



TUTEUR

Vivien Roussez

JURY

François Coquet
Éric Lesage

ÉLÈVE

Kim Antunez

Remerciements

À l'issue de ce travail, je tiens à remercier toutes les personnes qui ont contribué à la réalisation de ce rapport et celles qui consacreront de leur temps pour le lire.

Tout d'abord, je remercie tout particulièrement mes proches, Alain, Céleste, Céline et Romain qui, tout au long de l'élaboration de ce mémoire, m'ont été d'un appui certain, tant sur le plan technique que psychologique.

Mes remerciements vont également à Vivien Roussez, chef du Bureau de l'Observation des Territoires au CGET, pour avoir accepté d'être mon maître de stage et pour avoir assumé ce rôle tout en m'accordant confiance et autonomie et à Brigitte Baccaini, sous-directrice de l'observation et des analyses statistiques pour sa relecture précise.

Je remercie en outre, pour leur accueil chaleureux, tous les membres des deux structures entre lesquelles j'ai partagé mon temps ces derniers mois : la sous-direction de l'observation et des analyses statistiques au CGET et le Bureau de la statistique déconcentrée de la Construction au CGDD.

J'adresse enfin mes remerciements à Frédéric Minodier, chef du Bureau de la statistique déconcentrée de la construction, pour m'avoir accueillie au SOeS pour l'accès aux données de l'enquête EPTB. Je remercie également Benjamin Vignolles (chargé d'étude sur le logement) et Marie-Pierre Joubert (méthodologue en analyse spatiale), pour leurs multiples pistes d'études et de modélisations statistiques et Bruno Vermont, chargé d'études économiques au CGDD, pour les précisions qu'il m'a apportées sur sa récente publication concernant également le prix du logement.

Table des matières

Introduction	1
1 Sources et champ de l'étude	3
I Le champ de l'étude	3
II Les données exploitées	3
III Les zonages d'étude	4
2 Méthodes d'analyses et de statistiques spatiales	7
I Autocorrélation spatiale des prix	7
II Économétrie spatiale et déterminants du prix du logement	13
III Krigage de la carte des prix des terrains	16
3 Les déterminants spatiaux des prix du foncier	19
I Disparité des prix des terrains	19
II Distance aux grands espaces urbains, densité et prix du foncier	21
III Influence des spécificités des territoires sur les prix des terrains	28
Bilan de l'étude	35
Conclusion	37
A Inventaire des principales données sur le logement	39
B La méthode des prix hédoniques	41
C Extraits de code R	43

Table des figures

1.1	Le zonage en bassins de vie	5
1.2	Le zonage en aires urbaines	5
2.1	Diagramme de Moran des prix par aire urbaine par la methode des 10 plus proches voisins	10
2.2	Illustration du diagramme de Moran des prix par AU avec la méthode des 10 plus proches voisins	11
2.3	Illustration des diagrammes de Moran des prix par bassin de vie et aire urbaine avec la méthode des 10 plus proches voisins	12
2.4	Illustration des diagrammes de Moran des prix par bassin de vie avec différentes matrices de distances	12
2.5	Semi-variogramme des prix au mètre carré des terrains	17
3.1	Hétérogénéité des prix entre et au sein des aires urbaines	19
3.2	Distribution des prix par mètre carré des terrains	20
3.3	Prix lissés des terrains à bâtir en France en 2013	21
3.4	Prix des terrains et densité de population	22
3.5	Prix des terrains en fonction de la densité de la commune	22
3.6	Nombre de terrains à bâtir par bassin de vie en 2013	22
3.7	Répartition des prix des terrains dans les différents types de communes	23
3.8	Prix des terrains et logarithme du prix des terrains en fonction de la distance à la grande aire urbaine la plus proche	24
3.9	Prix et distance au centre dans quatre grandes aires urbaines	28
3.10	Prix de référence par aire urbaine et leur autocorrélation spatiale	29
3.11	Cercle de corrélation des spécificités des aires urbaines	32

Liste des tableaux

2.1	Effets directs et indirects dans les modèles spatiaux	15
3.1	Quantiles des prix des terrains dans les différents types de communes	23
3.2	Estimation des coefficients du modèle 3.1 appliqué à toutes les aires urbaines	25
3.3	Estimation des coefficients du modèle 3.1 appliqué par type d'aire urbaine	26
3.4	Gradients de prix par grande aire urbaine	27
3.5	Caractéristiques géographiques et socio-démographiques des aires urbaines	30
3.6	Estimations des modèles spatiaux classiques	33
3.7	Estimation du modèle SARAR	33
3.8	Estimations du modèle SARAR complet	34
B.1	Formes fonctionnelles des variables dans les modèles hédoniques	42

Introduction

Pour un ménage propriétaire, un logement représente souvent l'essentiel de son patrimoine. Ayant la particularité de jouer un double rôle économique d'investissement et de bien de consommation, il a des enjeux économiques et politiques majeurs. Ces enjeux sont d'autant plus importants que le prix des logements a augmenté de 60 % hors inflation de 2000 à 2012¹.

Les prix du logement, et donc du foncier, peuvent peser sur la consommation des ménages mais aussi sur l'investissement des entreprises et leurs emplois, en particulier dans des secteurs comme l'agriculture ou la construction. Une hausse des prix peut également toucher l'aménagement du territoire : elle peut générer une ségrégation socio-spatiale des individus, repoussant petit à petit les populations à faibles revenus en zone périurbaine [17]. À l'inverse, en période de baisse de prix, la situation devient préoccupante à la fois pour les promoteurs, les investisseurs et les ménages qui auront des scrupules à vendre leurs biens.

L'intervention publique est alors légitime et très fréquente pour rééquilibrer le marché de l'immobilier. L'État agit à la fois via la fiscalité immobilière et foncière mais également par le biais d'aides au financement et aux locataires, de constructions de logements sociaux ... De même, les collectivités locales s'intéressent beaucoup à la question du foncier (établissements publics fonciers et observatoire locaux, lancement de programmes d'étude par les collectivités territoriales, partenariats publics / privés sur le foncier...).

Dès lors, afin que les différents acteurs économiques disposent d'informations, il apparaît important de s'intéresser aux déterminants du prix du logement. Or, ces prix ne dépendent pas uniquement des caractéristiques physiques des biens. Ils sont également liés à leur situation géographique et aux caractéristiques des territoires qui les accueillent. L'objectif de ce mémoire est donc d'analyser les déterminants socio-économiques des différents territoires (densité de population, aménités...) qui influent sur la disparité géographique des prix du foncier.

Cette étude statistique vise à la fois à mieux connaître les zones en tension sur le plan de l'immobilier, ainsi que leurs origines, mais également à identifier les leviers les plus pertinents sur lesquels les politiques publiques peuvent agir pour limiter les inégalités résultant des disparités du prix du foncier.

Dans un premier chapitre seront évoqués, d'une part, les sources de données mobilisées dans ce mémoire (l'enquête EPTB sur les prix des terrains à bâtir ainsi que le recensement de la population) et d'autre part les zonages d'étude que sont les bassins de vie et aires urbaines.

Vous trouverez dans le deuxième chapitre une description de différentes méthodes d'analyses et de statistiques spatiales mises en regard avec la problématique de ce mémoire.

Enfin, dans le troisième et dernier chapitre, se trouve le cœur de l'étude sur les déterminants du prix du logement. Une première partie montrera que la distance aux grands espaces urbains est un déterminant majeur du prix du foncier. Il sera alors possible d'estimer un gradient de prix en fonction de la distance au centre de l'aire urbaine la plus proche. Puis, dans une seconde partie, nous montrerons par une modélisation d'économétrie spatiale que certaines caractéristiques socio-économiques des territoires, comme la densité de population, influent également sur le prix du foncier.

1. Bruno Vermont (2015) [20]

CHAPITRE 1

Sources et champ de l'étude

Pour mieux appréhender l'ensemble de l'étude présentée dans ce mémoire, il est nécessaire d'en établir le cadre. Nous verrons dans un premier temps pourquoi le champ de cette étude sur les prix du logement peut se restreindre aux prix des terrains. Nous nous intéresserons également aux données mobilisées : celles de l'enquête sur le prix des terrains à bâtir (EPTB) et du recensement de la population (RP). Ces données permettent à la fois d'appréhender les caractéristiques des terrains mais également celles des territoires qui les accueillent. Enfin, comme dans toute étude contenant une dimension spatiale, nous avons dû choisir des zonages d'étude. Nous avons ainsi utilisé comme mailles géographiques les bassins de vie et les aires urbaines.

I Le champ de l'étude

Afin d'étudier les déterminants du prix du logement, nous avons restreint notre champ aux terrains à bâtir.

En effet, travailler sur les prix des terrains permet de s'affranchir de nombreuses variables touchant les caractéristiques des logements (nombre de pièces, niveau de confort ...). Le focus est ainsi recentré sur la problématique stricte du mémoire : les déterminants *spatiaux* des prix¹

Le prix d'un logement est principalement déterminé par les paramètres de son bâti (surface, étage...) et par les caractéristiques socio-économiques de son territoire (densité de population, accessibilité aux équipements...). La conjoncture économique locale de l'industrie du bâtiment a peu d'influence sur ce prix. C'est pourquoi les prix des constructions neuves sont fortement « tirés » par les prix des terrains (cf. Joubert et Vignolles 2015 [10]). Les déterminants spatiaux des prix des logements peuvent donc être globalement résumés par ceux des prix des terrains.

Les sources de données sur le logement étant multiples et bien souvent impossibles à apparier (cf. annexe A sur la multiplicité des sources sur le logement), nous avons fait le choix de nous restreindre aux données de prix sur le foncier via l'enquête des terrains à bâtir (EPTB) du SOeS. Nous avons ensuite couplé ces données avec les caractéristiques socio-économiques des territoires présentes dans le recensement de la population.

II Les données exploitées

Pour croiser les données sur le prix des terrains et sur les spécificités des territoires qui les accueillent, deux sources distinctes sont exploitées dans ce mémoire. Pour les données sur les terrains, nous avons utilisé les enquêtes du SOeS sur le prix des terrains à bâtir (EPTB 2006 à 2013) et pour les caractéristiques socio-économiques des territoires, nous avons à notre disposition le recensement de la population 2012 (RP 2012) de l'Insee. Ces données étant disponibles à une

1. Toutefois, les terrains présentent également des caractéristiques intrinsèques : qualité du sol, pente, viabilisation... Seule la viabilisation est contrôlée dans nos données provenant d'EPTB.

échelle géographique fine (à la commune pour le recensement et au logement pour ETPB), nous avons pu les agréger à plusieurs mailles : celle des bassins de vie et celle des aires urbaines (cf. partie III).

II.1 L'enquête sur le prix des terrains à bâtir

Cette enquête, d'intérêt général et obligatoire telle que définie par le Cnis, est réalisée par le SOeS et traitée annuellement par le Bureau de la statistique déconcentrée de la construction. La collecte, mensuelle, est réalisée via un questionnaire envoyé par voie postale. Elle concerne les autorisations, faites aux particuliers, de construire une maison individuelle neuve. Elle est réalisée à un niveau national depuis 2006 mais n'est exhaustive sur son champ que depuis 2010. Cette exhaustivité n'empêche pas un redressement de la non-réponse et un calage sur marges² donnant lieu à une pondération individuelle des terrains. L'échantillon est réalisé d'après la base administrative Sit@del constituée à partir des formulaires de permis de construire.

L'enquête renseigne sur les prix et les caractéristiques du terrain (achat ou non du terrain, surface, commune, viabilisation, etc.) et de la construction de la maison (maître d'œuvre, surface, chauffage, etc.) mais donne aussi des informations sur le pétitionnaire (âge, CSP, etc.).

Les variables utilisées lors de ce projet sont le prix du terrain lorsque celui-ci a été acheté, sa surface, sa commune et sa viabilisation ou non. Un terrain est viabilisé s'il est raccordé aux réseaux d'eau, d'électricité, de gaz, de téléphone et d'assainissement. La commune est l'unique information géographique contenue dans la base EPTB. En effet, les données ne sont pas géolocalisées mais devraient l'être dans une prochaine édition.

EPTB est une enquête soumise à la loi de 1951³. Pour respecter le secret statistique, aucune statistique ne doit être publiée si elle se base sur moins de onze permis. Par exemple, les prix moyens des terrains par commune ne pourront être diffusés et cartographiés que pour les communes où auront été signés plus de onze permis.

II.2 Le recensement de la population

Le recensement de la population, mené par l'Insee, a pour objectif de dénombrer les logements et habitants sur le territoire français et d'en connaître leurs caractéristiques (âge, activité des ménages, type de logement, etc.). Ces informations, mises à la disposition de tous à l'échelle communale voire infra-communale, sont d'une aide précieuse pour de nombreux acteurs publics et privés. Elles permettent, par exemple, d'élaborer des politiques concernant le transport, le logement, les équipements scolaires ou encore d'accueil aux personnes âgées. Le recensement peut aussi servir aux projets d'implantation de commerces ou d'entreprises.

Le recensement existe depuis 1801. Il s'est déroulé tous les 5 à 9 ans jusqu'en 1999, puis a subi une refonte en 2004. Dès lors, les communes de moins de 10 000 habitants ont continué à être recensées exhaustivement tous les 5 ans mais celles de plus de 10 000 habitants font l'objet d'une enquête annuelle concernant 8 % de la population (soit 40 % de la population au bout de 5 ans).

Les variables issues du recensement caractérisent la situation socio-économique des territoires. On y trouve des informations sur l'état du parc des logements, la structure de la population, etc. Ces variables ont été utilisées pour la recherche des déterminants du prix du logement. Elles sont décrites en détail dans la partie III du chapitre 3.

III Les zonages d'étude

Afin de mener une analyse spatiale des prix, plusieurs zonages ont été utilisés en plus du zonage communal. Nous décrivons dans cette partie les zonages en bassins de vie et en aires urbaines.

2. Le calage sur marges permet de redresser un échantillon provenant d'une enquête. Les individus sont repondérés en utilisant une information auxiliaire disponible sur des variables dites de calage.

3. Cette loi traite de l'obligation, de la coordination et du secret en matière de statistiques.

III.1 Le zonage en bassins de vie

Le bassin de vie (abréviation BV) est le plus petit territoire sur lequel les habitants ont accès aux équipements et aux services les plus courants : services aux particuliers, commerces, enseignement, santé, sports, loisirs et culture, transport. En d'autres termes, ils correspondent aux espaces où les populations « vivent » la plupart de leur temps, d'où leurs noms. Les près de 1 700 bassins de vie découpent la totalité du territoire français.

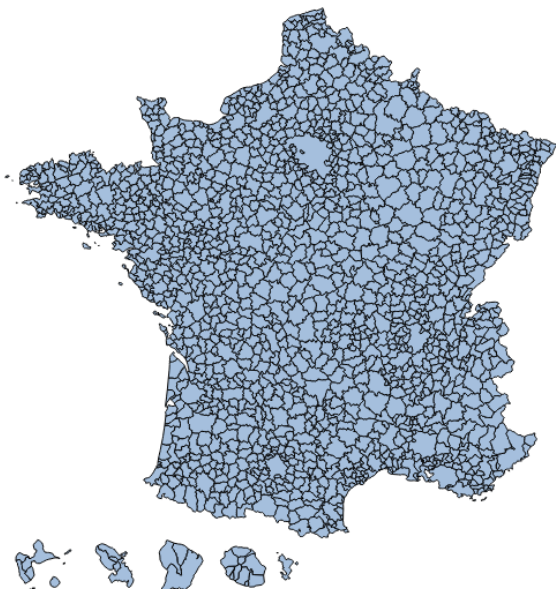


FIGURE 1.1– Le zonage en bassins de vie

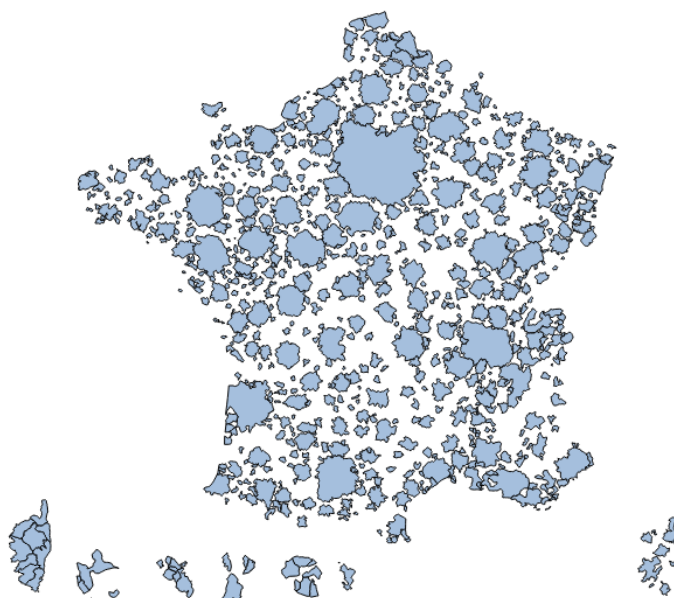


FIGURE 1.2– Le zonage en aires urbaines

III.2 Le zonage en aires urbaines

Une aire urbaine (AU) est un ensemble de communes, d'un seul tenant et sans enclave constitué par un pôle urbain (unité urbaine) accueillant plus de 1 500 emplois, et par des communes ou unités urbaines (couronne périurbaine) dont au moins 40 % de la population résidente ayant un emploi travaille dans le pôle ou dans des communes attirées par celui-ci. Selon le nombre d'emplois présents dans le pôle, on qualifie l'aire urbaine de petite, moyenne ou grande (pour respectivement de 1 500 à 5 000, de 5 000 à 10 000 et plus de 10 000 emplois).

