable
N1_SDB	Indique les biens immobiliers avec une salle de bains.	Binaire	Inédit
N2_SDB	Indique les biens immobiliers avec deux ou plus de deux salles de bains.	Binaire	Semblable
CHAUF_CENT	Indique la présence ou l'absence d'un chauffage central et/ou d'un climatiseur dans le bien immobilier.	Binaire	Inédit
NB_LOGEMENT	Indique le nombre de logements potentiels au sein du bien immobilier.	Quantitatif discret	Inédit
N1_LOGEMENT	Bien immobilier contenant un logement potentiel.	Binaire	Variante
N2_LOGEMENT	Bien immobilier contenant deux logements potentiels.	Binaire	Variante
N3_LOGEMENT	Bien immobilier contenant trois logements potentiels.	Binaire	Variante
N4_LOGEMENT	Bien immobilier contenant quatre ou plus de quatre logements potentiels.	Binaire	Variante
NB_PIECE	Nombre de pièces standards du bien immobilier (au sens du cadastre).	Quantitatif discret	Inédit
CAVE	Indique les biens immobiliers possédant une cave.	Binaire	Semblable
CAVE_VIE	Indique les biens immobiliers possédant une cave habitable.	Binaire	Inédit
CITE	Indique si la maison est de type : maison de cité.	Binaire	Inédit
VILLA	Indique si la maison est de type : villa.	Binaire	Inédit
BUNGALOW	Indique si la maison est de type : bungalow.	Binaire	Semblable
BEL_ETAGE	Indique si la maison est de type : bel-étage.	Binaire	Inédit
ACC_POT_EMPLOI	Accessibilité potentielle (en voiture) pondérée par le nombre d'emplois à partir des centres d'emplois principaux (secteur statistique avec au moins 1 000 emplois/km <sup>2</sup> ).	Quantitatif continu	Semblable
ISO_GUILLEMINS	Distance-temps en minute du bien immobilier à la gare des Guillemins.	Quantitatif continu	Semblable

ISO_STLAMBERT	Distance-temps en minute du bien immobilier au centre de Liège (Place Saint-Lambert).	Quantitatif continu	Semblable
ISO_NEV	Addition des distance-temps en minute du bien immobilier à la gare des Guillemins et à la Place Saint-Lambert.	Quantitatif continu	Inédit
ACC_BUS	Part modale du bus en pourcentage au lieu de résidence.	Quantitatif continu	Semblable
ACC_ML	Part modale des modes lents en pourcentage au lieu de résidence.	Quantitatif continu	Semblable
ACC_TRAIN	Part modale du train en pourcentage au lieu de résidence.	Quantitatif continu	Semblable
ACC_TOT	Addition des parts modales de l'ensemble des transports autres que la voiture en pourcentage.	Quantitatif continu	Inédit
ACC_POT_GARE	Accessibilité potentielle (en voiture) pondérée par nombre de montées à partir d'une gare de train en 2018.	Quantitatif continu	Inédit
ACC_POT_ARRET	Accessibilité potentielle (à pied) pondérée par nombre de montées à partir d'un arrêt de bus en 2015.	Quantitatif continu	Semblable
D300_GARE	Situation du bien à moins de 300 m d'une gare de train.	Binaire	Inédit
D500_GARE	Situation du bien entre 300 et 500 m d'une gare de train.	Binaire	Inédit
D700_GARE	Situation du bien entre 500 et 700 m d'une gare de train.	Binaire	Inédit
D1500_GARE	Situation du bien entre 700 et 1500 m d'une gare de train.	Binaire	Inédit
D100_BUS	Situation du bien à moins de 100 m d'un arrêt de bus.	Binaire	Inédit
D300_BUS	Situation du bien entre 100 et 300 m d'un arrêt de bus.	Binaire	Inédit
D500_BUS	Situation du bien entre 300 et 500 m d'un arrêt de bus.	Binaire	Inédit
D700_BUS	Situation du bien entre 500 et 700 m d'un arrêt de bus.	Binaire	Inédit
RMD_2011	Revenu médian du secteur statistique en 2011.	Quantitatif continu	Variante

TCH_2011	Taux de chômage du secteur statistique en 2011.	Quantitatif continu	Inédit
NINS_2011	Part des diplômés du supérieur du secteur statistique en 2011.	Quantitatif continu	Variante
DPOPSS_2011	Densité de population du secteur statistique en 2011 en habitant par kilomètre carré.	Quantitatif continu	Variante
DPOPPIX_2011	Nombre d'habitants par pixel d'un kilomètre carré en 2011.	Quantitatif continu	Variante
CAF81_2011	Part des logements construits après 1981 dans le secteur statistique en 2011.	Quantitatif continu	Variante
ACC_POT_COM	Accessibilité potentielle aux nodules commerciaux pondérée par la superficie commerciale en 2019.	Quantitatif continu	Semblable
D300_ALIM	Situation du bien à moins de 300 m d'un commerce alimentaire.	Binaire	Semblable
D500_ALIM	Situation du bien entre 300 à 500 m d'un commerce alimentaire.	Binaire	Semblable
D700_ALIM	Situation du bien entre 500 à 700 m d'un commerce alimentaire.	Binaire	Semblable
D1500_ALIM	Situation du bien entre 700 à 1 500 m d'un commerce alimentaire.	Binaire	Inédit
D300_BOUL	Situation du bien à moins de 300 m d'une boulangerie.	Binaire	Semblable
D500_BOUL	Situation du bien entre 300 à 500 m d'une boulangerie.	Binaire	Semblable
D700_BOUL	Situation du bien entre 500 à 700 m d'une boulangerie.	Binaire	Semblable
D1500_BOUL	Situation du bien entre 700 à 1 500 m d'une boulangerie.	Binaire	Inédit
D300_PHARM	Situation du bien à moins de 300 m d'une pharmacie.	Binaire	Semblable
D500_PHARM	Situation du bien entre 300 à 500 m d'une pharmacie.	Binaire	Semblable
D700_PHARM	Situation du bien entre 500 à 700 m d'une pharmacie.	Binaire	Semblable
D1500_PHARM	Situation du bien entre 700 à 1 500 m d'une pharmacie.	Binaire	Inédit

D_PT_PARC	Situation du bien à moins de 250 m d'un petit parc (<10 ha) (uniquement pour Liège).	Binaire	Inédit
D_MO_PARC	Situation du bien à moins de 500 m d'un parc de taille moyenne (10-30 ha) (uniquement pour Liège).	Binaire	Inédit
D_GD_PARC	Situation du bien à moins de 1000 m d'un grand parc (> 30 ha) (uniquement pour Liège).	Binaire	Inédit
D100_BLEU	Situation du bien à moins de 100 m d'un plan d'eau.	Binaire	Inédit
D300_BLEU	Situation du bien entre 100 et 300 m d'un plan d'eau.	Binaire	Inédit
D500_BLEU	Situation du bien entre 300 et 500 m d'un plan d'eau.	Binaire	Inédit
D700_BLEU	Situation du bien entre 500 et 700 m d'un plan d'eau.	Binaire	Inédit
D100_VERT	Situation du bien à moins de 100 m d'un espace vert.	Binaire	Semblable
D300_VERT	Situation du bien entre 100 à 300 m d'un espace vert.	Binaire	Inédit
D500_VERT	Situation du bien entre 300 à 500 m d'un espace vert.	Binaire	Semblable
D700_VERT	Situation du bien entre 500 à 700 m d'un espace vert.	Binaire	Semblable
D100_EXTRACT	Situation du bien à moins de 100 m d'un site d'extraction.	Binaire	Semblable
D300_EXTRACT	Situation du bien entre 100 à 300 m d'un site d'extraction.	Binaire	Semblable
D500_EXTRACT	Situation du bien entre 300 à 500 m d'un site d'extraction.	Binaire	Semblable
D700_EXTRACT	Situation du bien entre 500 à 700 m d'un site d'extraction.	Binaire	Inédit
D100_INDUST	Situation du bien à moins de 100 m d'un site industriel.	Binaire	Semblable
D300_INDUST	Situation du bien entre 100 à 300 m d'un site industriel.	Binaire	Semblable
D500_INDUST	Situation du bien entre 300 à 500 m d'un site industriel.	Binaire	Semblable



D700_INDUST	Situation du bien entre 500 à 700 m d'un site industriel.	Binaire	Inédit
D100_SAR	Situation du bien à moins de 100 m d'un SAR.	Binaire	Semblable
D300_SAR	Situation du bien entre 100 à 300 m d'un SAR.	Binaire	Semblable
D500_SAR	Situation du bien entre 300 à 500 m d'un SAR.	Binaire	Semblable
D700_SAR	Situation du bien entre 500 à 700 m d'un SAR.	Binaire	Inédit
BRUIT1_ROAD	Biens faiblement impactés par le bruit des routes principales (55-59 dB).	Binaire	Inédit
BRUIT2_ROAD	Biens moyennement impactés par le bruit des routes principales (60-64 dB).	Binaire	Inédit
BRUIT3_ROAD	Biens fortement impactés par le bruit des routes principales (65-69 dB).	Binaire	Inédit
BRUIT4_ROAD	Biens très fortement impactés par le bruit des routes principales (>70 dB).	Binaire	Inédit
BRUIT1_RAIL	Biens faiblement impactés par le bruit de la voie ferrée (55-59 dB).	Binaire	Inédit
BRUIT2_RAIL	Biens moyennement impactés par le bruit de la voie ferrée (60-64 dB).	Binaire	Inédit
BRUIT3_RAIL	Biens fortement impactés par le bruit de la voie ferrée (65-69 dB).	Binaire	Inédit
BRUIT4_RAIL	Biens très fortement impactés par le bruit de la voie ferrée (>70 dB).	Binaire	Inédit
BRUIT1_AVION	Biens faiblement impactés par le bruit des avions (55-59 dB) (uniquement à proximité de Liège Airport).	Binaire	Inédit
BRUIT2_AVION	Biens moyennement impactés par le bruit des avions (60-64 dB) (uniquement à proximité de Liège Airport).	Binaire	Inédit
BRUIT3_AVION	Biens fortement impactés par le bruit des avions (65-69 dB) (uniquement à proximité de Liège Airport).	Binaire	Inédit
BRUIT4_AVION	Biens très fortement impactés par le bruit des avions (>70 dB) (uniquement à proximité de Liège Airport).	Binaire	Inédit

300_SHANNON	Indice de Shannon évalué dans les 300 m autour du bien.	Quantitatif continu	Inédit
500_SHANNON	Indice de Shannon évalué dans les 500 m autour du bien.	Quantitatif continu	Semblable
KIEMSTEDT	Indice de valeur du paysage par Schmitz sur la base de Kiemstedt.	Quantitatif continu	Semblable

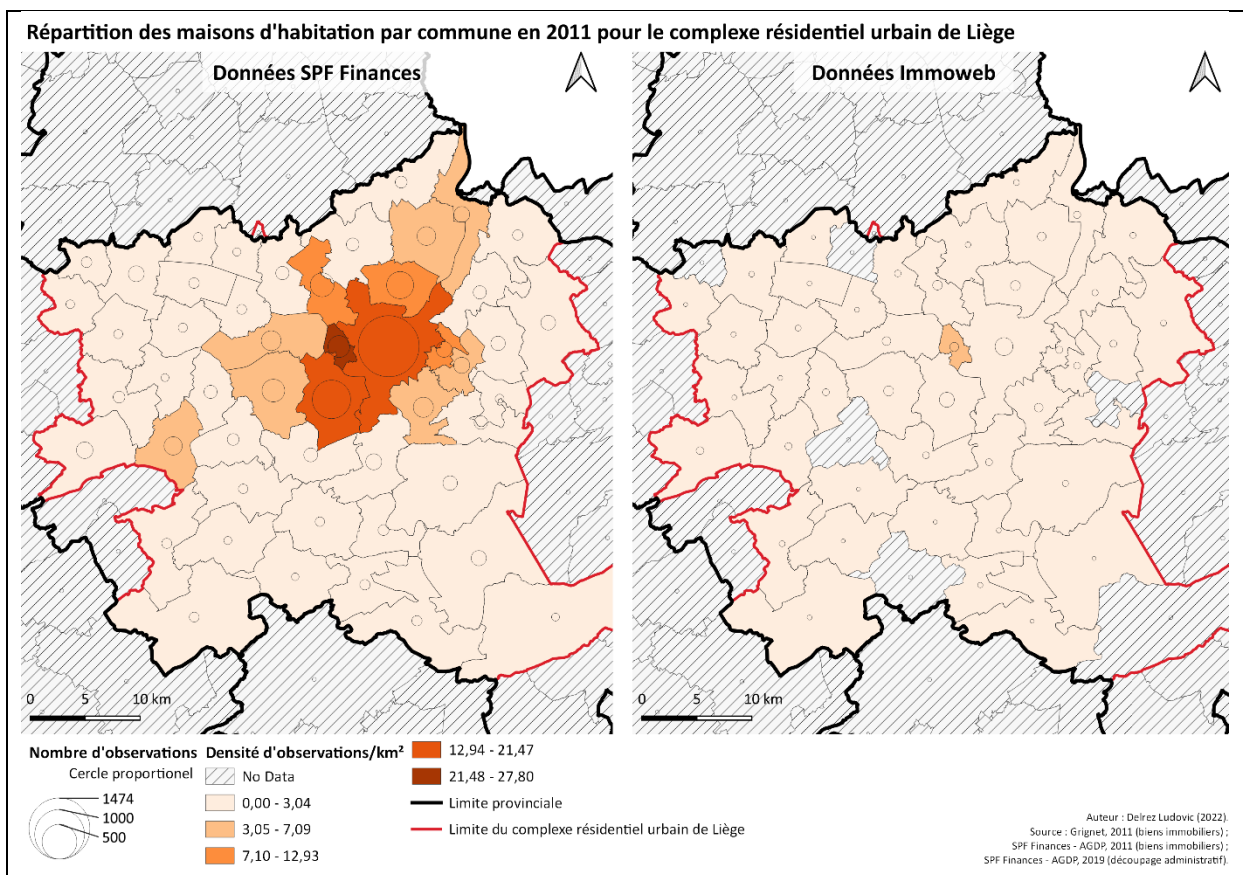
Les variables internes au set de données Immoweb ont été adaptées pour correspondre dans la mesure du possible aux variables internes du set de données du SPF Finances. Ceci est fait pour permettre une comparaison des modèles relatifs aux deux sets de données au terme de cette recherche. Les champs supplémentaires de la BD Immoweb, non disponibles dans la BD cadastrales, ne seront pas exploités (Tableau 2). De plus, l'ensemble des variables extrinsèques sont recalculées pour limiter les différences de variables d'un jeu de données à l'autre.

## Chapitre 3 Analyse par l'approche hédonique

### 3.1 Description et analyse de la base de données

#### 3.1.1 Répartition spatiale des biens immobiliers

La Carte 2 permet de comparer les données du SPF Finances et les données Immoweb. Pour les deux sets de données, aucune observation n'est disponible pour les communes d'Herstappe et de Voeren (les Fourons) situées dans la province flamande du Limbourg. Néanmoins, cet élément est peu important car Dethier réalisa également son modèle sans considérer ces communes. Les communes d'Engis, de Berloz, d'Olne, de Crisnée, de Stoumont et d'Ouffet ne présentent aucune donnée via Immoweb. Pour l'ensemble du territoire d'analyse, le nombre d'observations est bien supérieur à celui de Grignet (2011), de l'ordre de dix fois plus. En termes de répartition spatiale des biens, les densités les plus fortes sont centrales avec Liège et son agglomération morphologique (Saint-Nicolas présente la densité la plus importante). La commune d'Amay en périphérie de Huy présente également une densité plus importante.



Carte 2. Répartition des maisons d'habitation par commune en 2011 pour les deux jeux de données

De plus, la distribution des points est analysée par une analyse de voisinage. Le calcul du ratio du plus proche voisin ( $R(k)$ ) indique une concentration significative des points pour les deux jeux de données. L'hypothèse nulle ( $H_0$ ), qui indiquerait une répartition aléatoire des points, est rejetée car la p-valeur est très petite. La concentration de points est logiquement plus dense pour les données cadastrales étant donné le nombre d'enregistrements supérieur (Tableau 8).

**Tableau 8. Analyse de voisinage par le ratio du plus proche voisin**

	<b>Données SPF Finances</b>	<b>Données Immoweb</b>
Distance moyenne observée entre les points $r(k)$	133,08 m	539,11 m
Distance moyenne attendue entre les points $\varphi(k)$	335,36 m	936,84 m
<i>Nearest Neighbor Ratio</i> $R(k) = r(k)/\varphi(k)$	0,3968	0,5755
z-score	-85,5997	-18,3237
p-value	0,0000	0,0000
Résultat	$R(k) < 1$ (distribution concentrée)	$R(k) < 1$ (distribution concentrée)

### 3.1.2 L'application des statistiques descriptives

Pour chaque variable, des indicateurs statistiques ont été calculés (Annexe 6a et 6b). Cette étape permet notamment de mettre en évidence les valeurs extrêmes, les valeurs aberrantes, les erreurs de saisie ou encore les omissions. L'analyse statistique concerne les statistiques descriptives suivantes :

- 1) Indicateur de tendance centrale : moyenne et médiane ;
- 2) Indicateur de dispersion : l'écart-type, minimum, maximum et l'étendue ;
- 3) Indicateur de forme : coefficient d'asymétrie et d'aplatissement.

Les indices de formes sont calculés sur  $R$ . Lorsque les indices s'éloignent de zéro, alors la distribution est de moins en moins considérée comme normale. Pour le coefficient d'asymétrie, un coefficient négatif indique une asymétrie vers la droite, tandis qu'un coefficient positif indique une asymétrie vers la gauche (Équation 20a). Pour le coefficient d'aplatissement, un coefficient positif indique une distribution leptokurtique (signifie que la courbe s'élève fortement avant de retomber drastiquement), tandis qu'un coefficient négatif

présume une distribution platikurtique (une distribution aplatie par rapport à une forme gaussienne) (Équation 20b) (Farnir *et al.*, 2016).

$$I_{\text{Asymétrie}} = \frac{n}{(n-1)(n-2)} \sum \left( \frac{x - \bar{x}}{s} \right)^3 \quad (\text{a})$$

$$I_{\text{Aplatissement}} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left( \frac{x - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (\text{b})$$

Avec,

n : nombre d'observations ;

$\bar{x}$  : moyenne des valeurs de x ;

s : écart-type de x.

### Équation 20. Coefficient d'asymétrie et d'aplatissement

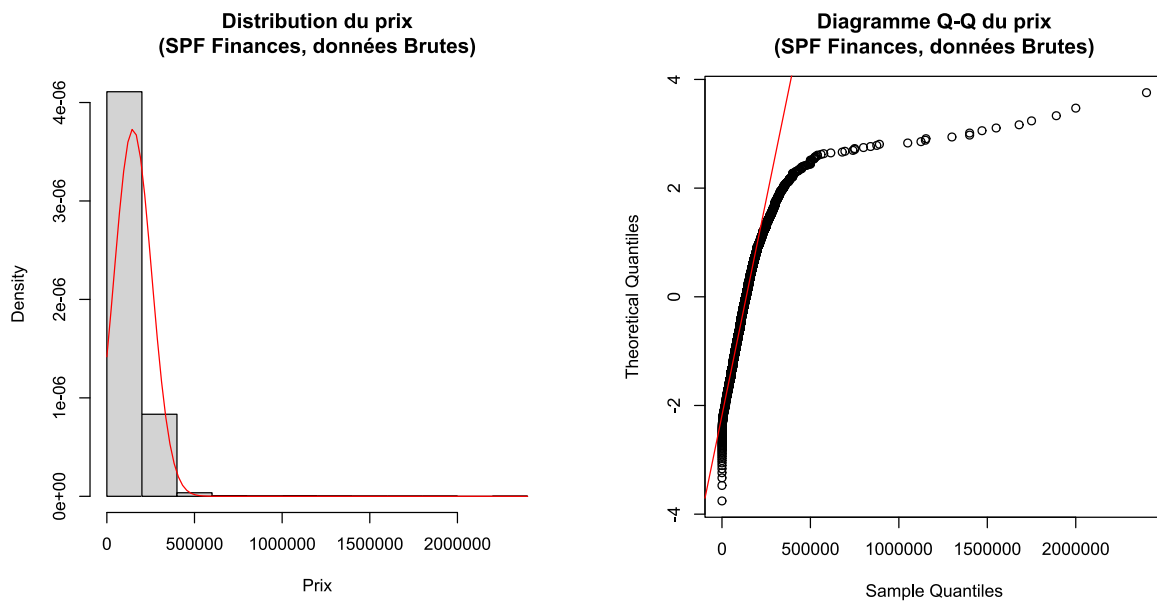
#### 3.1.2.1 Distribution et analyse descriptive du prix

La qualité du modèle de régression dépend fortement de la qualité de ces variables. Ainsi, certains prétraitements sont requis afin d'extraire les valeurs des biens immobiliers les plus aberrantes. Par exemple, les extrema peuvent être retirés de la gamme de prix de vente. Ceci permet d'exclure les ventes à 1 € et de considérer un marché unique (en excluant le bas de gamme et le haut de gamme). Dans la même logique, les maisons très anciennes peuvent être exclues car elles appartiennent à un marché à part entière. Finalement, les erreurs de saisie ne sont pas rares, celles-ci sont à localiser, à modifier ou, le cas échéant, à supprimer (Des Rosiers, 2001).

Pour l'analyse du prix, des représentations graphiques (histogramme et diagramme quantile-quantile) ont été réalisées (Annexe 5). L'utilisation d'une régression linéaire multiple nécessite le respect des hypothèses relatives à une distribution normale. Ainsi, la courbe normale doit être symétrique et la moyenne en est le centre. La moyenne est confondue avec la médiane et le mode. Deux paramètres décrivent la distribution : la moyenne ( $\mu$ ) et la variance ( $s^2$ ) (Des Rosiers, 2001).

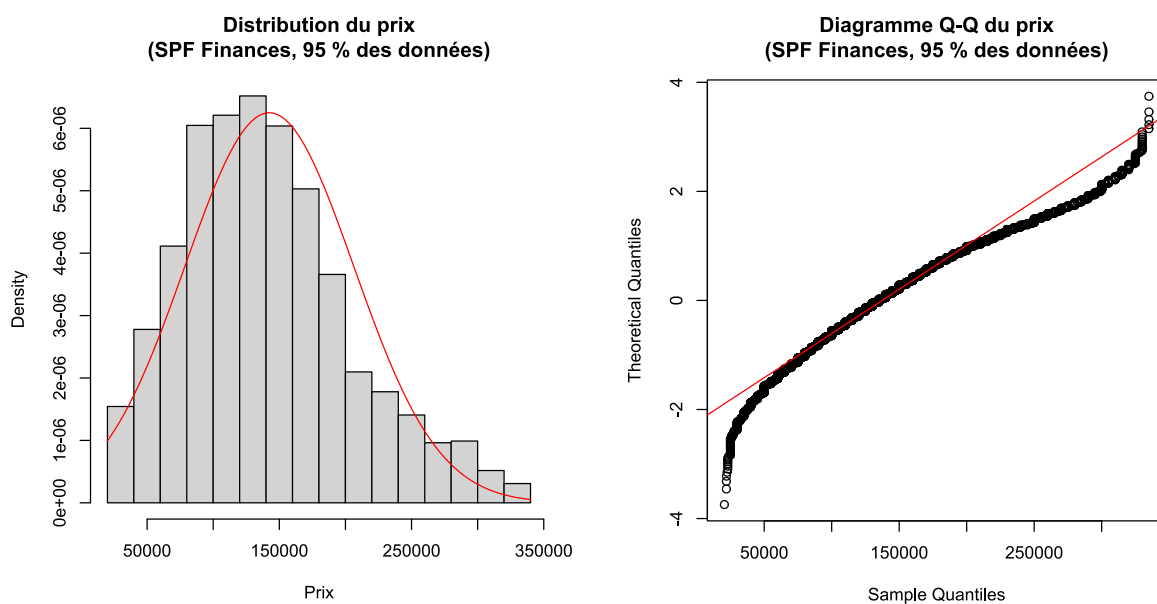
La segmentation du marché nécessite aussi une certaine homogénéité des valeurs de prix. Plus précisément, il faut éviter une variation de l'étendue de la distribution trop importante, ceci pour éviter de considérer simultanément des marchés distincts (haut de gamme et bas de gamme). Considérer ces paramètres permet de réduire le risque d'hétéroscédasticité (Des Rosiers, 2001). Pour cette étude, une coupure dans la distribution est envisagée afin de ne pas considérer le marché du haut de gamme et du bas de gamme. La distribution a été réduite à 99%, puis à 95 %. L'analyse de la distribution du prix brute montre que plusieurs valeurs sont

aberrantes, avec un minima commençant à zéro et un maxima à plus de deux millions d'euros. Ainsi, des prétraitements sur la distribution des données apparaissent nécessaires (Figure 10).



**Figure 10. Histogramme et diagramme Q-Q pour la distribution du prix du SPF Finances (brutes)**

Il est choisi d'écarter les valeurs les plus aberrantes en soustrayant 0,5 % et puis 2,5 % de part et d'autre de la distribution. Ceci a pour effet de nettement recentrer et aplatis la courbe de prix, qui était au départ désaxée vers la gauche et leptokurtique (Figure 11 et Annexe 5a).



**Figure 11. Histogramme et diagramme Q-Q pour la distribution du prix du SPF Finances (95 %)**

En parallèle, une analyse des données du prix pour la distribution des données Immoweb a été effectuée. Quelques valeurs aberrantes sont présentes à droite de la distribution Immoweb, indiquant des valeurs de prix plus hautes (Annexe 5b). Les variations entre les deux sets de données sont marquées avec des différences notables tant au niveau de la médiane, que de la moyenne. Néanmoins, le nombre d'observations est bien supérieur (dix fois plus) pour les données du SPF Finances en comparaison des données Immoweb (Tableau 9).

**Tableau 9. Statistique descriptive et indice de forme de la distribution du prix**

	Distribution du prix (SPF Finances, données brutes)	Distribution du prix (SPF Finances, 99 % des données)	Distribution du prix (SPF Finances, 95 % des données)	Distribution du prix (Immoweb, données brutes)
Effectif	5 806	5 746	5 507	548
Minimum	0	1 059	20 500	45 000
1 <sup>er</sup> quantile	95 000	95 000	95 850	134 375
Médiane	135 000	135 000	135 000	170 000
Moyenne	148 814	144 849	142 432	188 803
3 <sup>ème</sup> quantile	180 000	180 000	179 125	230 000
Maximum	2 400 000	525 000	335 000	550 000
Coefficient d'aplatissement	100,43	2,16	0,61	1,86
Coefficient d'asymétrie	7,01	1,06	0,08	1,19

Le choix se porte sur l'utilisation de la distribution à 95 % qui est la distribution qui se rapproche le plus d'une distribution normale parmi les trois distributions présentées. Le test de Kolmogorov-Smirnov permet de vérifier la normalité de la distribution. La p-valeur est inférieure à 0,05. Dès lors, l'hypothèse nulle (H0) est rejetée, il n'y a, *a priori*, pas de normalité (Tableau 10). D'autres transformations de la variable, comme l'application d'un logarithme, ont été envisagées sans grandes différences sur la normalité. Ainsi, même si la distribution n'est pas parfaitement gaussienne, il s'avère qu'une distribution normale est finalement rarement rencontrée pour les larges sets de données. De plus, une telle distribution n'empêche pas l'utilisation d'une régression linéaire (Lumley *et al.*, 2002).

**Tableau 10. Résultat du test de normalité de Kolmogorov-Smirnov sur le prix (95 % de la distribution)**

Lilliefors (Kolmogorov-Smirnov) normality test

Data: PRIX

d = 0,13981, p-value < 2,2e-16)

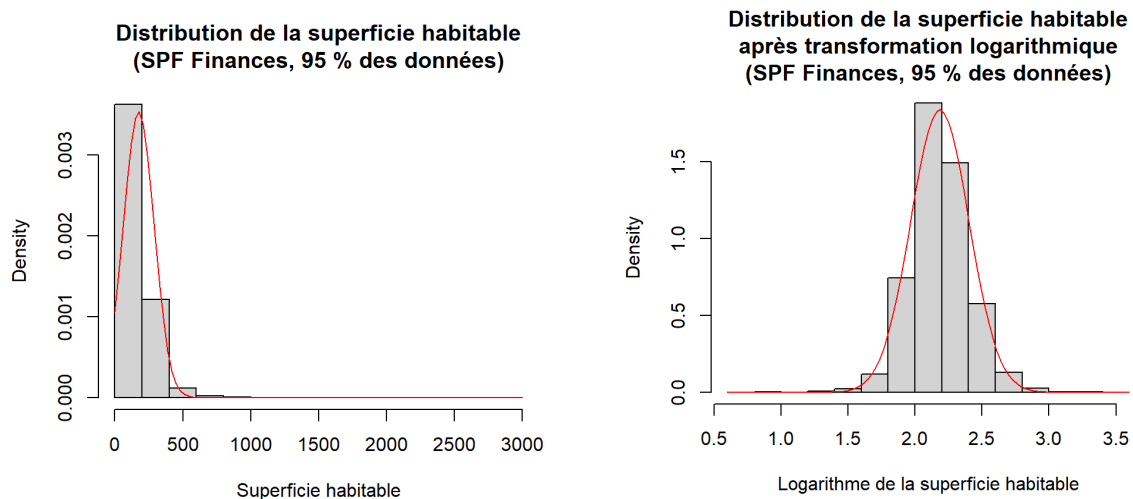
Dethier (2012), lors de la réalisation de son modèle final, arriva à la même conclusion et elle ne considéra pas les biens immobiliers dont le prix dépassait 350 000 €. Ceci revient à considérer uniquement les maisons d'habitation de standing standard. Pour comparer les deux jeux de données, la distribution Immoweb est ajustée selon la distribution du SPF Finances à 95 % avec des valeurs maximums à 335 000 €. Ainsi, trente-neuf habitations ont été exclues de la distribution Immoweb. Ceci permet d'éliminer les valeurs aberrantes, qui correspondent à des valeurs de prix hautes, repérées dans la distribution Immoweb auparavant (Annexe 5d).

### *3.1.2.2 Analyse descriptive et transformation des autres variables*

Au départ de la distribution de prix à 95 % (SPF Finances), les valeurs aberrantes des autres variables sont soustraites au set de données. Ces valeurs incohérentes ne trouvent pas d'explications particulières. L'exclusion est préférée à la modification des enregistrements pour éviter toute propagation d'erreurs lors de la modélisation hédonique. Par exemple, au niveau des indicateurs de mètres, des surfaces nulles sont rencontrées pour quatre biens immobiliers, ceux-ci sont éliminés. Un enregistrement concernant le nombre de pièces a été exclu du à un nombre de pièces trop importants (77 pièces).

Une transformation logarithmique est envisagée sur la superficie bâtie et la superficie de la parcelle cadastrale. Les indices de formes démontrent une forme désaxée vers la gauche et leptokurtique tout comme le prix l'était avant la suppression des valeurs les plus extrêmes (Annexe 6a). L'application du logarithme sur ces variables recentre les valeurs de la distribution (Figure 12).





Coefficient d'aplatissement : 90,6160

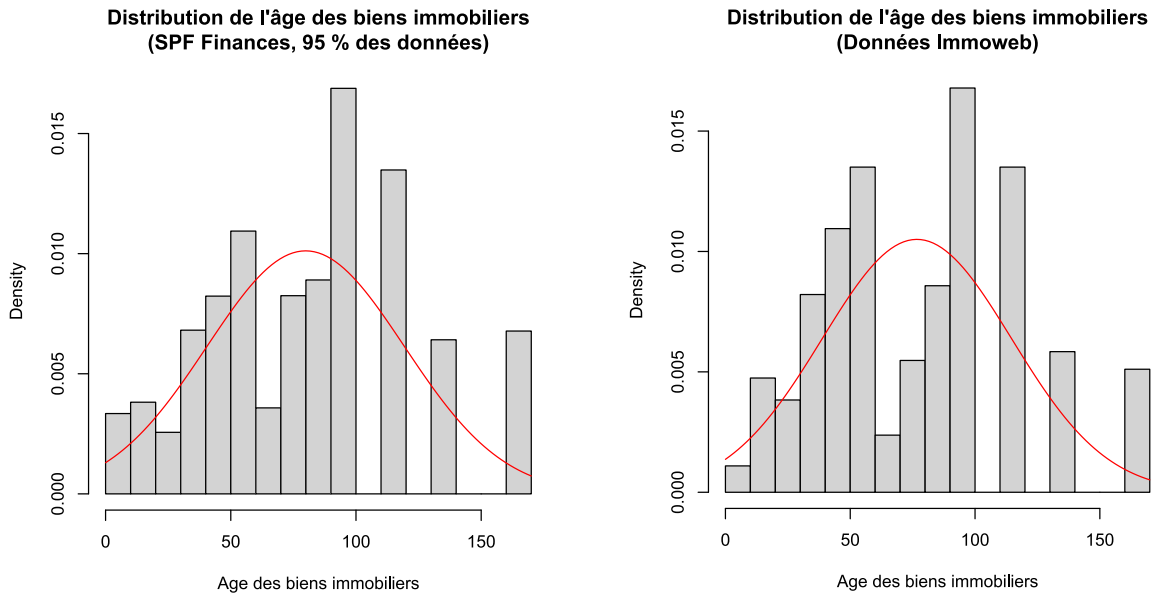
Coefficient d'asymétrie : 5,9645

Coefficient d'aplatissement : 2,0619

Coefficient d'asymétrie : 0,1160

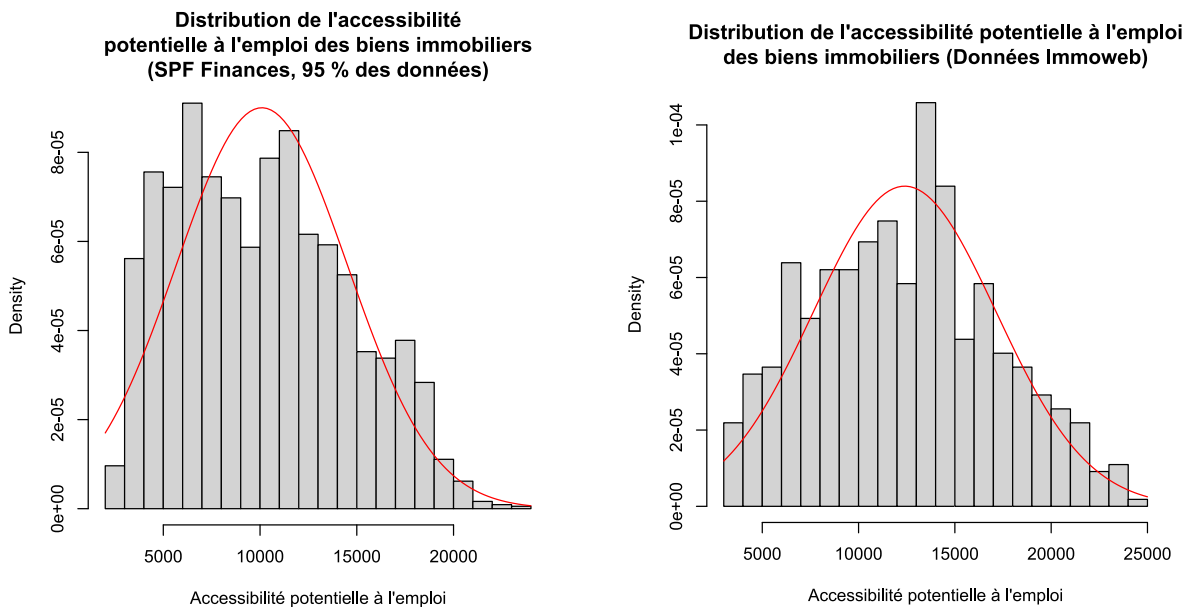
**Figure 12. Exemple de l'effet de la transformation logarithmique sur la distribution des superficies**

Malgré la source différente des données, l'allure des distributions des données Immoweb et des données du SPF Finances sont semblables. Pour rappel, l'âge des biens de plus de 81 ans est approximé, ce qui explique les coupures des deux distributions avec des pics plus importants et discontinus en fin de queue. Comme Dethier (2012) le soulignait déjà, la Seconde Guerre mondiale est visible dans la distribution avec une coupure pour les biens âgés de 70 ans, puis suivie par une phase de reconstructions où le nombre de biens est nettement plus représenté. Dans les années 1970, la crise économique induit une baisse du nombre de logements construits avec un minimum atteint dans les années 80 (Figure 13) (Dessouroux & Romainville, 2011).