Contrairement au zonage en bassins de vie, les 700 aires urbaines ne recouvrent pas l'intégralité du territoire français. En effet, certaines communes sont dites « multipolarisées » si elles sont influencées par deux pôles d'aires urbaines et les communes restantes sont qualifiées d'« hors influence des pôles ».

Méthodes d'analyses et de statistiques spatiales

“TOUT INTERAGIT AVEC TOUT, MAIS DEUX OBJETS PROCHES ONT PLUS DE CHANCES DE LE FAIRE QUE DEUX OBJETS ÉLOIGNÉS.” W. TOBLER

Cette phrase de Tobler illustre tout l'enjeu de ce chapitre. En effet, travailler sur des données géographiques implique que les caractéristiques des individus statistiques peuvent dépendre des caractéristiques de leur voisinage. Ainsi, le prix d'un terrain va être influencé par le prix des terrains voisins et les caractéristiques socio-démographiques d'un territoire vont être liées à celles des territoires limitrophes.

Ce chapitre présente des méthodes classiques d'analyse et de statistique spatiales mises en application sur les données de prix des terrains provenant d'EPTB. Ces analyses constituent un fondement théorique nécessaire à la réalisation d'une étude comme celle-ci, faisant appel à des modèles d'économétrie spatiale, où l'autocorrélation spatiale des prix décrite par Tobler doit être prise en compte.

I Autocorrélation spatiale des prix

Depuis une trentaine d'années, les progrès dans le traitement des données géographiques ont été considérables, notamment grâce aux SIG¹ et leurs outils de géolocalisation. Une étude comme celle-ci sur les prix du foncier implique de prendre en considération la structure spatiale des prix via deux phénomènes :

- l'*autocorrélation spatiale* : il y a dépendance entre le prix d'un terrain et le prix des terrains voisins ;
- l'*hétérogénéité spatiale* : les prix ont différents comportements suivant leur localisation (prix plus élevés dans certains types d'espaces, instabilité dans l'espace des paramètres explicatifs du prix...).

Dans cette partie, nous analysons l'autocorrélation spatiale des prix des terrains, c'est-à-dire l'interdépendance entre le prix d'un terrain et le prix des terrains voisins.

Nous commencerons par donner les définitions possibles du voisinage des terrains via la matrice de voisinage notée W , puis réaliserons un test d'autocorrélation spatiale : le test de Moran. Nous verrons ensuite, via le diagramme de Moran, comment observer localement l'autocorrélation spatiale des prix des terrains. Enfin, nous comparerons les résultats d'autocorrélation spatiale issus de l'utilisation de différentes matrices de distances et de différentes mailles géographiques.

1. Un système d'information géographique (SIG) est un système conçu pour traiter des données spatiales et géographiques. Les logiciels de cartographie sont des SIG.

I.1 Distances et matrices de poids

La matrice de poids W a pour rôle de définir le concept de proximité, de distances géographiques entre les terrains. La définition du voisinage des terrains est essentielle avant toute étude ou modélisation de l'autocorrélation spatiale des prix.

À l'inverse de l'autocorrélation temporelle, où seul le passé influe sur le futur, l'autocorrélation spatiale est multidimensionnelle : les prix des terrains situés à l'ouest comme au nord d'un terrain peuvent influencer sur son prix. Cette matrice est, en général, une hypothèse de modélisation appréciée par le statisticien, et n'est donc pas à estimer dans les modèles d'économétrie spatiale.

Nous avons comparé dans la partie I.3 les résultats d'autocorrélation spatiale des prix obtenus avec trois matrices de poids différentes : premièrement les 10 plus proches voisins, deuxièmement et troisièmement deux matrices de distances au carré avec comme seuils de distance respectifs \bar{d} 100 et 200 km (cf. encadré 2.1). En effet, ces choix permettent de prendre en compte un nombre raisonnable de voisins autour de l'aire urbaine étudiée en excluant toutefois les aires urbaines situées trop loin de celle-ci, n'ayant manifestement pas d'influence sur son prix. De plus, considérer un trop grand nombre de voisins rend la matrice W inutilement difficile à inverser².

Encadre 2.1 - La matrice W

La matrice de voisinage W peut prendre plusieurs formes dont les plus classiques :

1. Matrice de *contigüité*

$$W_{ij} = \begin{cases} 1 & \text{si les régions } i \text{ et } j \text{ sont contigües} \\ 0 & \text{sinon} \end{cases}$$

2. Matrice de *distance*

$$W_{ij} = \begin{cases} \frac{1}{d_{ij}^2} & \text{si } d_{ij} < \bar{d} \\ 0 & \text{sinon} \end{cases}$$

avec d_{ij} la distance entre i et j ^a et \bar{d} le seuil de distance à partir duquel on considère qu'il n'y a plus autocorrélation spatiale.

3. Matrice des *k plus proches voisins*

$$W_{ij} = \begin{cases} 1 & \text{si la région } j \text{ fait partie des } k \text{ voisins les plus proches de la région } i \\ 0 & \text{sinon} \end{cases}$$

Les paramètres \bar{d} et k sont choisis par le modélisateur. Plus ces paramètres sont élevés, plus des voisins lointains de i sont pris en compte.

La matrice W est souvent standardisée : la somme de chaque ligne vaut alors 1.

a. Nous utiliserons une distance euclidienne $d(A, B) = \sqrt{(X_B - X_A)^2 + (Y_B - Y_A)^2}$ mais d'autres types de distances existent, comme la distance de Manhattan $d(A, B) = |X_B - X_A| + |Y_B - Y_A|$ qui correspond à la distance entre deux points parcourue quand on se déplace dans une ville américaine où les rues sont agencées en quadrillages.

I.2 Test de Moran d'autocorrélation spatiale

La statistique de Moran, appelée indice de Moran et notée I (formule et interprétation en encadré 2.2), est le moyen le plus courant de tester et mesurer l'autocorrélation spatiale. En effet, quand il est centré et réduit, cet indice suit asymptotiquement une loi normale standard sous l'hypothèse nulle de non-autocorrélation spatiale.

2. Ce choix est également motivé par la valeur du palier obtenu lors de la représentation du variogramme du prix des terrains (cf. partie III)

Nous avons calculé et testé l'autocorrélation spatiale des prix des terrains en 2013 issus de l'enquête EPTB de la même année. Dans nos analyses et modèles d'économétrie spatiale, nous avons choisi de travailler surtout sur la maille géographique des aires urbaines³ avec la matrice des 10 plus proches voisins⁴. L'indice de Moran vaut alors 0,7. L'autocorrélation spatiale est bien positive et significative, c'est-à-dire que le prix moyen des terrains dans une aire urbaine (AU) influe sur (et est influencé par) le prix des terrains dans les AU voisines.

🔍 Encadre 2.2 - L'indice de Moran

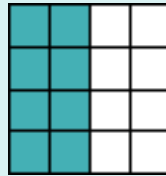
L'indice de Moran est défini comme suit :

$$I = \frac{\overbrace{\sum_{i,j} \omega_{ij} (x_i - \bar{x})(x_j - \bar{x})}^{\text{Covariance des unités voisines}}}{\underbrace{\sum_i \frac{(x_i - \bar{x})^2}{N}}_{\text{Variance totale}}} \quad \text{où} \quad \omega_{ij} = \frac{W_{ij}}{\sum_{k,l} W_{kl}}$$

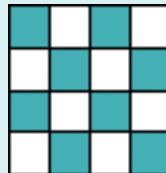
x_i correspond à la variable dont on teste l'autocorrélation spatiale : ici il correspond au prix moyen d'une aire urbaine i .

I est compris entre -1 et 1.

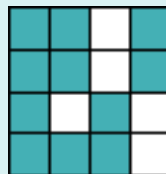
- S'il est proche de 1, il y a présence d'une autocorrélation positive. L'espace géographique ressemble à ses voisins en matière de prix des terrains. Schématiquement, les lieux où les prix des terrains sont élevés (carrés sombres) sont proches les uns des autres. De même pour les lieux dont les prix des terrains sont faibles (carré blancs).



- S'il est proche de -1, l'autocorrélation est spatialement négative. La dispersion est alors parfaite : des lieux proches ont des prix plus différents que des lieux éloignés.



- S'il est proche de 0, il y a absence d'autocorrélation spatiale. La distribution spatiale des prix est considérée comme complètement aléatoire.



L'indice I dépend de l'hypothèse de modélisation de la matrice de voisinage W et du zonage choisi. Cette dépendance à la maille est connue sous le nom de MAUP (*modifiable areal unit problem*). C'est-à-dire que les résultats d'autocorrélation spatiale vont dépendre de la manière dont les espaces sont découpés. Ils dépendent notamment de la taille des espaces choisis (effet d'échelle).

3. Le zonage en aires urbaine est le zonage que nous utiliserons également dans le modèle économétrique expliquant le prix des terrains des aires urbaines par les spécificités des territoires qui les accueillent (Chapitre 3)

4. En effet, la matrice de voisinage des 10 plus proches voisins est celle qui révèle le mieux l'autocorrélation spatiale du prix des terrains (cf. partie I.4)

Pour ces raisons, nous avons analysé en partie I.4 l'autocorrélation spatiale en utilisant plusieurs matrices de poids (distance et plus proches voisins) ainsi que deux échelles géographiques : bassins de vie et aires urbaines.

I.3 Diagramme de Moran

Le diagramme de Moran permet de visualiser l'autocorrélation spatiale globale. Le principe est simple : il s'agit ici de tracer un graphique avec, en abscisse, le prix moyen au mètre carré des terrains dans une aire urbaine et, en ordonnée, la moyenne pondérée, par la matrice des 10 plus proches voisins, des prix des terrains dans les aires urbaines voisines.

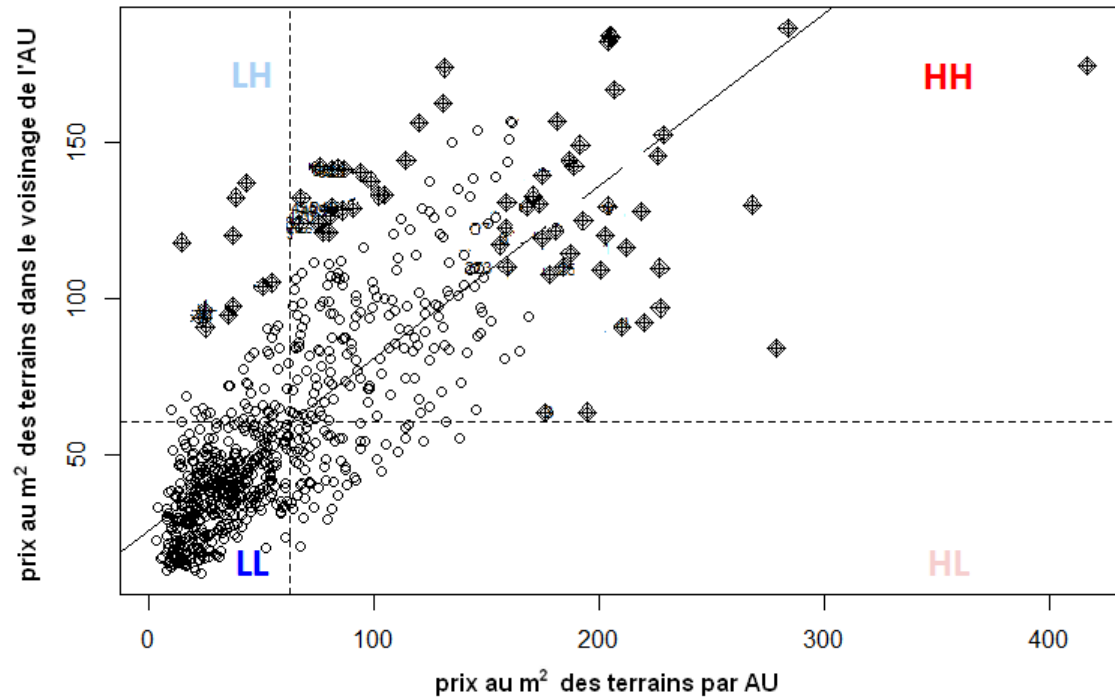


FIGURE 2.1— Diagramme de Moran des prix par AU par la méthode des 10 plus proches voisins
Source : EPTB 2013

Un point du diagramme représente une aire urbaine. Le diagramme de Moran comporte quatre cadrans délimités en abscisse par les moyennes des prix des terrains par AU et en ordonnée par les prix dans le voisinage de chaque AU :

1. high-high (HH) : quand les prix dans l'aire urbaine et chez ses voisins sont élevés par rapport à la moyenne (autocorrélation spatiale positive) ;
2. low-low (LL) : quand les prix dans l'aire urbaine et chez ses voisins sont faibles (autocorrélation spatiale positive) ;
3. high-low (HL) : quand les prix dans l'aire urbaine sont élevés alors qu'ils sont faibles chez ses voisines (autocorrélation spatiale négative) ;
4. low-high (LH) : quand les prix dans l'aire urbaine sont faible alors qu'ils sont élevés chez ses voisines (autocorrélation spatiale négative).

On retrouve l'indice de Moran, positif et significatif (0,7), qui correspond à la pente de la régression des points du diagramme.

La carte 2.2 indique pour chaque aire urbaine à quel cadran elle appartient. Elle illustre la nette autocorrélation spatiale positive des prix des terrains : des prix forts à Paris, sur les littoraux et dans les zones transfrontalières et leurs alentours, des prix faibles dans les communes situées dans la diagonale traversant la France du Nord- Est au Sud-Ouest. Toutefois, cette carte montre également une autocorrélation négative pour un certain nombre de grandes aires urbaines (celles de Toulouse, Bordeaux, Rennes...) où les aires urbaines voisines ont un prix faible.

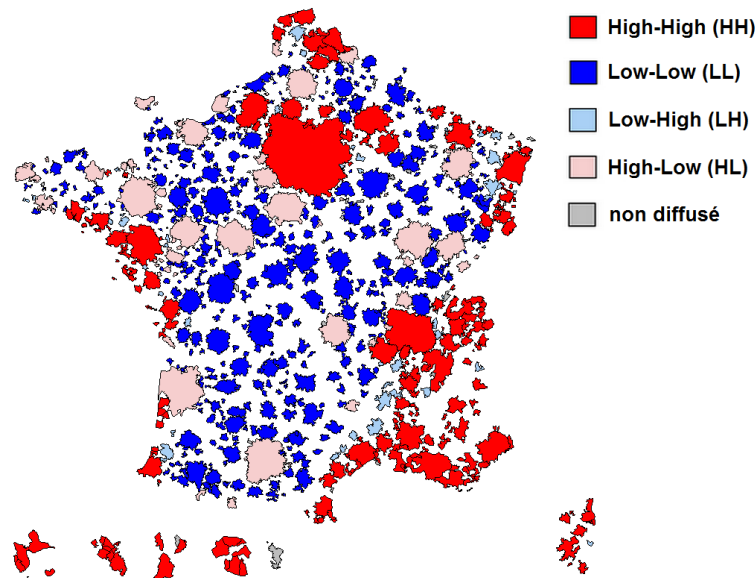


FIGURE 2.2— Illustration du diagramme de Moran des prix par AU avec la méthode des 10 plus proches voisins

Source : EPTB 2013

I.4 Modification du zonage et de la matrice de voisinage

Cette partie montre dans quelle mesure les choix du zonage (AU ou BV) et de la matrice W influencent les résultats d'autocorrélation spatiale du prix des terrains.