**Figure 13. Comparaison de la distribution de l'âge des biens immobiliers entre les deux sets de données**

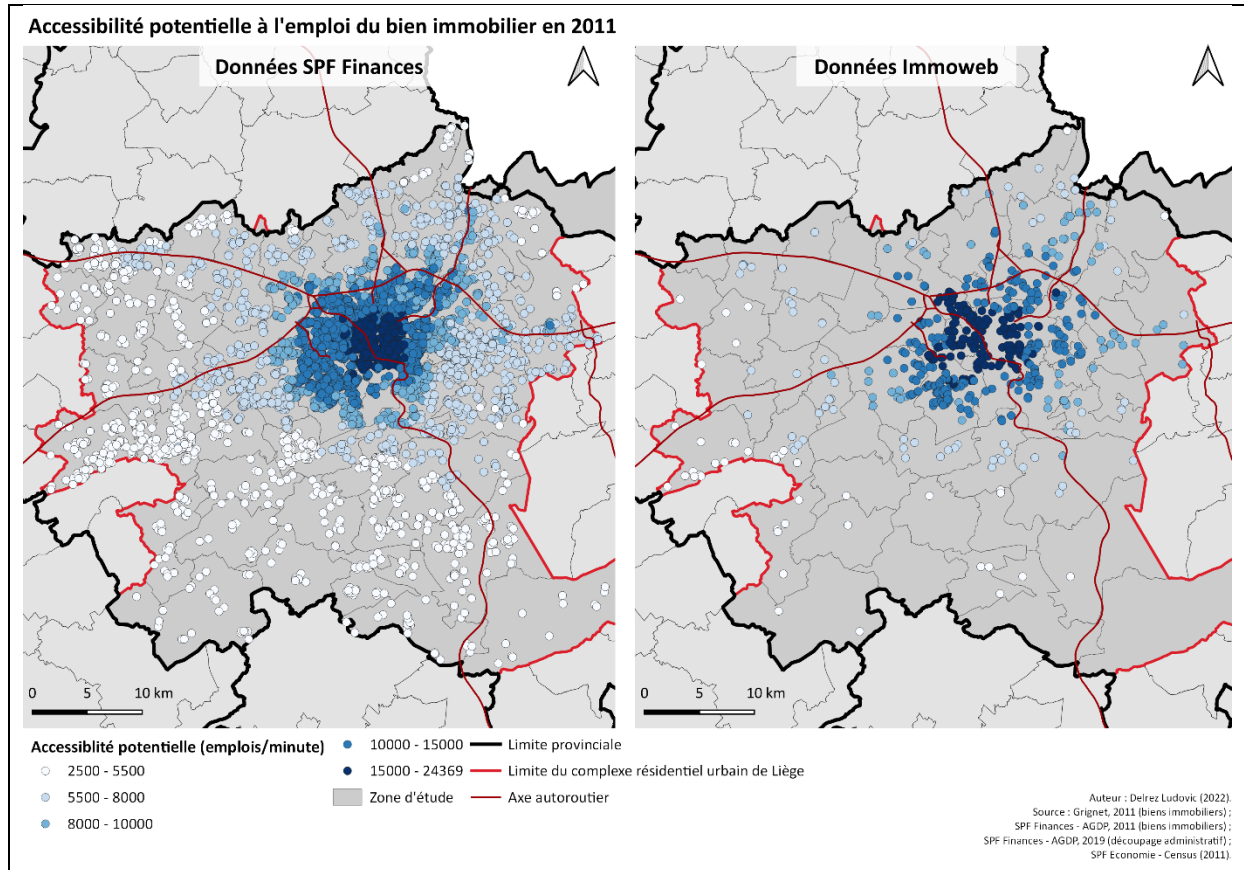
Pour l'accessibilité à l'emploi, les différences entre les deux sets de données sont plus marquées. Les distributions se rapprochent d'une forme gaussienne, mais il est plus complexe d'en tirer des conclusions (Figure 14).



**Figure 14. Comparaison de la distribution de l'accessibilité à l'emploi entre les deux sets de données**

Ainsi, une comparaison cartographique est plus adaptée pour mettre en avant d'éventuelles différences (ou ressemblances) pour cette variable. Pour les deux jeux de données, le potentiel

d'emploi le plus important se situe dans la ville-centrale, Liège. Ce potentiel diminue avec la distance au centre mais également avec l'éloignement des axes autoroutiers. Cela est cohérent étant donné l'utilisation d'une surface de coût pour le calcul de cet indice (Carte 3).



**Carte 3. Comparaison de l'accessibilité potentielle à l'emploi du bien immobilier en 2011**

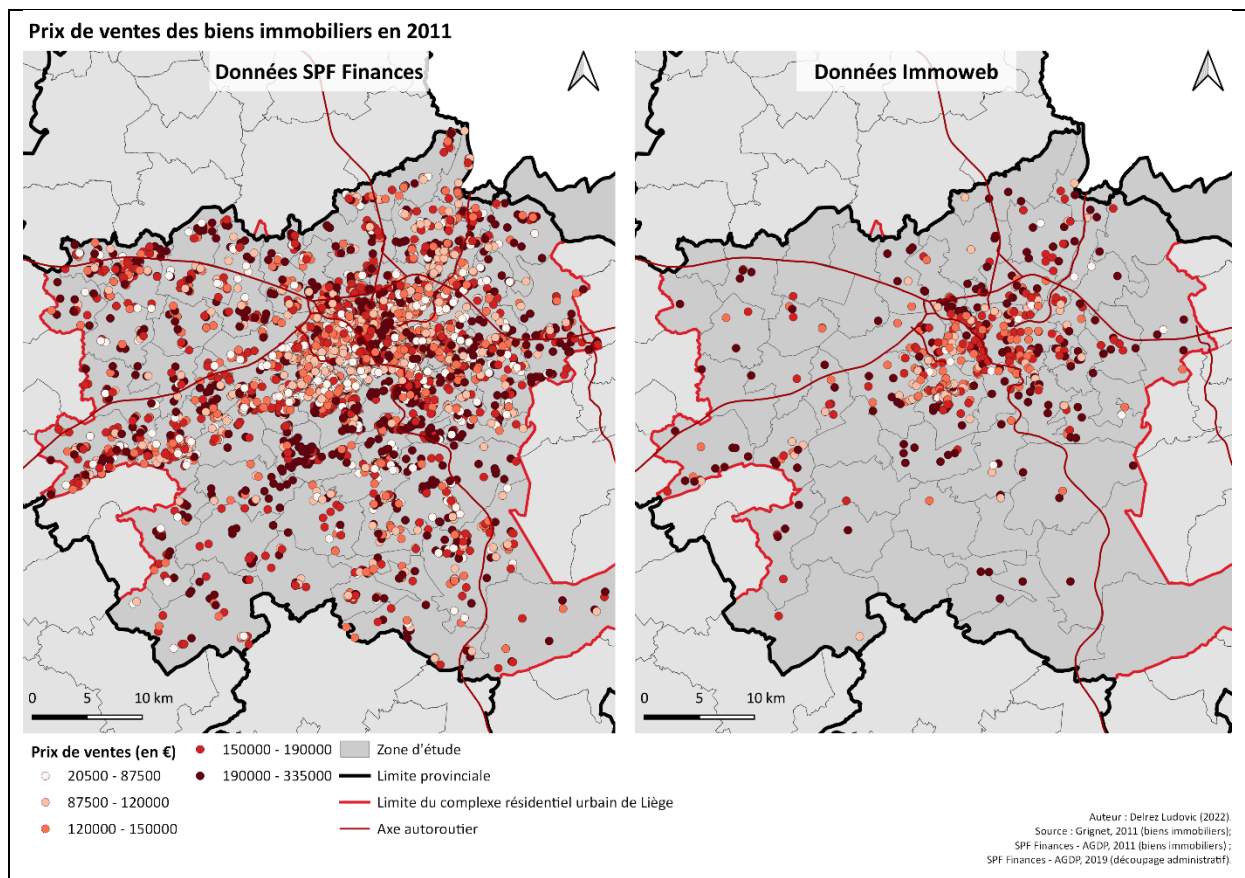
Les autres variables ne seront pas analysées ici car leur nombre est trop important pour en faire le détail complet dans ce document. Néanmoins, les tableaux en Annexe 6a et 6b reprennent les statistiques descriptives pour l'ensemble des variables (pour les deux jeux de données).

### 3.1.3 Distribution spatiale du prix des biens immobiliers

#### 3.1.3.1 Autocorrélation spatiale globale du prix

La Carte 4 compare la répartition du prix pour les données du SPF Finances et les données Immoweb. A priori, le prix est distribué de manière quelconque, les valeurs de prix ne semblent pas concentrées en considérant l'ensemble du territoire. Pour les données de Grignet, la distribution de points plus faible ne permet pas d'identifier visuellement des zones

de concentration. Pour les données cadastrales, localement, des concentrations de valeurs supérieures apparaissent au centre de Liège et au sud du territoire communal, notamment dans les communes de Neupré et Nandrin. Au contraire, la darse industrielle et ferroviaire, le long de la Meuse, présente des valeurs plus faibles. Les calculs de l'autocorrélation spatiale globale et locale permettent de confirmer ces hypothèses. Pour rappel, une absence d'autocorrélation spatiale est un prérequis pour l'application d'une régression (Le Gallo, 2002). Les phénomènes observés pouvant être influencé spatialement, une faible autocorrélation spatiale est tout de même toléré.



**Carte 4. Comparaison de la répartition du prix des biens immobiliers**

Avant de calculer l'indice I de Moran (Équation 9), il faut définir une matrice de poids qui considère la position relative de nos points. Fréquemment, la notion de contiguïté est considérée pour définir une matrice de poids. Cependant, l'utilisation de points envisage davantage l'utilisation d'une matrice de distances ou de voisinages (Le Gallo, 2002). La distance minimum pour que chaque point ait au moins un voisin est de 3 km, cette bande passante est importante et induit une grande variation en termes de voisinage (varie de 1 à 1101 voisins selon le point considéré). Ceci s'explique par l'écart plus important des

observations avec l'éloignement au centre de Liège. Ainsi, il est préféré de travailler avec une matrice de voisinages. Le nombre de voisins influençant les biens n'étant pas connu, plusieurs voisinages sont testés au 5<sup>ième</sup>, 10<sup>ième</sup> et 15<sup>ième</sup> voisin. Plus le voisinage augmente, plus les valeurs de prix sont harmonisées et plus l'indice I de Moran diminue. L'indice est positif mais il reste assez faible à 0,2 ainsi, il y a une légère autocorrélation spatiale globale du prix. Une simulation de Monte Carlo permute l'ensemble des points un grand nombre de fois (999 fois) ce qui permet de générer une distribution aléatoire en recalculant l'indice I de Moran pour chacune de ces permutations. Puis, la distribution aléatoire est comparée avec l'indice d'autocorrélation observé. Les p-valeurs déterminées pour l'ensemble des matrices de poids sont faibles, l'hypothèse nulle (H0) est rejetée, l'autocorrélation spatiale en présence n'est pas le fait du hasard (Tableau 11).

**Tableau 11. Résultat de l'indice I de Moran selon le choix de la matrice de poids pour les données du SPF Finances**

Matrice de poids	Indice de Moran	(pseudo) p-valeur	Z-value
5 <sup>ième</sup> voisins	0,2578	0,0010	32,1038
10 <sup>ième</sup> voisins	0,2320	0,0010	41,0591
15 <sup>ième</sup> voisins	0,2196	0,0010	47,6485
Distance minimale entre voisins (3 km)	0,122	0,0010	59,6757

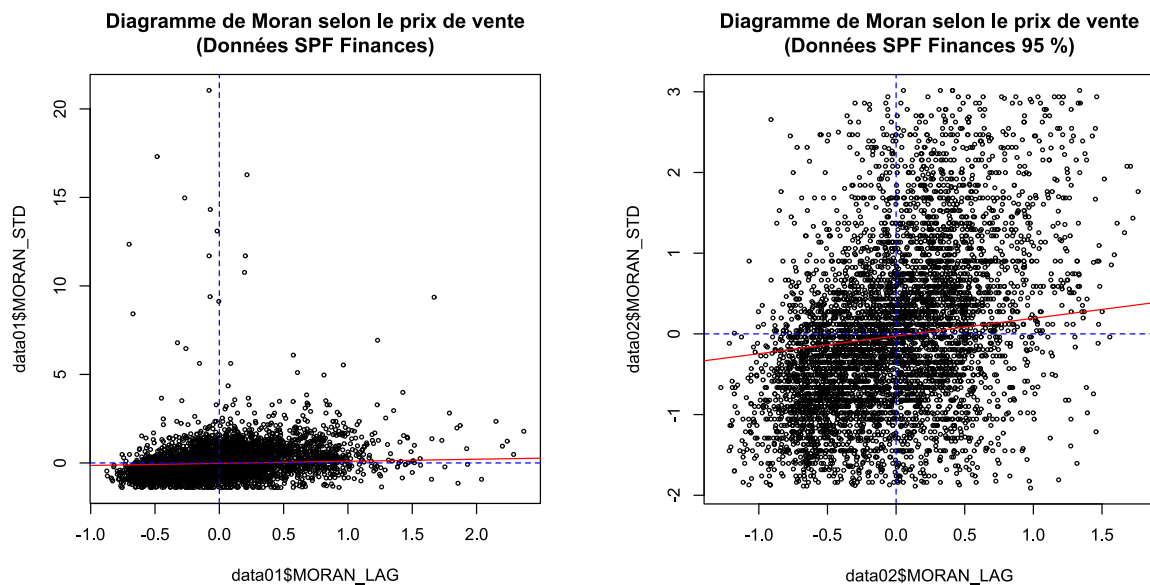
Les données Immoweb présentent également une très faible autocorrélation spatiale (autour de 0,14) lorsqu'un voisinage est considéré (Tableau 12).

**Tableau 12. Résultat de l'indice I de Moran selon le choix de la matrice de poids pour les données Immoweb**

Matrice de poids	Indice de Moran	(pseudo) p-valeur	Z-value
5 <sup>ième</sup> voisins	0,1745	0,0010	6,8387
10 <sup>ième</sup> voisins	0,1434	0,0010	7,9584
15 <sup>ième</sup> voisins	0,1423	0,0010	9,5060
Distance minimale entre voisins (7,5 km)	0,0509	0,0010	6,6279

### 3.1.3.2 Autocorrélation spatiale locale du prix

Une comparaison des diagrammes de Moran avant et après coupure dans la distribution du SPF Finances est reprise sur la Figure 15. En considérant les quinze premiers voisins, pour l'ensemble de la distribution l'indice de Moran est plus faible ( $I = 0,1173$ ), avec une droite de régression plus proche de la moyenne des valeurs de prix. Malgré la baisse de l'indice d'autocorrélation spatiale, le choix de travailler avec une distribution réduite prévient le risque de mélanger des marchés immobiliers distincts.

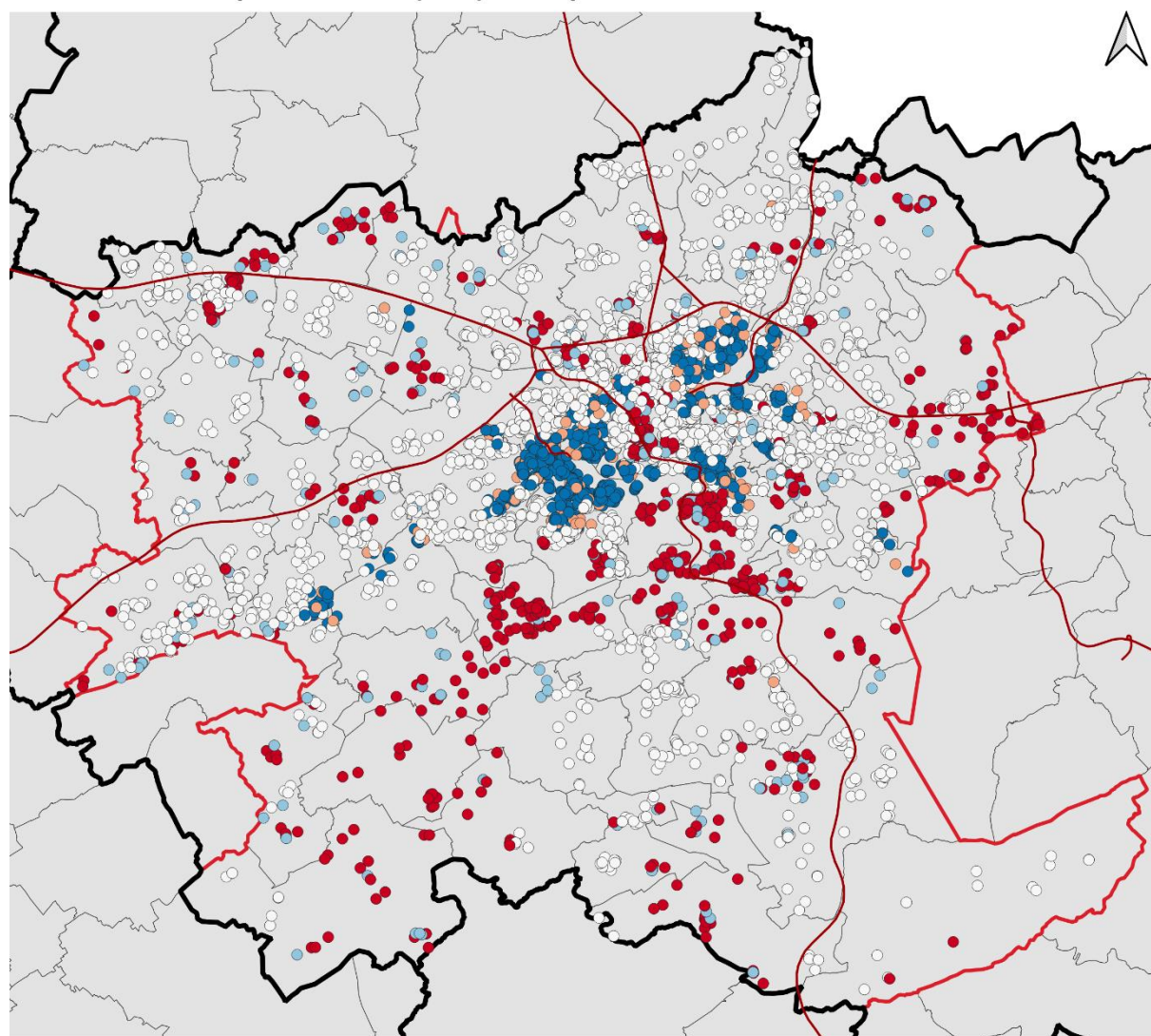


**Figure 15. Comparaison des diagrammes de Moran (15 voisins) pour les distributions OSIRIS**

Afin de mieux appréhender la présence d'autocorrélation spatiale, l'utilisation de l'indice d'autocorrélation spatiale locale (LISA) (Équation 10) est employée. L'indice est seulement analysé pour le jeu de données cadastrales (à 95 %) car la densité de points permet une meilleure compréhension visuelle des disparités spatiales observées. L'interprétation peut être combinée avec le diagramme de Moran. Les valeurs hautes dans un environnement haut (*high-high*) sont concentrées dans certains quartiers liégeois (des Guillemins, de Cointe, des Vennes, de Saint-Gilles ou encore de Saint-Laurent). Au sud de Liège, plusieurs localités ressortent également avec Neupré, Nandrin, Bonnelles, Beaufays ou encore Embourg. Les valeurs basses dans un environnement bas (*low-low*) se situent majoritairement dans les localités de Herstal, de Bressoux, de Grivegnée, de Chênée, d'Ougrée, de Seraing, de Saint-Nicolas ou encore d'Amay (Carte 5).



### Autocorrélation spatiale locale (LISA) sur le prix de vente en 2011



- |  |  |
|--|--|
| <b>Cluster selon le prix et l'environnement du prix (15 voisins)</b> | ○ Non-significatif (3123)                        |
| ● High-Hight (689)   | □ Zone d'étude                                   |
| ● High-Low (256)   | — Limite provinciale                             |
| ● Low-Hight (284)  | — Limite du complexe résidentiel urbain de Liège |
| ● Low-Low (1151)   | — Axe autoroutier                                |

Auteur : Delrez Ludovic (2022).  
 Source : SPF Finances - AGDP, 2011 (biens immobiliers) ;  
 SPF Finances - AGDP, 2019 (découpage administratif).

**Carte 5. Autocorrélation spatiale locale (LISA) sur le prix de vente en 2011**

### **3.1.4 Echantillonnage pour la validation-croisée**

Pour rappel des objectifs (2.1), une validation croisée est envisagée au terme de l'étude pour évaluer la robustesse des données du SPF Finances. L'échantillon de validation correspond approximativement à 10 % du set de données, soit à 550 observations. L'échantillon a été prélevé aléatoirement par la fonction « sélection aléatoire » dans QGIS. Cette étape n'est appliquée qu'au set de données du SPF Finances, car l'objectif n'est pas tellement de prouver l'efficacité des données Immoweb, lesquelles ont déjà été traitées dans d'autres études (Grignet, 2011 ; Dethier, 2012). De plus, le nombre d'observations Immoweb disponible est relativement faible. Les performances des modèles Immoweb seraient grandement réduites si un échantillon y était soustrait.

## **3.2 Analyse des corrélations**

L'analyse des corrélations permet de vérifier que la relation entre les différentes variables explicatives ne soit trop importante. Pour rappel, il faudra agréger ou faire un choix entre les variables interdépendantes qui présentent une corrélation supérieure à 0,8 (ou inférieure à -0,8) (Des Rosiers, 2001). De plus, la multiplication des variables explicatives ne génère pas nécessairement un meilleur modèle. Ainsi, il est conseillé de retirer les variables les moins corrélées à la variable dépendante (y) (Jawlik, 2016).

### **3.2.1 Analyse des corrélations pour la base de données du SPF Finances**

En ce qui concerne les variables quantitatives, le coefficient de corrélation de Spearman révèle des valeurs supérieures à 0,8 (ou inférieures à -0,8) pour un grand nombre de variables. Par exemple, la surface et le périmètre des parcelles cadastrales sont deux variables très corrélées. Il est choisi de conserver la surface car elle est davantage corrélée au prix. Ensuite, l'accessibilité potentielle à l'emploi est très corrélée avec les centres névralgiques de Saint-Lambert et des Guillemins, mais également avec toutes les variables d'accessibilité potentielle (aux arrêts de bus, aux gares de trains ou encore aux surfaces commerciales), ainsi qu'à la population par pixel. Étant donné que l'accessibilité potentielle à l'emploi est la variable la moins anachronique et qu'elle possède la relation la plus forte avec le prix, il est décidé de la conserver. Par ailleurs, d'autres variables peuvent rendre compte de l'effet des variables de potentielle non exploitées pour la régression. À l'inverse, Dethier (2012) décida de remplacer



la variable d'accessibilité à l'emploi par la distance-temps au centre de Liège (Saint-Lambert). Ces corrélations fortes entre ces différentes variables démontrent d'une part que les centres d'emploi les plus importants se situent également à proximité des superficies commerciales les plus grandes. Ces lieux étant généralement les mieux desservis en bus et en trains étant les plus centraux (par rapport à Liège). Somme toute, ce sont les centres qui présentent la part de population la plus importante. Parmi les variables du Lepur (Cornet *et al.*,2005), l'accessibilité en mode lent et l'accessibilité totale sont très corrélées. Afin de rendre compte de l'ensemble des variables d'accessibilité de la CPDT séparément, l'accessibilité totale n'est pas gardée. Finalement, une corrélation forte est rencontrée entre l'indice de Shannon évalué à 300 m et à 500 m. La corrélation avec le prix étant plus forte pour l'évaluation à 300 m, l'indice à 500 m n'est pas gardé (Annexe 7a).

Pour les corrélations bisériales seulement quelques corrélations sont supérieures à 0,8 (ou inférieures à -0,8). Sur les quatre corrélations mises en évidence, trois peuvent déjà être suspendues car elles ont été traitées au paragraphe précédent. Elles concernaient l'accessibilité potentielle au bus et la population dans le pixel concerné. Finalement, une corrélation forte est rencontrée entre l'âge de la rénovation (construction de l'indice au point 2.4.1) et l'indice de rénovation binarisée. L'âge de rénovation apporte davantage d'informations que la variable binaire. Ainsi, l'âge de rénovation est conservé pour la modélisation (Annexe 8a).

Une seconde matrice de corrélation, cette fois-ci entre les variables binaires, a permis de mettre en évidence d'autres relations trop intenses. Celles-ci concernent la configuration des maisons, le nombre d'étages, le nombre de logements, le nombre de garages et le nombre de salles de bains. Les variables non retenues pour ces groupements de variables sont :

- les maisons trois façades ;
- les maisons avec un étage ;
- les maisons avec un garage ;
- les maisons avec une salle de bains ;
- les maisons avec deux logements.

Ces variables auront, *a priori*, peu d'incidences sur la modélisation du prix. Leurs corrélations avec le prix est faible.

De plus, deux variables binaires sont corrélées. Il s'agit de la distance à un SAR ou d'une gare de trains dans un rayon de 300 à 500 m. La distance d'une gare de trains dans un rayon

de 300 à 500 m nous semble plus intéressante à conserver. Finalement, plusieurs variables ont une corrélation avec le prix trop faible telles que :

- l'indice de rénovation binaire ;
- la présence d'une cave habitable ;
- l'accessibilité au train (carte d'accessibilité de la CPDT) ;
- la proximité d'un arrêt de bus entre 100 et 300 m ;
- la proximité d'une gare de trains à moins de 300 m et entre 500 et 700 m ;
- la proximité des commerces (alimentaires, boulangeries et pharmacies) ;
- la distance à un petit, moyen ou grand parc ;
- la proximité à un cours d'eau en dessous de 100 m ;
- la proximité aux espaces verts au-delà de 100 m ;
- la proximité aux sites d'extraction au-delà de 100 m ;
- la proximité aux sites industriels entre 500 et 700 m ;
- le bruit des routes ;
- le bruit faible de la voie ferrée ;
- le bruit des avions ;
- la variable de Kiemstedt-Schmitz.

Plusieurs raisons peuvent expliquer ces faibles corrélations (le seuil de  $r \leq 0,10$  est considéré). Pour commencer, celles-ci concernent majoritairement des variables de qualité physique, celles-ci expliquent généralement moins bien le prix des biens immobiliers. Pour certaines variables, une mauvaise spécification est possible. Par exemple, la distance aux parcs n'est considérée que pour la commune de Liège. Les commerces sont anachroniques de la période de référence. Ainsi, seulement quarante-six variables sont exploitables pour la modélisation du prix sur les cent-onze variables produites.

### **3.2.2 Analyse des corrélations pour la base de données Immoweb**

La même procédure est appliquée à la base de données Immoweb. Les conclusions sont assez semblables au point précédent (3.2.1). Plus succinctement, les variables non retenues suite à l'analyse de la matrice de corrélation entre les variables quantitatives (Annexe 7b) sont :

- les accessibilités potentielles aux arrêts de bus, aux gares de trains et aux nœuds commerciaux ;
- les distances-temps de Saint-Lambert, des Guillemins et l'agrégation des deux variables ;

- l'accessibilité totale et l'accessibilité en train selon la CPDT ;
- l'indice de Shannon évalué dans un rayon de 500 m.

Pour les variables binaires, un seul coefficient de corrélation bisériale, entre le logarithme de la superficie foncière et la configuration quatre façades, est supérieur à 0,8. Le logarithme de la superficie foncière est conservé (Annexe 8b). Deux variables avec un coefficient de corrélation à 0,75 ont dû être retirées pour la modélisation. Il s'agit de la population par pixel d'un kilomètre carré et des maisons avec un seul garage. Finalement, les variables binaires dont les corrélations sont trop faibles sont exclues. Il s'agit des variables, ci-dessous :

- les maisons 3 façades ;
- les maisons avec 1 étage ;
- l'indice de rénovation binaire ;
- les maisons avec une salle de bains ;
- les maisons avec trois chambres ;
- les maisons avec une cave ;
- la proximité d'une gare de trains à moins de 300 m ;
- la proximité d'un arrêt de bus entre 100 à 300 m ;
- l'ensemble des variables liées aux commerces (excepté les biens situés entre 500 et 700 m d'une boulangerie ou entre 300 à 500 m d'une pharmacie) ;
- la distance à un parc de moyenne et de grande dimension ;
- la proximité d'un cours d'eau à moins de 300 m ;
- la proximité d'un espace vert (excepté les biens situés entre 300 et 500 m) ;
- la proximité d'un site d'extraction entre 100 et 300 m ;
- la proximité d'un site industriel entre 500 et 700 m ;
- le bruit induit par le réseau routier (sauf quand celui-ci est très fort) ;
- le bruit moyen induit par les avions de l'aéroport de Bierset.

Pour les données Immoweb, au-delà de la mauvaise spécification des variables, le manque d'échantillonnage peut ne pas rendre compte de certaines variables. Ainsi, cinquante-cinq variables ont été employées pour la modélisation.

### 3.3 Analyse de régression

Les modèles de régressions multiples sont soit effectués dans le logiciel GeoDa, soit dans ArcGIS via l'outil « *ordinary least squares (OLS)* ». La différence majeure entre les deux logiciels est que GeoDa ne calcule pas le VIF (Équation 8) mais un indice de multicollinéarité globale.

#### 3.3.1 Modèle A, inspiration du modèle de Dethier

Les modèles définis par Dethier (2012) sont des sources d'inspiration pour formuler un modèle dont la part explicative du prix ( $R^2$ ) devrait avoisiner les 80 à 90 %. Ainsi, en adaptant quelque peu les variables introduites, le coefficient de détermination atteint 42 % pour le set de données du SPF Finances. Les variables semblables sont utilisées pour le set de données Immoweb et le coefficient de détermination atteint les 70 %. Les seules différences dans les données sont la taille de l'échantillon et les variables intrinsèques aux biens immobiliers. En analysant les autres indicateurs d'ajustement, la somme des carrés des résidus est plus importante pour les données du SPF Finances. Néanmoins, l'échantillon plus important peut induire une erreur plus grande. Par contre, le modèle Immoweb présente une redondance des données plus marquées. Les deux modèles n'indiquent pas de problèmes liés à la normalité et à l'hétéroscédasticité. De plus, ils sont tous les deux significatifs lors du test F (avec un coefficient de significativité ( $\alpha = 0,05$ )) (Tableau 13 et Annexe 9a).

**Tableau 13. Résumé des indicateurs d'ajustements du modèle A**

	Données du SPF Finances	Données Immoweb
R-squared	0,4198	0,6982
Adjusted R-squared	0,4180	0,6890
Prix moyen	142 457 €	173 669 €
Erreur type	48 587,40 €	34 055 €
Erreur relative	34,11 %	19,61 %
F-statistic	F(16, 4937) = 238,0970 Prob(F-statistic) = 0,0000	F(16, 493) = 76,0258 Prob(F-statistic) = 0,0000
Multicollinearity condition number	33,8227	71,7870
Jarque-Bera test	JB(16, 2) = 444,2698 Prob(JB) = 0,0000	JB(16, 2) = 11,4419 Prob(JB) = 0,0033
Breusch-Pagan test	BP(16, 15) = 710,0141 Prob(BP) = 0,0000	BP(16, 15) = 61,6410 Prob(BP) = 0,0000

Koenker-Basset test	KB(16, 15) =410,2047 Prob(KP) = 0,0000	KB(16, 15) =45,3317 Prob(KP) = 0,0001
---------------------	---	--

Malgré un premier résultat peu concluant sur le devenir du jeu de données du SPF Finances, il y a une volonté d'analyser plus en détail ce dernier. Notamment, par l'analyse des résidus et par l'utilisation des variables les plus propices à la définition d'un modèle adéquat. En parallèle, les mêmes traitements ont été opérés sur la BD Immoweb comme points de comparaisons. Néanmoins, cette première modélisation est un peu naïve. Elle a l'avantage de montrer des résultats si la seule différence entre les modèles sont les variables internes aux biens immobiliers (et la taille d'échantillonnage). Pour les modèles suivants, l'ajustement des deux jeux de données induira des différences dans les choix de variables. Le modèle présenté ici, n'est pas à l'avantage du jeu de données du SPF Finances (en termes d'ajustement), ce qui conclut probablement à une performance moins importante du modèle. Les comparaisons sont plus délicates pour les modèles suivants, car les variables introduites ne sont plus semblables d'un jeu de données à l'autre. Les comparaisons sont dès lors faites sur le coefficient de détermination ( $R^2$ ) et sur l'erreur relative des modèles. Pour rappel, il est supposé que plus ce coefficient est important (en respectant l'ajustement du modèle) et plus le modèle est performant.

### 3.3.2 Modèle B, exploration en considérant l'ensemble des variables

Tout d'abord, la méthode proposée par Des Rosiers (2008) consiste à introduire l'ensemble des variables dans ce premier modèle. Le premier modèle du SPF Finances intègre ainsi les quarante-sept variables retenues après l'analyse des corrélations (point 3.2.1). Ceci permet tout d'abord de vérifier la significativité des variables et ensuite d'exclure les variables les moins intéressantes une à une dans les modèles suivants.

Le modèle B du SPF Finances est un peu plus explicite que le modèle A avec un coefficient de détermination ( $R^2$ ) qui atteint quasiment les 50 %. Cette augmentation résulte surtout de la multiplication des variables dans le modèle. Sur les quarante-quatre variables implémentées dans le modèle B, seulement dix-huit sont significatives au test t pour le modèle du SPF Finances ( $\alpha = 0,05$ ). Une redondance des données existe dans le modèle étant donné que l'ensemble des variables a été considéré. Finalement, l'erreur relative du modèle B (SPD Finances) est de 32 % (Tableau 14 et Annexe 9b).

Pour les données Immoweb, en introduisant les cinquante-quatre variables retenues suite à l'analyse des corrélations, la part explicative ( $R^2$ ) atteint 72 %. Parmi les variables retenues,

seulement huit variables sont significatives ( $\alpha = 0,05$ ). De plus, à nouveau, l'indice de multicollinéarité montre une grande redondance dans les données. Le test de Koenker-Basset réfute l'hypothèse nulle ( $H_0$ ), il y a probablement une hétéroscédasticité des résidus ( $\alpha = 0,05$ ). L'erreur relative est de 20 %, c'est moins que le modèle du SPF Finances (Tableau 14 et Annexe 9b).

**Tableau 14. Résumé des indicateurs d'ajustements du modèle B**

	<b>Données du SPF Finances</b>	<b>Données Immoweb</b>
R <sup>2</sup>	0,4928	0,7213
R <sup>2</sup> ajusté	0,4880	0,6881
Prix moyen	142 446 €	173 669 €
Erreur type	45 574,40 €	34 102,40 €
Erreur relative	31,99 %	19,64 %
Test F	F : 107,3840 Prob(>F), (48, 4904) : 0,0000*	F : 21,7564 Prob(>F), (55, 454) = 0,0000*
Indice de multicollinéarité	108,2379	140,4619
Test de Jarque-Bera	JB : 917,2379 Prob(> $\chi^2$ ), (2) : 0,0000*	JB(2) : 9,0452 Prob(> $\chi^2$ ), (2) = 0,0109*
Test de Breusch-Pagan	BP : 1161,2197 Prob(> $\chi^2$ ) (47) : 0,0000*	BP : 89,6966 Prob(> $\chi^2$ ), (54) : 0,0016*
Test de Koenker-Basset	KB : 572,1925 Prob(> $\chi^2$ ) (47) = 0,0000*	KB : 67,9776 Prob(> $\chi^2$ ), (54) : 0,0957

### 3.3.3 Modèle C, affinage du modèle B

Sur la base du modèle B, l'ensemble des variables ayant passé le test t de Student a été conservé. Vingt-trois variables sont utilisées pour la modélisation du modèle C. Le coefficient de détermination ( $R^2$ ) reste stable à 49 %. Le test de multicollinéarité montre une certaine redondance des données (Tableau 15 et Annexe 9c). Pour vérifier la redondance sur chaque variable, le modèle est également testé dans ArcGIS. Toutes les variables ont un VIF inférieur à 3. Les résultats sont meilleurs que le seuil de 7,5 attendu précédemment (Annexe 10a).

Pour le modèle Immoweb, la suppression des variables, suite au t test, a permis de conserver treize variables pour ce modèle. Malgré un nombre de variables presque deux fois moins important, le modèle Immoweb parvient à expliquer jusqu'à 71 % du prix. Néanmoins, la normalité des résidus n'est pas vérifiée par le test de Jarque-Bera, l'hypothèse nulle ( $H_0$ ) est rejetée ( $\alpha = 0,05$ ). L'indice de multicollinéarité indique également une redondance des données (Tableau 15 et Annexe 9c). Néanmoins, l'ensemble des VIF est inférieur à 3. La

redondance observée pourrait être liée aux variables socio-économiques (revenu médian et part des diplômés du supérieur) (Annexe 10b).

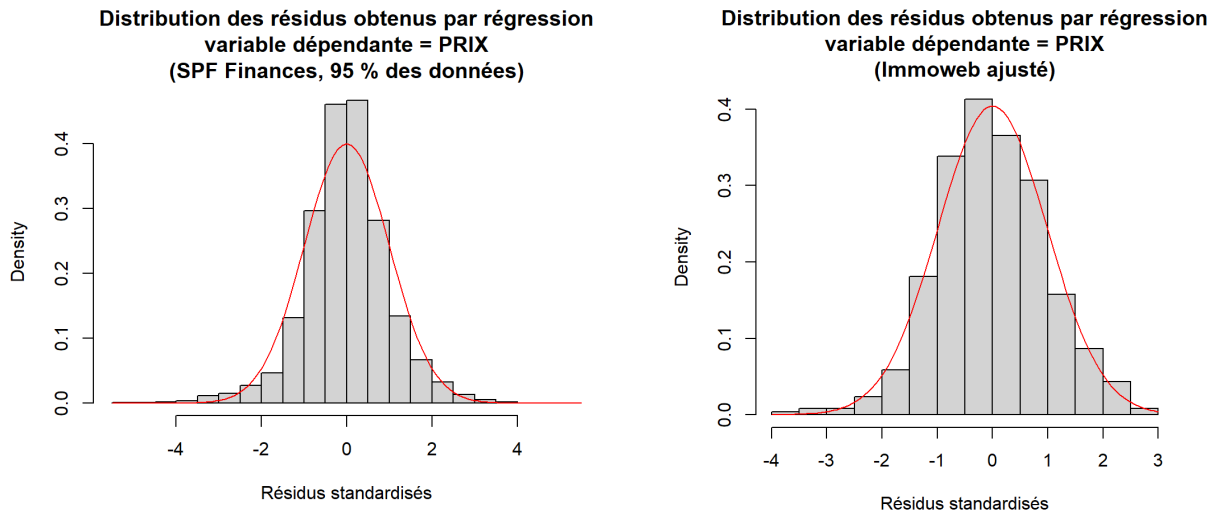
**Tableau 15. Résumé des indicateurs d'ajustements du modèle C**

	<b>Données du SPF Finances</b>	<b>Données Immoweb</b>
R <sup>2</sup>	0,4905	0,7076
R <sup>2</sup> ajusté	0,4884	0,6999
Prix moyen	142 446 €	173 669 €
Erreur type	45 552,70 €	33 451,80 €
Erreur relative	31,98 %	19,26 %
Test F	F : 237,3630 Prob(>F), (21, 4931) : 0,0000*	F : 92,1397 Prob(>F), (14, 495) = 0,0000*
Indice de multicollinéarité	71,6504	74,3467
Test de Jarque-Bera	JB : 961,0897 Prob(> $\chi^2$ ), (2) : 0,0000*	JB : 4,9798 Prob(> $\chi^2$ ), (2) : 0,0829
Test de Breusch-Pagan	BP : 1046,2237 Prob(> $\chi^2$ ) (20) : 0,0000*	BP : 45,2881 Prob(> $\chi^2$ ), (13) : 0,0000*
Test de Koenker-Basset	KB : 508,7810 Prob(> $\chi^2$ ) (20) = 0,0000*	KB : 36,6345 Prob(> $\chi^2$ ), (13) : 0,0047*

### 3.4 Analyse des résidus

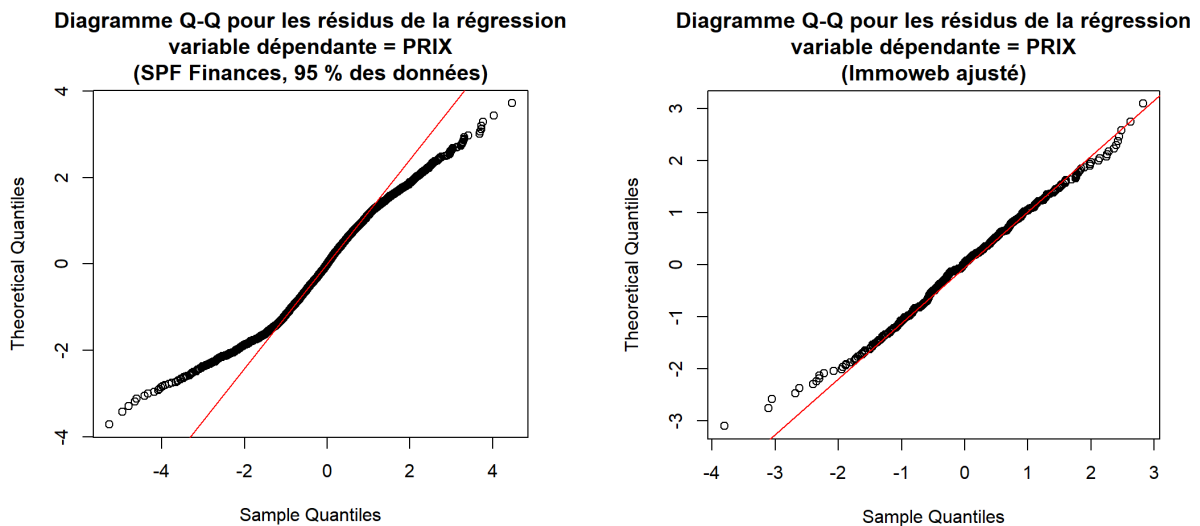
#### 3.4.1 Distribution des résidus obtenus par régression sur le prix

Les résidus sont analysés au départ du modèle C pour les données du SPF Finances. La distribution des résidus est centrée et réduite, car les résidus sont standardisés. Les indices de formes mettent tout de même en avant une distribution légèrement leptokurtique ( $I_{applatissement} = 2,11$ ), même si la distribution est centrée ( $I_{asymétrie} = -0,22$ ). Ensuite, les résidus sont analysés toujours au départ du modèle C mais cette fois-ci pour les données Immoweb. Les indices de formes démontrent une distribution relativement normale  $I_{applatissement} = 0,46$  et  $I_{asymétrie} = -0,05$  (Figure 16).



**Figure 16. Distribution des résidus obtenus par régression sur le prix**

Pour le jeu de données du SPF Finances, l'étendue est de 10,24, avec un minimum à -5,12 et un maximum à 5,12. Le diagramme quantile-quantile met en avant les résidus extrêmes qui ne suivent pas une distribution normale. Les valeurs marginales se situent au-delà de -2 et 2. Pour le set de données Immoweb, l'étendue est de 6,62. La valeur minimum est de -3,8 et la valeur maximum est de 2,82. Les valeurs suivent davantage la ligne rouge du diagramme quantile-quantile, ceci prouve à nouveau la distribution normale des résidus (Figure 17).



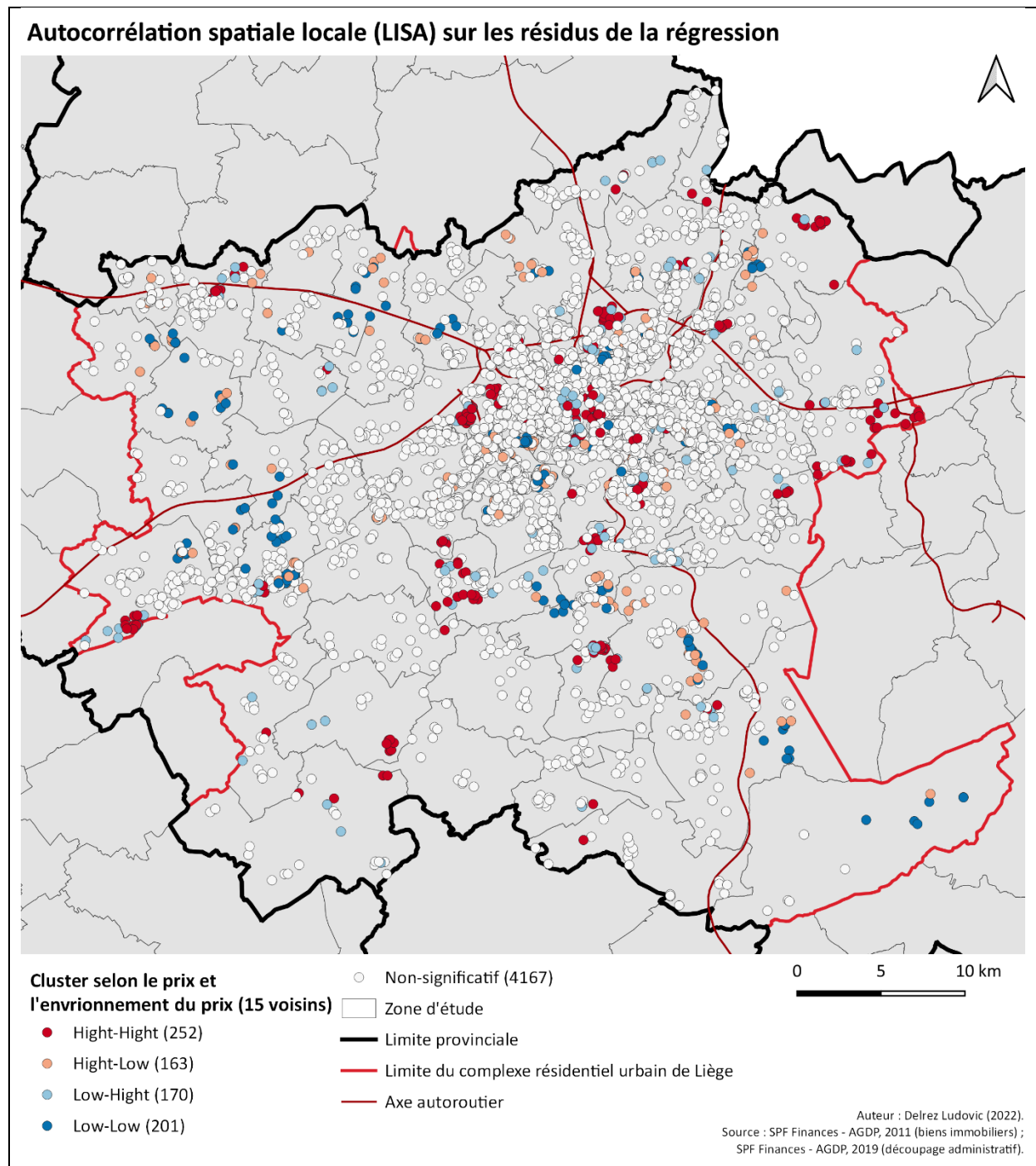
**Figure 17. Diagramme quantile-quantile pour les résidus obtenus par régression sur le prix**



### 3.4.2 Vérification de l'autocorrélation spatiale des résidus

Dans le cas où les phénomènes étudiés sont spatialement dépendants, la probabilité d'avoir une autocorrélation spatiale des résidus est grande, car si les données sont spatialement dépendantes, il existe certainement une redondance de l'information de manière localisée. Or, l'analyse de la structure spatiale des résidus peut permettre d'identifier une variable locale manquante au modèle (Feuillet *et al.*, 2019).

L'indice I de Moran (Équation 9), pour les résidus du SPF Finances, considère une autocorrélation spatiale globale aléatoire avec un  $I = 0,035$  (évaluation à partir des 15 premiers voisins). Sur la base du test de permutation de Monte Carlo (effectué 999 fois), l'hypothèse nulle ( $H_0$ ) est rejetée, une très faible autocorrélation spatiale globale existe. Ceci semble cohérent par rapport à la littérature (Feuillet *et al.*, 2019). Ainsi, l'indice d'autocorrélation spatiale locale (Équation 10) (LISA) est envisagé pour identifier les zones qui sont autocorrelées spatialement. Il permet de mettre en évidence une éventuelle omission de variables. Dans le cas présent, l'autocorrélation spatiale locale mise en évidence sur la Carte 6 permet difficilement d'établir le manque d'une ou de plusieurs variables. Les centralités de Verviers et de Huy pourraient également avoir un impact sur le modèle pour les résidus haut dans un environnement haut à la limite de la zone d'étude. Certaines valeurs résiduelles pourraient s'expliquer par les nuisances sonores induites par l'autoroute. Cependant, malgré l'agrégation des variables relatives aux bruits, elle n'est pas devenue significative au t test.



**Carte 6. Autocorrélation spatiale locale (LISA) sur les résidus de la régression**

L'indice I de Moran a également été calculé pour les résidus Immoweb. L'autocorrélation spatiale globale est aléatoire avec un  $I = -0,009$  (évaluation à partir des 15 premiers voisins). Le test de permutation de Monte Carlo (effectué 999 fois), ne réfute pas l'hypothèse nulle ( $H_0$ ), il semble que les résidus soient distribués de manière aléatoire (avec un coefficient de significativité  $\alpha = 0,001$ ) (Anselin, 2020).

### 3.4.3 L'exclusion des résidus marginaux

Une première étape consiste à exclure les résidus les plus extrêmes. En général, les erreurs standardisés (SSE) supérieures à 2 ou 3 sont considérées. Normalement, il est attendu de justifier chacune des exclusions (par exemple, vente au rabais, vente à un parent, etc.). Néanmoins, le manque d'informations sur les ventes ne le permet pas dans le cas présent (Des Rosiers, 2008). Pour commencer, il est envisagé d'exclure les résidus supérieurs à 3 (ou inférieurs à -3) du modèle. Puis, les résidus supérieurs à 2 (ou inférieurs à -2) sont supprimés. Pour les données du SPF Finances, cette suppression considère l'exclusion respective de 70 (1,4 %) et de 290 (5,6 %) enregistrements. Pour le jeu de données Immoweb, les résidus supérieurs à 2 (ou inférieurs à -2) sont directement envisagés étant donné le peu de valeurs de résidus supérieurs à 3 (ou inférieurs à -3). Ainsi, l'exclusion concerne 24 (4,7 %) observations (Figure 18).

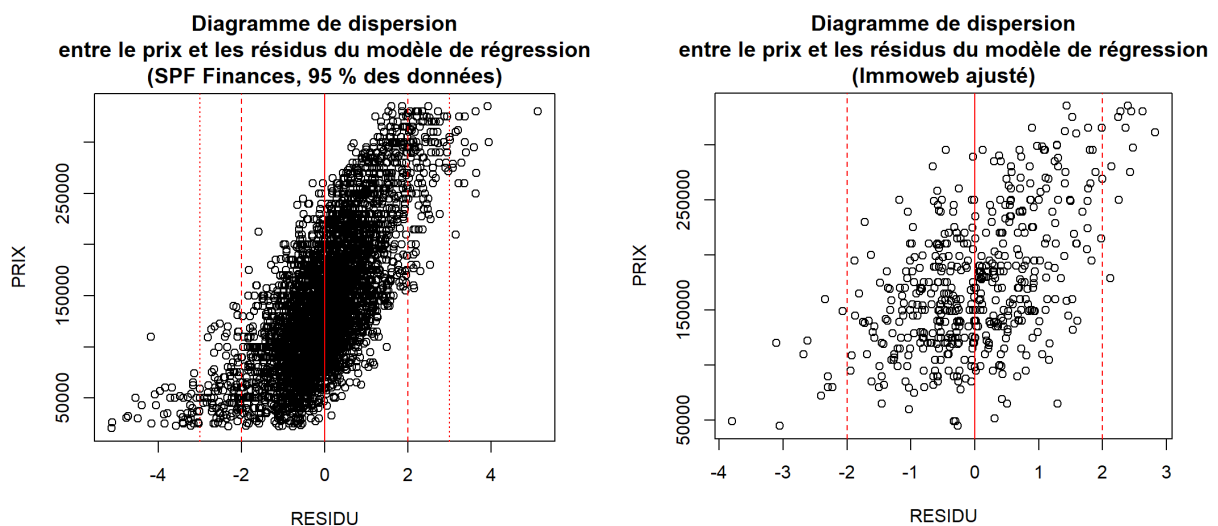


Figure 18. Diagramme de dispersion entre le prix et les résidus du modèle de régression

## 3.5 Mise au point du modèle final

### 3.5.1 Modèle final (Modèle D)

En supprimant, les résidus supérieurs à 3 (ou inférieurs à -3), le coefficient de détermination ( $R^2$ ) a pu atteindre les 56 % (Modèle D1). Ensuite, la suppression des résidus supérieurs à 2 (ou inférieurs à -2) a permis à nouveau une augmentation du coefficient de détermination ( $R^2$ ) jusqu'à 64 % (Modèle D2). Les paramètres statistiques des deux modèles sont validés. Excepté, pour le modèle D2, où le test de Jarque-Bera rejette l'hypothèse nulle ( $H_0$ ). La

distribution des résidus ne ressemble pas à une distribution normale ( $\alpha = 0,05$ ). Cependant, l'AIC et le BIC montrent que le modèle D2 est plus robuste que le modèle D1 (Tableau 16, Annexe 9d et 12).

Le modèle Immoweb atteint les 75 % de part explicative ( $R^2$ ) (Modèle D3). Ce résultat peut paraître faible en comparaison des résultats de Dethier (2012). L'explication principale réside dans le choix des variables et la non considération de certaines variables propres à la BD Immoweb. Pour rappel, ces variables n'ont pas été considérées pour s'ajuster aux variables disponibles à la BD cadastrale (Tableau 2). De plus, une redondance globale persiste dans les données malgré la suppression de la variable du revenu médian qui paraît redondante par rapport à la part des diplômés du supérieur. La première variable est tout de même exclue du modèle pour ne plus avoir d'hétéroscédasticité des résidus. Les valeurs de VIF sont toutes inférieures à 2 (Annexe 10). Tout comme le modèle D2 (SPF Finances), le test de Jarque-Bera n'est pas validé, l'hypothèse nulle ( $H_0$ ) est réfutée. La distribution des résidus ne ressemble pas à une distribution normale ( $\alpha = 0,05$ ) (Tableau 16 et Annexe 9d). Pour les trois modèles, l'indice de multicollinéarité est supérieur à 30. Cependant, l'analyse des VIF ne met pas en évidence une redondance exacerbée des données (Annexe 10c, 10d et 10e).

Outre le coefficient de détermination ( $R^2$ ), l'erreur relative est un élément indispensable pour conclure sur les performances des modèles. L'erreur relative est calculée pour le modèle D2 (SPF Finances) et le modèle D3 (Immoweb) (même si la distribution des résidus semble être anormale). Pour commencer, l'erreur relative sur le prix moyen est de 29,28 % pour le modèle D2, contre seulement 16,79 % pour le modèle D3. Ainsi l'erreur relative est proche des 10 % suggérée par Evans (1990) pour les données Immoweb. A contrario, les données du SPF Finances présentent une erreur moyenne relativement importante (Tableau 16).

**Tableau 16. Résumé des indicateurs d'ajustements des modèles D**

	<b>Données du SPF Finances – Modèle D1</b>	<b>Données du SPF Finances – Modèle D2</b>	<b>Données Immoweb – Modèle D3</b>
$R^2$	0,5649	0,6443	0,7539
$R^2$ ajusté	0,5631	0,6429	0,7476
Prix moyen	142 822 €	141 251 €	172 153 €
Erreur type	41 312,30 €	34 980,30 €	28 911,50 €
Erreur relative	28,93 %	24,76 %	16,79 %
AIC	117656	110802	/
BIC	117793	110924	/
Test F	F : 313,5400	F : 467,1790	F :120,4890

	Prob(>F) (21, 4861) : 0,0000*	Prob(>F) (19, 4861) : 0,0000*	Prob(>F),(13, 472) : 0,0000*
Indice de multicollinéarité	72,6604	73,1345	67,7445
Test de Jarque-Bera	JB : 103,0026 Prob(> $\chi^2$ ), (2) : 0,0000*	JB : 4,0375 Prob(> $\chi^2$ ), (2) : 0,13282	JB : 5,3471 Prob(> $\chi^2$ ), (2) : 0,0690
Test de Breusch-Pagan	BP : 588,3781 Prob(> $\chi^2$ ) (20) : 0,0000*	BP : 229,7941 Prob(> $\chi^2$ ) (18) : 0,0000*	BP : 21,8260 Prob(> $\chi^2$ ) (2) : 0,0395*
Test de Koenker-Basset	KB : 434,0621 Prob(> $\chi^2$ ) (20) = 0,0000*	KB : 246,4960 Prob(> $\chi^2$ ) (18) = 0,0000*	KB : 28,1030 Prob(> $\chi^2$ ) (2) : 0,0053*

### 3.5.2 Validation du modèle final

La validation croisée est envisagée uniquement pour les données du SPF Finances car la taille de l'échantillon le permet, contrairement aux données Immoweb. En reprenant l'échantillon prélevé au point 3.1.4, les modèles D1 et D2 sont évalués (sans réduction des résidus). Le modèle D1 est indéterminé, aucun paramètre n'a pu être calculé. Le modèle D2 est validé pour l'ensemble de ces paramètres statistiques. Néanmoins, un grand nombre de variables ne sont plus significatives au t test (Tableau 17 et Annexe 9e).

**Tableau 17. Résumé des indicateurs d'ajustements des modèles D pour la validation-croisée**

	SPF Finances, échantillon de la validation-croisée – Modèle D1	SPF Finances, échantillon de la validation-croisée – Modèle D2
R <sup>2</sup>	Néant	0,5277
R <sup>2</sup> ajusté	Néant	0,5117
Prix moyen	142 434 €	142 434 €
Erreur type	Néant	45 566,10 €
Erreur relative	Néant	32,02 %
AIC	Néant	13379
BIC	Néant	13461
Test F	Néant	F : 32,9612 Prob(>F), (19, 531) : 0,0000*
Indice de multicollinéarité	-999,0000	72,7803
Test de Jarque-Bera	Néant	JB : 33,3544 Prob(> $\chi^2$ ), (2) : 0,0000*
Test de Breusch-Pagan	Néant	BP : 123,5807 Prob(> $\chi^2$ ) (20) : 0,0000*
Test de Koenker-Basset	Néant	KB : 78,6498 Prob(> $\chi^2$ ) (20) = 0,0000*

### 3.5.3 Les valeurs de prix des variables indépendantes

En reprenant les coefficients de régressions par variable, il est possible de rendre compte de l'intensité des variables indépendantes (x) avec le prix. Étant donné l'erreur relative des modèles sur le prix, certaines variables ont une erreur trop importante pour être considérée. Cette erreur est proportionnelle à la significativité statistique des variables. Plus les p-valeur sont élevées et plus l'erreur relative de la variable considérée est grande. Ainsi, il est choisi de définir un seuil de significativité de 0,001 pour les variables à interpréter, les autres présentant une erreur trop grande (Annexe 11a et 11b).

Pour les quelques variables communes aux deux modèles, des comparaisons sont mises en avant. Un ratio est calculé, plus le ratio se rapproche de zéro et plus les valeurs des deux modèles sont proches. Ainsi, les variables les plus éloignées en termes de valeurs sont relatives à la superficie. Les définitions établies entre les deux modèles ne sont pas tout à fait semblables, ce qui explique certainement cette grande variation. Pour les données du cadastre, la superficie du terrain est estimée sur la base de la parcelle cadastrale, tandis que la superficie habitable est estimée par la multiplication de la superficie bâtie et du nombre d'étage. Les variables déterminées de manière identique ont des valeurs très proches (Tableau 18).

**Tableau 18. Coefficient de détermination et son écart entre les modèles D2 et D3**

<b>Variabes</b>	<b>Coefficient – Modèle D2 – SPF Finances</b>	<b>Coefficient – Modèle D3 – Immoweb</b>	<b>Ratio <math>ABS((D2-D3)/(D2+D3))</math></b>
LOG_SURF_H	45 464,60 €	138 355,00 €	0,51
LOG_SURF_F	12 947,20 €	41 372,00 €	0,52
AGE_BIEN	-362,43 €	-370,00 €	0,01
NIV_INSTRU	1 289,20 €	1 206,15 €	0,03

Dans les deux modèles, chaque pourcent de diplômés du supérieur du secteur statistique contribue à une augmentation marginale d'environ 1 200 €. L'âge du bien a une contribution marginale négative sur le prix d'environ 350 €. Dans le cas où le bien est rénové, la moins-value est un peu moins importante autour de 90 €. La relation définie ici est linéaire, il est possible que la relation soit plus complexe comme l'ont présenté Dethier (2012) et Grignet (2011) (Annexe 11b). Par ailleurs, il en est de même pour les autres variables considérées, le choix de la forme linéaire a été préféré pour simplifier le modèle, mais elle ne rend pas toujours compte de la réalité.

Pour les superficies foncières et immobilières, la relation logarithmique nécessite une transformation mathématique. Terra (2005) explique comment déterminer le prix implicite à travers le calcul d'une dérivée première. Pour le logarithme, la dérivée première du logarithme (base 10) est  $\log_{10}(x) = \frac{1}{x \ln(10)}$ , où x est ici la superficie (foncière et habitable). Ensuite, le résultat de la dérivée est multiplié par le coefficient de la régression afin de déterminer la contribution marginale de la superficie pour le bien considéré. Ainsi, pour un bien dont la superficie moyenne du terrain est de 500 m<sup>2</sup>, le prix implicite est autour de 11 €/m<sup>2</sup> pour les données du SPF Finances, tandis qu'il est autour de 36 €/m<sup>2</sup> pour les données Immoweb. De même, pour un bien dont la superficie moyenne habitable est de 150 m<sup>2</sup>, la contribution marginale est de 132 €/m<sup>2</sup> pour les données cadastrales, contre 400 €/m<sup>2</sup> pour les données Immoweb. À nouveau, les différences marquées du coefficient de régression influencent fortement les résultats obtenus pour cette variable.

En ce qui concerne les autres variables, pour le modèle D2 (SPF Finances), quinze variables restent significatives au seuil  $\alpha$  égale à 0,001. Plusieurs variables internes aux biens sont une plus-value pour le bien tel que la présence d'un chauffage central, la présence d'une cave, la présence d'un étage mansardé ou encore une configuration quatre façades.

Une maison sans étage présente une valeur supérieure aux maisons à deux étages. Néanmoins, avoir deux étages n'a pas une influence négative sur le prix du bien. Le nombre de pièce entraîne une augmentation du prix, cette variable est d'une certaine manière liée à la superficie du bien. À l'inverse, l'absence de garage entraîne une moins-value. En termes d'externalités environnementales, une contribution marginale est observée pour les biens à proximité de cours d'eau (entre 100 et 300 m). Il en est de même pour les maisons avec une bonne accessibilité en mode doux. La part de construction datant d'avant 1981 (au niveau du secteur statistique) évolue positivement avec le prix des biens immobiliers. Pour la seconde variable, l'explication supposée à cette augmentation du prix serait qu'une part de constructions plus anciennes, au niveau du secteur statistique, indique généralement les territoires les plus centraux (centre-historique). Finalement, les maisons localisées entre 700 et 1500 m d'une gare ont une perte de valeurs. Il est possible que cette diminution du prix soit liée à la présence récurrente de quartiers marginalisés à proximité des gares. (Annexe 11a).

Pour le modèle D3 (Immoweb), six variables restent significatives à un seuil de  $\alpha$  de 0,001. Les quatre variables communes au résultat du SPF Finances ont déjà été analysées. Pour les autres, une plus-value est observée pour les biens avec plus de deux salles de bains. L'indice du paysage de Kiemstedt affiche également une diminution du prix avec l'indice. Ce qui est

contraire à nos hypothèses supposant qu'un paysage de qualité aurait une plus-value sur le prix. Malgré une significativité de 0,001, cette variable atteint tout de même 30 % d'erreurs, il est possible que sa relation avec le prix ne soit pas claire (Annexe 11b).

Certaines hypothèses n'ont pas trouvé de réponses, le modèle final ne les ayant pas considérées. Cela ne signifie pas pour autant qu'elles n'ont pas d'impact sur le prix. Par exemple, une variable peut avoir un impact localisé, qui à l'échelle du CRU de Liège peut ne pas être significatif d'un point de vue statistique. De plus, la zone d'étude étant conséquente, elle intègre certainement plusieurs sous-marchés, ce qui explique d'une part les maigres résultats en termes prédictifs ainsi que la non-considération de certaines variables.



## Conclusion

Afin de répondre à la question de recherche qui pour rappel est : *La qualité de la base de données cadastrales de l'AGDP est-elle suffisante pour permettre des modélisations hédoniques relatives aux prix des logements ?* Tout d'abord, il apparaît que les données du SPF Finances, en comparaison des données Immoweb, manquent de fiabilité pour générer un modèle hédonique performant. Les modèles générés avec les données Immoweb ont toujours des parts explicatives ( $R^2$ ) plus importantes (Tableau 13, 14 et 15). En considérant les modèles finaux, la part explicative ( $R^2$ ) du modèle du SPF Finances est de 64 %, contre un coefficient de détermination ( $R^2$ ) de 74 % pour le modèle Immoweb (Tableau 16)

Ainsi, le modèle Immoweb est le plus performant en termes de prédiction du prix. Alors que l'ensemble des variables disponibles pour la BD Immoweb n'ont pas été employées dans la régression multiple (Tableau 2). De plus, en considérant l'erreur relative, à nouveau l'erreur est moindre pour les données Immoweb (17 %), là où les données du SPF Finances affichent une erreur relative importante (25 %) en comparaison de l'erreur proche des 10 % attendue par Evans (1990).

Finalement, l'hypothèse générale de la recherche qui supposait que l'échantillonnage des biens supérieurs de la BD du SPF Finances suffirait à produire un modèle plus performant n'est pas vérifiée. Néanmoins, il s'avère que le grand nombre d'enregistrements des données cadastrales permet de conserver une certaine stabilité des paramètres statistiques au cours de la modélisation. Les hypothèses d'hétéroscédasticité, de multicollinéarité ou de normalité des résidus ont rarement été rejetées au cours de la modélisation.

Pour conclure ce mémoire, quelques pistes de solutions peuvent être proposées pour approfondir cette recherche. Premièrement, les variables internes aux biens immobiliers pourraient dans une certaine mesure être complétées (télétection, croisement avec d'autres bases de données, etc.) pour augmenter la part prédictive des modèles. En ce qui concerne les variables externes aux biens immobiliers, d'autres propositions de variables pourraient amener davantage d'explications dans les modèles. Des transformations de variables auraient également pu être envisagées, celles-ci sont peu représentées dans les modèles. Finalement, certaines méthodes de modélisation n'ont pas été envisagées telles que la régression linéaire géographiquement pondérée (GWR), l'utilisation d'une analyse en composante principale (ACP) pour l'implémentation des variables ou encore un découpage de la zone d'étude en sous-zones d'étude davantage comparable pour la modélisation. Ce dernier point est particulièrement intéressant pour rendre compte d'un marché unique.