Effet d'un changement de maille géographique

Les cartes en figure 2.3 illustrent le diagramme de Moran des prix des terrains à deux échelles géographiques (bassins de vie à gauche et aires urbaines à droite) avec la méthode des 10 plus proches voisins.

Les résultats sont légèrement différents suivant l'échelle utilisée. En effet, certaines aires urbaines de grandes tailles aux prix des terrains élevés ont des voisines où les prix sont faibles (sont dans le cadran HL) alors que ces zones ont des voisines aux prix élevés (cadran HH) à l'échelle des bassins de vie. Cela est dû au fait que, ces aires urbaines étant d'une grande superficie, les 10 plus proches voisins sont à une distance plus élevée qu'à l'échelle des bassins de vie. Puisqu'ils sont plus éloignés des espaces denses et chers, les prix voisins y sont donc plus faibles.

Effet du changement de matrice de poids

Les cartes de la figure 2.4 illustrent le diagramme de Moran des prix des terrains dans les bassins de vie en utilisant trois méthodes différentes (les 10 plus proches voisins et l'inverse de la distance au carré ne prenant en compte que les espaces situés à moins de 100 et 200 km).

En général, plus on prend en compte des voisins lointains, plus le phénomène d'autocorrélation spatiale positive est gommé. Quand on augmente le seuil de distance, les prix dans le voisinage des bassins de vie est modifié : les bassins de vie où les prix sont élevés comportent plus de voisins où les prix sont faibles par rapport à la moyenne (certains espaces HH sont remplacés par des HL). À l'inverse, les bassins de vie aux prix modestes sont alors entourés par des bassins de vie aux prix élevés par rapport à la moyenne (LL remplacés par des LH).

On observe également le phénomène inverse dans quelques rares bassins de vie : des bassins de vie aux prix élevés classés en HL deviennent classés en HH quand on augmente le seuil de

distance. Ceci peut illustrer un phénomène de conurbation urbaine⁵ : prendre en compte des voisins plus éloignés peut dans ce cas augmenter les prix du voisinage d'une aire urbaine chère si un autre pôle urbain cher entre dans son voisinage.

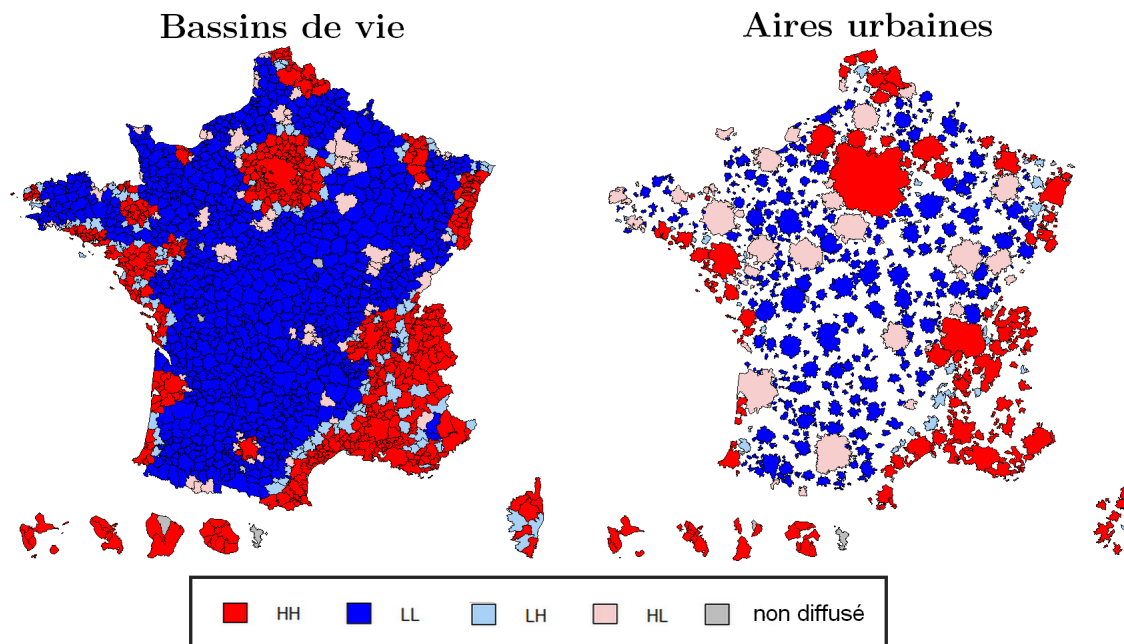


FIGURE 2.3— Illustration des diagrammes de Moran des prix par bassin de vie et aire urbaine avec la méthode des 10 plus proches voisins
Source : EPTB 2013

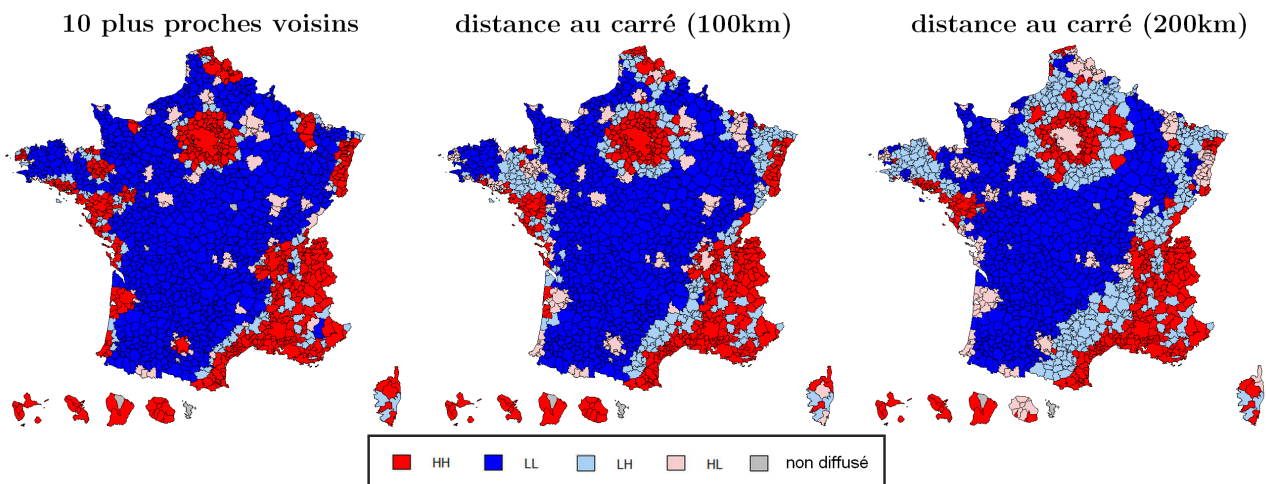


FIGURE 2.4— Illustration des diagrammes de Moran des prix par bassin de vie avec différentes matrices de distances
Source : EPTB 2013

Ainsi, travailler avec des mailles ou des seuils de distance plus importants gomme légèrement l'effet d'autocorrélation spatiale. En effet, prendre en compte des voisins plus éloignés va, par exemple, baisser globalement le prix du voisinage des grandes aires urbaines.

5. Une conurbation urbaine est un ensemble constitué de plusieurs noyaux urbains dont les banlieues finissent par se rejoindre.

II Économétrie spatiale et déterminants du prix du logement

L'économétrie spatiale est un moyen de réaliser des modèles prenant en compte l'autocorrélation spatiale, c'est-à-dire, dans notre cadre, la dépendance entre le prix des terrains et celui de leurs voisins [7] [14] [19].

En économétrie, l'autocorrélation spatiale peut provenir de plusieurs sources et donc être modélisée par plusieurs effets :

- un effet d'interaction endogène : la valeur du prix moyen des terrains d'une aire urbaine dépend directement de la valeur des prix des terrains dans les aires urbaines voisines ;
- un effet d'interaction exogène : le prix moyen des terrains dans une aire urbaine dépend des déterminants des prix du logement (densité, attractivité, aménités etc.) dans les aires urbaines voisines ;
- un effet d'interaction des inobservables : les effets inobservables qui affectent la valeur du prix moyen des terrains dans une aire urbaine sont autocorrélés avec ceux de son voisinage. Ce type de modélisation est approprié si l'on considère que l'on omet des variables explicatives du prix du logement pouvant suivre un processus d'autocorrélation spatiale.

II.1 Les modèles d'économétrie spatiale

Le modèle qui prend en compte les trois effets d'interaction cités précédemment est le modèle de Manski :

$$Y = \underbrace{\rho WY}_{\text{interaction endogène}} + X\beta + \underbrace{WX\theta}_{\text{interaction exogène}} + u \quad \text{avec } u = \underbrace{\lambda Wu}_{\text{interaction des inobservables}} + \varepsilon$$

Où :

- Y est le prix des terrains par aire urbaine expliqué par les déterminants des prix du logement, donc des terrains, présents dans X
- W est une même matrice de voisinage
- u est un bruit blanc⁶
- ρ , β , θ et λ sont des paramètres à estimer.

Néanmoins, ce modèle n'est pas identifiable. Pour qu'il le devienne, il existe deux possibilités. La première est d'utiliser trois matrices de voisinages différentes, une pour chaque effet d'interaction W_ρ , W_θ et W_λ . La seconde, privilégiée dans la littérature empirique, consiste à contraindre la nullité d'un ou deux effets d'interaction parmi ρ , θ et λ . Suivant les combinaisons de contraintes utilisées, on obtient différents modèles spatiaux. Les principaux modèles sont décrits en encadré 2.3.

II.2 Choix et estimation des modèles spatiaux

Le choix et la validation des modèles spatiaux peut s'effectuer au travers de tests de multiplicateurs de Lagrange (LM) qui permettent de comparer deux modèles emboîtés via leurs vraisemblances.

Les modèles spatiaux rendent les estimateurs par moindres carrés ordinaires non BLUE⁷ et doivent par conséquent être estimés par d'autres méthodes, dont :

- *par maximum de vraisemblance* : méthode la plus utilisée mais la maximisation de la fonction de vraisemblance est parfois très complexe à réaliser ;

6. Un bruit blanc gaussien en statistique suit une loi normale de moyenne et variance données.

7. Un estimateur BLUE (*Best Linear Unbiased Estimator*) est un estimateur non biaisé qui présente une variance minimale.

- par la méthode des moments généralisés (GMM) couplée avec la méthode des variables instrumentales. Cette méthode ne nécessite pas d'hypothèse sur la distribution des erreurs et est donc robuste à l'hétérogénéité spatiale. Elle admet également la possibilité de variables explicatives endogènes. C'est cette méthode d'estimation que nous allons utiliser dans le chapitre 3 grâce au modèle SARAR présenté ci-après.

🔍 Encadre 2.3 - Principaux modèles d'économétrie spatiale

- Le modèle SAR (Spatial AutoRegression) contient une interaction endogène ($\rho \neq 0$, $\theta = 0$ et $\lambda = 0$)

$$Y = \rho WY + X\beta + \varepsilon$$

- Le modèle SLX (Spatial Lag X) contient une interaction exogène ($\rho = 0$, $\theta \neq 0$ et $\lambda = 0$)

$$Y = X\beta + WX\theta + \varepsilon$$

- Le modèle SEM (Spatial Error Model) contient une interaction des inobservables ($\rho = 0$, $\theta = 0$ et $\lambda \neq 0$)

$$Y = X\beta + u \quad \text{avec } u = \lambda Wu + \varepsilon$$

- Le modèle de Durbin (SDM) contient des interactions exogène et endogène ($\rho \neq 0$ et $\theta \neq 0$) mais peut également prendre en compte des effets d'interaction des inobservables (en posant $\theta = -\rho\beta$ et en notant $u = Y - X\beta$)

$$Y = \rho WY + X\beta + WX\theta + \varepsilon$$

Ce modèle aura ainsi des statistiques de test valides même en cas d'effet d'interaction des inobservables. Néanmoins, nous n'avons pas trouvé dans la littérature de méthode d'estimation permettant de tenir compte de l'endogénéité^a de certaines caractéristiques socio-démographiques des territoires, comme la densité de population, dans les modèles de la partie III du chapitre 3 expliquant les prix des terrains.

- Le modèle de Kelejian-Prucha (SAC) contient des interactions endogène et des inobservables ($\rho \neq 0$, $\theta = 0$ et $\lambda \neq 0$)

$$Y = \rho WY + X\beta + u \quad \text{avec } u = \lambda Wu + \varepsilon$$

Les estimateurs de ce modèle présentent le défaut d'être biaisés et non convergents si le vrai modèle présente des interactions exogènes (LeSage et Pace 2009 [15]). Néanmoins, l'utilisation d'un modèle dit SARAR permet la prise en compte de phénomènes d'endogénéité (cf. partie II.3).

a. Dans un modèle économétrique, une variable explicative est endogène si ses valeurs sont également déterminées par la variable à expliquer. Par exemple, dans le modèle visant à expliquer le prix des terrains, certains déterminants de ces prix comme la densité de population vont également être influencés par ces derniers. Prendre en compte l'endogénéité de certaines variables dans un modèle économétrique nécessite des techniques particulières d'estimation.

Dans un modèle linéaire classique, les coefficients à estimer s'interprètent comme des effets marginaux. Ce n'est pas le cas pour les modèles spatiaux contenant un effet d'interaction endogène, pour lesquels l'effet marginal se décompose en deux effets (cf. tableau 2.1) :

- *un effet direct* : l'effet d'un changement d'une variable explicative X_{ki} de l'individu i sur la variable à expliquer Y_i
- *un effet indirect* : l'effet d'un changement d'une variable explicative X_{kj} d'une unité voisine j sur Y_i

Modèle	Effet direct	Effet indirect
Manski / SDM	$\frac{(3-\rho^2)}{3(1-\rho^2)}\beta_k + \frac{2\rho}{3(1-\rho^2)}\theta_k$	$\frac{(3\rho+\rho^2)}{3(1-\rho^2)}\beta_k + \frac{(3+\rho)}{3(1-\rho^2)}\theta_k$
SAR / SAC	$\frac{(3-\rho^2)}{3(1-\rho^2)}\beta_k$	$\frac{(3\rho+\rho^2)}{3(1-\rho^2)}\beta_k$
OLS / SEM	β_k	0

TABLEAU 2.1– Effets directs et indirects dans les modèles spatiaux

Note de lecture : Ces effets sont calculés pour une variable X_k qui représente un déterminant des prix des terrains. β_k et θ_k sont les coefficients associés à cette variable.

II.3 Un modèle adapté à l'endogénéité des variables explicatives : le modèle SARAR

Pour expliquer les prix des terrains en fonction des caractéristiques socio-démographiques des aires urbaines (cf. partie III du chapitre 3), la modélisation qui a été adoptée afin de pallier le problème d'endogénéité des variables explicatives est un modèle spatial SARAR. Ce modèle est un modèle SAC qui prend en compte un effet d'interaction endogène du prix des terrains dans les différentes aires urbaines ($\rho W_\rho Y$) et un effet d'interaction des inobservables ($\lambda W_\lambda u$). Pour être identifiable, il nécessite l'utilisation de deux matrices de voisinage, W_ρ (interaction endogène) et W_λ (interaction des inobservables).

$$Y = \rho W_\rho Y + X\beta + u \quad \text{avec } u = \lambda W_\lambda u + \varepsilon$$

En estimant ce modèle par la méthode GMM, via la fonction *sprege* du package *sphet* [16] de R, l'hétéroscédasticité⁸ des erreurs peut alors être prise en compte (cf. Kelejian et Prucha, 2007 [12]).

De plus, la fonction *sprege* permet de traiter l'endogénéité à deux niveaux :

1. La présence de WY dans les régresseurs introduit une forme d'endogénéité. Une solution est de l'instrumenter par les variables x retardées.
2. Parmi les régresseurs, il est également possible de trouver des variables endogènes, surtout lorsqu'il s'agit d'expliquer le prix des terrains. C'est le cas notamment de la densité de population. *sprege* permet alors d'instrumenter également ces variables.