## Références bibliographiques

- Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–722. <https://doi.org/10.1109/TAC.1974.1100705>
- Anselin, L. & Lozano-Gracia, N. (2009). Part IX Spatial econometrics. Spatial hedonic models. In Mills, T. C. & Patterson, K. (eds), *Palgrave handbook of econometrics. Volume 2: applied econometrics*. London : Palgrave Macmillan, 1213–1250. [https://doi.org/10.1057/9780230244405\\_26](https://doi.org/10.1057/9780230244405_26)
- Bezbakh, P. & Gherardi, S. (2000). *Dictionnaire de l'économie*. Paris : Larousse, "Le Monde", 646 p. <https://gallica.bnf.fr/ark:/12148/bpt6k1200505f/f1.item.zoom>, consulté le 01/02/2022.
- Bianchet, B., Claeys, D. & Hanin, Y. (2019). Atlas des gares de Wallonie. Belgium : CPDT, 804 p. <https://cpdt.wallonie.be/publications/hors-series/hors-series-1>, consulté le 17/08/2022.
- Bianchet, B., Blaffart, M., Claeys, D., Dupont, X., Lambotte, J.-M., Maldague, H. & Van Nngoc, H. (2018). Recherche 8 : urbanisation des nœuds et mixité des fonctions – Rapport final – Décembre 2018 – Rapport scientifique. Belgium : CPDT, 556 p. <https://hdl.handle.net/2268/238270>, consulté le 17/08/2022.
- Bruckmann, L. & Pepe, L. (2021). Document de travail. TP analyse du paysage par grille. Inédit.
- Brunet, R., Ferras, R. & Théry, H. (1993). *Les mots de la géographie, dictionnaire critique (3<sup>e</sup> édition)*. Montpellier-Paris : Reclus – La Documentation Française, 520 p.
- Cavailhès, J., Brossard, T., Foltête, J., Hilal, M., Joly, D., Tourneux, F., Tritz, C. & Wavresky, P. (2009). Chapitre 3 - Valeur des paysages ruraux et localisation résidentielle. In Aubert, F. (eds), *Politiques agricoles et territoires*. Versailles : Éditions Quæ, 69–92. <https://doi.org/10.3917/quae.aube.2009.01.0069>
- Cornet, Y., Daxhelet, D., Halleux, J.-M., Klinkenberg, A.-C., & Lambotte, J.-M. (2005). Cartographie de l'accessibilité par les alternatives à la voiture. In Witlox, F. & Van Acker, V. (eds), *Les Journées Géographiques Belges. Mobilité, Société et Environnement en Cartes*. Zeldate (Belgique) : DCL Print & Sign, 355–364. <https://hdl.handle.net/2268/3620>, consulté le 17/08/2022.
- Cornelis, E., Hubert, M., Huynen, P., Lebrun, K., Patriarche, G., De Witte, A., Creemers, L., Declercq, K., Janssens, D., Castaigne, M., Hollaert, L., & Walle, F. (2012). *La mobilité en Belgique en 2010 : résultats de l'enquête BELDAM*. SPF Mobilité & Transports, 352 p.
- Des Rosiers, F., Thériault, M., & Dubé, J. (2008). La modélisation hédonique des valeurs résidentielles : La mesure des externalités urbaines au Québec. In Des Rosiers, F. & Thériault, M. (eds.), *Information géographique et dynamiques urbaines*. Hermès Science Publishing, 2, 179–208. <https://www.researchgate.net/publication/313053087>, consulté le 17/08/2022.
- Delgado, & Wences, G. (2019). A hedonic approach to the valuation of the effect of criminal violence on housing prices in Acapulco City. *Empirical Economics*, 59(6), 2999–3018. <https://doi.org/10.1007/s00181-019-01804-3>
- Dessouroux, C. & Romainville, A. (2011). La production de logements en Belgique et à Bruxelles – Acteurs, dynamiques, géographie. *EchoGéo*, 15. <https://doi.org/10.4000/echogeo.12279>

- Dethier, P. (2012). Etude comparative entre les préférences révélées et déclarées. Les maisons d'habitation en région liégeoise. Mémoire de master en Sciences Géographiques. Université de Liège, Liège, inédit, 144 p.
- Devillet, G. (2021). Géomarketing, notes de cours. Université de Liège, Faculté des Sciences, Département de Géographie, Liège, inédit.
- Devillet, G., Jaspard, M., & Vazquez Parras, J. (2014). *Atlas du commerce en Wallonie : structures, dynamiques, comportements spatiaux des consommateurs*. Presses Universitaires de Liège.
- Evans, A. W. (1995). The Property Market: Ninety Per Cent Efficient? *Urban Studies (Edinburgh, Scotland)*, 32(1), 5–29. <https://doi.org/10.1080/00420989550013194>
- Esri (s.d.). *Autocorrélation spatiale (Global Moran's I)* [image]. <https://desktop.arcgis.com/fr/arcmap/10.5/tools/spatial-statistics-toolbox/spatial-autocorrelation.htm>, consulté le 08/07/2022.
- Farnir, F., Nyssen, L. & Moyses, E. (2016). *Biostatistique*. [http://www.biostat.ulg.ac.be/pages/Site\\_r/ABC\\_R.html](http://www.biostat.ulg.ac.be/pages/Site_r/ABC_R.html), consulté le 06/06/2021.
- Feuillet, T., Cossart, E. et Commenges, H. (2019). *Manuel de la géographie quantitative : concepts, outils, méthodes*. Armand Colin.
- Gibbons, Mourato, S., & Resende, G. M. (2013). The Amenity Value of English Nature: A Hedonic Price Approach. *Environmental & Resource Economics*, 57(2), 175–196. <https://doi.org/10.1007/s10640-013-9664-9>
- Grignet, Y. (2011). Approche hédonique du marché immobilier sur base d'annonces internet : les maisons d'habitation au sein de la région urbaine de Liège. Mémoire de master en Sciences Géographiques. Université de Liège, Liège, inédit.
- Halleux, J.M. (2019). Marchés fonciers et immobiliers. Notes de cours, Université de Liège, Faculté des Sciences, Département de Géographie, Liège, inédit.
- Halleux, J.-M. (2005). Valuing green structures. The use of hedonic models to assess the influence of green structure on residential property prices. In Werquin, A.-C., Duhem, B., Lindholm, G., Oppermann, B., Pauleit, S. & Tjallingi, S. (Eds.), *Green Structure and Urban Planning*. Bruxelles, Belgium: European Science Foundation, 267–273.
- Hendrickx S., Grandjean M., Marlier C., Zangl S., Ruelle C., Copée P., Defourny P. et Halleux J.-M. (2016). *Systèmes d'informations foncières et politiques publiques, rapport final de la subvention 2015-2016*. CPDT, Rapport de recherche, 188 p.
- Immoweb (2022). *Catégorie de biens immobiliers et fonciers sur Immoweb*. <https://www.immoweb.be/fr>, consulté le 30/01/2022.
- Jawlik, A. A. (2016). R, Multiple R,  $r^2$ ,  $R^2$ , R Square,  $R^2$  Adjusted. (2016). In *Statistics from A to Z: Confusing Concepts Clarified* (p. 274-319). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119272021.ch12>
- Justice belge (2020). 4 février 2020. - Loi portant le livre 3 " Les biens " du Code civil. Moniteur belge, le 17 mars 2020, 15753 p. [https://www.ejustice.just.fgov.be/cgi\\_loi/change\\_lg.pl?language=fr&la=F&cn=2020020416&table\\_name=loi](https://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=fr&la=F&cn=2020020416&table_name=loi), consulté le 28 janvier 2022.

- Kiemstedt, H. (1967). *Zur Bewertung natürlicher Landschaftselemente für die Planung von Erholungsgebieten*. Technische Hochschule Hannover.
- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *The Journal of Political Economy*, 74(2), 132–157. <https://doi.org/10.1086/259131>
- Lancelot, R. & Lesnoff, M. (2005). Sélection de modèles avec l’AIC et critères d’information dérivés. *CIRAD, Montpellier*.
- Le Gallo, J. (2002). Économétrie spatiale : l'autocorrélation spatiale dans les modèles de régression linéaire. *Économie & prévision*, 155, 139–157. <https://doi.org/10.3917/ecop.155.0139>
- Lieske, van den Nouwelant, R., Han, J. H., & Pettit, C. (2021). A novel hedonic price modelling approach for estimating the impact of transportation infrastructure on property prices. *Urban Studies (Edinburgh, Scotland)*, 58(1), 182–202. <https://doi.org/10.1177/0042098019879382>
- Lorquet, T. (2016). La formation de la valeur des biens sur les marchés des appartements à vendre et à louer de Liège. Mémoire de master en Sciences Géographiques. Université de Liège, Liège, inédit, 107 p.
- Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, 23(1), 151–169. <https://doi.org/10.1146/annurev.publhealth.23.100901.140546>
- Luttik, J. (2000). The value of trees, water and open space as reflected by house prices in the Netherlands. *Landscape and Urban Planning*, 48(3/4), 161–167. [https://doi.org/10.1016/S0169-2046\(00\)00039-6](https://doi.org/10.1016/S0169-2046(00)00039-6)
- Naets, T. (2015). Approche hédonique du marché des terrains à bâtir. Modélisation des prix dans la zone de marché de Bruxelles. Mémoire de master en Sciences Géographiques. Université de Liège, Liège, Inédit, 227 p.
- Maldague, H. (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 de master en Sciences Géographiques. Université de Liège, Liège, inédit, 161 p.
- Maslianskaïa-Pautrel, M. (2009). La valorisation de la qualité de l’air par l’approche hédonique : une revue de la littérature. *Revue française d’économie*, 23(3), 109–160. <https://doi.org/10.3406/rfeco.2009.1738>
- Mérenne, E. (2018). *Le dictionnaire de la géographie*. <https://www.fegepro.be/dico/>, consulté le 01/02/2022.
- Moreno, C. (2020). *Droit de cité: de la “ville-monde” à la “ville du quart d’heure”*. Paris : éditions de l’Observatoire.
- Ortiz-Burgos, S. (2016). Shannon-Weaver Diversity Index. In Kennish, M.J. (eds) *Encyclopedia of Estuaries. Encyclopedia of Earth Sciences Series*. Springer, Dordrecht. [https://doi.org/10.1007/978-94-017-8801-4\\_233](https://doi.org/10.1007/978-94-017-8801-4_233)
- Palmquist, R. B. (1992). Valuing localized externalities. *Journal of Urban Economics*, 31, 40–44. [https://doi.org/10.1016/0094-1190\(92\)90032-G](https://doi.org/10.1016/0094-1190(92)90032-G)
- Rodrigue, J.-P. (2020). *The geography of transport systems* (Fifth edition). Routledge. <https://doi.org/10.4324/9780429346323>
- Rahmatian M. & Cockerill, L. (2006). Airport noise and residential housing valuation in southern California: A hedonic pricing approach. *International Journal of Environmental Science and Technology (Tehran)*, 1(1). <https://doi.org/10.1007/BF03325812>

- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55. <https://doi.org/10.1086/260169>
- IGEAT (2010). OSIRIS : Organisation of Statistical Information for Research by Institutions and Scientists. Rapport de recherche – Juin 2010. 309 p.
- Shannon, C. E. & Weaver, W. (1949). The Mathematical Theory of Communication. *Urbana: University of Illinois Press*.
- Schmitz, S. (2020). Analyse et gestion des paysages. Notes de cours, Université de Liège, Faculté des Sciences, Département de Géographie, Liège, inédit.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- SPW (2022). Direction de l'Aménagement opérationnel. Les sites à réaménager (SAR). [http://lampspw.wallonie.be/dgo4/site\\_amenagement/index.php/site/directions/dao/sar](http://lampspw.wallonie.be/dgo4/site_amenagement/index.php/site/directions/dao/sar), consulté le 14/07/2022.
- Terra, S. (2005). Guide de bonnes pratiques pour la mise en œuvre de la méthode des prix hédoniques - Série méthode 05-M01. Paris : Ministère de l'Environnement, 35 p. <http://temis.documentation.developpement-durable.gouv.fr/document.html?id=Temis-0052985>, consulté le 30/01/2022.
- TGV Station Liège-Guillemins. (2010). News TGV Station Liege-Guillemins. In *Steel Construction : Design and Research*, 3(1), 33–33. Wiley. <https://doi.org/10.1002/stco.201090001>
- Tobler, W. (2004). On the First Law of Geography: a reply. *Annals of the Association of American Geographers*, 94(2), 304–310. <https://doi.org/10.1111/j.1467-8306.2004.09402009.x>
- Torre, A. (2011). Analyse économique du paysage et relations de proximités : de l'oubli au conflit. *Projets de Paysage*, 5(5). <https://doi.org/10.4000/paysage.22003>
- Travers, M., Appere, G., & Larue, S. (2013). Évaluation des aménités urbaines par la méthode des prix hédoniques : une application au cas de la ville d'Angers. *Economie & statistique*, 460(1), 145–163. <https://doi.org/10.3406/estat.2013.10203>
- Van Der Haegen, H., Van Hecke, E. & Juchtmans, G. (1996). Les Régions urbaines belges 1991. Etudes Statistiques. *Bruxelles, Institut National de Statistique*, 104, 3–42.
- Vanderstraeten, L. & Van Hecke, E. (2019). Les régions urbaines en Belgique. *BELGEO (Leuven)*, 1(1). <https://doi.org/10.4000/belgeo.32246>
- Van Dijk, D., Siber, R., Brouwer, R., Logar, I., & Sanadgol, D. (2016). Valuing water resources in Switzerland using a hedonic price model. *Water Resources Research*, 52(5), 3510–3526. <https://doi.org/10.1002/2015WR017534>
- Van Hecke, E., Halleux, J.-M., Decroly, J.-M., & Merenne-Schoumaker, B. (Eds.). (2009). *Noyaux d'habitat et régions urbaines dans une Belgique urbanisée. Enquête Socio-économique 2001. Monographies (n°9)*. Bruxelles, Belgium: SPF Economie, PME, Classes moyennes et Energie. <https://hdl.handle.net/2268/62469>
- Willis, A., Gjersoe, N., Havard, C., Kerridge, J., & Kukla, R. (2004). Human Movement Behaviour in Urban Spaces: Implications for the Design and Modelling of Effective *Pedestrian Environments*. *Environment and Planning. B, Planning & Design.*, 31(6), 805–828. <https://doi.org/10.1068/b3060>

Winke., T. (2017). The impact of aircraft noise on apartment prices: a differences-in-differences hedonic approach for Frankfurt, Germany. *Journal of Economic Geography*, 17(6), 1283–1300. <https://doi.org/10.1093/jeg/lbw040>

Wu, J. (2004). Effects of changing scale on landscape pattern analysis: Scaling relations. *Landscape Ecology*, 19(2), 125–138. <https://doi.org/10.1023/B:LAND.0000021711.40074.ae>

## Ressources statistiques et cartographiques

ACNAW (S.D.). *Limites de zones des PDLT et PEB*. <https://acnaw.be/home/politique-environnementale/introduction/limites-de-zones-des-pdlt-et-peb.html>, consulté le 16/07/2022.

Anselin, L. (2020). *GeoDa Documentation. An introduction to spatial data analysis*. <http://geodacenter.github.io/documentation.html>, consulté le 08/08/2022.

EEA (2020). *Corine Land Cover 2012*. <https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012?tab=metadata>, consulté le 14/07/2022.

Esri (2018). *Interprétation des résultats des moindres carrés ordinaires*. [https://desktop.arcgis.com/fr/arcmap/10.4/tools/spatial-statistics-toolbox/interpreting-ols-results.htm?fbclid=IwAR3sOI7SHx885x5A00Whmy37mGqAMgi1pm0kQYrC5udwsLXW\\_pn64n4i\\_EE](https://desktop.arcgis.com/fr/arcmap/10.4/tools/spatial-statistics-toolbox/interpreting-ols-results.htm?fbclid=IwAR3sOI7SHx885x5A00Whmy37mGqAMgi1pm0kQYrC5udwsLXW_pn64n4i_EE), consulté le 08/06/2022.

Geofabrik & OpenStreetMap Contributors (2021). *Download OpenStreetMap data for this region: Belgium*. <http://download.geofabrik.de/europe/belgium.html>, consulté le 01/02/2022.

QGIS Development Team (2021). *QGIS Geographic Information System. Open Source Geospatial Foundation Project*. <http://qgis.osgeo.org>, consulté le 13/08/2022.

IGN (2014). *CORINE Land Cover - Belgique – 2012*. <https://www.geo.be/catalog/details/bcd19aa9-c320-4116-971b-6e4376137f13?l=fr>, consulté le 10/07/2022.

Insee – Eurostat (2018). *Insee méthodes n°131 – Manuel d'analyse spatiale. Théorie et mise en œuvre pratique avec R*. <https://www.insee.fr/fr/information/3635442>, consulté le 13/08/2022.

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>, consulté le 13/08/2022.

SPF Economie (2018). *Secteurs statistiques 2011-2017*. <https://statbel.fgov.be/fr/open-data/secteurs-statistiques-2011-2017>, consulté le 07/07/2022.

SPF Economie (2013). *Répartition de la population par cellule carrée de 1km<sup>2</sup> en 2011*. <https://www.geo.be/catalog/details/ea430cb8-2805-4c67-8392-9e2a47c1ef55?l=fr>, consulté le 10/07/2022.

SPF Economie (2011). *Census 2011 – Belgique*. [https://census2011.fgov.be/download/statsect\\_fr.html](https://census2011.fgov.be/download/statsect_fr.html), consulté le 01/02/2022.

SPF Economie (2011). *Population par secteur statistique, 2011*. <https://statbel.fgov.be/fr/open-data/population-par-secteur-statistique>, consulté le 01/02/2022.

SPF Economie (2011). *Revenus fiscaux par secteur statistique en 2011*. <https://statbel.fgov.be/fr/open-data/statistique-fiscale-des-revenus-par-secteur-statistique>, consulté le 01/02/2022.

SPF Finances (2021). *Unité administrative et plan parcellaire cadastral*. <https://eservices.minfin.fgov.be/myminfin-web/pages/cadastral-plans>, consulté le 01/02/2022.

- SPW (2022). Site à réaménager de droit (SAR). <https://geoportail.wallonie.be/catalogue/8e2cf269-cedb-4fbe-943b-41a7587490e3.html>, consulté le 14/02/2022.
- SPW (2020). *Occupation du sol en Wallonie – WALOUS 2018*. <https://geoportail.wallonie.be/catalogue/a0ad23a1-1845-4bd5-8c2f-0f62d3f1ec75.html>, consulté le 01/02/2022.
- SPW (2018). *Bruit des axes ferroviaires dans les grandes agglomérations wallonnes - Rapportage 2012 – Série*. <https://geoportail.wallonie.be/catalogue/76941c77-2a18-4d0a-928a-994294e7847a.html>, consulté le 14/07/2022.
- SPW (2018). *Bruit des axes routiers dans les grandes agglomérations wallonnes - Rapportage 2012 – Série*. <https://geoportail.wallonie.be/catalogue/ed9077c2-d5b8-4f57-99db-51aedb5662fc.html>, consulté le 14/07/2022.
- SPW (2018). *Bruit de l'industrie dans les grandes agglomérations wallonnes - Rapportage 2012 – Série*. <https://geoportail.wallonie.be/catalogue/491c68e2-46d1-4e44-8602-8ad05b40fc49.html>, consulté le 14/07/2022.
- SPW (2015). *Relief de la Wallonie – Modèle Numérique de Terrain (MNT) 2013-2014*. <https://geoportail.wallonie.be/catalogue/6029e738-f828-438b-b10a-85e67f77af92.html>, consulté le 01/02/2022.
- SPW (2008). *Occupation et utilisation du sol en Wallonie – COSW 2007 - Série - Donnée historique*. <https://geoportail.wallonie.be/catalogue/290e1fe8-0d99-410e-967b-a02f389b931a.html>, consulté le 10/07/2022.
- SPW (2008). *Bruit des grands axes ferroviaires en 2006 en Wallonie – Série*. <https://geoportail.wallonie.be/catalogue/349a9e95-5cd5-49b0-8f7c-5d108140dbba.html>, consulté le 14/07/2022.
- SPW (2008). *Bruit des grands axes routiers en 2006 en Wallonie – Série*. <https://geoportail.wallonie.be/catalogue/031dd889-617f-4114-8ef3-58982a15ed18.html>, consulté le 14/07/2022.

## Annexes

### Annexe 1. Description des données contenues dans la BD cadastrales du SPF Finances

Nom	Description
ID_BIEN	Numéro d'identification unique d'un bien.
ID_SITUATION	Numéro d'identification de la situation du bien.
X_Lambert7	Coordonnée X en Lambert belge 1972, en mètre.
Y_Lambert7	Coordonnée Y en Lambert belge 1972, en mètre.
DIVISION_C	Champs correspondant à la division cadastrale, c'est-à-dire une partie du territoire qui correspondait aux territoires communaux (avant la fusion de 1977), puis a évolué de façon à devenir des éléments distincts du territoire.
SECTION	Champs correspondant aux sections cadastrales, cette portion du territoire est une subdivision des divisions cadastrales.
RADICAL	Champs correspondant au radical est un numéro unique attribué à chaque parcelle d'une section sur le plan cadastral primitif de 1835. -ex. : 328
NUMERO_BIS	Champs correspondant au numéro bis est destiné aux parcelles initialement non cadastrées car appartenant au domaine public. La nouvelle parcelle n'a pas de numéro au plan cadastral primitif de 1835. Au radical est ajouté un « / », puis un numéro nouvellement défini par rapport à la parcelle adjacente. -ex. : 328/02
EXPOSANT_L	Champs correspondant à la puissance Lettre, la lettre suit le radical et indique les modifications physiques successives d'un bien cadastral. -ex. : 328B est un bien qui a subi 2 modifications au cadastre.
EXPOSANT_C	Champs correspondant à l'Exposant Chiffre, un chiffre est ajouté à la puissance Lettre quand le nombre de modifications a dépassé le nombre de lettres de l'alphabet. -ex. : 328A2, la puissance lettre est remise à A de tel sorte que 328Z précède 328A2.
NUMERO_PAR	Numéro de la parcelle cadastrale.
CAPAKEY	Identifiant unique d'une parcelle cadastrale constitué du radical, du numéro bis, de la lettre (puissance) et d'un chiffre (exposant).
CODE_NATUR	Identifiant de la variable NATURE.
NATURE	Type d'usage prévu sur la parcelle. <u>Attention</u> : l'usage prévu peut différer de l'usage réel.
TYPE_NATUR	Indique si une parcelle cadastrale est bâtie ou non.
CATEGORIE_	Agrégation du champ NATURE en catégorie spécifique.
Shape_Leng	Périmètre de la parcelle cadastrale, en mètre.
Shape_Area	Superficie de la parcelle cadastrale, en mètre carré.
ID_ACTE	Numéro d'identification de l'acte de vente.
DATE_ACTE	Date de l'acte de vente (Année/Mois/Jour).



YEAR_ACTE	Année de l'acte de vente. <u>Rem</u> : les ventes varient entre 2007 et 2015.
MONTANT	Prix de vente (hors TVA et hors frais notariaux).
SURF_ACTE	Superficie renseignée dans l'acte de vente. <u>Rem</u> : Cette donnée manque de précision car elle renseigne parfois la superficie du bâti, parfois la superficie du terrain. Elle peut également considérer un ensemble de biens lors de la vente de lots.
INDICE	Identifiant de la variable Descr_indi.
DESCR_Indi	Classification du bâti.
TYPE*	Indique la configuration des maisons : mitoyenne (A), semi-mitoyenne (B) ou isolée (C).
NBRE_ETAGE	Nombre d'étages (inclut le rez-de-chaussée).
MANSARDE	Indique si le dernier niveau est mansardé ou non.
ANNEE_CONS	Année de construction. Disponible annuellement à partir de 1931, puis agrégation par période 1919-1930, 1900-1918, 1875-1899, 1850-1874 et antérieur à 1850.
ANNEE_MODI	Année de la dernière modification physique ayant nécessité une demande de permis.
NBRE_GARAG	Nombre de garages (manque de précision sur la considération des carports et/ou des emplacements privés)
CHAUF_CENT*	Indique la présence d'un chauffage central et/ou d'un système de climatisation.
NBRE_SDB*	Nombre de salles de bains.
NBRE_LOGEM	Nombre de logements au sens du cadastre. Correspond à l'ensemble des pièces pouvant servir de logements potentiels.
NBRE_PIECE	Nombre de pièces standards au sens du cadastre. Par exemple, les grandes pièces sont découpées en unités s'apparentant à plusieurs pièces standards.
SURF_BATI	Superficie construite au sol.

Les champs avec un astérisque sont ceux n'ayant pas été actualisés sur l'ensemble du parc de logements.

## Annexe 2. Source des données utilisées pour la construction des variables

Domaine	Description	Emprise géographique	Format de données	Fournisseur
Bâtiment	Localisation et relevé des caractéristiques des maisons d'habitation par Immoweb de Grignet (2011)	Tout le complexe résidentiel urbain de Liège	Fichier .shp, .dbf et .shx Format vecteur	Perinne Dethier (2022)
Administratif	Secteurs statistiques 2011-2017	Toute la Belgique	Fichier .shp, .dbf et .shx Format vecteur	Service Public Fédéral Belge – SPF Economie (2018)
Administratif	Sites à réaménager de droit (SAR)	Toute la province de Liège	Fichier .shp, .dbf et .shx Format vecteur	Service Public de Wallonie – SPW (2022)
Commerce	Localisation des points de ventes en 2019	Toute la Wallonie	Fichier .shp, .dbf et .shx Format vecteur	SEGEFA
Commerce	Localisation des nodules commerciaux en 2019	Toute la Wallonie	Fichier .shp, .dbf et .shx Format vecteur	SEGEFA
Occupation du sol	CORINE Land Cover 2012 – Résolution spatiale 100 m	Toute l'Europe	Fichier .tiff Format image	<i>European Environment Agency</i> – EEA (2020)
Population	Population par secteur statistique en 2011	Toute la Belgique	Fichier .xlsx Format tableau	Service Public Fédéral Belge – SPF Economie (2011)
Population	Répartition de la population par cellule carrée de 1km <sup>2</sup> en 2011	Toute la Belgique	Fichier .gml Format vecteur	Service Public Fédéral Belge – SPF Economie (2013)
Socio-économie	Revenus fiscaux par secteur statistique en 2011	Toute la Belgique	Fichier .xlsx Format tableau	Service Public Fédéral Belge – SPF Economie (2011)

Socio-économie	Census 2011 – Population scolaire et niveau d’instruction – Population de 18 ans et plus, ne suivant plus d’enseignement selon le lieu d’habitation, le sexe et le niveau d’instruction – Secteurs statistiques	Toute la Belgique	Fichier .xlsx Format tableau	Direction générale Statistique - Statistics Belgium – SPF Economie (2014)
Socio-économie	Census 2011 – Marché du travail – Population active occupée <sup>8</sup> selon le sexe, le lieu de travail et le statut professionnel – Population totale – Secteurs statistiques	Toute la Belgique	Fichier .xlsx Format tableau	Direction générale Statistique - Statistics Belgium – SPF Economie (2014)
Logement	Census 2011 – Logements – Logements classiques selon la période de construction – Secteurs statistiques	Toute la Belgique	Fichier .xlsx Format tableau	Direction générale Statistique - Statistics Belgium – SPF Economie (2014)
Topographie	Relief de la Wallonie – MNT 2013-2014	Toute la Wallonie	Fichier .tiff Format image	Service Public de Wallonie – SPW Géomatique (2015)
Transport	Part modale des modes lents au lieu de résidence et au lieu de travail	Toute la Wallonie	Fichier .adf Format image	Lepur (Cornet <i>et al.</i> , 2005)
Transport	Part modale du bus au lieu de résidence et au lieu de travail	Toute la Wallonie	Fichier .adf Format image	Lepur (Cornet <i>et al.</i> , 2005)
Transport	Part modale du train au lieu de résidence et au lieu de travail	Toute la Wallonie	Fichier .adf Format image	Lepur (Cornet <i>et al.</i> , 2005)
Transport	Localisation des arrêts TEC et leurs fréquences	Toute la Wallonie	Fichier .shp, .dbf et .shx Format vecteur	Lepur (SRTW, 2018 et 2015)

---

<sup>8</sup> Fréquence absolue de toutes les personnes disposant d’un emploi pour une aire géographique donnée.

Transport	Localisation des gares de trains et leurs fréquences	Toute la Wallonie	Fichier .shp, .dbf et .shx Format vecteur	Lepur (SRTW, 2018 et 2017)
Transport	Réseau routier rastérisé à une résolution de 100x100 m	Toute la Belgique	Fichier .tif Format image	SEGEFA
Bruit	Bruit des grands axes ferroviaires en 2006 en Wallonie - Série	Toute la Wallonie	Fichier .gml Fichier vecteur	Service Public de Wallonie – SPW (2008)
Bruit	Bruit des grands axes routiers en 2006 en Wallonie – Série	Toute la Wallonie	Fichier .gml Fichier vecteur	Service Public de Wallonie – SPW (2008)
Bruit	Bruit des axes ferroviaires dans les grandes agglomérations wallonnes - Rapportage 2012 – Série.	Toute la commune de Liège et de Charleroi	Fichier .gml Fichier vecteur	Service Public de Wallonie – SPW (2018)
Bruit	Bruit des axes routiers dans les grandes agglomérations wallonnes - Rapportage 2012 – Série.	Toute la commune de Liège et de Charleroi	Fichier .gml Fichier vecteur	Service Public de Wallonie – SPW (2018)
Bruit	Bruit de l'industrie dans les grandes agglomérations wallonnes - Rapportage 2012 – Série.	Toute la commune de Liège	Fichier .gml Fichier vecteur	Service Public de Wallonie – SPW (2018)
Bruit	Zone PDLT – Liège Airport	Toute la province de Liège	Fichier .shp, .dbf et .shx Fichier vecteur	ACNAW (S.D.)

### Annexe 3. Script python utilisé pour le calcul des variables

#### Annexe 3a. Script python utilisé pour le calcul de l'accessibilité potentielle

```
# Name : CostDistance
import arcpy
from arcpy import env
from arcpy.sa import *
arcpy.CheckOutExtension("Spatial")
workspace = "D:\\\\DATA\\ACC_EMPLOI_RASTER\\"
inCostRaster = "D:\\\\DATA\\MATRICE\\costmatrice2.tif"
maxDistance = 36000000
i = 0
while(i <= 324):
    inSourceData = workspace +str(i)+".shp"
    outCostDistance = CostDistance(inSourceData, inCostRaster, maxDistance)
    outCostDistance.save(workspace +str(i)+"_CD.tif")
    i += 1

# Division des distance-temps par 600 000
import arcpy
from arcpy import sa
from arcpy.sa import *
workspace = "D:\\\\DATA\\ACC_EMPLOI_RASTER\\"
arcpy.env.extent = "MAXOF"
arcpy.env.extent = arcpy.Extent(193223, 92468, 295133, 167660)
arcpy.env.extent = "193223, 92468, 295133, 167660"
i = 0
while(i <= 324):
    A = workspace+str(i)+"_CD.tif"
    B = 600 000
    outputRaster = arcpy.sa.Divide(A, B)
    outputRaster.save(workspace+str(i)+"POND_CD.tif")
    i += 1

# Name: Buffer
import arcpy
from arcpy import env
workspace = "D:\\\\DATA\\ACC_EMPLOI_RASTER\\"
i = 0
while(i <= 324):
```

```

inFeature = workspace +str(i)+".shp"
outFeature = workspace +str(i)+"_buff.shp"
distance = "500 Kilometers"
arcpy.Buffer_analysis(inFeature, outFeature, distance)
i += 1

```

#### **# Name: Conversion polygon To Raster**

```

import arcpy
from arcpy.sa import *
arcpy.CheckOutExtension("Spatial")
workspace = "D:\\DATA\\ACC_EMPLOI_RASTER\\"
arcpy.env.extent = "MAXOF"
arcpy.env.extent = arcpy.Extent(193223, 92468, 295133, 167660)
arcpy.env.extent = "193223, 92468, 295133, 167660"
cellSize = 100
valueField = "PopActiOcc"
i = 0
while(i <= 324):
    inFeature = workspace +str(i)+"_buff.shp"
    outRasterDataset = workspace + str(i)+"_CST_EMPLOI.tif"
    arcpy.PolygonToRaster_conversion(inFeature, valueField, outRasterDataset, "", "", cellSize)
    i += 1

```

#### **# Division du raster constant par la distance-temps**

```

import arcpy
from arcpy import sa
from arcpy.sa import *
arcpy.CheckOutExtension("spatial")
workspace = "D:\\DATA\\ACC_EMPLOI_RASTER\\"
arcpy.env.extent = "MAXOF"
arcpy.env.extent = arcpy.Extent(193223, 92468, 295133, 167660)
arcpy.env.extent = "193223, 92468, 295133, 167660"
i = 0
while(i <= 324):
    A = workspace+str(i)+"_CST_EMPLOI.tif"
    B = workspace+str(i)+"_CD.tif"
    outputRaster = arcpy.sa.Divide(A, B)
    outputRaster.save(workspace+str(i)+"POND_CD.tif")
    i += 1

```

### Annexe 3b. Script python utilisé dans le calcul de l'indice de diversité de Shannon

#### #Buffer

```
import arcpy
from arcpy import env
arcpy.env.parallelProcessingFactor = "100%"
workspace = "D:\\DATA\\ACC_EMPLOI_RASTER\\"
i = 0
while(i <= 5805):
    inFeature = "D:\\A\\"+str(i)+".shp"
    outFeature = "D:\\A\\BUFF_300m\\"+str(i)+"_buff300.shp"
    distance = "300 meters" #300, 500 et 700 m
    arcpy.Buffer_analysis(inFeature, outFeature, distance)
    i += 1
```

#### #Tabulate Area

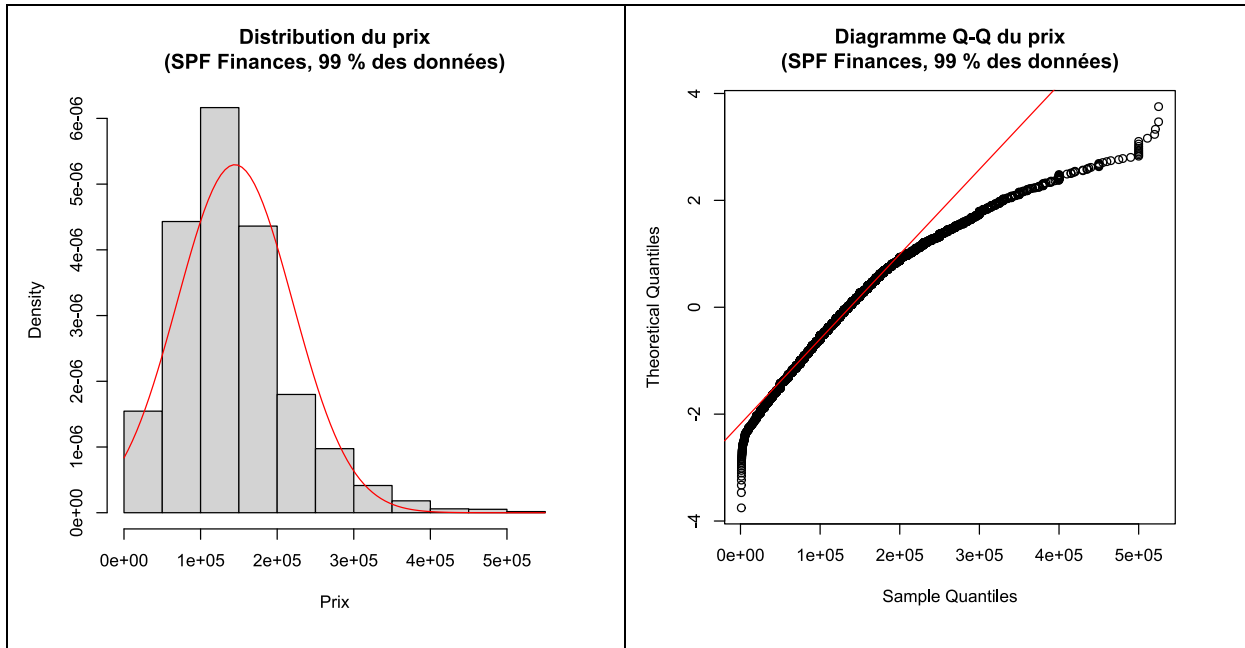
```
import arcpy
from arcpy import env
from arcpy.sa import *
workspace = "D:\\DATA\\SHANNON_INDEX\\"
i = 0
inClassData = workspace+"CLC_2012LG.tif"
classField = "Value"
zoneField = "ID_BIEN"
cellSize = 100
while(i <= 5805):
    inZoneData = "D:\\A\\BUFF_300m_G\\"+str(i)+".shp" #300, 500 et 700 m
    outTable = "D:\\A\\MER_300m_G\\"+str(i)+"_TAB300.dbf"
    TabulateArea(inZoneData, zoneField, inClassData, classField, outTable, cellSize)
    i += 1
```