Ce double traitement de l'endogénéité peut se faire grâce à une estimation FGS2SLS (*Feasible Generalized Spatial two-stage least squares*) suggérée par Fingleton et Le Gallo, 2008 [6]. Cette méthode est une extension de la méthode GS2SLS, proposée Kelejian et Prucha, 2010 [13], qui permet en plus de prendre en compte des variables explicatives endogènes. L'estimation du modèle par la méthode FGS2SLS consiste à combiner la méthode des variables instrumentales (pour prendre en compte l'autocorrélation spatiale de Y et des variables endogènes) et la méthode GMM (pour corriger l'autocorrélation spatiale des erreurs).

L'estimation est alors réalisée en trois étapes :

1. Un premier modèle (IV-SAR) est estimé par doubles moindres carrés (2SLS) faisant appel à des variables instrumentales à la fois pour la variable WY et les variables endogènes.
2. Les résidus obtenus à l'étape 1 permettent une seconde estimation du paramètre λ par la méthode des moments généralisés (Kelejian et Prucha, 1999 [11]).
3. Le modèle de l'étape 1 est réestimé par la méthode 2SLS en effectuant une transformation de type Cochrane-Orcutt⁹ pour prendre en compte la dépendance spatiale dans les résidus.

8. En statistique, on parle d'hétéroscédasticité lorsque les variances des variables étudiées sont distinctes, ce qui complique les procédés d'estimation.

9. L'estimation de Cochrane-Orcutt est une procédure économétrique permettant d'ajuster un modèle linéaire en cas d'autocorrélation dans le terme d'erreur.

III Krigeage de la carte des prix des terrains

Le lissage d'une carte par krigeage est une technique de géostatistique permettant, à partir de données dispersées, d'obtenir une représentation homogène des prix des terrains [1] [9]. Le krigeage, contrairement à un lissage par régression linéaire classique va prendre en compte le fait que les prix des terrains sont spatialement autocorrélés. Le produit fini du krigeage est une carte quadrillée. À chaque carreau est associée une valeur krigée, calculée à partir des prix des terrains dans ses communes voisines et des résultats d'autocorrélation spatiale.

Cette technique est en général utilisée pour des données ponctuelles comme par exemple des températures captées par des stations météo. Nous l'avons utilisée ici avec des données surfaciques (moyenne de prix au mètre carré par commune). En effet, les logements de la base n'étant pas géolocalisés à un niveau plus fin que communal, nous avons fait comme s'ils étaient tous localisés au centre de leur commune. Nous avons donc utilisé pour le krigeage le niveau géographique le plus fin dont nous disposions : la maille communale.

Le krigeage s'effectue en trois étapes. La première étape consiste à analyser la structure spatiale des données et vérifier l'autocorrélation spatiale via une représentation graphique appelée semi-variogramme (cf. encadré 2.4).

Encadre 2.4 - Le semi-variogramme

Le semi-variogramme est une représentation graphique de la semi-variance définie comme suit :

$$\gamma(h) = \frac{1}{2} \text{Var}(Z(s+h) - Z(s)) \quad \forall s \in D$$

avec $Z(s)$ le prix des terrains au lieu de coordonnées s , h un vecteur de distance et D le champ (ici le quadrillage France entière).

Ainsi, les lieux proches (distance h proche de 0) ont une semi-variance quasi-nulle. À l'inverse, des lieux éloignés ont une semi-variance plus élevée car les prix de leurs terrains sont moins semblables. La semi-variance se stabilise à partir d'une certaine valeur de h , à partir de laquelle il y a absence d'autocorrélation spatiale.

On fait l'hypothèse de modélisation que le semi-variogramme ne dépend pas de s (stationnarité).

La figure 2.5 présente le semi-variogramme obtenu pour le krigeage des données de prix des terrains par commune.

L'effet de pépité correspond à la limite du semi-variogramme en 0. Cette valeur devrait en théorie valoir 0. Elle correspond en pratique à une variation brutale des prix entre deux emplacements très proches.

La portée correspond à la distance à partir de laquelle le variogramme atteint un palier, c'est-à-dire la distance à partir de laquelle il n'y a plus dépendance spatiale entre les données (ici entre 200 et 250 km environ).

La seconde étape du krigeage consiste à estimer le semi-variogramme à l'aide d'une fonction « connue » (ici exponentielle). Cette estimation est nécessaire pour la troisième et dernière étape : le calcul d'un prix krigé (cf. encadré 2.5) en chaque carreau de la France.

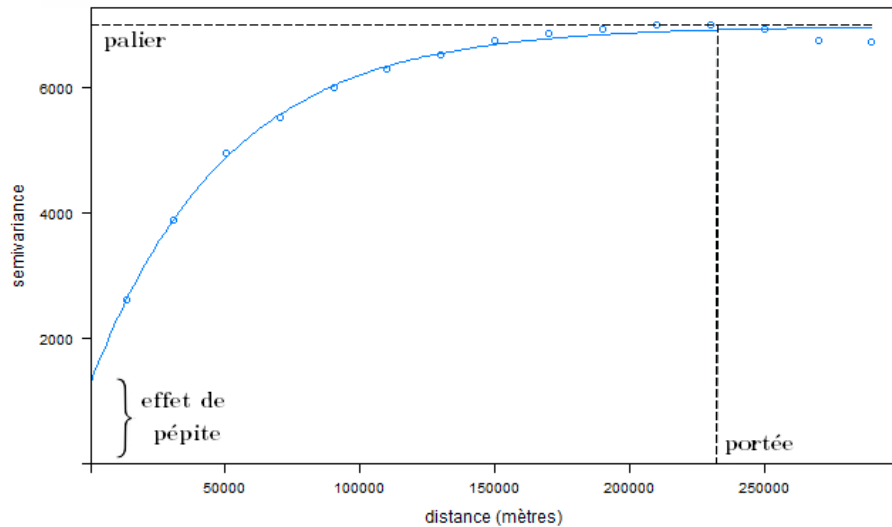


FIGURE 2.5– Semi-variogramme des prix au mètre carré des terrains

Source : EPTB 2013

Note de lecture : Les petits cercles sont les points du semi-variogramme. La courbe correspond au semi-variogramme estimé par une fonction exponentielle.

Encadre 2.5 - Principe général de l'estimation des prix par krigage

Le principe du krigage est d'estimer le prix en tous les points non échantillonnés (quadrillage France entière) à partir d'une combinaison linéaire de données adjacentes de prix par commune.

$$\hat{z}(s_0) = \sum_{i=1}^N \lambda_i \hat{z}(s_i)$$

où :

- $\hat{z}(s_i)$ correspond au prix relevé au centroïde d'une commune s_i
- λ_i est la pondération inconnue de la valeur relevée dans la commune s_i
- s_0 est l'emplacement de prévision (carreau)
- N est le nombre de prix relevés

Les poids λ_i sont estimés de manière à obtenir une prévision non biaisée et de variance minimale de Z . Ils dépendent à la fois de la localisation des observations et de leur structure de dépendance spatiale.

La carte obtenue après le krigage ordinaire¹⁰ des données de prix des terrains dans EPTB 2013 est présentée et analysée dans la partie I du chapitre 3.

10. Le krigage se décline en plusieurs variantes (simple, ordinaire, universel) qui utilisent toutes les mêmes principes. Leurs différences proviennent des hypothèses faites sur l'espérance (connue ou non) et sur la matrice de covariance de Z . Le krigage ordinaire est la méthode la plus utilisée : on considère que l'espérance est inconnue mais définie.

CHAPITRE 3

Les déterminants spatiaux des prix du foncier

En France, le marché du logement est marqué par une forte hétérogénéité spatiale des prix et ce à toutes les échelles géographiques. En effet, comme l'illustre la figure 3.1, les prix sont à la fois très différents entre les aires urbaines mais l'on remarque également au sein de celles-ci une forte hétérogénéité des prix. Par exemple, dans l'aire urbaine de Rouen, les prix diminuent fortement quand on s'éloigne de la capitale régionale jusqu'à être divisés par quatre.

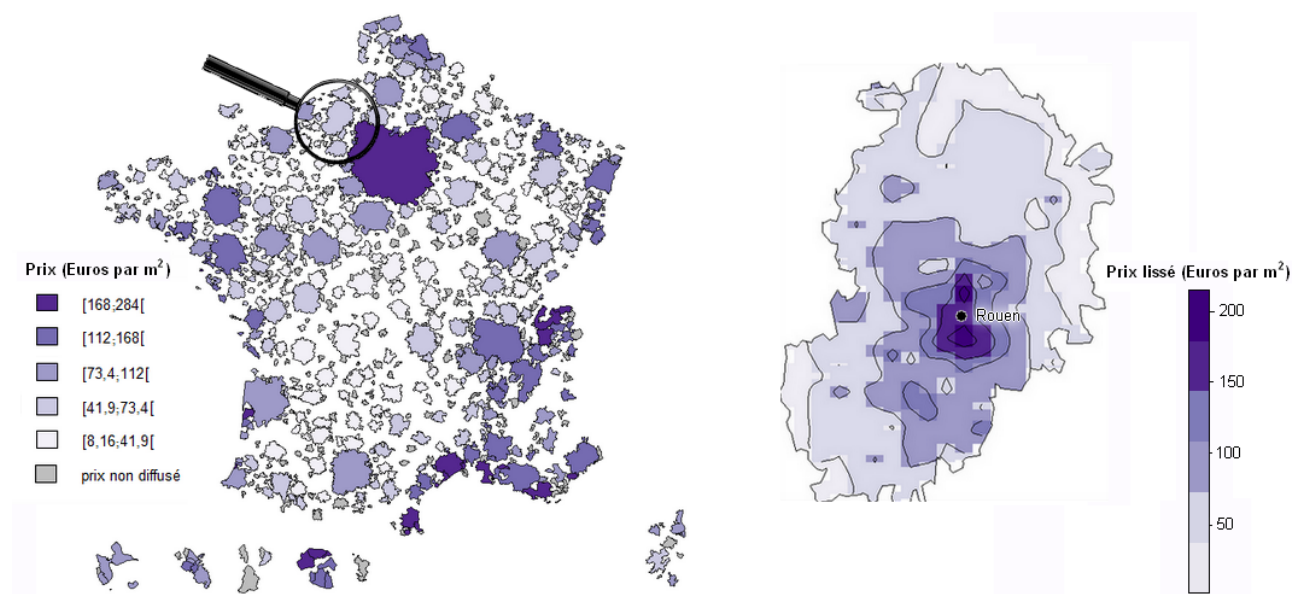


FIGURE 3.1– Hétérogénéité des prix entre et au sein des aires urbaines

Source : EPTB 2013

Pour comprendre ces disparités de prix, il apparaît nécessaire de s'intéresser aux déterminants des prix du logement, fortement liés à la situation géographique des terrains et aux spécificités des territoires qui les accueillent (cf. Combes et al., 2010 [3]). C'est l'objet de ce dernier chapitre.

I Disparité des prix des terrains

Les données d'EPTB montrent une grande diversité des prix des terrains. Ces prix varient de quelques euros à plus de 500 euros le mètre carré. Ils se concentrent pour la plupart aux alentours des 20 à 80 euros le mètre carré (cf. graphique 3.2).

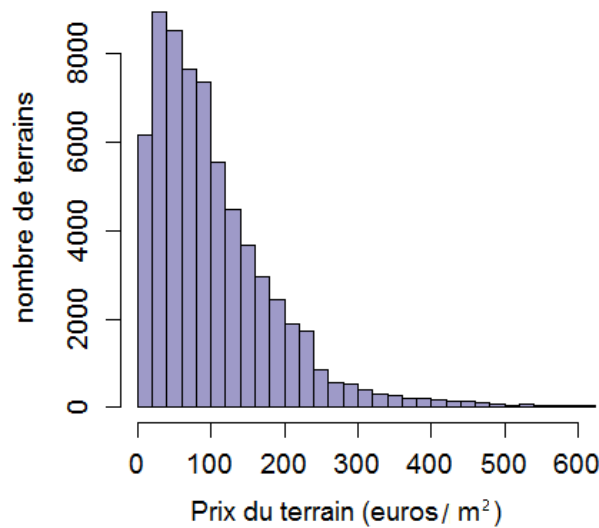


FIGURE 3.2– Distribution des prix par mètre carré des terrains
Source : EPTB 2013

La figure 3.3 présente une carte des prix moyens au mètre carré des terrains contenus dans EPTB 2013. Ces prix, initialement à la commune, ont été lissés sur R par une technique de krigeage (cf partie III et codes en annexe C).

Cette carte nous donne une première intuition de ce que peuvent être les déterminants du prix des terrains :

1. Les prix sont globalement d'autant plus élevés que les espaces urbains sont densément peuplés. Par exemple, la plupart des capitales régionales présentent des prix moyens au mètre carré élevés en 2013.
2. Les prix diminuent progressivement quand on s'éloigne du centre de ces grands espaces urbains. Cela laisse supposer qu'il est possible de mesurer un gradient de prix en fonction de la distance au centre des grands espaces urbains.
3. Les deux premiers points ne suffisent pas à expliquer l'hétérogénéité spatiale des prix des terrains. Certains espaces de densités comparables ne présentent pas les mêmes prix moyens au mètre carré. Ainsi, les zones des littoraux atlantique et méditerranéen et les communes frontalières ont un rapport de prix sur densité bien plus élevé que les capitales régionales. Cela laisse alors supposer qu'il existe d'autres déterminants des prix : des indicateurs socio-démographiques comme peut-être l'attractivité démographique du territoire ou encore sa localisation près d'une frontière ou sur un littoral.

Dans un premier temps, nous vérifierons statistiquement que la distance aux grands espaces urbains ainsi que la densité des espaces sont des déterminants majeurs du prix du foncier. Nous pourrons ainsi calculer des gradients de prix en fonction de la distance au centre des aires urbaines les plus proches. Puis, dans un second temps, nous chercherons les effets que peuvent avoir certaines caractéristiques socio-démographiques des territoires sur le prix du foncier.

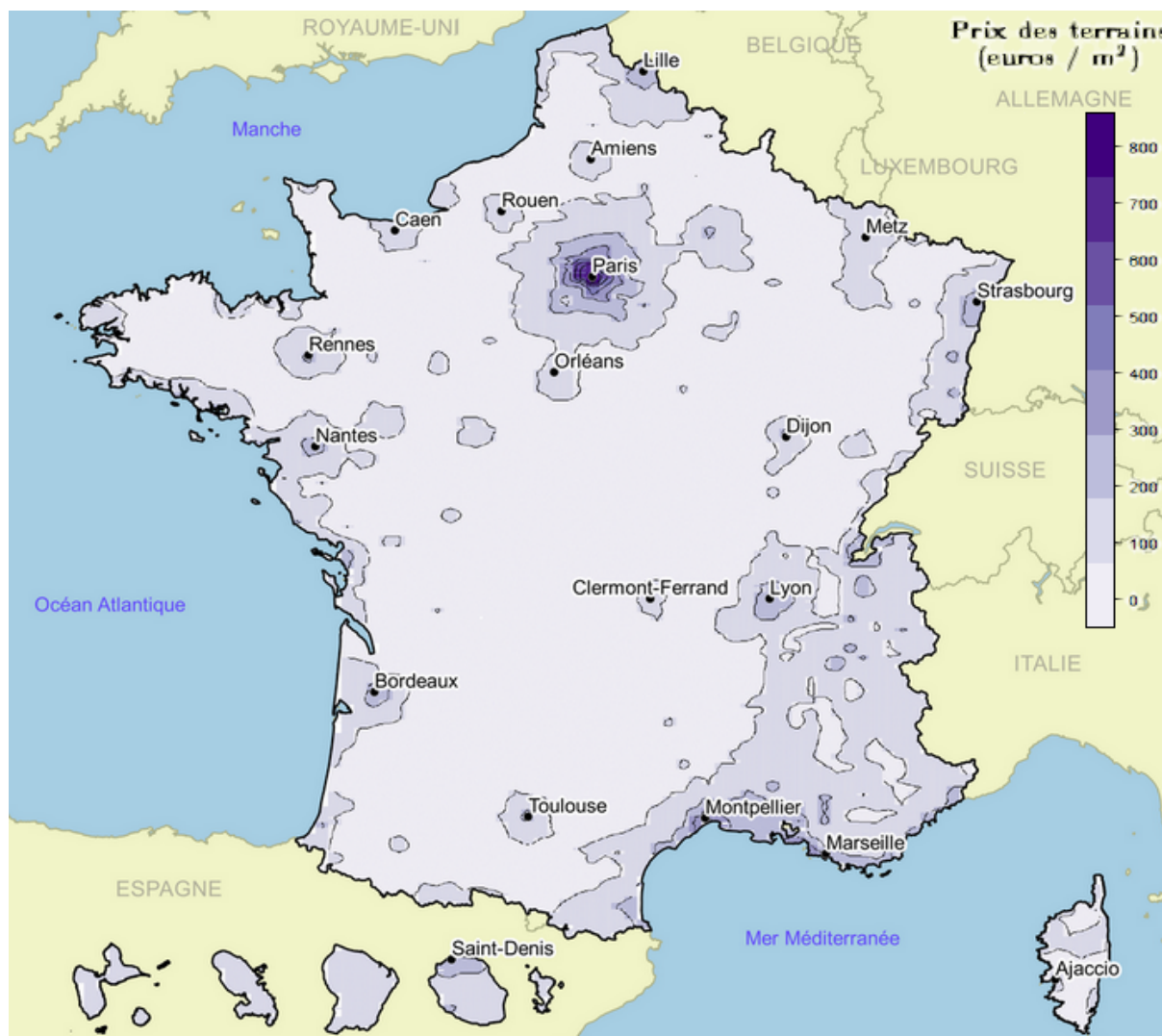


FIGURE 3.3– Prix lissés des terrains à bâtir en France en 2013

Source : EPTB 2013

II Distance aux grands espaces urbains, densité et prix du foncier

Cette partie est consacrée au lien entre les prix des terrains, issus des enquêtes EPTB, la densité des espaces et la distance au centre des aires urbaines les plus proches.