#### Annexe 4. Pondération et légende relative à la couche d'occupation CLC de 2012

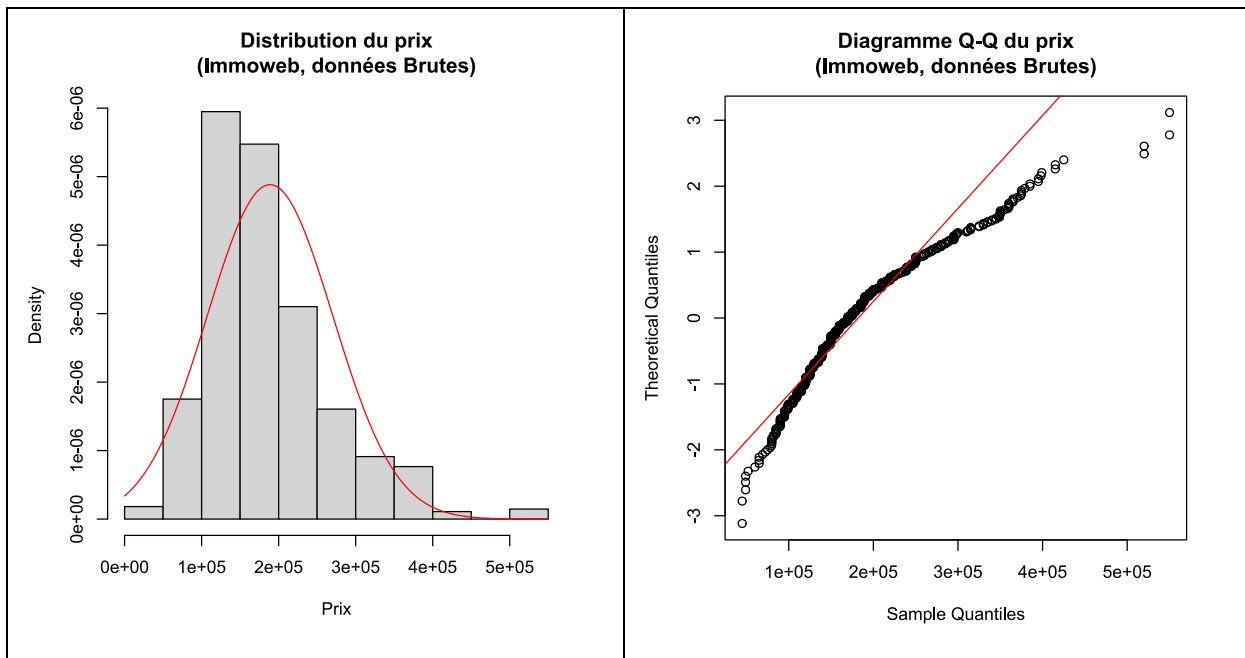
Type d'occupation du sol et description (IGN, 2014)	Poids attribués
111 – <i>Continuous urban fabric</i>	1
112 – <i>Discontinuous urban fabric</i>	0,5
121 – <i>Industrial or commercial units</i>	0
122 – <i>Road and rail networks and associated land</i>	0
123 – <i>Port areas</i>	0
124 – <i>Airports</i>	0
131 – <i>Mineral extraction sites</i>	1
132 – <i>Dump sites (décharge)</i>	0
141 – <i>Green urban areas</i>	1,5
142 – <i>Sport and leisure facilities</i>	0,5
211 – <i>Non-irrigated arable land</i>	0,5
222 – <i>Fruit trees and berry plantations</i>	2
231 – <i>Pastures</i>	1
242 – <i>Complex cultivation patterns</i>	2
243 – <i>Land principally occupied by agriculture</i>	1
311 – <i>Broad-leaved forest (forêt de feuillus)</i>	2,5
312 – <i>Coniferous forest</i>	1,5
313 – <i>Mixed forest</i>	3
321 – <i>Natural grassland</i>	2
322 – <i>Moors and heathland (landes et bruyères)</i>	1,5
324 – <i>Transitional woodland shrub</i>	2
412 – <i>Peat bogs (tourbière)</i>	2
511 – <i>Water courses</i>	5
512 – <i>Water bodies</i>	5



**Annexe 5. Représentations graphiques de la distribution du prix**



**Annexe 5a. Histogramme et diagramme Q-Q du prix pour les données à 99 % du SPF Finances**



**Annexe 5b. Histogramme et diagramme Q-Q du prix pour les données brutes d'Immoweb**

## Annexe 6. Statistiques descriptives de l'ensemble des variables

### Annexe 6a. Statistiques descriptives de l'ensemble des variables (SPF Finances 95 %)

	n	mean	sd	median	min	max	range	skew	kurtosis	se
PRIX	5502	142444,4680	63836,6623	135000,0000	20500,0000	335000,0000	314500,0000	0,6125	0,0798	860,6169
SURF_FONC	5502	483,2131	666,1320	251,3743	4,8929	12915,9937	12911,1008	5,1866	53,8589	8,9805
LOG_SURF_F	5502	2,4444	0,4454	2,4003	0,6896	4,1111	3,4216	0,2368	-0,2032	0,0060
LONG_FONC	5502	97,0057	55,8399	84,5922	8,9432	495,1696	486,2264	1,8233	5,5805	0,7528
SURF_BATI	5502	94,5482	54,0131	84,0000	4,0000	1440,0000	1436,0000	5,4706	90,3863	0,7282
LOG_SURF_B	5502	1,9250	0,2075	1,9243	0,6021	3,1584	2,5563	-0,0556	1,6919	0,0028
SURF_HAB	5502	174,8595	112,5819	148,0000	4,0000	2880,0000	2876,0000	5,9645	90,6160	1,5178
LOG_SURF_HAB	5502	2,1851	0,2172	2,1703	0,6021	3,4594	2,8573	0,1160	2,0619	0,0029
NB_FACADE	5502	2,7288	0,7996	3,0000	2,0000	4,0000	2,0000	0,5258	-1,2463	0,0108
N2_FACADE	5502	0,4920	0,5000	0,0000	0,0000	1,0000	1,0000	0,0320	-1,9993	0,0067
N3_FACADE	5502	0,2872	0,4525	0,0000	0,0000	1,0000	1,0000	0,9406	-1,1156	0,0061
N4_FACADE	5502	0,2208	0,4148	0,0000	0,0000	1,0000	1,0000	1,3457	-0,1892	0,0056
NB_ETAGE	5502	1,9160	0,5789	2,0000	1,0000	4,0000	3,0000	0,2245	0,8080	0,0078
N0_ETAGE	5502	0,2059	0,4044	0,0000	0,0000	1,0000	1,0000	1,4541	0,1143	0,0055
N1_ETAGE	5502	0,6792	0,4668	1,0000	0,0000	1,0000	1,0000	-0,7676	-1,4110	0,0063
N2_ETAGE	5502	0,1149	0,3189	0,0000	0,0000	1,0000	1,0000	2,4150	3,8330	0,0043
MANSARDE	5502	0,2534	0,4350	0,0000	0,0000	1,0000	1,0000	1,1338	-0,7146	0,0059
AGE_BIEN	5502	79,9093	39,4390	81,0000	1,0000	161,0000	160,0000	0,2164	-0,4659	0,5317
AGE_RENO	5502	61,3119	43,5026	55,0000	0,0000	161,0000	161,0000	0,5180	-0,6203	0,5865
RENOVATION	5502	0,2530	0,4348	0,0000	0,0000	1,0000	1,0000	1,1360	-0,7096	0,0059

NB_GARAGE	5502	0,5834	0,6634	1,0000	0,0000	11,0000	11,0000	2,2467	19,6933	0,0089
N0_GARAGE	5502	0,4826	0,4997	0,0000	0,0000	1,0000	1,0000	0,0698	-1,9955	0,0067
N1_GARAGE	5502	0,4660	0,4989	0,0000	0,0000	1,0000	1,0000	0,1362	-1,9818	0,0067
N2_GARAGE	5502	0,0514	0,2209	0,0000	0,0000	1,0000	1,0000	4,0604	14,4896	0,0030
NB_SDB	5502	0,7732	0,5693	1,0000	0,0000	11,0000	11,0000	2,2217	30,0253	0,0077
N0_SDB	5502	0,2706	0,4443	0,0000	0,0000	1,0000	1,0000	1,0323	-0,9346	0,0060
N1_SDB	5502	0,7003	0,4582	1,0000	0,0000	1,0000	1,0000	-0,8741	-1,2361	0,0062
N2_SDB	5502	0,0291	0,1680	0,0000	0,0000	1,0000	1,0000	5,6036	29,4057	0,0023
CHAUF_CENT	5502	0,5256	0,4994	1,0000	0,0000	1,0000	1,0000	-0,1026	-1,9898	0,0067
NB_LOGEMEN	5502	1,0892	0,5086	1,0000	1,0000	14,0000	13,0000	9,9750	156,2029	0,0069
N1_LOGEMEN	5502	0,9527	0,2122	1,0000	0,0000	1,0000	1,0000	-4,2663	16,2042	0,0029
N2_LOGEMEN	5502	0,0242	0,1536	0,0000	0,0000	1,0000	1,0000	6,1945	36,3789	0,0021
N3_LOGEMEN	5502	0,0140	0,1175	0,0000	0,0000	1,0000	1,0000	8,2723	66,4435	0,0016
N4_LOGEMEN	5502	0,0091	0,0949	0,0000	0,0000	1,0000	1,0000	10,3436	105,0099	0,0013
NB_PIECE	5502	5,7323	2,0651	5,0000	1,0000	40,0000	39,0000	2,5952	20,8497	0,0278
CAVE	5502	0,7477	0,4344	1,0000	0,0000	1,0000	1,0000	-1,1405	-0,6995	0,0059
CAVE_VIE	5502	0,0200	0,1400	0,0000	0,0000	1,0000	1,0000	6,8566	45,0211	0,0019
CITE	5502	0,0605	0,2385	0,0000	0,0000	1,0000	1,0000	3,6850	11,5816	0,0032
VILLA	5502	0,0327	0,1779	0,0000	0,0000	1,0000	1,0000	5,2522	25,5901	0,0024
BUNGALOW	5502	0,0800	0,2713	0,0000	0,0000	1,0000	1,0000	3,0962	7,5876	0,0037
BEL_ETAGE	5502	0,0522	0,2224	0,0000	0,0000	1,0000	1,0000	4,0270	14,2195	0,0030
ACC_EMPLOI	5502	10092,5871	4432,2350	9897,9900	2477,0400	23408,1000	20931,0600	0,3387	-0,7758	59,7534
ISO_GUILLE	5502	12,0067	7,0563	10,5822	0,2200	38,1715	37,9515	0,8752	0,3930	0,0951

ISO_STLAMB	5502	12,8940	7,4971	11,1151	0,6422	41,9112	41,2690	1,0122	0,8346	0,1011
ISO_NEV	5502	24,9008	14,3083	21,3892	4,6436	80,0663	75,4227	1,0121	0,7439	0,1929
ACC_BUS	5502	5,3835	3,0800	4,3195	2,4317	23,2662	20,8345	1,4396	2,2570	0,0415
ACC_ML	5502	9,4222	4,1035	9,3281	0,2315	22,3933	22,1618	0,3000	-0,1121	0,0553
ACC_TRAIN	5502	4,7684	2,6742	3,8534	2,1992	19,4765	17,2773	1,8709	4,3817	0,0361
ACC_TOT	5502	19,5742	7,4069	19,0002	5,4163	52,0326	46,6163	0,7244	0,6471	0,0999
ACC_POT_TR	5502	293,8454	113,7276	271,8600	103,4400	2062,8000	1959,3600	2,4413	16,3232	1,5332
ACC_POT_BU	5502	24,4886	16,0385	21,3285	4,6764	135,2190	130,5426	1,2636	1,9366	0,2162
D300_GARE	5502	0,0418	0,2002	0,0000	0,0000	1,0000	1,0000	4,5775	18,9574	0,0027
D500_GARE	5502	0,1529	0,3599	0,0000	0,0000	1,0000	1,0000	1,9289	1,7209	0,0049
D700_GARE	5502	0,0509	0,2198	0,0000	0,0000	1,0000	1,0000	4,0859	14,6972	0,0030
D1500_GARE	5502	0,2572	0,4371	0,0000	0,0000	1,0000	1,0000	1,1108	-0,7663	0,0059
GARE_ED	5502	2329,8682	1930,4541	1769,1800	0,0000	14560,2000	14560,2000	1,8085	4,4159	26,0255
D100_BUS	5502	0,3929	0,4884	0,0000	0,0000	1,0000	1,0000	0,4383	-1,8083	0,0066
D300_BUS	5502	0,4384	0,4962	0,0000	0,0000	1,0000	1,0000	0,2483	-1,9387	0,0067
D500_BUS	5502	0,1118	0,3151	0,0000	0,0000	1,0000	1,0000	2,4635	4,0696	0,0042
D700_BUS	5502	0,0262	0,1597	0,0000	0,0000	1,0000	1,0000	5,9343	33,2220	0,0022
BUS_ED	5502	208,9235	245,2543	141,4210	0,0000	3337,6600	3337,6600	5,4392	45,1665	3,3064
RMD_2011	5502	20516,6011	4209,3913	20147,0000	0,0000	42765,0000	42765,0000	0,2321	2,5798	56,7491
TAUX_CHOMA	5502	14,4394	7,2041	13,5294	0,0000	66,6667	66,6667	0,7539	0,8725	0,0971
NIV_INSTRU	5502	22,4321	10,1488	20,5782	0,0000	64,1975	64,1975	0,7464	0,2688	0,1368
D_POP	5502	3366,9623	2860,6705	2642,1079	3,5014	18851,9758	18848,4744	1,8236	5,7344	38,5663
POP_PIXEL	5502	2465,0534	2244,3360	1780,0000	0,0000	10752,0000	10752,0000	1,3201	1,6815	30,2571

CONST_AF81	5502	16,6668	16,8906	10,6020	0,0000	98,2906	98,2906	1,5457	2,4851	0,2277
ACC_POT_CO	5502	140866,9387	43034,0970	135580,6200	50885,8800	363988,4400	313102,5600	0,7095	0,9491	580,1662
D300_ALIM	5502	0,1107	0,3138	0,0000	0,0000	1,0000	1,0000	2,4810	4,1563	0,0042
D500_ALIM	5502	0,0994	0,2993	0,0000	0,0000	1,0000	1,0000	2,6767	5,1659	0,0040
D700_ALIM	5502	0,0941	0,2921	0,0000	0,0000	1,0000	1,0000	2,7787	5,7224	0,0039
D1500_ALIM	5502	0,2750	0,4466	0,0000	0,0000	1,0000	1,0000	1,0076	-0,9850	0,0060
SUPER_ED	5502	1733,8194	1625,0214	1208,3000	0,0000	10863,2000	10863,2000	1,7793	3,8965	21,9078
D300_BOUL	5502	0,1141	0,3180	0,0000	0,0000	1,0000	1,0000	2,4263	3,8875	0,0043
D500_BOUL	5502	0,0983	0,2978	0,0000	0,0000	1,0000	1,0000	2,6972	5,2761	0,0040
D700_BOUL	5502	0,0965	0,2953	0,0000	0,0000	1,0000	1,0000	2,7321	5,4653	0,0040
D1500_BOUL	5502	0,2744	0,4463	0,0000	0,0000	1,0000	1,0000	1,0106	-0,9788	0,0060
BOUL_ED	5502	1702,8264	1595,3213	1204,1600	0,0000	10638,1000	10638,1000	1,7953	3,8842	21,5074
D300_PHARM	5502	0,1203	0,3254	0,0000	0,0000	1,0000	1,0000	2,3334	3,4456	0,0044
D500_PHARM	5502	0,1069	0,3090	0,0000	0,0000	1,0000	1,0000	2,5443	4,4741	0,0042
D700_PHARM	5502	0,1072	0,3094	0,0000	0,0000	1,0000	1,0000	2,5381	4,4428	0,0042
D1500_PHAR	5502	0,2694	0,4437	0,0000	0,0000	1,0000	1,0000	1,0395	-0,9195	0,0060
PHARM_ED	5502	1679,6572	1641,2343	1104,5400	0,0000	10837,4000	10837,4000	1,8107	3,6691	22,1264
D_PT_PARC	5502	0,0362	0,1867	0,0000	0,0000	1,0000	1,0000	4,9671	22,6764	0,0025
D_MO_PARC	5502	0,0505	0,2190	0,0000	0,0000	1,0000	1,0000	4,1031	14,8381	0,0030
D_GD_PARC	5502	0,1036	0,3048	0,0000	0,0000	1,0000	1,0000	2,6009	4,7654	0,0041
D100_BLEU	5502	0,0171	0,1296	0,0000	0,0000	1,0000	1,0000	7,4511	53,5287	0,0017
D300_BLEU	5502	0,0414	0,1993	0,0000	0,0000	1,0000	1,0000	4,6004	19,1667	0,0027
D500_BLEU	5502	0,0574	0,2327	0,0000	0,0000	1,0000	1,0000	3,8032	12,4667	0,0031

D700_BLEU	5502	0,0509	0,2198	0,0000	0,0000	1,0000	1,0000	4,0859	14,6972	0,0030
DIST_BLEU	5502	3987,5202	4385,9209	2163,3300	0,0000	21323,5000	21323,5000	1,6294	1,8652	59,1290
D100_VERT	5502	0,0458	0,2091	0,0000	0,0000	1,0000	1,0000	4,3441	16,8741	0,0028
D300_VERT	5502	0,0978	0,2970	0,0000	0,0000	1,0000	1,0000	2,7076	5,3321	0,0040
D500_VERT	5502	0,0863	0,2809	0,0000	0,0000	1,0000	1,0000	2,9450	6,6741	0,0038
D700_VERT	5502	0,0645	0,2457	0,0000	0,0000	1,0000	1,0000	3,5441	10,5626	0,0033
DIST_VERT	5502	1938,8745	1621,2438	1552,4200	0,0000	7762,0900	7762,0900	0,9960	0,5503	21,8569
D100_EXTRA	5502	0,0024	0,0486	0,0000	0,0000	1,0000	1,0000	20,4940	418,0800	0,0007
D300_EXTRA	5502	0,0109	0,1039	0,0000	0,0000	1,0000	1,0000	9,4161	86,6784	0,0014
D500_EXTRA	5502	0,0120	0,1089	0,0000	0,0000	1,0000	1,0000	8,9628	78,3462	0,0015
D700_EXTRA	5502	0,0200	0,1400	0,0000	0,0000	1,0000	1,0000	6,8566	45,0211	0,0019
DIST_EXTRA	5502	4620,1452	3061,2040	4301,1600	0,0000	16944,6000	16944,6000	1,0434	1,3728	41,2698
D100_INDUS	5502	0,0709	0,2567	0,0000	0,0000	1,0000	1,0000	3,3433	9,1796	0,0035
D300_INDUS	5502	0,0958	0,2943	0,0000	0,0000	1,0000	1,0000	2,7463	5,5431	0,0040
D500_INDUS	5502	0,1054	0,3071	0,0000	0,0000	1,0000	1,0000	2,5691	4,6013	0,0041
D700_INDUS	5502	0,0736	0,2612	0,0000	0,0000	1,0000	1,0000	3,2648	8,6604	0,0035
DIST_INDUS	5502	1872,2968	2328,0297	1100,0000	0,0000	16822,0000	16822,0000	2,7367	9,3906	31,3854
D100_SAR	5502	0,0616	0,2405	0,0000	0,0000	1,0000	1,0000	3,6453	11,2906	0,0032
D300_SAR	5502	0,1381	0,3451	0,0000	0,0000	1,0000	1,0000	2,0970	2,3978	0,0047
D500_SAR	5502	0,1529	0,3599	0,0000	0,0000	1,0000	1,0000	1,9289	1,7209	0,0049
D700_SAR	5502	0,1056	0,3074	0,0000	0,0000	1,0000	1,0000	2,5660	4,5852	0,0041
SAR_ED	5502	1577,1736	1987,5377	781,0250	0,0000	14182,4000	14182,4000	2,5945	8,6528	26,7951
BFA_ROAD	5502	0,1443	0,3514	0,0000	0,0000	1,0000	1,0000	2,0238	2,0963	0,0047

BM_ROAD	5502	0,0596	0,2368	0,0000	0,0000	1,0000	1,0000	3,7189	11,8324	0,0032
BFO_ROAD	5502	0,0167	0,1282	0,0000	0,0000	1,0000	1,0000	7,5359	54,8003	0,0017
BTFO_ROAD	5502	0,0036	0,0602	0,0000	0,0000	1,0000	1,0000	16,4911	270,0044	0,0008
BFA_TRAIN	5502	0,0087	0,0930	0,0000	0,0000	1,0000	1,0000	10,5628	109,5929	0,0013
BM_TRAIN	5502	0,0055	0,0736	0,0000	0,0000	1,0000	1,0000	13,4278	178,3395	0,0010
BFO_TRAIN	5502	0,0040	0,0631	0,0000	0,0000	1,0000	1,0000	15,7150	245,0047	0,0009
BTFO_TRAIN	5502	0,0016	0,0404	0,0000	0,0000	1,0000	1,0000	24,6577	606,1135	0,0005
BFA_AVION	5502	0,0454	0,2083	0,0000	0,0000	1,0000	1,0000	4,3641	17,0483	0,0028
BM_AVION	5502	0,0262	0,1597	0,0000	0,0000	1,0000	1,0000	5,9343	33,2220	0,0022
BFO_AVION	5502	0,0120	0,1089	0,0000	0,0000	1,0000	1,0000	8,9628	78,3462	0,0015
BTFO_AVION	5502	0,0016	0,0404	0,0000	0,0000	1,0000	1,0000	24,6577	606,1135	0,0005
SHANNON_300	5502	0,4837	0,3684	0,5285	0,0000	1,6842	1,6842	0,1691	-0,9606	0,0050

**Annexe 6b. Statistiques descriptives de l'ensemble des variables (Immoweb ajustée)**

	n	mean	sd	median	min	max	range	skew	kurtosis	se
PRIX	509	173669,4499	61065,0922	165000,0000	45000,0000	335000,0000	290000,0000	0,4963	-0,2668	2706,6626
SURF_FONCI	509	546,8487	765,2563	286,0000	25,0000	5800,0000	5775,0000	3,4929	14,8383	33,9194
LOG_SURF_F	509	2,4889	0,4483	2,4564	1,3979	3,7634	2,3655	0,3093	-0,1158	0,0199
SURF_HAB	509	144,5108	51,3184	135,0000	50,0000	400,0000	350,0000	1,1663	2,1078	2,2746
LOG_SURF_H	509	2,1346	0,1479	2,1303	1,6990	2,6021	0,9031	0,0594	0,1187	0,0066
NB_FACADE	509	2,7701	0,8536	3,0000	1,0000	4,0000	3,0000	0,2287	-1,1907	0,0378
N2_FACADE	509	0,4342	0,4961	0,0000	0,0000	1,0000	1,0000	0,2648	-1,9337	0,0220
N3_FACADE	509	0,2908	0,4546	0,0000	0,0000	1,0000	1,0000	0,9188	-1,1581	0,0201
N4_FACADE	509	0,2515	0,4343	0,0000	0,0000	1,0000	1,0000	1,1423	-0,6966	0,0192
NB_ETAGE	509	1,3320	0,6226	1,0000	0,0000	3,0000	3,0000	0,2138	-0,0479	0,0276
N0_ETAGE	509	0,0589	0,2357	0,0000	0,0000	1,0000	1,0000	3,7345	11,9703	0,0104
N1_ETAGE	509	0,5737	0,4950	1,0000	0,0000	1,0000	1,0000	-0,2971	-1,9155	0,0219
N2_ETAGE	509	0,3674	0,4826	0,0000	0,0000	1,0000	1,0000	0,5485	-1,7024	0,0214
MANSARDE	509	0,2338	0,4237	0,0000	0,0000	1,0000	1,0000	1,2542	-0,4277	0,0188
ANNEE	509	1931,4872	36,8212	1930,0000	1850,0000	2005,0000	155,0000	-0,3421	-0,4729	1,6321
AGE_BIEN	509	79,5128	36,8212	81,0000	6,0000	161,0000	155,0000	0,3421	-0,4729	1,6321
RENOVATION	509	0,1081	0,3108	0,0000	0,0000	1,0000	1,0000	2,5176	4,3467	0,0138
NB_GARAGE	509	0,6582	0,8163	0,0000	0,0000	6,0000	6,0000	1,5006	3,9631	0,0362
N0_GARAGE	509	0,5128	0,5003	1,0000	0,0000	1,0000	1,0000	-0,0509	-2,0013	0,0222
N1_GARAGE	509	0,3497	0,4773	0,0000	0,0000	1,0000	1,0000	0,6285	-1,6082	0,0212
N2_GARAGE	509	0,1375	0,3447	0,0000	0,0000	1,0000	1,0000	2,0988	2,4096	0,0153
NB_SDB	509	1,0039	0,3658	1,0000	0,0000	3,0000	3,0000	0,2888	5,7606	0,0162
N0_SDB	509	0,0629	0,2430	0,0000	0,0000	1,0000	1,0000	3,5912	10,9185	0,0108
N1_SDB	509	0,8723	0,3341	1,0000	0,0000	1,0000	1,0000	-2,2244	2,9537	0,0148
N2_SDB	509	0,0648	0,2465	0,0000	0,0000	1,0000	1,0000	3,5242	10,4406	0,0109
NB_CHB	509	3,0786	0,9770	3,0000	0,0000	7,0000	7,0000	0,6646	1,5524	0,0433
N1_CHB	509	0,0255	0,1579	0,0000	0,0000	1,0000	1,0000	5,9973	34,0341	0,0070



N2_CHB	509	0,2279	0,4199	0,0000	0,0000	1,0000	1,0000	1,2935	-0,3274	0,0186
N3_CHB	509	0,4735	0,4998	0,0000	0,0000	1,0000	1,0000	0,1059	-1,9927	0,0222
N4_CHB	509	0,2711	0,4450	0,0000	0,0000	1,0000	1,0000	1,0267	-0,9477	0,0197
NB_PIECE	509	5,6405	1,7963	5,0000	0,0000	18,0000	18,0000	1,6658	6,8029	0,0796
CAVE	509	0,7485	0,4343	1,0000	0,0000	1,0000	1,0000	-1,1423	-0,6966	0,0192
A_RENOVER	509	0,1022	0,3032	0,0000	0,0000	1,0000	1,0000	2,6195	4,8712	0,0134
BON	509	0,3438	0,4754	0,0000	0,0000	1,0000	1,0000	0,6557	-1,5731	0,0211
EXCELLENT	509	0,2181	0,4133	0,0000	0,0000	1,0000	1,0000	1,3614	-0,1468	0,0183
BON_OU_EXC	509	0,5619	0,4966	1,0000	0,0000	1,0000	1,0000	-0,2487	-1,9419	0,0220
S_EQ	509	0,2141	0,4106	0,0000	0,0000	1,0000	1,0000	1,3895	-0,0693	0,0182
EQ	509	0,6424	0,4798	1,0000	0,0000	1,0000	1,0000	-0,5926	-1,6520	0,0213
GREN	509	0,3851	0,4871	0,0000	0,0000	1,0000	1,0000	0,4710	-1,7817	0,0216
TERR	509	0,5933	0,4917	1,0000	0,0000	1,0000	1,0000	-0,3788	-1,8601	0,0218
JARD	509	0,8212	0,3835	1,0000	0,0000	1,0000	1,0000	-1,6717	0,7961	0,0170
VERANDA	509	0,0314	0,1747	0,0000	0,0000	1,0000	1,0000	5,3549	26,7278	0,0077
PISCINE	509	0,0059	0,0766	0,0000	0,0000	1,0000	1,0000	12,8721	164,0144	0,0034
BUANDERIE	509	0,4558	0,4985	0,0000	0,0000	1,0000	1,0000	0,1770	-1,9725	0,0221
NB_SDD	509	0,2633	0,4583	0,0000	0,0000	2,0000	2,0000	1,3145	0,3534	0,0203
ACC_POT_EM	509	12573,6389	4790,4107	12810,9000	3291,7800	24368,6400	21076,8600	0,1205	-0,6958	212,3312
ACC_GUILLEM	509	11,5085	6,7345	10,0629	1,1399	36,5253	35,3854	1,1743	1,3130	0,2985
ACC_STLAMBE	509	12,5618	6,9031	11,0899	1,2908	39,9457	38,6549	1,2958	1,9316	0,3060
ACC_NEV	509	24,0702	13,3674	21,0315	5,6343	76,4710	70,8367	1,3259	1,8735	0,5925
ACC_BUS	509	0,0524	0,0265	0,0449	0,0243	0,1437	0,1193	1,1112	0,5675	0,0012
ACC_ML	509	0,0938	0,0388	0,0935	0,0095	0,2236	0,2141	0,2750	0,0065	0,0017
ACC_TRAIN	509	0,0473	0,0258	0,0377	0,0223	0,1878	0,1655	1,8727	4,7376	0,0011
ACC_TOTAL	509	0,1935	0,0658	0,1925	0,0637	0,4438	0,3801	0,5065	0,3474	0,0029
ACC_POT_BU	509	23,8096	13,4173	21,6230	4,9359	79,9064	74,9705	1,1348	1,6820	0,5947
ACC_POT_TR	509	4220,6062	2124,5645	3754,9800	1555,6200	17807,3400	16251,7200	2,4537	8,5866	94,1697
D300_GARE	509	0,0413	0,1991	0,0000	0,0000	1,0000	1,0000	4,5996	19,1937	0,0088

D500_GARE	509	0,0648	0,2465	0,0000	0,0000	1,0000	1,0000	3,5242	10,4406	0,0109
D700_GARE	509	0,0452	0,2079	0,0000	0,0000	1,0000	1,0000	4,3663	17,0986	0,0092
D1500_GARE	509	0,2613	0,4398	0,0000	0,0000	1,0000	1,0000	1,0834	-0,8278	0,0195
GARE_ED	509	2214,5672	1761,6528	1802,7800	0,0000	13346,2000	13346,2000	1,8443	5,0255	78,0839
D100_BUS	509	0,4126	0,4928	0,0000	0,0000	1,0000	1,0000	0,3541	-1,8783	0,0218
D300_BUS	509	0,4165	0,4935	0,0000	0,0000	1,0000	1,0000	0,3377	-1,8896	0,0219
D500_BUS	509	0,1139	0,3181	0,0000	0,0000	1,0000	1,0000	2,4228	3,8774	0,0141
D700_BUS	509	0,0413	0,1991	0,0000	0,0000	1,0000	1,0000	4,5996	19,1937	0,0088
ARRET_ED	509	190,1238	163,4469	141,4210	0,0000	1204,1600	1204,1600	1,9665	6,0617	7,2447
RMD_2011	509	20648,7623	3927,8394	20021,0000	13379,0000	36545,0000	23166,0000	0,7745	1,2167	174,0984
TAUX_CHOMA	509	13,9725	6,8427	13,0000	0,0000	34,0000	34,0000	0,5177	-0,2473	0,3033
NIV_INSTRU	509	20,7407	9,7278	19,0000	5,0000	55,0000	50,0000	0,8872	0,5616	0,4312
D_POP_2011	509	3364,1982	2633,0744	3011,2652	12,5381	18851,9758	18839,4376	1,8780	6,7627	116,7090
DPOPIX	509	2450,8959	2016,9584	1848,0000	52,0000	10752,0000	10700,0000	1,2251	1,9089	89,4001
CONST_AF81	509	14,8802	14,8598	10,0000	0,0000	71,0000	71,0000	1,2368	0,9134	0,6587
ACC_POT_CO	509	142134,2239	37807,0953	138171,4800	58933,6200	285441,3600	226507,7400	0,5887	0,3368	1675,7700
D300_ALIM	509	0,0904	0,2870	0,0000	0,0000	1,0000	1,0000	2,8490	6,1286	0,0127
D500_ALIM	509	0,1022	0,3032	0,0000	0,0000	1,0000	1,0000	2,6195	4,8712	0,0134
D700_ALIM	509	0,1022	0,3032	0,0000	0,0000	1,0000	1,0000	2,6195	4,8712	0,0134
D1500_ALIM	509	0,3006	0,4590	0,0000	0,0000	1,0000	1,0000	0,8672	-1,2503	0,0203
SUPER_ED	509	1638,8167	1466,3269	1204,1600	0,0000	9544,1100	9544,1100	1,8194	4,2438	64,9938
D300_BOUL	509	0,0943	0,2925	0,0000	0,0000	1,0000	1,0000	2,7682	5,6741	0,0130
D500_BOUL	509	0,1022	0,3032	0,0000	0,0000	1,0000	1,0000	2,6195	4,8712	0,0134
D700_BOUL	509	0,1022	0,3032	0,0000	0,0000	1,0000	1,0000	2,6195	4,8712	0,0134
D1500_BOUL	509	0,2986	0,4581	0,0000	0,0000	1,0000	1,0000	0,8774	-1,2325	0,0203
BOUL_ED	509	1592,5061	1429,3517	1200,0000	0,0000	9730,8800	9730,8800	1,8615	4,5240	63,3549
D300_PHARM	509	0,1041	0,3057	0,0000	0,0000	1,0000	1,0000	2,5847	4,6897	0,0136
D500_PHARM	509	0,1041	0,3057	0,0000	0,0000	1,0000	1,0000	2,5847	4,6897	0,0136
D700_PHARM	509	0,1139	0,3181	0,0000	0,0000	1,0000	1,0000	2,4228	3,8774	0,0141