II.1 Des prix plus élevés dans les grands espaces urbains

Les deux cartes en figure 3.4 mettent en regard la densité de la population au dernier recensement (RP 2012) et les prix moyens des terrains par mètre carré (EPTB 2013). Elles laissent pressentir une corrélation positive entre prix et densité. En effet, une densité de population importante correspond à une demande plus forte qui fait augmenter les prix. Les espaces denses comme les capitales régionales et leurs agglomérations ou encore le littoral méditerranéen sont donc les lieux où les prix des terrains sont en général les plus élevés.

Ce fait est confirmé par le graphique 3.5 qui représente les prix au mètre carré des terrains en fonction de la densité des communes qui les accueillent. Là encore, on observe une corrélation positive entre prix et densité de population. D'ailleurs, le coefficient de corrélation linéaire de Pearson indique bien une corrélation significative positive de 0,57.

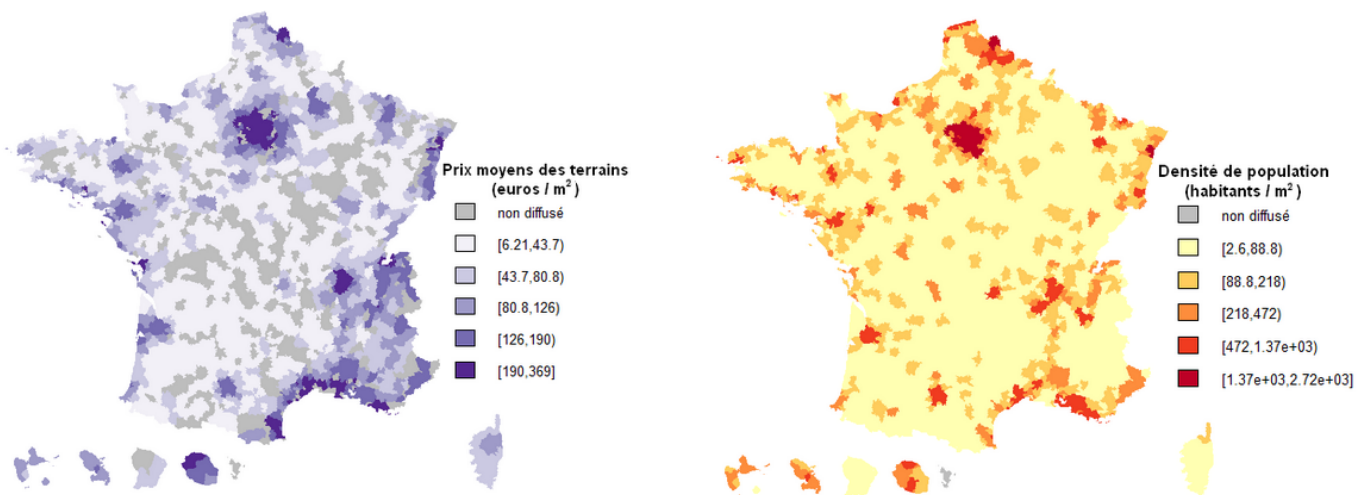


FIGURE 3.4 – Prix des terrains et densité de population

Sources : EPTB 2013, RP 2012

Les terrains vendus sont moins nombreux dans les villages que dans les villes, plus peuplées, où la demande est plus forte. Néanmoins, les très grandes villes françaises comme Lyon ou Marseille contiennent très peu de terrains à bâtir. Quant à la capitale, aucun terrain vendu n'est recensé dans les bases EPTB dans Paris intra-muros.

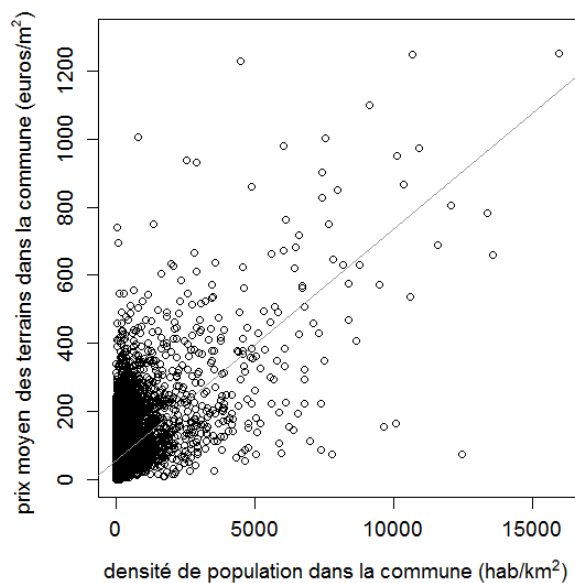


FIGURE 3.5– Prix des terrains en fonction de la densité de la commune

Sources : EPTB 2013, RP 2012

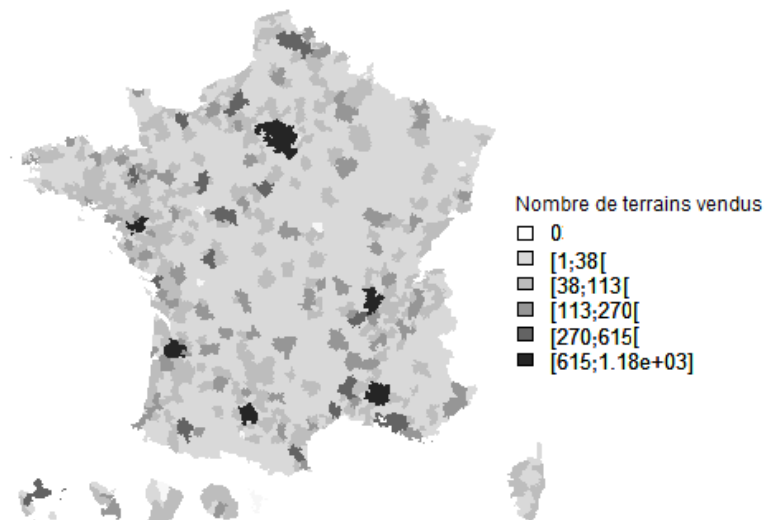
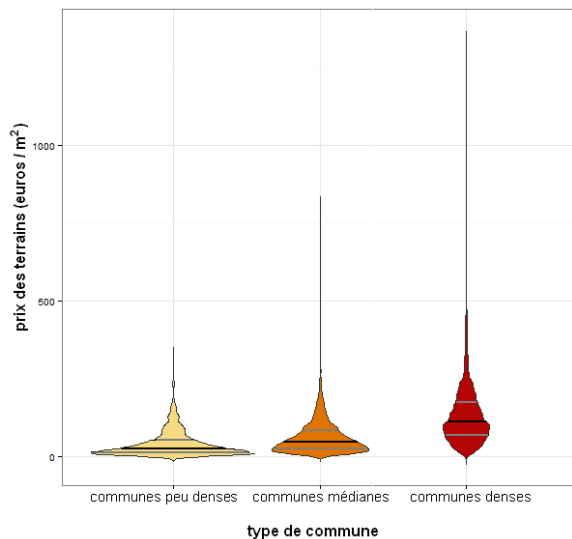


FIGURE 3.6– Nombre de terrains à bâtir par bassin de vie en 2013

Source : EPTB 2013

Afin d'entrer un peu plus dans le détail dans la relation entre prix et densité, nous avons observé la répartition des prix au mètre carré des terrains dans trois catégories de communes réparties selon leur densité :

- les communes peu denses : le premier quart des communes les moins densément peuplées (2 239 terrains en 2013) ;
- les communes médianes : les communes autour de la médiane (22 282 terrains en 2013) ;
- les communes denses : le quart des communes les plus densément peuplées (41 396 terrains en 2013).



Communes	Q1	Médiane	Moyenne	Q3
peu denses	12,5	26,4	41,4	53,2
médianes	25,8	48,0	63,0	84,0
denses	69,5	110,6	136,3	175,0

FIGURE 3.7– Répartition des prix des terrains dans les différents types de communes

Source : EPTB 2013

TABLEAU 3.1– Quantiles des prix des terrains dans les différents types de communes

Source : EPTB 2013

Note : Le diagramme en violon ou en haricot est le mélange d'une boîte à moustaches et d'un histogramme. Il permet ici de voir à la fois la répartition et la fréquence (largeur du violon) des prix au mètre carré des terrains selon la densité des communes qui les accueillent. Comme dans une boîte à moustache, le grand trait noir correspond à la médiane et les traits gris aux premier et troisième quartiles.

Comme attendu au vu de la carte 3.3 des prix des terrains, les communes les moins densément peuplées présentent des prix moins élevés que les communes denses. Ces dernières sont en effet plus attractives grâce à la présence des nombreux équipements et emplois qu'elles accueillent. En effet, les quantiles, médianes et moyennes de prix augmentent en fonction de la densité de population des communes. Par exemple, les médianes valent 26, 48 et 111 euros le mètre carré pour respectivement les communes peu, moyennement et plutôt denses. On remarque aussi grâce au diagramme en violon que les prix sont plus dispersés pour les communes les plus denses alors que les prix dans les communes peu denses se concentrent autour de la médiane. Cela est également en partie dû au nombre d'observations qui est plus élevé dans les communes denses (voir carte 3.6).

II.2 Gradients de prix du logement en fonction de la distance au centre des aires urbaines

Après quelques statistiques descriptives montrant l'influence de la densité de population sur les prix du foncier, nous cherchons, dans cette partie, à calculer un gradient de prix des terrains en fonction de la distance aux grands espaces urbains.

Pour ce faire, nous avons calculé pour chaque terrain une distance au centre urbain le plus proche. Nous avons, dans un premier temps, calculé la distance entre le chef-lieu de la commune accueillant le terrain et la ville-centre du pôle de la grande aire urbaine la plus proche¹.

Dans un second temps, nous avons restreint le champ aux communes appartenant à une aire urbaine puis séparé l'échantillon en trois sous-échantillons : les terrains appartenant aux

1. Ce calcul a été effectué pour les grandes aires urbaines portant le nom d'une ville, comme Paris, Lyon, etc. Pour les aires urbaines ayant plusieurs villes influentes, comme l'aire urbaine de Douai-Lens, a été fait le choix de prendre les coordonnées du centroïde de l'aire urbaine.

communes des petites, moyennes et grandes aires urbaines. La distance au centre urbain, calculée de la même manière que précédemment, est la distance du chef-lieu de la commune du terrain à la ville-centre de l'aire urbaine à laquelle la commune est rattachée.

Ensuite, nous avons régressé le prix de chaque terrain présent dans les enquêtes EPTB de 2006 à 2013 à l'aide d'un modèle hédonique² de ce type :

$$\ln(P_i) = \beta_0 + \beta_S \ln(S_i) + \beta_V \mathbb{1}_{nonV} + \beta_D D_i + \beta_{T_1} \mathbb{1}_{2007} + \dots + \beta_{T_7} \mathbb{1}_{2013} + \beta_{AU_{X_1}} \mathbb{1}_{AU_{X_1}} + \dots + \beta_{AU_{X_N}} \mathbb{1}_{AU_{X_N}} + \varepsilon_i \quad (3.1)$$

avec :

- P_i , le prix du terrain i s'il a été acheté ;
- S_i , sa surface ;
- $\mathbb{1}_{nonV}$, une indicatrice de « non-viabilisation du terrain » (qui vaut 1 si le terrain n'est pas viabilisé)³ ;
- les $\mathbb{1}_{20XX}$, les sept indicatrices temporelles permettant, en isolant l'inflation, de comparer des terrains vendus à des années différentes, l'année de référence choisie étant 2006 ;
- D_i , la distance au centre de l'aire urbaine (AU) calculée comme expliqué précédemment. Le coefficient associé à cette variable correspond à un gradient de prix en fonction de la distance aux grands espaces urbains. On intuite que plus la distance au centre de l'AU sera élevée, plus les prix seront faibles (cf. graphique 3.8) ;
- les $\mathbb{1}_{AU_{XXX}}$, une indicatrice par AU permettant de capter les effets propres des aires urbaines et de limiter ainsi le biais de variables omises. L'aire urbaine que nous choisirons comme référence correspondra à l'aire urbaine médiane en terme de prix moyen par mètre carré des terrains par aire urbaine. Ces coefficients serviront, dans la partie III, à calculer un prix de référence par AU d'un terrain viabilisé de 1000 m² en 2006 placé au centre du pôle de chaque aire urbaine.

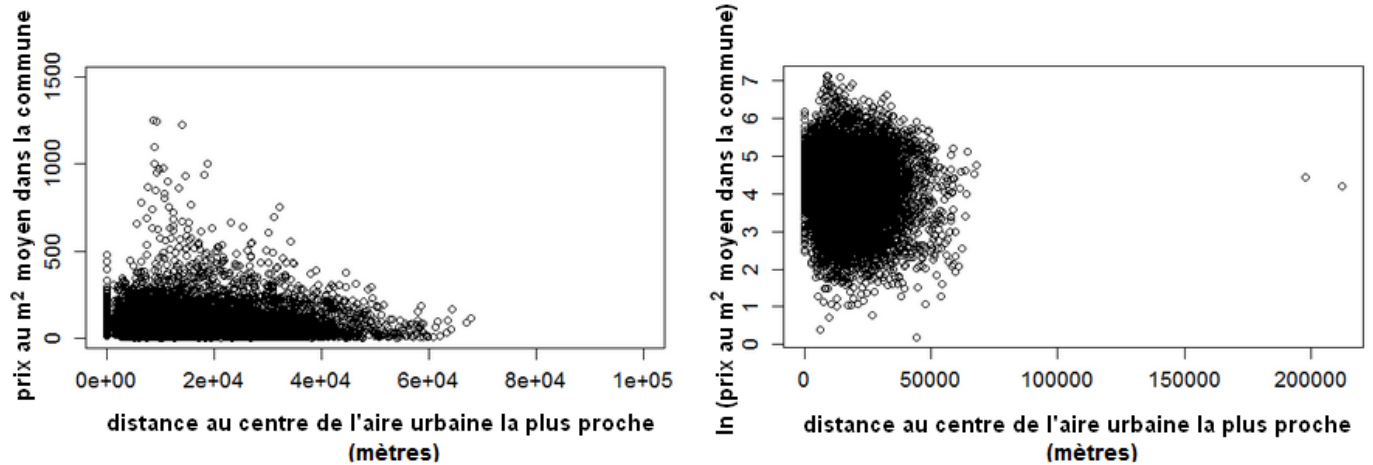


FIGURE 3.8— Prix des terrains et logarithme du prix des terrains en fonction de la distance à la grande aire urbaine la plus proche

Sources : EPTB 2013

Nous avons choisi de réaliser des transformations logarithmiques pour les variables P_i et S_i . En effet, cette transformation permet d'interpréter le coefficient devant S_i comme une élasticité. De même, concernant le gradient de prix (coefficient devant D_i), il sera alors possible de dire de combien de pourcents chute le prix quand on s'éloigne d'un km du centre de l'aire urbaine⁴. De

2. La définition ainsi que les fondements théoriques de la modélisation hédonique du prix du logement sont expliqués en annexe B.