D1500_PHAR	509	0,2947	0,4564	0,0000	0,0000	1,0000	1,0000	0,8980	-1,1959	0,0202
PHARM_ED	509	1590,4454	1498,4612	1104,5400	0,0000	8994,4400	8994,4400	1,9118	4,3094	66,4181
D_PT_PARC	509	0,0236	0,1519	0,0000	0,0000	1,0000	1,0000	6,2617	37,2821	0,0067
D_MO_PARC	509	0,0373	0,1898	0,0000	0,0000	1,0000	1,0000	4,8670	21,7308	0,0084
D_GD_PARC	509	0,0825	0,2754	0,0000	0,0000	1,0000	1,0000	3,0257	7,1689	0,0122
D100_BLEU	509	0,0138	0,1166	0,0000	0,0000	1,0000	1,0000	8,3257	67,4506	0,0052
D300_BLEU	509	0,0530	0,2243	0,0000	0,0000	1,0000	1,0000	3,9767	13,8415	0,0099
D500_BLEU	509	0,0472	0,2122	0,0000	0,0000	1,0000	1,0000	4,2603	16,1822	0,0094
D700_BLEU	509	0,0452	0,2079	0,0000	0,0000	1,0000	1,0000	4,3663	17,0986	0,0092
ED_BLEU	509	3302,8799	3495,6293	2012,4600	0,0000	19160,4000	19160,4000	1,9304	3,7917	154,9410
D100_VERT	509	0,0491	0,2163	0,0000	0,0000	1,0000	1,0000	4,1604	15,3394	0,0096
D300_VERT	509	0,1022	0,3032	0,0000	0,0000	1,0000	1,0000	2,6195	4,8712	0,0134
D500_VERT	509	0,0923	0,2898	0,0000	0,0000	1,0000	1,0000	2,8080	5,8965	0,0128
D700_VERT	509	0,0688	0,2533	0,0000	0,0000	1,0000	1,0000	3,3983	9,5672	0,0112
ED_VERT	509	1927,0991	1687,7924	1476,4800	0,0000	7669,4200	7669,4200	1,1047	0,8204	74,8101
D100_EXTRA	509	0,0039	0,0626	0,0000	0,0000	1,0000	1,0000	15,8122	248,5128	0,0028
D300_EXTRA	509	0,0157	0,1245	0,0000	0,0000	1,0000	1,0000	7,7643	58,3990	0,0055
D500_EXTRA	509	0,0236	0,1519	0,0000	0,0000	1,0000	1,0000	6,2617	37,2821	0,0067
D700_EXTRA	509	0,0177	0,1319	0,0000	0,0000	1,0000	1,0000	7,2978	51,3593	0,0058
EXTRACT_ED	509	4025,6372	2637,0471	3452,5400	0,0000	13946,3000	13946,3000	0,7938	0,4633	116,8851
D100_INDUS	509	0,0707	0,2566	0,0000	0,0000	1,0000	1,0000	3,3390	9,1670	0,0114
D300_INDUS	509	0,1022	0,3032	0,0000	0,0000	1,0000	1,0000	2,6195	4,8712	0,0134
D500_INDUS	509	0,1375	0,3447	0,0000	0,0000	1,0000	1,0000	2,0988	2,4096	0,0153
D700_INDUS	509	0,0786	0,2694	0,0000	0,0000	1,0000	1,0000	3,1229	7,7679	0,0119
INDUST_ED	509	1572,7583	1847,1411	1019,8000	0,0000	14303,1000	14303,1000	2,8398	11,2433	81,8731
D100_SAR	509	0,0766	0,2663	0,0000	0,0000	1,0000	1,0000	3,1741	8,0906	0,0118
D300_SAR	509	0,1454	0,3528	0,0000	0,0000	1,0000	1,0000	2,0062	2,0287	0,0156
D500_SAR	509	0,1670	0,3733	0,0000	0,0000	1,0000	1,0000	1,7804	1,1723	0,0165
D700_SAR	509	0,1139	0,3181	0,0000	0,0000	1,0000	1,0000	2,4228	3,8774	0,0141

SAR_ED	509	1304,8504	1592,4690	700,0000	0,0000	13013,8000	13013,8000	2,6404	9,9148	70,5849
BFA_ROAD	509	0,1454	0,3528	0,0000	0,0000	1,0000	1,0000	2,0062	2,0287	0,0156
BMO_ROAD	509	0,0530	0,2243	0,0000	0,0000	1,0000	1,0000	3,9767	13,8415	0,0099
BFO_ROAD	509	0,0295	0,1693	0,0000	0,0000	1,0000	1,0000	5,5481	28,8382	0,0075
BTFO_ROAD	509	0,0039	0,0626	0,0000	0,0000	1,0000	1,0000	15,8122	248,5128	0,0028
BFA_TRAIN	509	0,0295	0,1693	0,0000	0,0000	1,0000	1,0000	5,5481	28,8382	0,0075
BMO_TRAIN	509	0,0059	0,0766	0,0000	0,0000	1,0000	1,0000	12,8721	164,0144	0,0034
BFO_TRAIN	509	0,0059	0,0766	0,0000	0,0000	1,0000	1,0000	12,8721	164,0144	0,0034
BTFO_TRAIN	509	0,0039	0,0626	0,0000	0,0000	1,0000	1,0000	15,8122	248,5128	0,0028
BFA_AVION	509	0,0354	0,1849	0,0000	0,0000	1,0000	1,0000	5,0165	23,2111	0,0082
BMO_AVION	509	0,0373	0,1898	0,0000	0,0000	1,0000	1,0000	4,8670	21,7308	0,0084
BFO_AVION	509	0,0118	0,1080	0,0000	0,0000	1,0000	1,0000	9,0202	79,5201	0,0048
BTFO_AVION	/	/	/	/	/	/	/	/	/	/
SHANNON_300	509	0,5807	0,3426	0,6672	0,1237	1,3280	1,2043	-0,1934	-1,3646	0,0152
SHANNON_500	509	0,6458	0,4040	0,6483	0,0000	1,6440	1,6440	0,0526	-0,8748	0,0179
KIEMSTEDT	509	69,2283	13,7088	71,6350	31,1470	113,5050	82,3580	-0,2586	-0,1584	0,6076

**Annexe 7. Matrice de corrélation de Spearman pour les variables quantitatives**

**Annexe 7a. Matrice de corrélation de Spearman pour les variables quantitatives (SPF Finances 95 %)**

	PRIX	LOG_SURF_FONC	LOG_SURF_HAB	AGE_BIEN	AGE_RENO	NB_PIECE	ACC_POT_EMPLOI	ISO_GUILLEMIN	ISO_STLAMBERT	ISO_NEV	ACC_BUS	ACC_ML	ACC_TRAIN	ACC_TOT	ACC_POT_GARE	ACC_POT_ARRET	RMD_2011	TCH_2011	NINS_2011	D_POPSS_2011	DPOPIX_2011	CAFT81_2011	ACC_POT_COM	SHANNON_300	SHANNON_500	KIEMSTEDT
PRIX	1,00	0,49	0,29	-0,45	-0,38	0,31	-0,22	0,18	0,16	0,17	-0,24	-0,17	-0,02	-0,19	-0,19	-0,20	0,34	-0,35	0,41	-0,26	-0,21	0,29	-0,16	0,10	0,09	-0,07
LOG_SURF_FONC	0,49	1,00	0,31	-0,40	-0,32	0,16	-0,55	0,49	0,49	0,50	-0,54	-0,43	-0,06	-0,48	-0,47	-0,54	0,52	-0,54	0,35	-0,57	-0,53	0,51	-0,46	0,24	0,23	-0,01
LOG_SURF_HAB	0,29	0,31	1,00	0,10	-0,01	0,52	-0,08	0,08	0,04	0,06	-0,04	0,03	0,06	0,02	-0,06	-0,08	0,07	-0,09	0,14	-0,06	-0,06	0,04	-0,04	0,04	0,04	-0,09
AGE_BIEN	-0,45	-0,40	0,10	1,00	0,53	-0,01	0,18	-0,15	-0,14	-0,15	0,23	0,20	0,10	0,24	0,17	0,17	-0,30	0,29	-0,19	0,22	0,18	-0,31	0,15	-0,02	-0,01	0,06
AGE_RENO	-0,38	-0,32	-0,01	0,53	1,00	-0,04	0,16	-0,13	-0,13	-0,13	0,19	0,17	0,05	0,19	0,14	0,15	-0,22	0,22	-0,14	0,20	0,16	-0,25	0,13	-0,04	-0,03	0,03
NB_PIECE	0,31	0,16	0,52	-0,01	-0,04	1,00	0,10	-0,08	-0,11	-0,10	0,11	0,13	0,09	0,16	0,10	0,10	-0,06	0,05	0,13	0,10	0,09	-0,09	0,12	0,00	-0,01	-0,09
ACC_POT_EMPLOI	-0,22	-0,55	-0,08	0,18	0,16	0,10	1,00	-0,95	-0,91	-0,96	0,73	0,43	0,06	0,54	0,89	0,94	-0,60	0,68	-0,32	0,77	0,82	-0,58	0,91	-0,41	-0,45	-0,06
ISO_GUILLEMIN	0,18	0,49	0,08	-0,15	-0,13	-0,08	-0,95	1,00	0,89	0,97	-0,69	-0,32	-0,11	-0,46	-0,92	-0,91	0,51	-0,59	0,23	-0,68	-0,74	0,53	-0,88	0,36	0,40	0,05
ISO_STLAMBERT	0,16	0,49	0,04	-0,14	-0,13	-0,11	-0,91	0,89	1,00	0,97	-0,68	-0,32	-0,03	-0,44	-0,84	-0,94	0,52	-0,60	0,22	-0,67	-0,74	0,48	-0,92	0,34	0,39	0,12
ISO_NEV	0,17	0,50	0,06	-0,15	-0,13	-0,10	-0,96	0,97	0,97	1,00	-0,70	-0,34	-0,08	-0,47	-0,91	-0,95	0,53	-0,61	0,22	-0,70	-0,76	0,52	-0,93	0,36	0,41	0,08
ACC_BUS	-0,24	-0,54	-0,04	0,23	0,19	0,11	0,73	-0,69	-0,68	-0,70	1,00	0,57	0,06	0,73	0,66	0,75	-0,63	0,66	-0,38	0,72	0,71	-0,62	0,68	-0,39	-0,39	-0,17
ACC_ML	-0,17	-0,43	0,03	0,20	0,17	0,13	0,43	-0,32	-0,32	-0,34	0,57	1,00	0,22	0,89	0,33	0,41	-0,53	0,53	-0,31	0,70	0,70	-0,52	0,35	-0,47	-0,47	-0,21
ACC_TRAIN	-0,02	-0,06	0,06	0,10	0,05	0,09	0,06	-0,11	-0,03	-0,08	0,06	0,22	1,00	0,47	0,30	-0,01	-0,06	0,07	0,01	0,09	0,07	-0,16	0,14	-0,01	-0,01	-0,12

	PRIX	LOG_SURF_FONC	LOG_SURF_HAB	AGE_BIEN	AGE_RENO	NB_PIECE	ACC_POT_EMPLOI	ISO_GUILLEMIN	ISO_STLAMBERT	ISO_NEV	ACC_BUS	ACC_ML	ACC_TRAIN	ACC_TOT	ACC_POT_GARE	ACC_POT_ARRET	RMD_2011	TCH_2011	NINS_2011	D_POPSS_2011	DPOPIX_2011	CAFT81_2011	ACC_POT_COM	SHANNON_300	SHANNON_500	KIEMSTEDT
ACC_TOT	-0,19	-0,48	0,02	0,24	0,19	0,16	0,54	-0,46	-0,44	-0,47	0,73	0,89	0,47	1,00	0,53	0,51	-0,57	0,58	-0,32	0,72	0,69	-0,59	0,49	-0,42	-0,42	-0,23
ACC_POT_GARE	-0,19	-0,47	-0,06	0,17	0,14	0,10	0,89	-0,92	-0,84	-0,91	0,66	0,33	0,30	0,53	1,00	0,84	-0,53	0,60	-0,27	0,64	0,70	-0,54	0,90	-0,33	-0,37	-0,03
ACC_POT_ARRET	-0,20	-0,54	-0,08	0,17	0,15	0,10	0,94	-0,91	-0,94	-0,95	0,75	0,41	-0,01	0,51	0,84	1,00	-0,59	0,67	-0,27	0,76	0,82	-0,59	0,87	-0,40	-0,45	-0,02
RMD_2011	0,34	0,52	0,07	-0,30	-0,22	-0,06	-0,60	0,51	0,52	0,53	-0,63	-0,53	-0,06	-0,57	-0,53	-0,59	1,00	-0,89	0,70	-0,68	-0,60	0,69	-0,53	0,25	0,24	0,01
TCH_2011	-0,35	-0,54	-0,09	0,29	0,22	0,05	0,68	-0,59	-0,60	-0,61	0,66	0,53	0,07	0,58	0,60	0,67	-0,89	1,00	-0,69	0,72	0,67	-0,68	0,61	-0,30	-0,30	0,00
NINS_2011	0,41	0,35	0,14	-0,19	-0,14	0,13	-0,32	0,23	0,22	0,22	-0,38	-0,31	0,01	-0,32	-0,27	-0,27	0,70	-0,69	1,00	-0,41	-0,35	0,43	-0,27	0,19	0,17	0,04
D_POPSS_2011	-0,26	-0,57	-0,06	0,22	0,20	0,10	0,77	-0,68	-0,67	-0,70	0,72	0,70	0,09	0,72	0,64	0,76	-0,68	0,72	-0,41	1,00	0,83	-0,68	0,65	-0,45	-0,46	-0,08
DPOPIX_2011	-0,21	-0,53	-0,06	0,18	0,16	0,09	0,82	-0,74	-0,74	-0,76	0,71	0,70	0,07	0,69	0,70	0,82	-0,60	0,67	-0,35	0,83	1,00	-0,57	0,72	-0,55	-0,59	-0,07
CAFT81_2011	0,29	0,51	0,04	-0,31	-0,25	-0,09	-0,58	0,53	0,48	0,52	-0,62	-0,52	-0,16	-0,59	-0,54	-0,59	0,69	-0,68	0,43	-0,68	-0,57	1,00	-0,48	0,24	0,22	-0,09
ACC_POT_COM	-0,16	-0,46	-0,04	0,15	0,13	0,12	0,91	-0,88	-0,92	-0,93	0,68	0,35	0,14	0,49	0,90	0,87	-0,53	0,61	-0,27	0,65	0,72	-0,48	1,00	-0,34	-0,38	-0,16
SHANNON_300	0,10	0,24	0,04	-0,02	-0,04	0,00	-0,41	0,36	0,34	0,36	-0,39	-0,47	-0,01	-0,42	-0,33	-0,40	0,25	-0,30	0,19	-0,45	-0,55	0,24	-0,34	1,00	0,90	0,12
SHANNON_500	0,09	0,23	0,04	-0,01	-0,03	-0,01	-0,45	0,40	0,39	0,41	-0,39	-0,47	-0,01	-0,42	-0,37	-0,45	0,24	-0,30	0,17	-0,46	-0,59	0,22	-0,38	0,90	1,00	0,15
KIEMSTEDT	-0,07	-0,01	-0,09	0,06	0,03	-0,09	-0,06	0,05	0,12	0,08	-0,17	-0,21	-0,12	-0,23	-0,03	-0,02	0,01	0,00	0,04	-0,08	-0,07	-0,09	-0,16	0,12	0,15	1,00

Annexe7b. Matrice de corrélation de Spearman pour les variables quantitatives (Immoweb ajusté)

	PRIX	LOG_SURF_FONC	LOG_SURF_BATI	AGE_BIEN	NB_PIECE	ACC_POT_EMPLOI	ISO_GUILLEMINIS	ISO_STLAMBERT	ISO_NEV	ACC_BUS	ACC_ML	ACC_TRAIN	ACC_TOT	ACC_POT_ARRET	ACC_POT_GARE	RMD_2011	TCH_2011	NINS_2011	D_POPSS_2011	DPOPIX_2011	CAFT81_2011	ACC_POT_COM	SHANNON_300	SHANNON_500	KIEMSTEDT
PRIX	1,00	0,60	0,64	-0,48	0,23	-0,26	0,25	0,15	0,21	-0,24	-0,21	-0,07	-0,24	-0,25	-0,24	0,39	-0,40	0,43	-0,31	-0,25	0,37	-0,18	0,18	0,09	-0,17
LOG_SURF_FONC	0,60	1,00	0,34	-0,29	0,04	-0,55	0,50	0,45	0,50	-0,51	-0,44	-0,14	-0,50	-0,55	-0,51	0,51	-0,51	0,35	-0,58	-0,52	0,53	-0,44	0,40	0,24	-0,04
LOG_SURF_BATI	0,64	0,34	1,00	-0,21	0,41	-0,09	0,08	0,05	0,07	0,01	-0,02	-0,03	0,00	-0,09	-0,06	0,15	-0,19	0,22	-0,09	-0,09	0,17	-0,02	0,13	0,07	-0,13
AGE_BIEN	-0,48	-0,29	-0,21	1,00	0,05	0,11	-0,11	-0,06	-0,09	0,16	0,18	0,12	0,21	0,11	0,12	-0,28	0,26	-0,18	0,18	0,11	-0,26	0,07	-0,10	0,01	0,11
NB_PIECE	0,23	0,04	0,41	0,05	1,00	0,11	-0,08	-0,10	-0,10	0,13	0,11	0,07	0,15	0,10	0,10	-0,05	0,03	0,10	0,13	0,13	-0,07	0,13	-0,08	0,00	-0,09
ACC_POT_EMPLOI	-0,26	-0,55	-0,09	0,11	0,11	1,00	-0,95	-0,91	-0,96	0,66	0,37	0,08	0,48	0,94	0,94	-0,57	0,61	-0,32	0,72	0,77	-0,47	0,91	-0,64	-0,40	-0,06
ISO_GUILLEMINIS	0,25	0,50	0,08	-0,11	-0,08	-0,95	1,00	0,86	0,96	-0,62	-0,26	-0,09	-0,40	-0,91	-0,97	0,51	-0,54	0,24	-0,64	-0,69	0,44	-0,86	0,57	0,35	0,00
ISO_STLAMBERT	0,15	0,45	0,05	-0,06	-0,10	-0,91	0,86	1,00	0,96	-0,56	-0,20	-0,04	-0,33	-0,91	-0,88	0,46	-0,50	0,18	-0,57	-0,64	0,32	-0,90	0,59	0,31	0,15
ISO_NEV	0,21	0,50	0,07	-0,09	-0,10	-0,96	0,96	0,96	1,00	-0,61	-0,24	-0,08	-0,38	-0,94	-0,96	0,51	-0,54	0,22	-0,63	-0,69	0,40	-0,91	0,61	0,34	0,07
ACC_BUS	-0,24	-0,51	0,01	0,16	0,13	0,66	-0,62	-0,56	-0,61	1,00	0,53	0,04	0,72	0,68	0,61	-0,56	0,54	-0,32	0,67	0,61	-0,57	0,56	-0,47	-0,35	-0,09
ACC_ML	-0,21	-0,44	-0,02	0,18	0,11	0,37	-0,26	-0,20	-0,24	0,53	1,00	0,13	0,86	0,36	0,26	-0,50	0,49	-0,34	0,71	0,70	-0,49	0,24	-0,48	-0,51	-0,17
ACC_TRAIN	-0,07	-0,14	-0,03	0,12	0,07	0,08	-0,09	-0,04	-0,08	0,04	0,13	1,00	0,43	0,00	0,20	-0,04	0,03	0,04	0,07	0,01	-0,15	0,17	-0,01	0,04	-0,12

	PRIX	LOG_SURF_FONC	LOG_SURF_BATI	AGE_BIEN	NB_PIECE	ACC_POT_EMPLOI	ISO_GUILLEMIN	ISO_STLAMBERT	ISO_NEV	ACC_BUS	ACC_ML	ACC_TRAIN	ACC_TOT	ACC_POT_ARRET	ACC_POT_GARE	RMD_2011	TCH_2011	NINS_2011	D_POPSS_2011	DPOPIX_2011	CAFT81_2011	ACC_POT_COM	SHANNON_300	SHANNON_500	KIEMSTEDT
ACC_TOT	-0,24	-0,50	0,00	0,21	0,15	0,48	-0,40	-0,33	-0,38	0,72	0,86	0,43	1,00	0,46	0,45	-0,54	0,51	-0,30	0,71	0,65	-0,57	0,39	-0,46	-0,41	-0,17
ACC_POT_ARRET	-0,25	-0,55	-0,09	0,11	0,10	0,94	-0,91	-0,91	-0,94	0,68	0,36	0,00	0,46	1,00	0,88	-0,58	0,62	-0,29	0,72	0,79	-0,50	0,83	-0,69	-0,41	0,00
ACC_POT_GARE	-0,24	-0,51	-0,06	0,12	0,10	0,94	-0,97	-0,88	-0,96	0,61	0,26	0,20	0,45	0,88	1,00	-0,52	0,54	-0,25	0,63	0,68	-0,42	0,91	-0,55	-0,34	-0,06
RMD_2011	0,39	0,51	0,15	-0,28	-0,05	-0,57	0,51	0,46	0,51	-0,56	-0,50	-0,04	-0,54	-0,58	-0,52	1,00	-0,89	0,75	-0,69	-0,58	0,66	-0,48	0,46	0,23	-0,02
TCH_2011	-0,40	-0,51	-0,19	0,26	0,03	0,61	-0,54	-0,50	-0,54	0,54	0,49	0,03	0,51	0,62	0,54	-0,89	1,00	-0,72	0,68	0,60	-0,66	0,51	-0,51	-0,27	0,04
NINS_2011	0,43	0,35	0,22	-0,18	0,10	-0,32	0,24	0,18	0,22	-0,32	-0,34	0,04	-0,30	-0,29	-0,25	0,75	-0,72	1,00	-0,45	-0,36	0,46	-0,24	0,22	0,16	-0,02
D_POPSS_2011	-0,31	-0,58	-0,09	0,18	0,13	0,72	-0,64	-0,57	-0,63	0,67	0,71	0,07	0,71	0,72	0,63	-0,69	0,68	-0,45	1,00	0,81	-0,61	0,56	-0,61	-0,48	-0,09
DPOPIX_2011	-0,25	-0,52	-0,09	0,11	0,13	0,77	-0,69	-0,64	-0,69	0,61	0,70	0,01	0,65	0,79	0,68	-0,58	0,60	-0,36	0,81	1,00	-0,49	0,63	-0,67	-0,59	-0,05
CAFT81_2011	0,37	0,53	0,17	-0,26	-0,07	-0,47	0,44	0,32	0,40	-0,57	-0,49	-0,15	-0,57	-0,50	-0,42	0,66	-0,66	0,46	-0,61	-0,49	1,00	-0,33	0,39	0,19	-0,18
ACC_POT_COM	-0,18	-0,44	-0,02	0,07	0,13	0,91	-0,86	-0,90	-0,91	0,56	0,24	0,17	0,39	0,83	0,91	-0,48	0,51	-0,24	0,56	0,63	-0,33	1,00	-0,48	-0,31	-0,19
SHANNON_300	0,18	0,40	0,13	-0,10	-0,08	-0,64	0,57	0,59	0,61	-0,47	-0,48	-0,01	-0,46	-0,69	-0,55	0,46	-0,51	0,22	-0,61	-0,67	0,39	-0,48	1,00	0,65	-0,01
SHANNON_500	0,09	0,24	0,07	0,01	0,00	-0,40	0,35	0,31	0,34	-0,35	-0,51	0,04	-0,41	-0,41	-0,34	0,23	-0,27	0,16	-0,48	-0,59	0,19	-0,31	0,65	1,00	0,13
KIEMSTEDT	-0,17	-0,04	-0,13	0,11	-0,09	-0,06	0,00	0,15	0,07	-0,09	-0,17	-0,12	-0,17	0,00	-0,06	-0,02	0,04	-0,02	-0,09	-0,05	-0,18	-0,19	-0,01	0,13	1,00



**Annexe 8. Matrice de corrélation bisériale**

**Annexe 8a. Matrice de corrélation bisériale (SPF Finances 95 %)**

	PRIX	LOG_SURF_FONC	LOG_SURF_HAB	AGE_BIEN	AGE_RENO	NB_PIECE	ACC_POT_EMPLOI	ISO_GUILLEMIN	ISO_STLAMBERT	ISO_NEV	ACC_BUS	ACC_ML	ACC_TRAIN	ACC_TOT	ACC_POT_GARE	ACC_POT_ARRET	RMD_2011	TCH_2011	NINS_2011	D_POPSS_2011	DPOPIX_2011	CAFT81_2011	ACC_POT_COM	SHANNON_300	SHANNON_500	KIEMSTEDT
N2_FACADE	-0,36	-0,70	-0,12	0,31	0,25	0,05	0,66	-0,56	-0,56	-0,57	0,57	0,49	0,10	0,54	0,50	0,62	-0,52	0,57	-0,33	0,60	0,60	-0,52	0,58	-0,29	-0,29	-0,05
N3_FACADE	-0,08	0,12	0,16	0,10	0,01	-0,07	-0,24	0,17	0,15	0,16	-0,22	-0,11	-0,02	-0,16	-0,19	-0,26	0,11	-0,18	-0,06	-0,23	-0,24	0,08	-0,21	0,04	0,05	0,03
N4_FACADE	0,58	0,80	-0,02	-0,53	-0,34	0,01	-0,61	0,56	0,57	0,58	-0,51	-0,53	-0,11	-0,54	-0,46	-0,53	0,57	-0,56	0,51	-0,55	-0,53	0,61	-0,53	0,34	0,32	0,03
N0_ETAGE	0,48	0,64	-0,47	-0,62	-0,42	-0,17	-0,51	0,44	0,44	0,45	-0,44	-0,45	-0,11	-0,47	-0,39	-0,46	0,51	-0,48	0,39	-0,47	-0,46	0,58	-0,43	0,26	0,25	0,00
N1_ETAGE	-0,45	-0,33	0,11	0,46	0,29	-0,16	0,08	-0,11	-0,08	-0,09	0,04	0,11	0,01	0,08	0,03	-0,03	-0,22	0,15	-0,35	0,04	0,02	-0,28	0,02	-0,12	-0,10	0,06
N2_ETAGE	0,13	-0,32	0,48	0,05	0,07	0,55	0,61	-0,45	-0,50	-0,48	0,57	0,47	0,14	0,55	0,53	0,73	-0,35	0,43	0,06	0,62	0,65	-0,33	0,61	-0,17	-0,18	-0,11
MANSARDE	0,25	0,06	-0,20	-0,08	-0,14	0,21	0,03	0,02	0,03	0,03	0,11	0,04	0,08	0,10	0,09	0,08	0,01	0,01	0,15	0,04	0,04	0,09	0,06	0,06	0,06	-0,04
RENOVATION	0,05	0,06	0,15	0,18	-0,84	0,04	-0,02	0,03	0,05	0,04	-0,02	-0,02	0,03	-0,01	0,00	-0,01	-0,02	0,00	-0,01	-0,04	-0,01	-0,01	-0,02	0,01	0,00	0,03
N0_GARAGE	-0,51	-0,64	-0,35	0,55	0,38	-0,04	0,46	-0,37	-0,36	-0,37	0,42	0,33	0,07	0,38	0,35	0,43	-0,44	0,47	-0,30	0,44	0,41	-0,43	0,38	-0,16	-0,16	0,03
N1_GARAGE	0,42	0,53	0,26	-0,47	-0,30	-0,02	-0,42	0,34	0,34	0,34	-0,39	-0,28	-0,05	-0,34	-0,32	-0,40	0,40	-0,43	0,27	-0,39	-0,36	0,36	-0,35	0,14	0,13	-0,02
N2_GARAGE	0,32	0,39	0,34	-0,31	-0,31	0,23	-0,15	0,13	0,09	0,11	-0,14	-0,19	-0,05	-0,18	-0,14	-0,12	0,16	-0,16	0,12	-0,18	-0,16	0,26	-0,11	0,07	0,09	-0,05
N0_SDB	-0,49	-0,43	-0,25	0,51	0,66	-0,16	0,15	-0,10	-0,12	-0,11	0,20	0,16	0,00	0,17	0,09	0,17	-0,25	0,25	-0,22	0,20	0,16	-0,27	0,12	0,00	0,02	0,04
N1_SDB	0,40	0,39	0,16	-0,47	-0,58	0,04	-0,19	0,13	0,15	0,14	-0,24	-0,19	-0,01	-0,21	-0,12	-0,23	0,25	-0,26	0,18	-0,23	-0,21	0,26	-0,16	0,01	0,00	-0,03
N2_SDB	0,32	0,07	0,44	-0,07	-0,27	0,62	0,24	-0,15	-0,17	-0,17	0,27	0,19	0,06	0,24	0,21	0,32	-0,07	0,09	0,16	0,18	0,28	-0,02	0,25	-0,07	-0,07	-0,05
CHAUF_CENT	0,57	0,41	0,17	-0,54	-0,62	0,11	-0,17	0,12	0,12	0,12	-0,16	-0,15	0,01	-0,15	-0,09	-0,16	0,28	-0,28	0,27	-0,20	-0,17	0,29	-0,12	0,03	0,02	-0,03
N1_LOGEMEN	-0,18	0,23	-0,51	-0,17	-0,10	-0,87	-0,61	0,41	0,47	0,45	-0,70	-0,53	-0,10	-0,62	-0,50	-0,79	0,40	-0,47	-0,06	-0,62	-0,70	0,33	-0,65	0,15	0,15	0,15
N2_LOGEMEN	0,10	-0,16	0,38	0,11	0,08	0,57	0,44	-0,29	-0,33	-0,32	0,48	0,36	0,05	0,42	0,32	0,54	-0,27	0,31	0,03	0,42	0,47	-0,25	0,43	-0,12	-0,11	-0,08
N3_LOGEMEN	0,17	-0,21	0,50	0,14	0,16	0,85	0,55	-0,38	-0,44	-0,42	0,66	0,49	0,12	0,59	0,49	0,72	-0,38	0,45	0,05	0,56	0,64	-0,30	0,61	-0,16	-0,13	-0,18
N4_LOGEMEN	0,24	-0,27	0,42	0,24	0,00	0,92	0,65	-0,42	-0,48	-0,45	0,75	0,60	0,12	0,68	0,55	0,89	-0,42	0,50	0,08	0,71	0,81	-0,31	0,73	-0,09	-0,15	-0,17
CAVE	-0,27	-0,31	0,12	0,57	0,32	0,11	0,12	-0,07	-0,10	-0,09	0,22	0,17	0,06	0,21	0,09	0,13	-0,22	0,19	-0,17	0,13	0,14	-0,22	0,12	-0,02	-0,02	-0,04
CAVE_VIE	-0,02	-0,18	-0,09	0,04	0,01	0,08	0,32	-0,29	-0,22	-0,26	0,07	0,13	0,16	0,16	0,40	0,32	-0,11	0,12	0,13	0,22	0,27	-0,19	0,27	-0,16	-0,20	0,19
CITE	-0,13	-0,04	-0,20	-0,19	-0,08	-0,09	0,01	-0,03	0,00	-0,02	-0,11	0,13	-0,02	0,02	-0,04	-0,07	-0,13	0,09	-0,27	0,09	-0,02	-0,07	-0,06	-0,15	-0,11	0,03

VILLA	0,63	0,58	-0,07	-0,56	-0,37	0,13	-0,35	0,30	0,32	0,31	-0,28	-0,37	-0,07	-0,35	-0,23	-0,29	0,41	-0,41	0,41	-0,35	-0,32	0,52	-0,24	0,19	0,18	-0,08
BUNGALOW	0,30	0,54	-0,24	-0,54	-0,32	-0,23	-0,40	0,32	0,34	0,34	-0,37	-0,42	-0,14	-0,44	-0,31	-0,36	0,45	-0,38	0,34	-0,39	-0,38	0,40	-0,35	0,24	0,24	0,10
BEL_ETAGE	0,09	-0,10	0,22	-0,38	-0,20	-0,07	0,21	-0,23	-0,24	-0,24	0,06	0,07	0,00	0,07	0,15	0,19	-0,02	0,03	0,00	0,13	0,16	-0,07	0,20	-0,15	-0,17	-0,04
D300_GARE	-0,08	-0,17	0,02	0,11	0,08	0,07	0,02	0,00	0,07	0,04	0,22	0,23	0,60	0,44	0,65	0,01	-0,23	0,17	-0,19	0,09	-0,05	-0,24	0,08	0,02	0,06	0,08
D500_GARE	-0,13	-0,24	-0,01	0,07	0,06	0,06	0,33	-0,28	-0,30	-0,30	0,32	0,27	0,05	0,30	0,27	0,31	-0,31	0,35	-0,27	0,33	0,28	-0,24	0,33	-0,16	-0,16	-0,07
D700_GARE	0,00	-0,15	-0,03	0,06	0,04	0,03	0,22	-0,23	-0,16	-0,20	0,20	0,23	0,42	0,36	0,37	0,23	-0,10	0,10	0,08	0,17	0,17	-0,22	0,25	-0,06	-0,06	0,01
D1500_GARE	-0,10	-0,21	0,00	0,07	0,05	0,11	0,43	-0,34	-0,33	-0,34	0,29	0,30	0,30	0,39	0,28	0,43	-0,27	0,35	-0,10	0,38	0,42	-0,25	0,38	-0,19	-0,21	-0,09
D100_BUS	-0,12	-0,27	-0,01	0,16	0,10	0,08	0,31	-0,24	-0,24	-0,24	0,39	0,25	0,01	0,30	0,24	0,30	-0,25	0,28	-0,14	0,32	0,27	-0,26	0,25	-0,15	-0,13	-0,05
D300_BUS	-0,03	0,00	-0,02	-0,05	-0,02	-0,05	-0,04	0,02	0,02	0,02	-0,05	0,06	0,02	0,02	-0,03	-0,07	-0,03	-0,01	-0,07	-0,03	-0,04	-0,06	-0,05	-0,05	-0,03	-0,01
D500_BUS	0,15	0,28	0,03	-0,13	-0,08	-0,06	-0,27	0,21	0,20	0,21	-0,40	-0,31	-0,01	-0,34	-0,20	-0,23	0,24	-0,26	0,17	-0,32	-0,24	0,33	-0,18	0,17	0,11	0,05
D700_BUS	0,21	0,31	0,03	-0,14	-0,10	-0,06	-0,27	0,20	0,19	0,20	-0,41	-0,40	-0,09	-0,43	-0,21	-0,23	0,41	-0,36	0,30	-0,30	-0,26	0,50	-0,19	0,26	0,24	0,11
D300_ALIM	-0,03	-0,36	0,06	0,23	0,18	0,26	0,50	-0,32	-0,35	-0,34	0,71	0,66	0,14	0,71	0,40	0,62	-0,45	0,54	-0,08	0,60	0,65	-0,34	0,61	-0,22	-0,20	-0,18
D500_ALIM	-0,03	-0,21	0,03	0,09	0,09	0,13	0,35	-0,24	-0,21	-0,23	0,32	0,49	0,15	0,46	0,28	0,35	-0,26	0,29	-0,04	0,45	0,44	-0,24	0,33	-0,33	-0,36	-0,10
D700_ALIM	-0,05	-0,15	-0,09	-0,06	-0,03	0,00	0,24	-0,23	-0,20	-0,22	0,12	0,25	0,09	0,22	0,20	0,19	-0,13	0,12	-0,04	0,21	0,23	-0,14	0,22	-0,23	-0,26	-0,06
D1500_ALIM	-0,04	-0,05	-0,07	-0,14	-0,08	-0,07	0,17	-0,24	-0,19	-0,22	-0,03	-0,02	-0,02	-0,03	0,10	0,10	0,02	0,02	-0,06	-0,02	0,07	0,00	0,08	-0,19	-0,24	0,02
D300_BOUL	-0,03	-0,32	0,08	0,20	0,17	0,24	0,34	-0,22	-0,27	-0,25	0,65	0,64	0,06	0,65	0,29	0,51	-0,40	0,44	-0,11	0,52	0,53	-0,29	0,45	-0,22	-0,19	-0,15
D500_BOUL	-0,03	-0,22	0,02	0,06	0,07	0,10	0,27	-0,23	-0,22	-0,23	0,27	0,42	0,10	0,38	0,24	0,30	-0,26	0,27	-0,06	0,37	0,34	-0,24	0,26	-0,25	-0,28	-0,09
D700_BOUL	-0,07	-0,16	-0,07	-0,02	-0,03	-0,01	0,22	-0,24	-0,20	-0,23	0,14	0,26	0,05	0,22	0,15	0,19	-0,16	0,13	-0,10	0,16	0,23	-0,19	0,18	-0,23	-0,27	-0,06
D1500_BOUL	-0,07	-0,09	-0,08	-0,11	-0,06	-0,06	0,24	-0,27	-0,25	-0,26	0,01	0,03	0,04	0,03	0,14	0,16	-0,03	0,09	-0,09	0,03	0,16	-0,03	0,17	-0,20	-0,25	0,02
D300_PHARM	-0,07	-0,35	0,05	0,22	0,18	0,22	0,37	-0,23	-0,28	-0,26	0,69	0,69	0,12	0,71	0,26	0,49	-0,43	0,46	-0,18	0,57	0,56	-0,31	0,47	-0,27	-0,24	-0,22
D500_PHARM	-0,02	-0,20	-0,01	0,04	0,06	0,08	0,26	-0,22	-0,20	-0,22	0,28	0,43	0,09	0,39	0,23	0,30	-0,26	0,26	-0,09	0,32	0,39	-0,22	0,30	-0,27	-0,31	-0,10
D700_PHARM	-0,07	-0,17	-0,06	-0,02	0,02	0,02	0,24	-0,24	-0,20	-0,22	0,15	0,30	0,04	0,24	0,15	0,23	-0,16	0,15	-0,07	0,21	0,27	-0,20	0,18	-0,27	-0,30	-0,04
D1500_PHAR	-0,04	-0,07	-0,06	-0,12	-0,09	-0,07	0,21	-0,26	-0,23	-0,25	-0,05	-0,03	0,02	-0,03	0,16	0,13	0,00	0,05	-0,04	0,00	0,09	-0,03	0,12	-0,17	-0,22	0,03
D_PT_PARC	0,01	-0,34	0,09	0,23	0,13	0,25	0,64	-0,43	-0,57	-0,51	0,62	0,50	-0,10	0,50	0,46	0,95	-0,45	0,57	0,10	0,73	0,76	-0,37	0,65	-0,22	-0,26	0,01
D_MO_PARC	0,09	-0,21	0,09	0,08	0,03	0,20	0,58	-0,48	-0,48	-0,49	0,29	0,17	0,30	0,32	0,63	0,66	-0,17	0,20	0,36	0,32	0,45	-0,28	0,41	-0,24	-0,30	0,21
D_GD_PARC	-0,06	-0,34	-0,02	0,05	0,08	0,16	0,68	-0,50	-0,51	-0,51	0,47	0,39	0,00	0,41	0,42	0,63	-0,29	0,45	-0,04	0,64	0,60	-0,32	0,56	-0,05	-0,03	-0,14
D100_BLEU	-0,06	-0,25	0,10	0,14	0,16	0,21	0,29	-0,17	-0,19	-0,18	0,36	0,30	0,16	0,38	0,28	0,28	-0,33	0,36	-0,15	0,23	0,27	-0,24	0,36	0,58	0,47	-0,11
D300_BLEU	-0,12	-0,22	0,05	0,21	0,21	0,12	0,13	-0,06	-0,08	-0,07	0,24	0,07	0,18	0,20	0,15	0,09	-0,31	0,34	-0,27	0,12	0,04	-0,28	0,14	0,29	0,42	-0,02
D500_BLEU	-0,12	-0,26	0,05	0,15	0,17	0,19	0,32	-0,21	-0,23	-0,23	0,54	0,33	0,25	0,50	0,31	0,31	-0,38	0,49	-0,25	0,41	0,26	-0,30	0,29	0,00	0,13	-0,19