3. Voir la définition de la viabilisation en partie II du chapitre 1

4. Cette approximation est possible pour les petites valeurs de β . En effet, le développement limité de l'exponentielle nous donne $\% \Delta_{P_i} = 100 \cdot (\exp \beta_D - 1) \cdot \Delta_{D_i} \simeq 100 \cdot \beta_D \cdot \Delta_{D_i}$

plus, une transformation logarithmique permet de limiter l'influence de certains outliers et les problèmes d'hétéroscédasticité.

Le test de White indiquant la présence d'hétéroscédasticité dans les modèles suivants, nous avons corrigé les valeurs des statistiques de tests, et des p-valeurs associées, concernant la significativité des différents coefficients.

Les coefficients estimés dans les modèles suivants sont néanmoins à prendre avec précaution en raison du biais de variables omises. En effet, le fait d'omettre des variables explicatives du prix du logement, comme les caractéristiques des territoires qui les accueillent⁵, peut fausser l'estimation des modèles économétriques.

Gradient de prix sur l'ensemble des communes appartenant à une aire urbaine

Nous avons, dans un premier temps, appliqué le modèle 3.1 à l'ensemble des terrains i appartenant à une aire urbaine. D_i correspond donc à la distance entre le chef-lieu de la commune accueillant le terrain et le centre de la grande aire urbaine la plus proche. Les indicatrices $\mathbb{1}_{AU_{xxx}}$ non présentes dans le tableau 3.2 par souci de visibilité, sont au nombre de 791 avec en plus en référence l'aire urbaine médiane de Saintes (132) en Charente-Maritime.

Variable	Coefficient
Constante	8,92*
$\ln(S_i)$	0,27*
$\mathbb{1}_{nonV}$	-0,17*
$\mathbb{1}_{2007}$	0,12*
$\mathbb{1}_{2008}$	0,16*
$\mathbb{1}_{2009}$	0,13*
$\mathbb{1}_{2010}$	0,13*
$\mathbb{1}_{2011}$	0,18*
$\mathbb{1}_{2012}$	0,21*
$\mathbb{1}_{2013}$	0,21*
D_i	$-2,03^* \cdot 10^{-5}$
R^2	0,59
Observations	330 649

TABLEAU 3.2– Estimation des coefficients du modèle 3.1 appliqué à toutes les aires urbaines

Le modèle 3.1, appliqué à l'ensemble des communes appartenant à une AU, semble expliquer en grande partie les prix des terrains. En effet, tous les coefficients (hors indicatrices par aire urbaine) sont significatifs (*) à 1 % et le R^2 atteint 0,59. Les estimations de ce modèle, présentes en tableau 3.2 nous confirment plusieurs éléments :

1. Le prix d'un terrain est d'autant plus élevé que sa surface l'est. Le coefficient de la variable D_i étant inférieur à 1, cela signifie que, toutes choses égales par ailleurs, le prix augmente moins fortement que la surface augmente. En d'autres termes, ici, si on « augmente » la surface du terrain d'un pourcent, son prix n'augmente que de 0,27 %.
2. Le fait qu'un terrain ne soit pas viabilisé va jouer en défaveur de son prix.
3. L'effet des indicatrices temporelles sur les prix augmente dans le temps, avec l'exception des années 2009 et 2010 post-crises où les coefficients chutent au niveau de celui de 2007.
4. Le fait d'être éloigné du centre de la grande aire urbaine la plus proche a un effet négatif significatif sur les prix. Ici, on peut dire que le fait de s'éloigner qu'un kilomètre du centre de l'aire urbaine fait chuter le niveau de prix de 2,03 %⁶.

5. Les caractéristiques socio-démographiques feront partie des variables explicatives des prix des terrains en partie III.

6. La conversion de kilomètres en mètres est possible grâce au développement limité $(1 + 100 \cdot \beta_D)^{1000} \simeq (1 + 1000 \times 100 \cdot \beta_D) \simeq (1 + 100\,000 \cdot \beta_D)$ car $100 \cdot \beta$ est petit ($2,03 \cdot 10^{-3}$). Attention, ces coefficients ne sont valables qu'autour du centre de l'aire urbaine, sinon, les prix finiraient pas être gratuits en s'éloignant.

Ainsi, le coefficient associé à la variable $\ln(S_i)$ ($-2,03 \cdot 10^{-5}$) correspond au gradient de prix selon l'éloignement des terrains aux grands espaces urbains. Il confirme l'intuition selon laquelle les prix sont de plus en plus faibles quand on s'éloigne des centres urbains (diminution de 2,03 % par kilomètre). Il reste maintenant à examiner si ce gradient a une valeur équivalente selon les différents types d'aires urbaines et également à repérer les aires urbaines où ce gradient est particulièrement faible ou élevé.

Gradient par type d'aire urbaine

Dans cette partie, le modèle 3.1 est appliqué successivement à trois champs différents : les terrains appartenant aux petites, moyennes, puis grandes aires urbaines. D_i correspond cette fois-ci à la distance entre le chef-lieu de la commune du terrain et le centre de l'aire urbaine auquel il appartient. Les variables sont donc identiques dans les trois modèles à l'exception des indicatrices par aire urbaine $\mathbb{1}_{AU_{xxx}}$:

- Les indicatrices des petites aires urbaines sont au nombre de 418 avec en plus en référence l'aire urbaine médiane de Montoire-sur-le-Loir (590) dans le Loir-et-Cher.
- Le modèle sur les moyennes aires urbaines comporte 129 indicatrices par aire urbaine avec en plus en référence l'aire urbaine médiane de Oloron-Sainte-Marie (232) en Pyrénées-Atlantiques.
- Le modèle sur les grandes aires urbaines contient 239 indicatrices avec en plus en référence l'aire urbaine de Saint-Brieuc (051) en Côtes-d'Armor.

Variable	Petites AU	Moyennes AU	Grandes AU	Toutes AU
Constante	8,04*	8,59*	8,86*	8,92*
$\ln(S_i)$	0,35*	0,33*	0,30*	0,27*
$\mathbb{1}_{nonV}$	-0,18*	-0,15*	-0,18*	-0,17*
$\mathbb{1}_{2007}$	0,13*	0,13*	0,12*	0,12*
$\mathbb{1}_{2008}$	0,18*	0,21*	0,16*	0,16*
$\mathbb{1}_{2009}$	0,19*	0,20*	0,12*	0,13*
$\mathbb{1}_{2010}$	0,20*	0,21*	0,13*	0,13*
$\mathbb{1}_{2011}$	0,22*	0,23*	0,18*	0,18*
$\mathbb{1}_{2012}$	0,24*	0,26*	0,20*	0,21*
$\mathbb{1}_{2013}$	0,22*	0,25*	0,20*	0,21*
D_i	$-2,16^* \cdot 10^{-5}$	$-2,96^* \cdot 10^{-5}$	$-2,34^* \cdot 10^{-5}$	$-2,03^* \cdot 10^{-5}$
R^2	0,63	0,58	0,61	0,59
Observations	24 142	19 947	286 560	330 649

TABLEAU 3.3— Estimation des coefficients du modèle 3.1 appliqué par type d'aire urbaine

Pour chacun de ces modèles, des remarques identiques à précédemment peuvent être faites à savoir une élasticité inférieure à 1 du prix par rapport à la surface et l'influence négative de la non-viabilisation et de la distance au centre de l'AU sur les prix. Néanmoins, on remarque quelques spécificités suivant les champs étudiés.

Tout d'abord, les indicatrices temporelles n'ont pas le même effet selon les tailles des aires urbaines. Une baisse des prix se ressent dans les grandes aires urbaines après la crise de 2008 mais pas dans les petites et moyennes aires où la baisse des prix a plutôt lieu en 2013⁷.

Enfin, le gradient de prix en fonction de la distance est plus élevé, en valeur absolue, pour les moyennes que pour les petites et grandes aires urbaines. En d'autres termes, les moyennes aires urbaines sont les lieux où les prix décroissent le plus quand on s'éloigne de leurs pôles. En effet, les prix dans les petites aires urbaines étant déjà les moins chers, la possibilité de décroissance

7. Le fait que les coefficients soient sensiblement les mêmes à l'échelle des grandes aires et pour l'ensemble des terrains appartenant à une AU vient probablement du fait qu'une grande partie de l'échantillon EPTB concerne les terrains de grandes aires urbaines, plus denses et donc riches en terrains à bâtir.

des prix en est donc rendue plus faible. Le fait que le gradient soit moins fort pour les grandes que pour les moyennes aires peut, quant à lui, provenir du fait que les grandes aires urbaines possèdent une plus grande centralité, qui implique des prix forts dans leur espace périurbain. Toutefois, les gradients restent sensiblement les mêmes selon les échelles, d'une décroissance valant entre 2 et 3 % par km.

Gradients par aire urbaine

Dans cette partie, nous avons une nouvelle fois utilisé le modèle 3.1 mais cette fois-ci par grande aire urbaine. Nous avons donc restreint le champ à chacune des grandes AU et retiré les variables indicatrices par AU du modèle. Le modèle suivant est appliqué à chaque ensemble de terrains appartenant à une grande aire urbaine fixée :

$$\ln(P_i) = \beta_0 + \beta_S \ln(S_i) + \beta_V \mathbf{1}_{nonV} + \beta_D D_i + \beta_{T_1} \mathbf{1}_{2007} + \dots + \beta_{T_7} \mathbf{1}_{2013} + \beta_{AU_{X_1}} + \varepsilon_i \quad (3.2)$$

Ces modélisations permettent d'observer les grandes aires urbaines qui ressortent avec un gradient particulièrement faible ou élevé. Nous avons résumé dans le tableau 3.4 les gradients de prix des aires urbaines des capitales régionales ainsi que leur rang dans le classement par ordre décroissant des gradients (en valeur absolue) des 240 grandes aires urbaines présentes dans la base EPTB. Nous avons également tracé en graphique 3.9 les prix dans quatre de ces AU en fonction de leur distance au centre.

Rang	Aire urbaine	Gradient de prix
36	Reims	-4,59 .10 ⁻⁵
46	Dijon	-4,44 .10 ⁻⁵
53	Amiens	-4,31 .10 ⁻⁵
69	Poitiers	-3,98 .10 ⁻⁵
80	Limoges	-3,82 .10 ⁻⁵
84	Clermont-Ferrand	-3,77 .10 ⁻⁵
106	Besançon	-3,18 .10 ⁻⁵
107	Caen	-3,17 .10 ⁻⁵
112	Metz	-3,08 .10 ⁻⁵
118	Strasbourg	-3,03 .10 ⁻⁵
121	Châlons-en-Champagne	-2,99 .10 ⁻⁵
122	Rennes	-2,99 .10 ⁻⁵
130	Nantes	-2,75 .10 ⁻⁵
132	Orléans	-2,75 .10 ⁻⁵
140	Rouen	-2,60 .10 ⁻⁵
143	Toulouse	-2,52 .10 ⁻⁵
144	Lille	-2,49 .10 ⁻⁵
150	Lyon	-2,22 .10 ⁻⁵
151	Paris	-2,11 .10 ⁻⁵
168	Montpellier	-1,79 .10 ⁻⁵
187	Bordeaux	-1,20 .10 ⁻⁵
213	Marseille - Aix-en-Provence	-0,55 .10 ⁻⁵

TABLEAU 3.4– Gradients de prix par grande aire urbaine

Les grandes aires urbaines les plus étendues et denses (Paris, Lyon et Marseille notamment) ressortent avec un gradient plus faible en valeur absolue. Cela suit la même logique que le fait que les grandes aires urbaines ont un gradient global plus faible que les moyennes : leur centralité en termes d'emploi, d'équipements et de transports rendent la décroissance des prix moins rapide. Ces résultats s'observent non seulement dans le tableau 3.4 des coefficients mais également dans le graphique 3.9 du nuage de points de la répartition des prix des terrains en fonction de la distance au centre dans quatre grandes aires urbaines : celles de Paris et Marseille où les gradients sont moins forts qu'à Reims et Dijon.

Marseille est une ville qui ressort avec un gradient particulièrement faible, ce qui n'est pas étonnant car cette ville est réputée pour avoir une banlieue plus chère que ses arrondissements les plus excentrés.

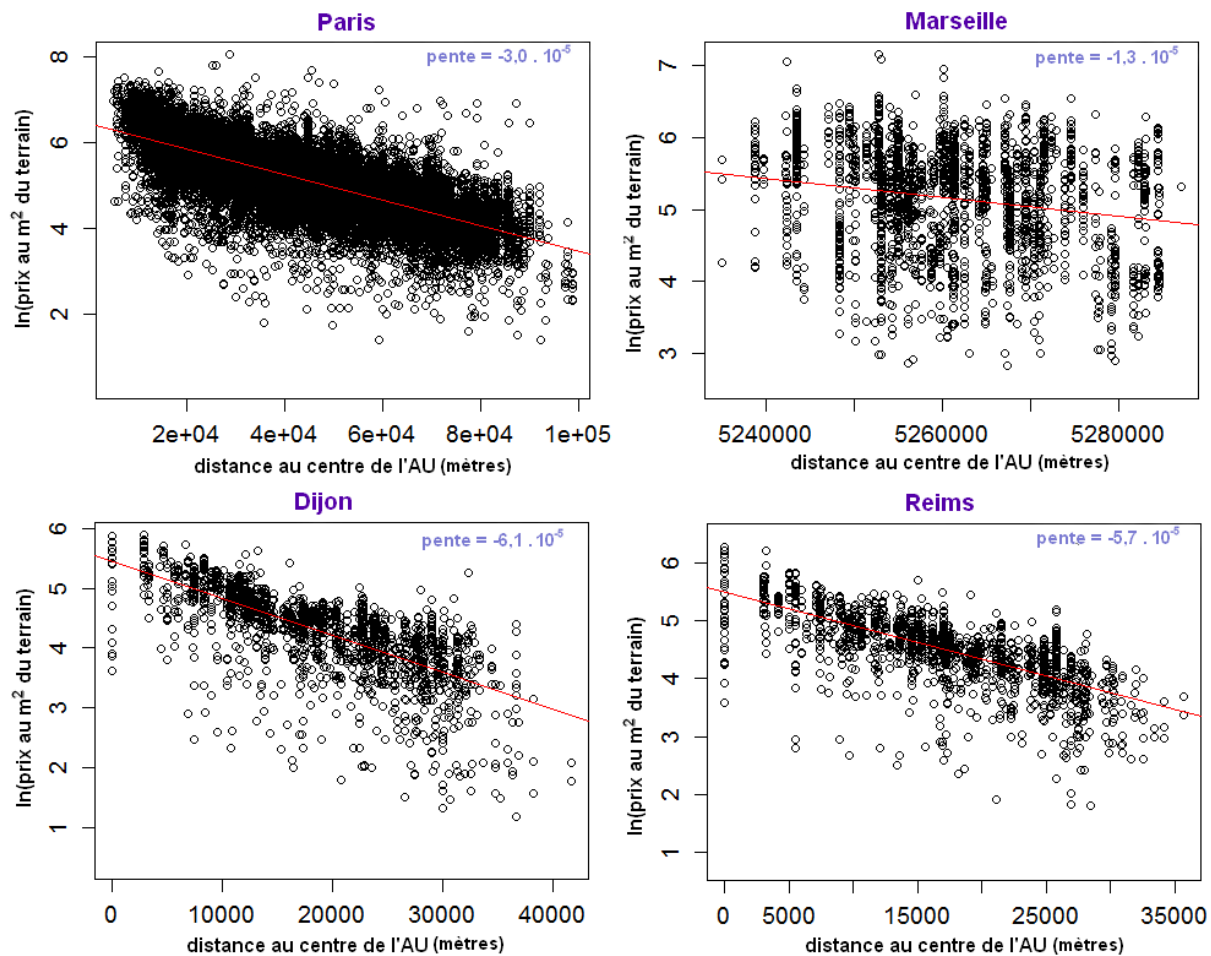


FIGURE 3.9– Prix et distance au centre dans quatre grandes aires urbaines

Source : EPTB 2013

III Influence des spécificités des territoires sur les prix des terrains

Après avoir mis en évidence l'effet de la distance aux grands espaces urbains, densément peuplés, sur les prix des terrains, nous allons analyser, dans cette partie, l'effet d'autres caractéristiques socio-démographiques des aires urbaines.

Dans les modélisations économétriques qui suivent, nous prendrons comme champ d'étude l'ensemble des aires urbaines de France métropolitaine⁸ et utiliserons comme définition de proximité des aires urbaines la matrice des 10 plus proches voisins⁹.

Afin de parfaire nos modèles, au lieu de prendre comme variable à expliquer le prix moyen au mètre carré des terrains dans chaque aire urbaine, nous avons utilisé les résultats de la régression de la partie II.2. En effet, à partir des coefficients associés à chaque aire urbaine ($1_{AU_{XXX}}$), nous avons pu calculer pour chacune d'entre elles le prix de référence d'un terrain de 1000 m^2 viabilisé,

8. Nous avons exclu les DOM de l'étude en raison de certaines variables, comme le revenu médian des habitants, qui ne sont pas disponibles dans ces territoires. De plus, les coordonnées géographiques des DOM étant usuellement dans des projections différentes que celles du reste de la France, nous avons préféré les retirer pour éviter les erreurs dans les calculs des voisinages.

9. cf. explications et justifications en partie I du chapitre 1.

vendu en 2006, placé en son centre. Ces prix de référence ont l'avantage de résumer les prix dans chaque aire urbaine en mettant de côté l'effet de l'inflation, celui des caractéristiques propres aux terrains (viabilisation) et la proximité géographique des terrains au centre de l'aire urbaine.

La figure 3.10 présente la carte des prix de référence des aires urbaines et celle de leur autocorrélation spatiale. Ces cartes sont semblables à celles qui concernent les prix moyens par mètre carré des terrains. Leurs interprétations sont alors identiques à celles effectuées dans la partie I du chapitre 3 sur l'autocorrélation spatiale des prix des terrains et dans la partie introductive de ce chapitre.

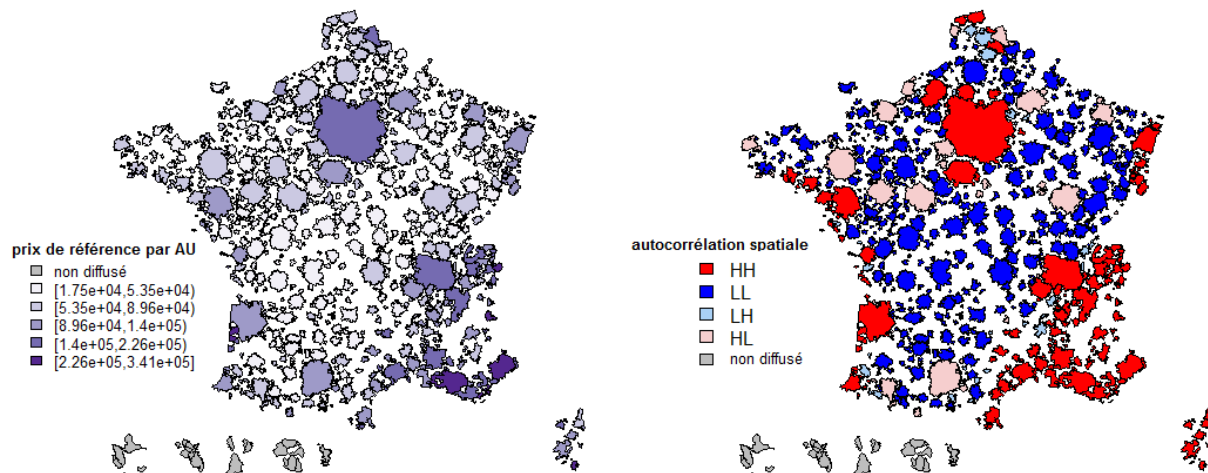


FIGURE 3.10– Prix de référence par aire urbaine et leur autocorrélation spatiale

Source : EPTB 2006-2013, RP 2012

Endogénéité de certaines variables explicatives

Les variables explicatives des modèles correspondent à des caractéristiques sociodémographiques et géographiques permettant de décrire les aires urbaines. Ces variables, détaillées dans le tableau 3.5 concernant à la fois la population présente dans chaque aire urbaine (âge, revenu, nombre d'emplois dans l'aire urbaine...), la situation géographique du territoire (proximité d'une mer ou d'un océan, d'une frontière, zones montagneuses et climat) l'accessibilité (équipements et population utilisant les transports en commun) et enfin le parc des logements (part de logements sociaux et résidences secondaires, âge et occupation du parc).

Comme dans beaucoup de modélisations du prix de l'immobilier, de nombreuses variables explicatives sont endogènes¹⁰. Les variables potentiellement concernées sont dans la catégorie « endogènes » du tableau 3.5.

La densité de population et son évolution entre 1999 et 2012 (*densite* et *densite_evol*) dans chaque aire urbaine jouent sur le prix des terrains par effet de demande. Réciproquement, les prix vont également influencer la population présente sur le territoire : des prix faibles peuvent notamment inciter la population à s'y installer. Nous avons donc choisi d'instrumenter la densité et son évolution par respectivement la part des jeunes de moins de 20 ans dans l'aire urbaine et l'évolution de cette part (*partm20ans* et *partm20ans_evol* dans la partie « instruments »¹¹ du tableau 3.5). En effet, ces variables sont très corrélées avec les variables endogènes et ne le sont pas avec le prix (cf. partie III.1 sur les corrélations entre variables). Les jeunes ne sont effectivement a priori pas concernés par le prix des terrains puisque l'accession à la propriété se fait très rarement avant l'âge de 20 ans.

10. Pour rappel, une variable explicative est endogène si ses valeurs sont également déterminées par la variable à expliquer. Prendre en compte l'endogénéité de certaines variables dans un modèle économétrique nécessite des techniques particulières d'estimation.

11. En économétrie, les instruments des variables endogènes sont des variables qui sont très corrélées à celles-ci mais qui ne sont pas corrélées au terme d'erreur et donc ici au prix. Ces variables instrumentales permettent d'estimer le modèle économétrique en tenant compte de l'endogénéité de certaines variables.

	Code	Variable	Sources
<i>exogènes</i>	<i>hlm</i>	Part des logements ¹² sociaux parmi les résidences principales	Insee, RP 2012
	<i>resid2</i>	Part des résidences secondaires parmi l'ensemble des logements	Insee, RP 2012
	<i>mont</i>	AU comprenant des communes de montagne	IGN
	<i>litto</i>	AU littorale	IGN
	<i>trans</i>	AU transfrontalière	IGN
	<i>climat</i>	Zone climatique ¹³ en fonction des températures hivernales (H1, H2, H3)	Météo France
<i>endogènes</i>	<i>densite</i> ⁽¹⁾	Densité de la population	Insee, RP 2012
	<i>densite_evol</i> ⁽²⁾	Évolution de la densité de population	Insee, RP 1999-2012
	<i>emplois</i>	Nombre d'emplois de long terme pour 100 habitants	Insee, RP 2012
	<i>logvac</i>	Part des logements vacants sur l'ensemble des résidences principales	Insee, RP 2012
	<i>emmen10p</i>	Part de la population ayant emménagé il y a plus de 10 ans	Insee, RP 2012
	<i>acht1946</i>	Part des résidences principales datant d'avant 1946	Insee, RP 2012
	<i>chom</i>	Part des demandeurs d'emplois en fin de mois sur la population active	Insee, RP 2012
	<i>equip</i>	Nombre d'équipements pour 1000 habitants	Insee, BPE ¹⁴ 2013 - RP 2012
	<i>revmed</i>	Revenu disponible médian ¹⁵	Insee, Filosofi 2012
	<i>deplt_tc</i>	Part de la population ¹⁶ utilisant les transports en commun pour aller travailler	Insee, RP 2012
<i>instruments</i>	<i>Partm20ans</i> ⁽¹⁾	Part des moins de 20 ans	Insee, RP 2012
	<i>Evol_partm20ans</i> ⁽²⁾	Evolution de la part des moins de 20 ans	Insee, RP 1999-2012

TABLEAU 3.5– Caractéristiques géographiques et socio-démographiques des aires urbaines

De nombreuses autres variables peuvent également être considérées comme endogènes aux prix des terrains, c'est-à-dire être influencées par celui-ci :

- Le prix du foncier a un effet certain sur le niveau des revenus de la population résidente (*revmed*). En effet, si un terrain a un prix trop élevé, le revenu médian dans l'aire urbaine en sera donc plus élevé par effet d'éviction des ménages à bas revenus, ces derniers ne pouvant accéder à la propriété.
- Le prix du foncier peut jouer sur le nombre d'emplois (*emplois*) présents dans l'aire urbaine. Un prix élevé peut en effet dissuader les entreprises et les petits commerces (*equip*) de s'y installer et par conséquent jouer sur le taux de chômage (*chom*). Un prix élevé peut également, dans certains cas, inciter les commerces à s'implanter si ces lieux sont attractifs

12. Un logement social est un logement destiné, à la suite d'une initiative principalement publique, exclusivement à des personnes dont les revenus sont modestes.

13. La répartition des départements en zones climatiques, élaborée notamment par Météo France, a été effectuée en fonction des températures en période hivernale (douces pour H1, moyennes pour H2, froides pour H3).

14. La Base Permanente des Équipements (BPE) est une base Insee, en partie géolocalisée, destinée à renseigner sur le niveau d'équipements et de services présents sur un territoire.

15. Le revenu disponible comprend les revenus d'activité et du patrimoine, les transferts en provenance d'autres ménages et les prestations sociales nets des impôts directs

16. Il s'agit de la population au lieu de travail.

avec des populations à haut revenus.

- Les variables concernant le parc du logement vont également être touchées par le niveau des prix des terrains dans l'aire urbaine. En effet, des prix faibles peuvent, par exemple, diminuer la part des logements vacants (*logvac*) ou augmenter la part des personnes ayant emménagé depuis longtemps (*emme10p*), ces personnes pouvant, grâce aux prix faibles, avoir un niveau de vie raisonnable.
- Le secteur de la construction est également très lié au niveau des prix. Ces derniers peuvent influencer à la hausse ou à la baisse la construction et jouer ainsi sur l'âge global du parc (*acht1946*). Le secteur des transports va également être sensible aux prix. Une hausse des prix des terrains pouvant influencer le coût des transports, cela peut également jouer sur la part des personnes utilisant les transports en commun (*deplt_tc*).

Ainsi, les spécifications des modèles économétriques de la partie III devront tenir compte de plusieurs phénomènes :

1. l'endogénéité de nombreuses caractéristiques socio-démographiques des aires urbaines dans l'explication du prix des terrains ;
2. l'hétéroscédasticité dans les résidus ;
3. l'autocorrélation spatiale des prix des terrains ;
4. le choix nécessaire parmi la multitude de variables explicatives à notre disposition (pour éviter la présence de multicollinéarité), ainsi que leurs formes fonctionnelles (par exemple transformations logarithmiques cf. annexe B sur la méthode hédonique) ;

Ces nombreux paramètres, surtout s'ils sont à prendre en compte simultanément, rendent complexe la recherche de modèles économétriques valides.

III.1 Analyse multivariée des corrélations entre les caractéristiques socio-démographiques des territoires

Le cercle¹⁷ en figure 3.11, issu d'une ACP¹⁸ illustre les corrélations des caractéristiques socio-démographiques des territoires, corrélations entre elles mais également avec le prix de référence par aire urbaine (*prxm2*).

Les variables explicatives qui semblent les plus corrélées positivement avec le prix des terrains sont le revenu médian des habitants de l'aire urbaine (*revmed*), la densité de population (*densite*) et son évolution (*densite_evol*). La part des personnes utilisant les transports en commun (*deplt_tc*) pour aller travailler, la part des résidences secondaires (*resid2*) et le nombre d'équipements pour 1000 habitants (*equip*) sont également corrélées positivement au prix. Cela n'est pas étonnant car ces variables correspondent aux caractéristiques des territoires urbains denses, où les prix sont plus élevés (voir partie II.1).

Les variables les plus corrélées négativement au prix des terrains sont les variables décrivant un âge du parc élevé (*acht1946*), une grande part de logements vacants (*logvac*) et de populations ayant emménagé depuis plus de 10 ans (*emmen10p*, potentiellement des personnes plutôt âgées vivant donc dans des communes plutôt rurales où les prix sont globalement plus faibles). De même, un fort taux de chômage (*chom*) influe négativement sur les prix.

La part de logements sociaux, quant à elle, apparaît non corrélée aux prix des terrains. En effet, le choix d'allouer des appartements pour les populations à faibles revenus est politique, il ne répond pas au système d'offre et demande. Le nombre d'emplois, bien que mal représenté¹⁹

17. Dans un cercle de corrélations, les variables dont les droites forment un angle proche de 0° sont très corrélées positivement, un angle proche de 180° sont très corrélées négativement et un angle proche de 90° ne sont pas corrélées.

18. Une Analyse en Composantes principales (ACP) est une technique d'analyse statistique multivariée permettant de représenter sous forme graphique le plus d'informations possibles concernant les corrélations entre variables ou entre individus statistiques.

19. Le nombre d'emplois est une variable mal représentée dans cette ACP car la flèche qui la représente dans le cercle de corrélation est de petite taille.

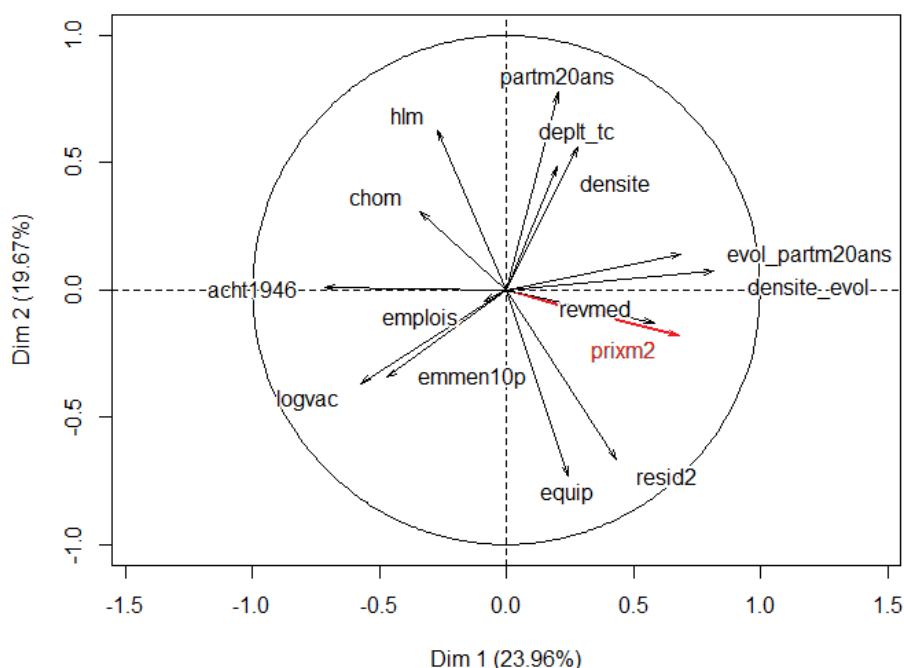


FIGURE 3.11– Cercle de corrélation des spécificités des aires urbaines

Sources : RP 2012, BPE 2013, et EPTB 2006-2013

dans cette ACP, semble également peu corrélé avec les prix des terrains. Cela peut s'expliquer par le fait que certains emplois sont localisés dans des lieux où les locaux sont relativement peu chers, et d'autres, nécessitant d'être dans des espaces denses (commerces) sont, au contraire, localisés dans des lieux où les prix sont élevés.

Enfin, comme indiqué dans la partie dédiée au problème d'endogénéité, les variables *densite* et *densite_evol* sont très corrélées à leurs instruments respectifs : la part des moins de 20 ans (*partm20ans*) et son évolution (*partm20ans_evol*).