D700_BLEU	-0,12	-0,26	0,01	0,12	0,10	0,13	0,29	-0,17	-0,21	-0,20	0,33	0,30	0,11	0,34	0,17	0,28	-0,38	0,45	-0,30	0,38	0,33	-0,30	0,26	0,01	-0,01	-0,04
D100_VERT	0,11	0,22	-0,04	-0,04	-0,05	-0,05	-0,35	0,31	0,42	0,37	-0,32	-0,39	-0,05	-0,37	-0,22	-0,30	0,20	-0,20	0,15	-0,32	-0,31	0,05	-0,40	0,52	0,47	0,62
D300_VERT	0,03	0,14	-0,05	-0,05	-0,01	-0,10	-0,30	0,25	0,38	0,32	-0,30	-0,25	-0,09	-0,30	-0,13	-0,30	0,20	-0,18	0,11	-0,22	-0,30	0,09	-0,35	0,30	0,41	0,49
D500_VERT	0,00	0,12	-0,03	0,00	-0,01	-0,11	-0,21	0,18	0,32	0,26	-0,21	-0,12	-0,12	-0,20	-0,16	-0,27	0,11	-0,12	0,05	-0,17	-0,21	0,06	-0,28	-0,05	0,10	0,24
D700_VERT	0,00	0,12	-0,01	-0,01	0,00	-0,04	-0,14	0,15	0,27	0,21	-0,12	0,01	-0,08	-0,07	-0,13	-0,22	0,05	-0,08	0,02	-0,07	-0,11	0,02	-0,20	-0,07	-0,12	0,06
D100_EXTRA	-0,20	0,14	0,06	0,13	-0,07	0,03	-0,17	0,16	0,20	0,18	-0,26	-0,26	0,00	-0,25	-0,11	-0,19	-0,28	0,07	-0,31	-0,21	-0,20	-0,06	-0,14	0,41	0,37	0,33
D300_EXTRA	-0,03	0,09	-0,02	0,01	0,02	-0,05	-0,16	0,14	0,19	0,17	-0,23	-0,06	-0,06	-0,15	-0,12	-0,19	-0,04	-0,06	-0,15	-0,17	-0,15	0,07	-0,19	0,18	0,27	0,19
D500_EXTRA	-0,06	0,05	0,02	0,04	0,07	-0,05	-0,11	0,08	0,16	0,12	-0,22	-0,04	-0,06	-0,13	-0,06	-0,16	-0,01	-0,06	-0,16	-0,14	-0,09	-0,01	-0,15	-0,08	0,05	0,26
D700_EXTRA	-0,04	0,11	0,00	0,00	0,02	-0,06	-0,05	0,03	0,12	0,08	-0,16	-0,09	-0,07	-0,14	-0,01	-0,13	-0,03	0,04	-0,20	-0,12	-0,09	-0,06	-0,09	0,02	0,07	0,24
D100_INDUS	-0,25	-0,28	-0,10	0,14	0,15	-0,05	0,30	-0,36	-0,28	-0,32	0,30	-0,01	0,13	0,16	0,34	0,15	-0,34	0,39	-0,38	0,11	0,05	-0,23	0,29	0,35	0,31	-0,21
D300_INDUS	-0,18	-0,26	-0,04	0,08	0,08	0,00	0,24	-0,27	-0,20	-0,24	0,36	0,18	0,15	0,30	0,21	0,13	-0,28	0,32	-0,31	0,26	0,08	-0,31	0,22	0,14	0,21	-0,14
D500_INDUS	-0,13	-0,21	-0,03	0,06	0,09	0,05	0,22	-0,21	-0,21	-0,21	0,29	0,22	0,16	0,30	0,15	0,13	-0,22	0,24	-0,25	0,23	0,18	-0,23	0,24	-0,31	-0,15	-0,08
D700_INDUS	-0,08	-0,11	-0,02	-0,01	0,05	0,01	0,21	-0,16	-0,19	-0,18	0,09	0,21	0,15	0,21	0,17	0,17	-0,15	0,16	-0,14	0,25	0,23	-0,15	0,17	-0,27	-0,27	0,03
D100_SAR	-0,15	-0,19	-0,01	0,12	0,08	0,06	0,22	-0,21	-0,23	-0,22	0,17	0,16	0,00	0,16	0,14	0,23	-0,32	0,34	-0,33	0,16	0,22	-0,19	0,23	-0,02	-0,06	0,04
D300_SAR	-0,18	-0,32	-0,04	0,15	0,15	0,07	0,37	-0,30	-0,32	-0,31	0,40	0,30	0,07	0,36	0,31	0,36	-0,40	0,45	-0,33	0,32	0,34	-0,28	0,37	-0,11	-0,10	-0,06
D500_SAR	-0,13	-0,24	-0,01	0,07	0,06	0,06	0,33	-0,28	-0,30	-0,30	0,32	0,27	0,05	0,30	0,27	0,31	-0,31	0,35	-0,27	0,33	0,28	-0,24	0,33	-0,16	-0,16	-0,07
D700_SAR	-0,10	-0,19	-0,04	0,04	0,03	-0,01	0,24	-0,21	-0,23	-0,23	0,16	0,21	0,04	0,20	0,15	0,20	-0,18	0,22	-0,20	0,25	0,23	-0,16	0,23	-0,23	-0,25	-0,01
BFA_ROAD	-0,03	-0,13	0,02	-0,01	0,00	0,08	0,32	-0,34	-0,37	-0,36	0,25	0,05	0,10	0,17	0,37	0,28	-0,14	0,17	-0,04	0,16	0,16	-0,10	0,42	-0,04	-0,02	-0,15
BM_ROAD	0,00	-0,10	-0,06	-0,06	-0,07	0,02	0,21	-0,26	-0,29	-0,28	0,11	-0,02	0,05	0,05	0,24	0,12	-0,09	0,07	-0,09	0,01	0,00	0,06	0,39	0,05	0,03	-0,16
BFO_ROAD	-0,03	-0,11	-0,04	-0,04	-0,01	0,04	0,09	-0,13	-0,13	-0,13	0,15	-0,02	0,00	0,05	0,08	0,03	-0,06	0,10	-0,10	-0,07	-0,02	0,00	0,19	0,07	0,07	-0,07
BTFO_ROAD	-0,08	0,00	0,04	0,10	0,06	-0,04	0,00	0,03	-0,02	0,00	-0,02	-0,11	0,07	-0,05	-0,01	0,00	0,11	0,01	0,02	-0,03	-0,08	0,14	0,16	0,27	0,26	-0,01
BFA_TRAIN	0,00	0,02	0,01	0,06	0,05	0,00	-0,13	0,13	0,08	0,10	0,00	0,02	0,25	0,10	0,01	-0,14	-0,14	0,03	-0,17	-0,17	-0,14	0,01	-0,12	0,00	0,04	0,01
BM_TRAIN	-0,11	0,09	0,05	0,14	0,03	-0,05	-0,17	0,15	0,15	0,15	-0,03	-0,01	0,22	0,06	-0,03	-0,16	-0,04	0,02	-0,17	-0,12	-0,19	-0,12	-0,09	0,04	0,13	0,15
BFO_TRAIN	-0,10	-0,03	0,02	0,08	0,19	-0,07	-0,22	0,18	0,22	0,20	-0,13	-0,02	0,30	0,04	-0,10	-0,22	-0,06	-0,11	-0,18	-0,13	-0,21	-0,11	-0,18	0,15	0,27	0,28
BTFO_TRAIN	-0,16	0,14	0,18	0,18	-0,01	0,02	-0,31	0,37	0,38	0,38	-0,22	0,02	0,32	0,03	-0,14	-0,28	-0,25	0,21	-0,24	-0,17	-0,18	-0,05	-0,22	0,28	0,25	0,23
BFA_AVION	0,04	0,21	0,05	-0,05	-0,06	-0,05	-0,23	0,10	0,06	0,08	-0,17	-0,19	0,10	-0,14	-0,10	-0,28	0,20	-0,23	-0,02	-0,22	-0,27	0,15	0,00	0,07	0,04	-0,26
BM_AVION	0,09	0,18	0,06	-0,07	-0,06	-0,01	-0,15	-0,03	-0,05	-0,04	-0,14	-0,27	0,06	-0,18	-0,03	-0,23	0,19	-0,23	0,01	-0,24	-0,24	0,18	0,11	0,05	0,12	-0,25
BFO_AVION	0,09	0,32	0,07	-0,01	-0,06	-0,03	-0,23	0,05	0,04	0,05	-0,28	-0,28	-0,02	-0,28	-0,11	-0,26	0,20	-0,24	0,04	-0,30	-0,28	0,27	-0,01	0,20	0,16	-0,20
BTFO_AVION	0,04	0,17	0,12	-0,06	0,08	-0,12	-0,15	-0,02	-0,03	-0,03	-0,24	-0,22	-0,03	-0,23	-0,06	-0,18	0,09	-0,11	0,04	-0,20	-0,19	0,20	0,03	0,15	0,10	-0,24

Annexe 8b. Matrice de corrélation bisériale (Immoweb ajusté)

	PRIX	LOG_SURF_FONC	LOG_SURF_BATI	AGE_BIEN	NB_PIECE	ACC_POT_EMPLOI	ISO_GUILLEMIN	ISO_STLAMBERT	ISO_NEV	ACC_BUS	ACC_ML	ACC_TRAIN	ACC_TOT	ACC_POT_ARRET	ACC_POT_GARE	RMD_2011	TCH_2011	NINS_2011	D_POPSS_2011	DPOPIX_2011	CAFT81_2011	ACC_POT_COM	SHANNON_300	SHANNON_500	KIEMSTEDT
N2_FACADE	-0,54	-0,69	-0,23	0,32	0,06	0,57	-0,50	-0,47	-0,50	0,38	0,33	0,04	0,37	0,51	0,41	-0,39	0,39	-0,24	0,46	0,49	-0,34	0,50	-0,40	-0,27	0,00
N3_FACADE	0,04	0,04	0,11	-0,01	-0,01	-0,10	0,07	0,08	0,07	-0,03	0,03	0,09	0,04	-0,14	-0,05	-0,04	-0,02	-0,09	-0,07	-0,13	0,00	-0,11	0,08	-0,02	-0,05
N4_FACADE	0,67	0,85	0,23	-0,41	-0,06	-0,63	0,58	0,55	0,58	-0,46	-0,46	-0,12	-0,51	-0,53	-0,46	0,58	-0,52	0,42	-0,52	-0,50	0,44	-0,53	0,45	0,34	0,04
N0_ETAGE	0,19	0,39	-0,12	-0,43	-0,23	-0,31	0,24	0,21	0,23	-0,33	-0,39	-0,22	-0,45	-0,30	-0,25	0,58	-0,35	0,34	-0,33	-0,28	0,48	-0,27	0,20	0,13	0,01
N1_ETAGE	0,10	0,23	-0,13	-0,15	-0,19	-0,26	0,19	0,16	0,18	-0,26	-0,18	-0,11	-0,25	-0,29	-0,31	0,14	-0,17	-0,03	-0,27	-0,27	0,09	-0,23	0,15	-0,02	-0,01
N2_ETAGE	-0,16	-0,36	0,18	0,29	0,27	0,37	-0,27	-0,24	-0,26	0,37	0,31	0,18	0,40	0,40	0,40	-0,33	0,29	-0,08	0,38	0,37	-0,25	0,32	-0,22	-0,02	0,01
MANSARDE	0,15	-0,02	0,18	-0,02	0,17	0,05	-0,03	-0,01	-0,02	0,11	0,05	0,11	0,11	0,09	0,09	-0,03	0,04	0,10	0,05	0,07	-0,05	0,08	0,02	0,12	-0,11
RENOVATION	-0,08	-0,13	-0,14	0,26	-0,14	0,09	-0,09	-0,08	-0,09	0,14	0,01	-0,11	0,02	0,12	0,05	-0,20	0,22	-0,14	0,05	0,02	-0,20	0,10	-0,04	-0,02	-0,03
N0_GARAGE	-0,56	-0,55	-0,34	0,53	0,04	0,37	-0,31	-0,26	-0,29	0,34	0,29	0,06	0,33	0,38	0,27	-0,39	0,41	-0,20	0,35	0,35	-0,31	0,29	-0,27	-0,17	0,12
N1_GARAGE	0,35	0,34	0,16	-0,42	-0,05	-0,21	0,15	0,14	0,15	-0,20	-0,19	-0,07	-0,22	-0,23	-0,15	0,29	-0,28	0,17	-0,22	-0,23	0,22	-0,17	0,15	0,04	-0,14
N2_GARAGE	0,43	0,43	0,35	-0,26	0,01	-0,32	0,32	0,25	0,29	-0,29	-0,21	0,00	-0,24	-0,31	-0,24	0,21	-0,28	0,08	-0,27	-0,25	0,20	-0,23	0,23	0,23	0,01
N0_SDB	-0,35	-0,16	-0,31	0,23	0,09	0,04	-0,10	-0,07	-0,09	-0,02	-0,08	-0,11	-0,10	0,02	-0,04	0,00	0,05	-0,07	-0,03	-0,02	-0,02	-0,04	-0,11	-0,01	0,09
N1_SDB	0,04	0,10	0,00	-0,07	-0,09	-0,14	0,16	0,17	0,17	-0,02	0,05	0,10	0,06	-0,15	-0,08	0,06	-0,05	0,05	-0,04	-0,08	-0,02	-0,11	0,15	0,03	-0,02
N2_SDB	0,28	-0,01	0,31	-0,11	0,06	0,20	-0,17	-0,21	-0,19	0,05	0,00	-0,06	0,00	0,23	0,17	-0,10	0,03	-0,01	0,10	0,15	0,05	0,22	-0,14	-0,04	-0,05
N1_CHB	-0,53	-0,17	-0,54	0,30	-0,39	0,00	-0,06	-0,07	-0,07	-0,07	-0,11	0,05	-0,07	-0,01	0,00	-0,11	0,17	-0,21	0,00	-0,05	-0,02	-0,06	0,05	0,08	0,21
N2_CHB	-0,42	-0,20	-0,61	0,10	-0,35	0,05	-0,09	-0,03	-0,06	-0,04	0,03	0,00	0,00	0,04	0,01	-0,05	0,06	-0,07	-0,01	0,02	-0,10	0,01	-0,13	-0,11	0,19
N3_CHB	0,03	-0,04	0,01	-0,09	-0,12	-0,07	0,07	0,06	0,07	-0,14	-0,04	0,03	-0,07	-0,14	-0,08	0,07	-0,09	-0,04	-0,07	-0,10	0,05	-0,06	0,08	0,09	-0,07
N4_CHB	0,44	0,25	0,64	-0,04	0,53	0,05	0,00	-0,04	-0,02	0,22	0,05	-0,05	0,10	0,14	0,09	-0,03	0,03	0,15	0,10	0,11	0,03	0,08	0,00	-0,04	-0,11
CAVE	-0,10	-0,15	0,11	0,14	0,26	0,22	-0,21	-0,20	-0,21	0,18	0,19	0,10	0,22	0,17	0,18	-0,22	0,17	-0,17	0,19	0,15	-0,27	0,17	-0,17	-0,12	-0,11

D300_GARE	0,00	-0,20	0,17	0,02	0,12	0,08	0,00	0,07	0,03	0,22	0,27	0,61	0,49	0,02	0,38	-0,34	0,29	-0,22	0,24	-0,02	-0,27	0,09	0,04	0,02	-0,03
D500_GARE	-0,29	-0,32	-0,14	0,18	0,16	-0,09	0,13	0,12	0,13	0,14	0,22	0,62	0,43	-0,02	0,06	-0,18	0,12	-0,04	0,08	-0,05	-0,25	-0,01	0,00	0,01	0,19
D700_GARE	-0,13	-0,20	-0,05	0,14	0,07	0,30	-0,22	-0,19	-0,21	0,26	0,23	0,45	0,41	0,35	0,53	-0,20	0,19	0,09	0,22	0,24	-0,20	0,32	-0,22	-0,10	0,01
D1500_GARE	-0,12	-0,15	-0,07	0,08	-0,01	0,36	-0,29	-0,24	-0,27	0,16	0,15	0,39	0,30	0,32	0,30	-0,18	0,21	-0,06	0,23	0,28	-0,20	0,35	-0,29	-0,04	0,01
D100_BUS	-0,21	-0,19	-0,03	0,17	0,11	0,18	-0,16	-0,12	-0,14	0,27	0,20	-0,04	0,21	0,16	0,14	-0,16	0,15	-0,04	0,24	0,12	-0,12	0,09	-0,09	-0,07	0,07
D300_BUS	0,05	-0,04	0,02	-0,06	-0,02	0,02	0,00	-0,01	-0,01	0,08	0,05	0,04	0,08	0,05	0,04	-0,03	0,02	-0,05	0,02	0,06	-0,17	0,02	-0,05	0,02	-0,05
D500_BUS	0,21	0,25	0,03	-0,17	-0,17	-0,18	0,11	0,08	0,10	-0,39	-0,22	0,03	-0,27	-0,21	-0,17	0,16	-0,19	0,08	-0,29	-0,16	0,23	-0,05	0,13	-0,05	-0,11
D700_BUS	0,17	0,25	0,02	-0,06	0,03	-0,26	0,25	0,16	0,21	-0,48	-0,42	-0,11	-0,49	-0,25	-0,24	0,31	-0,28	0,16	-0,34	-0,26	0,62	-0,15	0,23	0,26	0,08
D300_ALIM	-0,10	-0,32	0,13	0,25	0,26	0,39	-0,23	-0,24	-0,24	0,75	0,68	0,14	0,76	0,49	0,22	-0,50	0,54	-0,19	0,60	0,60	-0,37	0,51	-0,37	-0,36	-0,21
D500_ALIM	-0,10	-0,22	0,03	0,03	0,21	0,36	-0,30	-0,25	-0,28	0,24	0,41	0,12	0,38	0,39	0,35	-0,29	0,30	0,02	0,40	0,47	-0,22	0,38	-0,26	-0,23	-0,05
D700_ALIM	-0,10	-0,28	-0,03	-0,07	0,08	0,23	-0,20	-0,14	-0,18	0,17	0,28	0,07	0,26	0,15	0,23	-0,12	0,06	-0,08	0,28	0,22	-0,19	0,19	-0,17	-0,22	-0,04
D1500_ALIM	0,00	-0,05	-0,05	-0,15	-0,10	0,23	-0,26	-0,23	-0,25	-0,03	0,01	-0,11	-0,05	0,17	0,13	0,02	0,03	-0,04	-0,01	0,12	-0,02	0,11	-0,25	-0,24	0,07
D300_BOUL	-0,08	-0,32	0,18	0,23	0,26	0,26	-0,15	-0,20	-0,18	0,63	0,66	0,08	0,67	0,38	0,15	-0,47	0,45	-0,16	0,55	0,52	-0,29	0,34	-0,33	-0,32	-0,18
D500_BOUL	-0,08	-0,17	0,02	0,06	0,10	0,22	-0,20	-0,19	-0,20	0,19	0,29	-0,09	0,22	0,29	0,16	-0,24	0,21	-0,08	0,25	0,28	-0,21	0,24	-0,14	-0,17	-0,03
D700_BOUL	-0,11	-0,24	-0,05	-0,10	0,01	0,18	-0,17	-0,13	-0,15	0,13	0,24	0,01	0,20	0,11	0,12	-0,06	0,04	-0,06	0,12	0,19	-0,23	0,10	-0,09	-0,27	0,04
D1500_BOUL	0,03	-0,04	-0,08	-0,18	-0,07	0,30	-0,30	-0,30	-0,30	0,01	0,04	-0,03	0,02	0,25	0,21	-0,06	0,08	-0,03	0,08	0,23	-0,06	0,23	-0,31	-0,26	0,04
D300_PHARM	-0,07	-0,28	0,19	0,23	0,22	0,27	-0,15	-0,18	-0,17	0,69	0,73	0,09	0,74	0,37	0,10	-0,51	0,51	-0,29	0,56	0,57	-0,34	0,35	-0,32	-0,34	-0,24
D500_PHARM	-0,12	-0,26	-0,01	0,02	0,08	0,29	-0,21	-0,24	-0,23	0,27	0,41	0,03	0,36	0,38	0,17	-0,23	0,20	-0,04	0,31	0,40	-0,16	0,34	-0,19	-0,35	-0,10
D700_PHARM	-0,03	-0,13	-0,03	-0,15	0,06	0,15	-0,13	-0,07	-0,10	0,09	0,35	0,04	0,25	0,10	0,14	-0,13	0,08	-0,07	0,24	0,24	-0,23	0,04	-0,16	-0,27	0,08
D1500_PHAR	-0,03	-0,12	-0,08	-0,14	-0,08	0,32	-0,36	-0,34	-0,36	-0,01	-0,07	-0,08	-0,08	0,25	0,24	-0,04	0,09	-0,04	0,00	0,13	-0,09	0,23	-0,32	-0,19	0,09
D_PT_PARC	-0,12	-0,36	-0,05	0,41	0,12	0,66	-0,41	-0,56	-0,49	0,72	0,46	-0,06	0,54	1,00	0,57	-0,37	0,45	0,20	0,48	0,70	-0,36	0,72	-0,54	-0,15	0,07
D_MO_PARC	-0,02	-0,29	0,02	0,00	0,25	0,55	-0,46	-0,47	-0,47	0,16	0,11	0,31	0,25	0,67	0,91	-0,12	0,14	0,42	0,25	0,53	-0,27	0,39	-0,59	-0,28	0,17
D_GD_PARC	-0,05	-0,31	-0,06	0,05	0,18	0,61	-0,48	-0,48	-0,49	0,44	0,31	0,04	0,38	0,64	0,49	-0,30	0,40	0,06	0,67	0,54	-0,29	0,53	-0,62	0,02	-0,17
D100_BLEU	-0,08	-0,09	-0,08	0,27	0,05	0,00	-0,02	0,07	0,02	0,04	-0,09	0,29	0,08	-0,06	0,13	-0,22	0,23	-0,30	0,05	-0,20	-0,27	-0,01	0,27	0,72	0,19
D300_BLEU	-0,03	-0,12	0,08	0,13	0,25	0,03	0,00	-0,03	-0,01	0,23	0,06	0,41	0,29	0,05	0,11	-0,23	0,19	-0,14	0,05	-0,01	-0,19	0,08	0,08	0,48	0,02
D500_BLEU	-0,20	-0,10	0,06	0,15	0,01	0,06	-0,03	-0,02	-0,03	0,25	0,11	0,27	0,27	0,01	0,02	-0,22	0,26	-0,22	0,11	-0,06	-0,26	0,08	0,03	0,22	-0,08
D700_BLEU	-0,17	-0,10	-0,26	0,07	0,16	0,03	0,08	0,01	0,04	0,19	0,14	0,14	0,21	0,03	-0,04	-0,33	0,40	-0,40	0,20	0,12	-0,31	0,03	-0,03	0,13	-0,05
D100_VERT	-0,06	0,08	-0,18	-0,06	-0,14	-0,30	0,14	0,26	0,21	-0,28	-0,45	0,00	-0,38	-0,26	-0,20	0,12	-0,08	0,05	-0,35	-0,35	-0,10	-0,32	0,35	0,60	0,62
D300_VERT	0,05	0,12	0,02	0,07	-0,15	-0,29	0,20	0,35	0,28	-0,25	-0,16	-0,03	-0,21	-0,28	-0,13	0,08	-0,18	0,10	-0,16	-0,25	0,02	-0,32	0,23	0,35	0,44



D500_VERT	-0,21	-0,17	-0,09	-0,07	-0,05	0,00	-0,05	0,15	0,05	0,01	0,12	-0,17	0,01	-0,10	-0,07	-0,08	0,04	-0,18	0,08	0,00	-0,14	-0,17	0,00	-0,09	0,27
D700_VERT	-0,02	0,12	-0,13	0,03	0,06	-0,10	0,10	0,25	0,18	0,00	0,05	-0,01	0,03	-0,16	-0,06	0,03	0,07	0,02	0,04	-0,02	-0,06	-0,15	0,11	-0,17	0,08
D100_EXTRA	-0,38	-0,06	-0,48	0,05	-0,21	0,23	-0,23	-0,15	-0,20	-0,28	-0,13	-0,18	-0,26	0,11	0,05	-0,20	0,25	-0,30	0,02	0,05	0,14	0,06	-0,41	0,24	0,26
D300_EXTRA	-0,05	0,25	-0,12	0,04	-0,03	-0,04	-0,01	0,05	0,02	-0,24	-0,13	-0,10	-0,21	-0,09	-0,05	-0,11	0,05	-0,28	-0,07	-0,04	0,15	-0,13	0,01	0,25	0,17
D500_EXTRA	-0,14	0,07	-0,14	-0,03	-0,09	-0,11	0,04	0,14	0,09	-0,21	0,01	-0,17	-0,14	-0,14	-0,10	0,00	-0,03	-0,14	-0,13	-0,09	0,17	-0,17	-0,04	0,03	0,16
D700_EXTRA	-0,21	-0,11	-0,06	0,09	0,31	-0,04	0,06	0,15	0,10	-0,05	-0,07	-0,06	-0,08	-0,10	-0,02	-0,05	0,09	-0,09	-0,16	-0,12	-0,03	-0,02	-0,11	0,02	0,15
D100_INDUS	-0,15	-0,14	0,06	-0,02	-0,02	0,21	-0,25	-0,19	-0,22	0,13	-0,01	0,17	0,11	0,06	0,27	-0,22	0,17	-0,28	0,09	0,01	-0,15	0,22	0,11	0,35	-0,22
D300_INDUS	-0,12	-0,27	0,04	0,05	0,10	0,18	-0,24	-0,16	-0,20	0,30	0,10	0,15	0,24	0,09	0,11	-0,27	0,27	-0,23	0,14	0,02	-0,28	0,17	0,05	0,35	-0,06
D500_INDUS	-0,18	-0,21	-0,14	0,02	0,04	0,14	-0,11	-0,08	-0,10	0,24	0,18	0,25	0,30	0,03	0,04	-0,31	0,35	-0,33	0,20	0,08	-0,27	0,13	-0,09	-0,11	-0,01
D700_INDUS	-0,07	-0,13	-0,14	-0,01	0,06	0,24	-0,17	-0,24	-0,21	0,06	0,15	0,10	0,15	0,26	0,17	-0,07	0,05	-0,06	0,19	0,25	0,00	0,25	-0,20	-0,26	-0,14
D100_SAR	-0,21	-0,10	-0,15	0,08	-0,10	0,12	-0,14	-0,17	-0,16	0,00	0,00	0,03	0,01	0,12	-0,02	-0,29	0,32	-0,35	0,03	0,04	-0,15	0,12	-0,08	0,00	0,05
D300_SAR	-0,23	-0,29	-0,13	0,13	0,17	0,32	-0,25	-0,25	-0,26	0,30	0,25	0,02	0,28	0,33	0,20	-0,34	0,35	-0,29	0,32	0,31	-0,20	0,34	-0,26	-0,04	0,10
D500_SAR	-0,11	-0,25	0,04	0,08	0,09	0,28	-0,22	-0,22	-0,23	0,24	0,29	0,11	0,31	0,23	0,18	-0,28	0,29	-0,23	0,28	0,16	-0,15	0,23	-0,20	-0,12	-0,15
D700_SAR	-0,19	-0,20	-0,10	0,11	0,02	0,25	-0,19	-0,16	-0,18	0,27	0,22	0,10	0,28	0,19	0,19	-0,23	0,22	-0,20	0,24	0,23	-0,22	0,19	-0,11	-0,17	0,01
BFA_ROAD	0,01	-0,17	0,06	0,05	0,07	0,51	-0,42	-0,50	-0,47	0,33	0,06	0,07	0,20	0,52	0,50	-0,23	0,24	0,03	0,24	0,30	-0,05	0,54	-0,36	-0,05	-0,16
BMO_ROAD	-0,09	-0,18	0,00	-0,17	0,02	0,29	-0,28	-0,30	-0,30	0,15	-0,03	0,12	0,09	0,15	0,29	-0,06	0,03	-0,01	0,04	-0,01	-0,10	0,45	-0,11	0,01	-0,18
BFO_ROAD	0,07	-0,04	0,17	-0,03	0,22	0,15	-0,17	-0,18	-0,18	0,19	0,04	0,00	0,10	0,13	0,09	-0,03	0,06	-0,05	0,06	0,05	-0,04	0,27	0,10	0,15	-0,03
BTFO_ROAD	0,23	0,10	-0,28	-0,18	-0,03	0,11	-0,07	-0,29	-0,18	-0,29	-0,47	0,06	-0,37	0,00	0,01	0,29	-0,24	0,04	-0,29	-0,25	0,17	0,44	0,44	0,37	-0,22
BFA_TRAIN	-0,26	-0,29	-0,17	0,21	0,27	0,22	-0,15	-0,14	-0,15	0,30	0,18	0,28	0,34	0,29	0,22	-0,32	0,46	-0,19	0,25	0,17	-0,34	0,16	-0,18	0,17	0,06
BMO_TRAIN	-0,20	-0,21	-0,37	-0,02	-0,19	-0,14	0,06	-0,03	0,01	-0,07	-0,19	-0,02	-0,15	-0,15	-0,13	0,09	-0,24	-0,06	-0,26	-0,22	-0,01	-0,10	0,21	0,10	0,13
BFO_TRAIN	-0,24	-0,18	0,16	0,20	0,33	0,15	-0,18	-0,06	-0,13	0,35	-0,02	0,29	0,24	0,00	0,32	-0,36	0,26	-0,23	-0,14	-0,19	-0,25	0,17	0,04	0,11	0,29
BTFO_TRAIN	-0,24	-0,39	-0,28	0,12	0,35	0,23	-0,17	-0,35	-0,27	0,38	0,04	0,14	0,23	0,43	0,00	-0,34	0,32	-0,22	0,12	0,11	-0,30	0,10	-0,42	-0,03	0,36
BFA_AVION	0,19	0,23	0,27	-0,08	0,02	-0,05	-0,02	-0,07	-0,04	-0,17	-0,23	0,04	-0,19	-0,20	-0,08	0,19	-0,17	0,04	-0,18	-0,21	0,22	0,20	0,34	0,05	-0,34
BMO_AVION	-0,01	0,13	0,07	0,20	0,02	-0,11	0,02	-0,04	-0,01	-0,09	-0,23	0,04	-0,16	-0,27	-0,09	0,26	-0,36	0,14	-0,28	-0,28	0,35	0,16	0,39	0,17	-0,41
BFO_AVION	0,16	0,37	0,13	0,20	-0,07	-0,35	0,19	0,17	0,18	-0,28	-0,14	-0,04	-0,21	-0,37	-0,22	-0,01	-0,21	0,06	-0,33	-0,38	0,18	-0,17	0,46	0,30	-0,11
BTFO_AVION	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

## Annexe 9. Modélisation hédonique du prix

### Annexe 9a. Modèle A, inspiration du modèle de Dethier

#### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : SPF\_95\_TRAITEMENT  
 Dependent Variable : PRIX Number of Observations: 4953  
 Mean dependent var : 142457 Number of Variables : 16  
 S.D. dependent var : 63681.7 Degrees of Freedom : 4937

R-squared : 0.419754 F-statistic : 238.097  
 Adjusted R-squared : 0.417991 Prob(F-statistic) : 0\*  
 Sum squared residual: 1.16549e+13 Log likelihood : -60468.4  
 Sigma-square : 2.36073e+09 Akaike info criterion : 120969  
 S.E. of regression : 48587.4 Schwarz criterion : 121073  
 Sigma-square ML : 2.35311e+09  
 S.E of regression ML: 48508.8

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	76370.5	7339.7	10.4051	0.00000*
RENOVATION	8839.61	1621.7	5.45083	0.00000*
LOG_SURF_F	16610.9	1602.66	10.3646	0.00000*
LOG_SURF_B	23464.4	1733.76	13.5338	0.00000*
N4_FACADE	16144.1	2392.13	6.74885	0.00000*
N0_ETAGE	413.124	2296.67	0.179879	0.85723
N2_SDB	14599.9	4734.71	3.08359	0.00206*
N1_LOGEMEN	-18176.9	4141.79	-4.38866	0.00001*
N4_LOGEMEN	19837.4	8102.16	2.4484	0.01438*
AGE_BIEN	-515	20.2393	-25.4455	0.00000*
ACC_EMPLOI	0.57143	0.198081	2.88483	0.00393*
NIV_INSTRU	1641.66	76.4579	21.4715	0.00000*
_ACC_ML	1193.42	203.007	5.87873	0.00000*
D700_BUS	9109.94	4238.36	2.1494	0.03165*
D300_INDUS	-1161.52	2398.15	-0.484341	0.62820
SHANNON_30	6003.3	2456.59	2.44375	0.01457*

#### REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 33.822692

#### TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	444.2698	0.00000*

#### DIAGNOSTICS FOR HETEROSKEDASTICITY

##### RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	15	710.0141	0.00000*
Koenker-Bassett test	15	410.2047	0.00000*

**SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION**

Data set : IMMOWEB\_95\_TRAITEMENT  
 Dependent Variable : PRIX Number of Observations: 509  
 Mean dependent var : 173669 Number of Variables : 16  
 S.D. dependent var : 61005.1 Degrees of Freedom : 493

R-squared : 0.698173 F-statistic : 76.0258  
 Adjusted R-squared : 0.688990 Prob(F-statistic) : 0\*  
 Sum squared residual: 5.71752e+11 Log likelihood : -6025.9  
 Sigma-square : 1.15974e+09 Akaike info criterion : 12083.8  
 S.E. of regression : 34055 Schwarz criterion : 12151.5  
 Sigma-square ML : 1.12328e+09  
 S.E of regression ML: 33515.4

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-279336	29335.3	-9.52216	0.00000*
RENOVATION	13335.8	4965.9	2.68547	0.00749*
Log_SURF_F	47784.1	5108.27	9.35427	0.00000*
Log_SURF_H	148460	12826.9	11.5741	0.00000*
N4_FACADE	21041.8	4844.7	4.34327	0.00002*
N0_ETAGE	-14864.1	7008.69	-2.12081	0.03444*
N2_SDB	16301.6	6280.07	2.59577	0.00972*
N1_CHB	-20141.1	9883.64	-2.03783	0.04210*
N4_CHB	3513.11	3958.15	0.887564	0.37521
AGE_BIEN	-403.905	45.2568	-8.92474	0.00000*
ACC_POT_EM	0.793088	0.47153	1.68195	0.09321
NIV_INSTRU	1231.42	174.475	7.05789	0.00000*
ACC_ML	72289.5	47684.7	1.51599	0.13016
D700_BUS	8113.19	7804.3	1.03958	0.29904
D300_INDUS	382.701	5166.21	0.0740778	0.94094
SHANNON_30	-1248.92	6073.62	-0.20563	0.83716

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 71.787002

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	11.4419	0.00328*

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	15	61.6410	0.00000*
Koenker-Bassett test	15	45.3317	0.00007*



## Annexe 9b. Modèle B, exploration en considérant l'ensemble des variables

### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : SPF\_95\_TRAITEMENT  
 Dependent Variable : PRIX Number of Observations: 4952  
 Mean dependent var : 142446 Number of Variables : 48  
 S.D. dependent var : 63682.9 Degrees of Freedom : 4904

R-squared : 0.492815 F-statistic : 101.384  
 Adjusted R-squared : 0.487954 Prob(F-statistic) : 0\*  
 Sum squared residual: 1.01857e+13 Log likelihood : -60123.1  
 Sigma-square : 2.07703e+09 Akaike info criterion : 120342  
 S.E. of regression : 45574.4 Schwarz criterion : 120655  
 Sigma-square ML : 2.05689e+09  
 S.E of regression ML: 45353

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-13014.8	13885.3	-0.937313	0.34865
LOG_SURF_F	8607.64	1623.99	5.30031	0.00000*
LOG_SURF_H	43051.8	4926.97	8.73798	0.00000*
N2_FACADE	-1277.73	1743.55	-0.732837	0.46371
N4_FACADE	15636.7	2306.27	6.78007	0.00000*
N0_ETAGE	13223.7	2689.1	4.9175	0.00000*
N2_ETAGE	5188.21	2620.66	1.97974	0.04779*
MANSARDE	11854.3	1732.78	6.8412	0.00000*
AGE_BIEN	-330.042	24.4955	-13.4736	0.00000*
AGE_RENO	-78.5421	20.6342	-3.8064	0.00014*
N0_GARAGE	-7372.16	1776.64	-4.1495	0.00003*
N2_GARAGE	-628.271	3071.51	-0.204548	0.83801
N0_SDB	-2127.49	1897.06	-1.12147	0.26212
N2_SDB	4835.93	4508.57	1.07261	0.28350
CHAUF_CENT	19239.8	1708.32	11.2624	0.00000*
N1_LOGEMEN	1440.32	4886.01	0.294784	0.76828
N3_LOGEMEN	7031.68	7000.95	1.00439	0.31521
N4_LOGEMEN	4272.47	8150.54	0.524194	0.60017
NB_PIECE	3919.93	470.952	8.32342	0.00000*
CAVE	3572.6	1769.92	2.01851	0.04360*
ACC_EMPLOI	0.388492	0.285647	1.36004	0.17387
ACC_BUS	743.381	343.069	2.16685	0.03030*
ACC_ML	683.062	238.995	2.85806	0.00428*
D500_GARE	-1865.77	2139.85	-0.871918	0.38334
D1500_GARE	-4338.18	1600.25	-2.71094	0.00673*
D500_BUS	3684.52	2176.88	1.69257	0.09060
D700_BUS	8223.72	4072.46	2.01935	0.04351
RMD_2011	-0.428557	0.289939	-1.4781	0.13945
TAUX_CHOMA	-181.445	189.758	-0.956194	0.33905
NIV_INSTRU	1282.69	104.194	12.3105	0.00000*
D_POP	0.0374727	0.425094	0.0881517	0.92921
CONST_AF81	191.236	54.9321	3.48131	0.00050*
D300_BLEU	9816.32	3559.1	2.75809	0.00584*
D500_BLEU	1460.53	3111.03	0.469469	0.63875
D700_BLEU	2976.45	3179.39	0.936169	0.34922
D100_VERT	5313.21	3307.7	1.60632	0.10826
D300_VERT	-2109.2	2297.91	-0.91788	0.35871
D500_VERT	-305.493	2391.56	-0.127738	0.89820
D100_INDUS	-5055.04	3017.39	-1.6753	0.09394
D300_INDUS	-3563.7	2488.21	-1.43224	0.15213

D500_INDUS	-2864.86	2283.67	-1.2545	0.20973
D100_SAR	-1470.04	2995.11	-0.490814	0.62360
D300_SAR	-1281.97	2271.01	-0.564491	0.57246
D700_SAR	-899.518	2378.27	-0.378224	0.70539
BM_TRAIN	1423.89	8712.61	0.163429	0.87045
BFO_TRAIN	4643.4	10350.2	0.448631	0.65361
BTFO_TRAIN	-28442.6	15493.5	-1.83578	0.06644
SHANNON_30	2721.26	2327.43	1.16921	0.24235
-----				
REGRESSION DIAGNOSTICS				
MULTICOLLINEARITY CONDITION NUMBER 108.237913				
TEST ON NORMALITY OF ERRORS				
TEST	DF	VALUE	PROB	
Jarque-Bera	2	917.8741	0.00000*	
DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	47	1161.2197	0.00000*	
Koenker-Bassett test	47	572.1925	0.00000*	
-----				
<b>SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION</b>				
Data set	: IMMOWEB_95_TRAITEMENT			
Dependent Variable	: PRIX	Number of Observations:	509	
Mean dependent var	: 173669	Number of Variables	: 55	
S.D. dependent var	: 61005.1	Degrees of Freedom	: 454	
R-squared	: 0.721275	F-statistic	: 21.7564	
Adjusted R-squared	: 0.688122	Prob(F-statistic)	: 0*	
Sum squared residual	: 5.2799e+11	Log likelihood	: -6005.63	
Sigma-square	: 1.16297e+09	Akaike info criterion	: 12121.3	
S.E. of regression	: 34102.4	Schwarz criterion	: 12354.1	
Sigma-square ML	: 1.03731e+09			
S.E of regression ML	: 32207.3			
-----				
Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-161412	43047.6	-3.74961	0.00020*
Log_SURF_F	46998.4	5346.75	8.7901	0.00000*
Log_SURF_H	137547	15099	9.10967	0.00000*
N2_FACADE	-11469	4012.9	-2.85803	0.00446*
N0_ETAGE	-5637.16	7524.31	-0.749193	0.45413
N2_ETAGE	-5552.85	3851.19	-1.44185	0.15003
MANSARDE	5312.38	3840.19	1.38336	0.16723
AGE_BIEN	-352.544	50.7573	-6.94568	0.00000*
N0_GARAGE	-3310.77	4113.13	-0.804927	0.42128
N2_GARAGE	5297.92	5065.73	1.04584	0.29620
N0_SDB	-8193.71	6753.1	-1.21332	0.22563
N2_SDB	15770.5	6560.28	2.40394	0.01662*
N1_CHB	-15807.7	10629.6	-1.48713	0.13767
N2_CHB	-7218.92	4664.06	-1.54778	0.12237
N4_CHB	635.106	4374.44	0.145186	0.88462
NB_PIECE	-278.795	1058.48	-0.263391	0.79238
ACC_POT_EM	0.282621	0.650577	0.434415	0.66419
ACC_BUS	41592.8	91981.6	0.452186	0.65134
ACC_ML	12351.8	63501.7	0.194511	0.84586
D500_GARE	-9037.79	7120.54	-1.26926	0.20500

D700_GARE	-5024	7954.23	-0.631613	0.52796
D1500_GARE	-2611.12	4243.88	-0.615267	0.53869
D100_BUS	-8807.83	3559.42	-2.47451	0.01371*
D500_BUS	2471.64	5589.61	0.442185	0.65858
D700_BUS	6145.94	8621.54	0.712859	0.47630
RMD_2011	-2.66489	0.99814	-2.66985	0.00786*
NIV_INSTRU	1821.29	293.153	6.21278	0.00000*
TAUX_CHOMA	-414.962	495.6	-0.837292	0.40287
D_POP_2011	0.511948	1.10254	0.464336	0.64263
CONST_AF81	-40.4148	158.018	-0.255761	0.79825
D700_BOUL	-807.528	5461.33	-0.147863	0.88251
D500_PHARM	-3301.36	5424.78	-0.608571	0.54312
D_PT_PARC	2695.39	11772.7	0.228953	0.81900
D500_BLEU	-8593.23	7952.42	-1.08058	0.28046
D700_BLEU	3270.67	8585.8	0.380939	0.70342
D500_VERT	-4224.9	5889.09	-0.717412	0.47349
D100_EXTRA	-8720.99	25722.6	-0.339039	0.73473
D500_EXTRA	-13451	10464.7	-1.28536	0.19932
D700_EXTRA	-8107.45	12431.3	-0.652178	0.51462
D100_INDUS	-9910.72	6748.4	-1.4686	0.14263
D300_INDUS	-1842.6	6286.66	-0.293098	0.76958
D500_INDUS	1767	5170.12	0.341772	0.73269
D100_SAR	-1986.93	6705.14	-0.296329	0.76713
D300_SAR	3704.22	5326.47	0.695436	0.48714
D500_SAR	3564.1	5010.45	0.711333	0.47724
D700_SAR	-139.551	5576.44	-0.025025	0.98011
BTFO_ROAD	52730.8	27578	1.91206	0.05650
BFA_TRAIN	-3422.69	9936.25	-0.344465	0.73065
BMO_TRAIN	-19105.5	22392.9	-0.853196	0.39400
BFO_TRAIN	-19132.5	21166.8	-0.903892	0.36653
BTFO_TRAIN	154.125	27082.7	0.00569092	0.99585
BFA_AVION	-1698.46	8873.02	-0.191419	0.84828
BFO_AVION	468.533	14749.7	0.0317656	0.97477
SHANNON_30	47.3604	6697.46	0.00707139	0.99413
KIEMSTEDTI	-238.06	132.458	-1.79726	0.07296

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 140.461916

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	9.0452	0.01086*

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	54	89.6966	0.00163*
Koenker-Bassett test	54	67.9776	0.09565

**Annexe 9c. Modèle C, affinage du modèle B**

**SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION**

Data set : SPF\_95\_TRAITEMENT  
 Dependent Variable : PRIX Number of Observations: 4952  
 Mean dependent var : 142446 Number of Variables : 21  
 S.D. dependent var : 63682.9 Degrees of Freedom : 4931

R-squared : 0.490508 F-statistic : 237.363  
 Adjusted R-squared : 0.488441 Prob(F-statistic) : 0\*  
 Sum squared residual: 1.02321e+13 Log likelihood : -60134.3  
 Sigma-square : 2.07505e+09 Akaike info criterion : 120311  
 S.E. of regression : 45552.7 Schwarz criterion : 120447  
 Sigma-square ML : 2.06625e+09  
 S.E of regression ML: 45456

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-23288.5	9550.95	-2.43834	0.01479*
LOG_SURF_F	8518.51	1609.76	5.2918	0.00000*
LOG_SURF_H	43792.1	4781.03	9.15955	0.00000*
N4_FACADE	15805.4	2219.85	7.12005	0.00000*
N0_ETAGE	13651	2611.73	5.2268	0.00000*
N2_ETAGE	6496.04	2484.86	2.61425	0.00897*
MANSARDE	11819.4	1716	6.88776	0.00000*
AGE_BIEN	-325.703	24.0948	-13.5175	0.00000*
AGE_RENO	-88.7068	19.7558	-4.49017	0.00001*
N0_GARAGE	-7319.53	1736.24	-4.21574	0.00003*
CHAUF_CENT	19693.6	1589.73	12.3881	0.00000*
NB_PIECE	4254.21	442.915	9.60501	0.00000*
CAVE	3344.37	1755.65	1.90492	0.05685*
ACC_BUS	713.937	278.948	2.55939	0.01051*
ACC_ML	642.907	201.891	3.18442	0.00146*
D1500_GARE	-3935.6	1540.59	-2.5546	0.01066*
NIV_INSTRU	1315.05	75.6261	17.3889	0.00000*
CONST_AF81	192.131	51.1337	3.75742	0.00017*
D300_BLEU	8264.07	3364.78	2.45605	0.01408*
D100_VERT	6503.2	3163.51	2.05569	0.03987*
BTFO_TRAIN	-28734.3	15392.4	-1.86679	0.06198

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 71.650352

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	961.0897	0.00000*

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	20	1046.2237	0.00000*
Koenker-Bassett test	20	508.7810	0.00000*

**SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION**

```

Data set           :  IMMOWEB_95_TRAITEMENT
Dependent Variable :      PRIX  Number of Observations:  509
Mean dependent var :      173669  Number of Variables   :   14
S.D. dependent var :      61005.1  Degrees of Freedom   :   495

R-squared          :      0.707588  F-statistic          :      92.1397
Adjusted R-squared :      0.699908  Prob(F-statistic)   :           0*
Sum squared residual: 5.53917e+11  Log likelihood       :     -6017.83
Sigma-square       : 1.11903e+09  Akaike info criterion :    12063.7
S.E. of regression :      33451.8  Schwarz criterion    :    12122.9
Sigma-square ML    : 1.08825e+09
S.E of regression ML:      32988.6
    
```

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-173332	30314.7	-5.71775	0.00000*
Log_SURF_F	49364.1	4471.1	11.0407	0.00000*
Log_SURF_H	141147	12509.2	11.2834	0.00000*
N2_FACADE	-11248.4	3627.74	-3.10065	0.00204*
AGE_BIEN	-380.944	43.3952	-8.77848	0.00000*
N2_SDB	16756.5	6133.17	2.73211	0.00652*
N1_CHB	-20404.4	9913.68	-2.05821	0.04009*
N2_CHB	-8625.22	4037.18	-2.13644	0.03313*
RMD_2011	-2.64257	0.627453	-4.21158	0.00003*
BTFO_ROAD	48240.8	23873.8	2.02066	0.04385*
NIV_INSTRU	1884.84	233.527	8.07117	0.00000*
D100_BUS	-8907.05	3087.09	-2.88526	0.00408*
D500_BLEU	-13496.1	7097.01	-1.90166	0.05780
KIEMSTEDTI	-303.042	110.879	-2.73308	0.00650*

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 74.346730

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	4.9798	0.08292

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	13	45.2881	0.00002*
Koenker-Bassett test	13	36.6345	0.00047*

## Annexe 9d. Modèle D, exclusion des résidus

### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : SPF\_95\_TRAITEMENT\_RESIDU3  
 Dependent Variable : PRIX Number of Observations: 4882  
 Mean dependent var : 142822 Number of Variables : 21  
 S.D. dependent var : 62494.5 Degrees of Freedom : 4861

R-squared : 0.564886 F-statistic : 315.54  
 Adjusted R-squared : 0.563096 Prob(F-statistic) : 0\*  
 Sum squared residual: 8.2963e+12 Log likelihood : -58807.1  
 Sigma-square : 1.70671e+09 Akaike info criterion : 117656  
 S.E. of regression : 41312.3 Schwarz criterion : 117793  
 Sigma-square ML : 1.69937e+09  
 S.E of regression ML: 41223.4

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-36395.3	8850	-4.11246	0.00004*
LOG_SURF_F	11614.1	1477.34	7.8615	0.00000*
LOG_SURF_H	46985.4	4424.79	10.6187	0.00000*
N4_FACADE	17627.3	2035.03	8.66194	0.00000*
N0_ETAGE	13151.7	2388.48	5.50631	0.00000*
N2_ETAGE	5406.61	2270.65	2.38108	0.01730*
MANSARDE	12084	1570.18	7.69598	0.00000*
AGE_BIEN	-354.124	22.1129	-16.0144	0.00000*
AGE_RENO	-100.513	18.0403	-5.57156	0.00000*
N0_GARAGE	-7101.99	1588.73	-4.47023	0.00001*
CHAUF_CENT	18873.1	1452.03	12.9977	0.00000*
NB_PIECE	4754.46	415.291	11.4485	0.00000*
CAVE	3866.48	1608.49	2.40379	0.01627*
ACC_BUS	674.95	255.169	2.64511	0.00819*
ACC_ML	674.146	184.583	3.65227	0.00026*
D1500_GARE	-3408.23	1405.29	-2.42529	0.01533*
NIV_INSTRU	1331.39	69.478	19.1628	0.00000*
CONST_AF81	206.516	46.7952	4.41318	0.00001*
D300_BLEU	10860.5	3069.44	3.53828	0.00041*
D100_VERT	5121.5	2890.3	1.77196	0.07646
BTFO_TRAIN	-31521	13963.2	-2.25743	0.02402*

#### REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 72.660379

#### TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	103.0026	0.00000*

#### DIAGNOSTICS FOR HETEROSKEDASTICITY

##### RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	20	588.3781	0.00000*
Koenker-Bassett test	20	434.0621	0.00000*

**SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION**

Data set : SPF\_95\_TRAITEMENT\_RESIDU2  
 Dependent Variable : PRIX Number of Observations: 4655  
 Mean dependent var : 141023 Number of Variables : 21  
 S.D. dependent var : 58452.9 Degrees of Freedom : 4634

R-squared : 0.642748 F-statistic : 416.862  
 Adjusted R-squared : 0.641206 Prob(F-statistic) : 0\*  
 Sum squared residual: 5.68207e+12 Log likelihood : -55302.6  
 Sigma-square : 1.22617e+09 Akaike info criterion : 110647  
 S.E. of regression : 35016.7 Schwarz criterion : 110783  
 Sigma-square ML : 1.22064e+09  
 S.E of regression ML: 34937.6

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	42295.6	3928.04	10.7676	0.00000*
LOG_SURF_F	16300.2	1234.71	13.2016	0.00000*
LOG_SURF_B	11624.8	1401.98	8.29166	0.00000*
N4_FACADE	18811	1657.09	11.3518	0.00000*
N2_ETAGE	12939.3	1971.89	6.56188	0.00000*
MANSARDE	10000.8	1269.03	7.88068	0.00000*
AGE_BIEN	-347.119	19.0466	-18.2247	0.00000*
AGE_RENO	-88.9374	16.2683	-5.4669	0.00000*
N0_GARAGE	-10213.3	1343.41	-7.60253	0.00000*
N0_SDB	-4832.92	1481.61	-3.26194	0.00111*
CHAUF_CENT	18289.1	1340.97	13.6387	0.00000*
NB_PIECE	5198.25	336.071	15.4677	0.00000*
CAVE	6416.84	1367.61	4.69203	0.00000*
ACC_BUS	550.353	225.947	2.43576	0.01490*
ACC_ML	662.565	158.196	4.18826	0.00003*
D1500_GARE	-4339.51	1218.32	-3.56189	0.00037*
D500_BUS	3353.17	1720.87	1.94853	0.05141
D700_BUS	9023.36	3236.43	2.78806	0.00532*
NIV_INSTRU	1277.28	61.2045	20.8691	0.00000*
CONST_AF81	196.603	40.6831	4.83255	0.00000*
D300_BLEU	7290.27	2671.36	2.72905	0.00638*

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 28.362526

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	4.7455	0.09322

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	20	194.3361	0.00000*
Koenker-Bassett test	20	209.0825	0.00000*

**SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION**

Data set : IMMOWEB\_95\_TRAITEMENT\_RESIDU2  
 Dependent Variable : PRIX Number of Observations: 485  
 Mean dependent var : 172153 Number of Variables : 13  
 S.D. dependent var : 57492.3 Degrees of Freedom : 472

R-squared : 0.753893 F-statistic : 120.489  
 Adjusted R-squared : 0.747636 Prob(F-statistic) : 0\*  
 Sum squared residual: 3.94534e+11 Log likelihood : -5663.51  
 Sigma-square : 8.35877e+08 Akaike info criterion : 11353  
 S.E. of regression : 28911.5 Schwarz criterion : 11407.4  
 Sigma-square ML : 8.13472e+08  
 S.E of regression ML: 28521.4

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-189082	25461.6	-7.42617	0.00000*
Log_SURF_F	41372	3863.87	10.7074	0.00000*
Log_SURF_H	138355	11088.8	12.4769	0.00000*
N2_FACADE	-9664.73	3177.78	-3.04135	0.00249*
AGE_BIEN	-369.996	38.7496	-9.54839	0.00000*
N2_SDB	18129.2	5478.12	3.30937	0.00101*
N1_CHB	-24062.9	8584.85	-2.80295	0.00527*
N2_CHB	-11365.8	3542.35	-3.20856	0.00142*
D100_BUS	-5020.63	2713.15	-1.85048	0.06487
NIV_INSTRU	1206.15	144.974	8.31981	0.00000*
D500_BLEU	-12682.7	6143.97	-2.06425	0.03954*
BTFO_ROAD	43312.6	20609.2	2.10161	0.03612*
KIEMSTEDTI	-327.847	99.0853	-3.30874	0.00101*

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 67.744461

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	5.3471	0.06901

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	12	21.8260	0.03952*
Koenker-Bassett test	12	28.1030	0.00534*



## Annexe 9e. Modèle D appliqué à l'échantillon de validation-croisée

### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

```

Data set           : SPF_ECHANT_VALIDATION
Dependent Variable :      PRIX   Number of Observations: 550
Mean dependent var :    142434   Number of Variables   : 21
S.D. dependent var :    65148.2   Degrees of Freedom   : 529

R-squared          : -nan(ind)   F-statistic          : -nan(ind)
Adjusted R-squared : -nan(ind)   Prob(F-statistic)   : -nan(ind)
Sum squared residual: -nan(ind)   Log likelihood       : -nan(ind)
Sigma-square       : -nan(ind)   Akaike info criterion : -nan(ind)
S.E. of regression : -nan(ind)   Schwarz criterion    : -nan(ind)
Sigma-square ML    : -nan(ind)
S.E of regression ML: -nan(ind)
  
```

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
LOG_SURF_H	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
LOG_SURF_F	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
N4_FACADE	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
N0_ETAGE	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
N2_ETAGE	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
MANSARDE	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
AGE_BIEN	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
AGE_RENO	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
N0_GARAGE	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
CHAUF_CENT	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
NB_PIECE	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
CAVE	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
ACC_BUS	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
ACC_ML	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
D1500_GARE	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
NIV_INSTRU	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
CONST_AF81	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
D300_BLEU	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
D100_VERT	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)
BTFO_TRAIN	-nan(ind)	-nan(ind)	-nan(ind)	-nan(ind)

#### REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER -999.000000

#### TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	-nan(ind)	0.00000*

#### DIAGNOSTICS FOR HETEROSKEDASTICITY

##### RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	20	-nan(ind)	N/A
Koenker-Bassett test	20	-nan(ind)	N/A

**SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION**

Data set : SPF\_ECHANT\_VALIDATION  
 Dependent Variable : PRIX Number of Observations: 550  
 Mean dependent var : 142434 Number of Variables : 19  
 S.D. dependent var : 65148.2 Degrees of Freedom : 531

R-squared : 0.527707 F-statistic : 32.9612  
 Adjusted R-squared : 0.511697 Prob(F-statistic) : 0\*  
 Sum squared residual: 1.1025e+12 Log likelihood : -6670.55  
 Sigma-square : 2.07627e+09 Akaike info criterion : 13379.1  
 S.E. of regression : 45566.1 Schwarz criterion :  
 Sigma-square ML : 2.00455e+09  
 S.E of regression ML: 44772.1

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-135090	28783.8	-4.69327	0.00000*
LOG_SURF_F	13042.3	5276.95	2.47156	0.01377*
LOG_SURF_H	94639.9	14808.5	6.39093	0.00000*
N4_FACADE	26243.1	6620.47	3.96394	0.00008*
N0_ETAGE	15027.2	8037.29	1.86968	0.06208
N2_ETAGE	7025.43	7350.85	0.955731	0.33965
MAN SARDE	19798.9	5205.36	3.80355	0.00016*
AGE BIEN	-213.861	76.4527	-2.79729	0.00534
AGE RENO	-69.9107	62.5346	-1.11795	0.26409
N0_GARAGE	-3447.98	5255.76	-0.656039	0.51209
CHAUF_CENT	26411.1	5115.59	5.16288	0.00000*
NB_PIECE	-743.772	984.738	-0.755299	0.45041
CAVE	1476.8	5368.71	0.275075	0.78338
ACC_BUS	1555	819.845	1.8967	0.05841
ACC_ML	878.243	633.114	1.38718	0.16597
D1500_GARE	-449.221	4840.76	-0.0927997	0.92611
NIV_INSTRU	1263.76	239.115	5.28515	0.00000*
CONST_AF81	204.048	155.873	1.30906	0.19108
D300_BLEU	5185.49	9422.35	0.550339	0.58232

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 72.780278

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	33.3544	0.00000*

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	18	123.5807	0.00000*
Koenker-Bassett test	18	78.6498	0.00000*

## Annexe 10. Résultat du VIF

### Annexe 10a. Résultat du VIF pour le modèle C1 (SPF Finances)

Variables	VIF
LOG_SURF_F	1,7336
LOG_SURF_H	2,5677
N4_FACADE	2,0477
N0_ETAGE	2,6835
N2_ETAGE	1,4802
MANSARDE	1,3268
AGE_BIEN	2,1546
AGE_RENO	1,7614
N0_GARAGE	1,7966
CHAUF_CENT	1,5023
NB_PIECE	1,8506
CAVE	1,3961
ACC_BUS	1,7582
ACC_ML	1,6625
D1500_GARE	1,0883
NIV_INSTRU	1,4311
CONSTR_AF81	1,8026
D300_BLEU	1,0571
D100_VERT	1,0664
BTFO_TRAIN	1,0257

### Annexe 10b. Résultat du VIF pour le modèle C2 (Immoweb)

Variables	VIF
LOG_SURF_F	1,8240
LOG_SURF_H	1,5538
N2_FACADE	1,4711
AGE_BIEN	1,1591
N2_SDB	1,0374
N1_CHB	1,1126
N2_CHB	1,3045
D100_BUS	1,0506
RMD_2011	2,7573
NIV_INSTRU	2,3428
D500_BLEU	1,0293
BTFO_ROAD	1,0147
KIEMSTEDI	1,0489

**Annexe 10c. Résultat du VIF pour le modèle D1 (SPF Finances)**

<b>Variables</b>	<b>VIF</b>
LOG_SURF_F	1,7080
LOG_SURF_H	2,5510
N4_FACADE	2,0382
N0_ETAGE	2,6697
N2_ETAGE	1,4767
MANSARDE	1,3277
AGE_BIEN	2,1601
AGE_RENO	1,7590
N0_GARAGE	1,8036
CHAUF_CENT	1,5031
NB_PIECE	1,8505
CAVE	1,4001
ACC_BUS	1,7553
ACC_ML	1,6598
D1500_GARE	1,0884
NIV_INSTRU	1,4326
CONST_AF81	1,8005
D300_BLEU	1,0538
D100_VERT	1,0639
BTFO_TRAIN	1,0262

**Annexe 10d. Résultat du VIF pour le modèle D2 (SPF Finances)**

<b>Variables</b>	<b>VIF</b>
LOG_SURF_F	1,6964
LOG_SURF_H	2,5232
N4_FACADE	2,0563
N0_ETAGE	2,6977
N2_ETAGE	1,4573
MANSARDE	1,3434
AGE_BIEN	2,1841
AGE_RENO	1,7725
N0_GARAGE	1,8152
CHAUF_CENT	1,4998
NB_PIECE	1,8426
CAVE	1,3974
ACC_BUS	1,7318
ACC_ML	1,6314
D1500_GARE	1,0876
NIV_INSTRU	1,4459
CONST_AF81	1,8021
D300_BLEU	1,0546
D100_VERT	1,0647
BTFO_TRAIN	1,0289

**Annexe 10e. Résultat du VIF pour le modèle D3 (Immoweb)**

<b>Variables</b>	<b>VIF</b>
LOG_SURF_F	1,6880
LOG_SURF_H	1,5451
N2_FACADE	1,4431
AGE_BIEN	1,1566
N2_SDB	1,0418
N1_CHB	1,1155
N2_CHB	1,3011
D100_BUS	1,0365
NIV_INSTRU	1,1564
D500_BLEU	1,0302
BTFO_ROAD	1,0121
KIEMSTEDTI	1,0541

## Annexe 11. Coefficient et erreur des modèles finaux

### Annexe 11a. Coefficient et erreur selon les variables du modèle D2 (SPF Finances)

Variables	Coefficient	Erreur type	Erreur relative	Probabilité
LOG_SURF_H	45 464,60 €	3 875,71 €	8,524676342	0,0000*
CHAUF_CENT	19 551,00 €	1 255,88 €	6,423610046	0,0000*
N4_FACADE	19 266,00 €	1 789,89 €	9,290407973	0,0000*
LOG_SURF_F	12 947,20 €	1 285,41 €	9,928092561	0,0000*
NO_ETAGE	12 343,90 €	2 083,28 €	16,87699998	0,0000*
MANSARDE	10 625,70 €	1 371,37 €	12,90616148	0,0000*
D300_BLEU	7 676,58 €	2 668,41 €	34,76040112	0,0040*
CAVE	4 953,06 €	1 394,29 €	28,15007288	0,0004*
N2_ETAGE	4 518,56 €	1 981,34 €	43,84892532	0,0226
NB_PIECE	4 503,86 €	369,70 €	8,208603287	0,0000*
NIV_INSTRU	1 289,20 €	60,96 €	4,728862861	0,0000*
ACC_ML	612,66 €	157,66 €	25,7345068	0,0001*
ACC_BUS	502,57 €	222,74 €	44,31961988	0,0241
CONST_AF81	214,62 €	40,26 €	18,76033828	0,0000*
AGE_RENO	-86,95 €	15,73 €	18,08910971	0,0000*
AGE_BIEN	-362,43 €	19,33 €	5,333351728	0,0000*
D1500_GARE	-3 797,17 €	1 212,61 €	31,93457233	0,0018
NO_GARAGE	-7 290,48 €	1 380,35 €	18,93359559	0,0000*
CONSTANT	-32 814,60 €	7 740,89 €	23,58977406	0,0000*

**Annexe 11b. Coefficient et erreur selon les variables du modèle D3 (Immoweb)**

<b>Variables</b>	<b>Coefficient</b>	<b>Erreur type</b>	<b>Erreur relative</b>	<b>Probability</b>
Log_SURF_H	138 355,00 €	11 088,80 €	8,01 %	0,0000*
BTFO_ROAD	43 312,60 €	20609,20 €	47,58 %	0,0361
Log_SURF_F	41 372,00 €	3863,87 €	9,34 %	0,0000*
N2_SDB	18 129,20 €	5478,12 €	30,22 %	0,0010*
NIV_INSTRU	1 206,15 €	144,97 €	12,02 %	0,0000*
KIEMSTEDTI	-327,85 €	99,09 €	30,22 %	0,0010*
AGE_BIEN	-370,00 €	38,75 €	10,47 %	0,0000*
D100_BUS	-5 020,63 €	2713,15 €	54,04 %	0,0649
N2_FACADE	-9 664,73 €	3177,78 €	32,88 %	0,0025
N2_CHB	-11 365,80 €	3542,35 €	31,17 %	0,0014
D500_BLEU	-12 682,70 €	6143,97 €	48,44 %	0,0395
N1_CHB	-24 062,90 €	8584,85 €	35,68 %	0,0053
CONSTANT	-189 082,00 €	25461,60 €	13,47 %	0,0000*

## Annexe 12. Les critères de choix entre les modèles hédoniques

En plus du  $R^2$  et de l'erreur relative du modèle, plusieurs indicateurs sont employés pour définir la comparaison de plusieurs modèles dont la composition des variables est proche. Les deux indicateurs sont calculés dans le logiciel GeoDa. Pour commencer, l'AIC (*Akaike information criterion*) offre un compromis entre le biais du modèle et le nombre de variables introduites. Lors de la comparaison de modèles similaires, le plus petit AIC indique le meilleur modèle (Lancelot & Lesnoff, 2005).

$$AIC = -2 \log(L) + 2k$$

Avec,

$L$  = maximum de vraisemblance ;

$k$  = nombre de paramètres introduits dans le modèle.

### Équation 21. Critère d'information d'Akaike (AIC) (Akaike, 1974)

Un deuxième indice est le BIC (*Bayesian information criterion*), tout comme l'AIC, il permet un arbitrage pour le choix entre plusieurs modèles semblables. Le BIC est mieux adapté aux grands sets de données (plusieurs milliers d'échantillons). De plus, le BIC aboutit à des modèles généralement plus parcimonieux que l'AIC (Lancelot & Lesnoff, 2005).

$$BIC = -2 \ln(L) + k \log(n)$$

Avec,

$L$  = maximum de vraisemblance ;

$k$  = nombre de paramètres introduits dans le modèle ;

$n$  = nombre d'observations.

### Équation 22. Bayesian Information Criterion (BIC) (Schwarz, 1978)