III.2 Focus sur la densité de population et l'attractivité des territoires

Dans cette partie, nous avons restreint les variables explicatives du prix des terrains dans les aires urbaines à la densité de population (*densite*), son évolution (*densite_evol*) ainsi qu'aux spécificités géographiques des territoires (*litto*, *mont*, *trans* et *climat*²⁰). Nous nous sommes en effet cantonnés dans un premier temps aux deux principales variables explicatives endogènes pour lesquelles nous avons des instruments que sont la part des moins de 20 ans (*partm20ans*) et son évolution (*partm20ans_evol*).

Avec ces variables, nous avons estimé les paramètres de plusieurs modèles d'économétrie spatiale décrits dans la partie II du chapitre 2. Ces estimations sont présentées dans le tableau 3.6.

Nous constatons dans les estimations des divers modèles spatiaux que la densité de population (*densite*), l'attractivité démographique (*densite_evol*) et le fait d'être en zones littorale et transfrontalière (*litto* et *trans*) ont toujours un effet significatif positif sur le prix des terrains dans les aires urbaines. À l'inverse, être dans une zone de montagne (*mont*) ou dans la zone climatique chaude (*climat*) ne semble pas avoir d'effet significatif.

Parmi les modèles spatiaux du tableau 3.6, les tests de multiplicateurs de Lagrange *LM*, qui comparent la vraisemblance des modèles, nous conduisent à retenir les modèles SDM et SAC. La minimisation du critère d'information de l'AIC²¹ nous amène à préférer le modèle SAC au modèle SDM.

20. Comme variable renseignant sur le climat, nous avons simplement créé une indicatrice valant 1 si l'aire urbaine appartient à la zone climatique chaude H3. En effet, la zone climatique H2 ne semblait pas avoir d'effet significatif dans le modèle estimé par moindres carrés ordinaires (OLS).

21. Le critère d'information d'Akaike (AIC) mesure de la qualité d'un modèle statistique. En le minimisant, on maximise la vraisemblance du modèle tout en prenant en compte la parcimonie du modèle.

	OLS	SAR	SEM	SLX	SDM	SAC
<i>constante</i>	9,90*	1,21*	10,56*	9,25*	1,45*	2,57*
<i>ln(densite)</i>	0,18*	0,10*	0,12*	0,12*	0,11*	0,11*
<i>densite_evol</i>	1,25*	0,50*	0,39*	0,63*	0,42*	0,46*
<i>litto</i>	0,31*	0,20*	0,26*	0,26*	0,24*	0,23*
<i>trans</i>	0,31*	0,18*	0,20*	0,21*	0,18*	0,18*
<i>mont</i>	0,24*	0,04	-0,02	0,00	-0,03	0,02
<i>climat</i>	0,55*	-0,03	0,09	-0,05	0,02	0,06
<i>AIC</i>	888	371	398	777	370	353

TABLEAU 3.6– Estimations des modèles spatiaux classiques

Source : EPTB 2006-2013, RP 2012, Météo France

Note de lecture : les coefficients significatifs au seuil de 1 % sont marqués d'un astérisque. Le modèle OLS correspond au modèle estimé par la méthode des moindres carrés ordinaires. Les autres modèles ont été présentés en partie II du chapitre 2. Aucun de ces modèles ne fait appel à la méthode des variables instrumentales.

Le modèle SAC est le modèle duquel découle le modèle SARAR estimé par la méthode FGS2SLS. Cette méthode d'estimation, également décrite en partie II du chapitre 2 permet de tenir compte simultanément de l'endogénéité de certaines variables explicatives et de l'autocorrélation spatiale. Ses résultats sont présentés en tableau 3.7.

	SARAR
<i>constante</i>	-0,31
<i>ln(densite)</i>	0,04
<i>densite_evol</i>	0,31*
<i>litto</i>	0,19*
<i>trans</i>	0,16*
<i>mont</i>	-0,02
<i>climat</i>	-0,13*
ρ	1,01*
λ	-0,16

TABLEAU 3.7– Estimations du modèle SARAR

Source : EPTB 2006-2013, RP 2012, Météo France

Note de lecture : les coefficients significatifs au seuil de 1 % sont marqués d'un astérisque. Le modèle SARAR a été présenté en partie II du chapitre 2. Il fait appel à la méthode des variables instrumentales pour instrumenter les deux variables endogènes *densite* et *densite_evol* par respectivement *partm20ans* et *partm20ans_evol*. W_ρ correspond à la matrice des 10 plus proches voisins et W_λ à celle de l'inverse de la distance au carré avec comme seuil 100km.

Ainsi, les variables faisant augmenter les prix des terrains dans les aires urbaines semblent être l'évolution de la densité de population entre 1999 et 2012 et les appartenances aux zones littorales et transfrontalières²². À l'inverse, la densité de population ainsi que le climat et le fait d'appartenir à une zone de montagne ne semblent pas avoir d'effet sur les prix dans ce modèle. Ces résultats sont à prendre avec précaution en raison de la complexité des modèles et de leur robustesse qui n'est pas assurée.

III.3 Un exemple de modèle plus complet des déterminants spatiaux des prix des terrains

Cette dernière partie a pour objectif de déterminer, parmi les multiples variables socio-démographiques et géographiques permettant de décrire les aires urbaines, celles qui ont le plus

22. Nous n'avons malheureusement pas calculé les effets directs et indirects associés aux coefficients. En effet, le ρ obtenu est supérieur à 1 alors que le calcul suppose qu'il soit inférieur à 1.

d'influence sur le prix des terrains.

Les coefficients d'un exemple²³ de modèle SARAR où tous les coefficients sont significatifs figurent en tableau 3.8. ρ étant inférieur à 1, nous avons cette fois-ci pu calculer les effets directs et indirects des différentes variables (voir ligne SAC du tableau II.2 du chapitre 2)

	SARAR	effet direct	effet indirect
<i>constante</i>	-10,55*		
<i>hlm</i>	0,47*	0,91*	1,09*
<i>ln(emplois)</i>	-0,22*	-0,42*	-0,50*
<i>logvac</i>	-0,75*	-1,44*	-1,71*
<i>ln(equip)</i>	0,56*	1,07*	1,28*
<i>deplt_tc</i>	2,09*	4,03*	4,79*
<i>ln(revmed)</i>	1,13*	2,19*	2,60*
<i>ln(densite)</i>	0,15*	0,30*	0,35*
<i>densite_evol</i>	0,29*	0,57*	0,67*
ρ	0,76*		
λ	0,07		

TABLEAU 3.8— Estimations du modèle SARAR complet

Source : EPTB 2006-2013, RP 2012, BPE 2013, Météo France

Note de lecture : les coefficients significatifs au seuil de 1 % sont marqués d'un astérisque. Le modèle SARAR a été présenté en partie II du chapitre 2. Il fait appel à la méthode des variables instrumentales pour instrumenter les deux variables endogènes *densite* et *densite_evol* par respectivement *partm20ans* et *partm20ans_evol*. W_ρ correspond à la matrice des 10 plus proches voisins et W_λ à celle de l'inverse de la distance au carré avec comme seuil 100km.

Dans les variables significatives de ce modèle, nous retrouvons l'effet positif sur le prix de la densité de population et de son évolution. Par exemple, une hausse de 1 % de la densité fait augmenter, toutes choses égales par ailleurs, de 0,3 % le prix de référence des terrains dans l'aire urbaine. De même, une hausse de prix des terrains de 1 % dans le voisinage de cette AU fait augmenter son prix de référence de 0,35 %. Les variables géographiques (climat, présence de littoral, montagne ou frontière) ne sont plus présentes dans le modèle. Ces variables sont remplacées par l'effet à la hausse sur le prix de la présence d'équipements et de logements sociaux, le haut revenu des habitants et la grande part des personnes utilisant les transports en commun. Une hausse de la part des logements vacants fait, quant à elle, baisser les prix des terrains car augmente l'offre de logements. Enfin, la densité d'emplois a également un effet négatif sur les prix : ceci n'est pas étonnant au vu de sa non corrélation apparente dans le cercle de corrélation en figure 3.11.

Ainsi, les modèles économétriques confirment l'intuition que nous avons concernant les déterminants des prix des terrains. En effet, les variables caractérisant des territoires urbanisés font jouer les prix à la hausse : la densité de population, l'attractivité du territoire, les hauts revenus de la population, la densité en équipements ou encore l'utilisation par les travailleurs de transports en commun. Toutefois, la robustesse des modèles et l'interprétation des coefficients sont à prendre avec beaucoup de précaution. En effet, par exemple, la variable *litto* qui était significative dans le modèle précédent ne l'est plus dans les modèles comprenant de plus nombreuses variables. De plus, seule l'endogénéité de deux variables (*densite* et *densite_evol*) a été traitée. Les variables endogènes comme les revenus médians de l'aire urbaine (*revmed*) devraient également être instrumentées pour obtenir des résultats plus précis.

23. Pour sélectionner ce modèle, nous avons gardé uniquement les coefficients significatifs dans le modèle comprenant toutes les variables du tableau 3.5. Une méthode plus adaptée aurait été d'effectuer un algorithme de sélection de variables, du type *stepwise*, que nous n'avons pas encore implémenté sous R pour les modèles spatiaux.

Bilan et Prolongements

Après une analyse de l'autocorrélation spatiale des prix des terrains, ce mémoire a abouti, grâce à des modélisations d'économétrie spatiale, à une première analyse des déterminants possibles des prix du foncier. Les modèles économétriques obtenus sont certes imparfaits, mais la complexité méthodologique d'une telle étude (endogénéité, multicollinéarité, autocorrélation spatiale...) rend la recherche d'un modèle complet, statistiquement valide, très difficile. Des choix doivent donc être effectués. Dans sa présentation des différents types de modèles hédoniques, Sheppard (1999, [17]) indique que la spécification du modèle choisi doit, avant tout, dépendre des données utilisées et de l'objectif de l'analyse. Il faut notamment arbitrer, selon les cas, entre un modèle avec un bon pouvoir prédictif et un modèle dont les coefficients sont plus stables. Par exemple, Bruno Vermont (2015, [20]), souhaitant rechercher les déterminants spatiaux des prix du logement, a passé en revue de nombreuses variables sans pouvoir raisonnablement traiter de l'endogénéité de chacune d'elle. Quant à Combes *et al.* (2012, [3]), ils se sont concentrés sur un nombre plus restreint de variables explicatives et ont ainsi pu explorer les manières les plus rigoureuses de les instrumenter. Ces deux études, pourtant très complètes, ont mis de côté le traitement de l'autocorrélation spatiale des données.

Les méthodes présentées dans ce mémoire pourraient toutefois être améliorées par la suite. Il serait en effet intéressant de rechercher des instruments pour chacune des variables explicatives endogènes et de tester leur validité à l'aide de tests adaptés. Nous pourrions notamment vérifier l'exogénéité de certaines variables grâce à un test d'Hausman et la non-corrélation entre les instruments des variables endogènes choisis et le terme d'erreur via un test de Sargan. Rechercher dans la littérature un modèle spatial tenant compte de l'endogénéité a été chronophage. De plus, un seul modèle permettant de le faire a été trouvé (SARAR) : intégrer des variables endogènes dans tous les autres modèles spatiaux, notamment le modèle de Durbin, permettrait de comparer et choisir le modèle le plus adapté. Pour affiner la recherche d'un modèle valide, il pourrait être judicieux d'implémenter, pour les modèles spatiaux, une procédure de sélection du meilleur modèle explicatif possible (du type *forward*, *backward* ou *stepwise*) ou encore d'examiner plus en profondeur les liens de multicollinéarité entre certaines variables pouvant parfois masquer leurs effets explicatifs. En introduisant dans les modèles des coefficients qui varient dans l'espace, nous pourrions aussi explorer l'hétérogénéité spatiale des données, c'est-à-dire les variations locales de l'effet des déterminants des prix sur ces derniers. Enfin, au lieu de choisir la maille géographique des aires urbaines et de leur pôle comme symbole de centralité, nous pourrions utiliser la théorie des graphes, outil permettant d'analyser plus en détail les relations entre territoires.

Au-delà du perfectionnement des modèles et de leur validité, nous pouvons imaginer un prolongement de cette étude et un affinement des résultats grâce à la version future d'EPTB qui sera géolocalisée. Cette géolocalisation permettra alors d'être beaucoup plus précis dans les calculs de distance aux centres, déterminant majeur des prix du foncier.

Conclusion

En somme, les disparités des prix des logements s’observent à tous les nouveaux géographiques : entre mais également au sein des aires urbaines.

Ces différences de prix ne s’arrêtent pas aux caractéristiques physiques des logements. En se focalisant sur le champ des terrains à bâtir issus de l’enquête EPTB, les caractéristiques sociodémographiques et géographiques des espaces ont un impact significatif sur les prix du foncier. En outre, ces prix sont non seulement déterminés par les caractéristiques des territoires qui les accueillent, mais également par les prix des terrains et leurs déterminants dans des espaces voisins.

Les prix des terrains sont influencés par la densité de population de la commune dans laquelle ils se trouvent. Les espaces denses, urbanisés, sont en effet des lieux où la demande en logement est forte : les équipements et emplois y sont nombreux et les rendent attractifs. Cette constatation conduit à examiner l’effet, à caractéristiques physiques des terrains fixées, de la proximité des terrains aux grands espaces urbains grâce à un gradient de prix en fonction de la distance au centre de l’aire urbaine la plus proche. Comme attendu, les prix apparaissent d’autant plus faibles que les terrains sont loin des grands centres urbains, d’une décroissance globale de 2 % par kilomètre parcouru loin du centre.

D’autres caractéristiques propres aux différents territoires, différents de leur densités et éloignements aux espaces denses, influent sur le prix des terrains. En effet, les spécificités des territoires urbains font jouer les prix à la hausse : l’attractivité démographique, les hauts revenus de la population, la densité en équipements ou encore l’utilisation par les travailleurs de transports en commun. Toutefois, la caractérisation urbaine n’est pas l’unique facteur explicatif des prix qui ressort de cette étude : des informations plus géographiques comme la proximité au littoral et aux frontières font également augmenter les prix.

De multiples voies de prolongation de cette étude sont possibles. D’un point de vue méthodologique, l’utilisation de nouveaux instruments, de tests concernant leur validité, ainsi que de procédures de sélection de modèles, pourraient améliorer la robustesse des modèles et affiner l’interprétation de leurs coefficients. De même, les futures données d’EPTB géolocalisées apporteront une meilleure précision dans le calcul des distances aux centres qui pourront même être analysées, à destination des collectivités locales, à une échelle infracommunale.

En prenant connaissance des principaux déterminants des prix du foncier, les décideurs publics peuvent identifier les leviers pertinents sur lesquels les politiques publiques peuvent agir pour limiter les inégalités résultant des disparités du prix du foncier. Grâce aux résultats obtenus sur les prix des terrains, ils pourront observer les zones de tension sur le marché de l’immobilier auxquelles il est possible de venir en aide via la fiscalité immobilière et foncière locales ou via diverses aides au financement comme, par exemple, le prêt à taux zéro en Hauts-de-Seine ou le Prêt Paris Logement.

ANNEXE A

Inventaire des principales données sur le logement

Avant de débiter cette étude portant sur le logement, nous avons effectué une revue des données de la statistique publique concernant le logement et le foncier. Vous trouverez dans cette annexe un inventaire des principales sources (administratives et enquêtes) ainsi que des différents producteurs de ces données. La documentation trouvée à ce sujet a pu être largement complétée grâce aux renseignements recueillis lors de la commission CNIS « démographie et questions sociales » du 7 mai 2015.

Lors de cette commission, un rapport sur l'organisation du service statistique dans le domaine du logement (J. Friggit A. Chappert, Juin 2014 [8]) a été présenté. Ce document dresse un bilan des différentes sources et de leurs limites. Il est également composé de recommandations pour que ces sources soient plus facilement mobilisables. Par exemple, ce rapport recommande que l'enquête logement, actuellement gérée par l'Insee, soit sous la responsabilité du SOeS, ou encore que l'Insee prévoie sur son site internet un portail dédié aux différentes sources sur le logement. En effet, ce rapport recense et décrit 79 sources différentes dans le domaine du logement. Nous ne présenterons bien entendu pas ces 79 sources mais les principales, et celles sur lesquelles nous nous sommes particulièrement renseignés. D'autres sources, comme le recensement de la population, le dispositif SRCV (statistiques sur les ressources et conditions de vie), les enquêtes famille et logement, patrimoine et sans domicile et bien d'autres, ne doivent pas être oubliées.

L'enquête Logement de l'Insee, une référence

L'enquête Logement est une enquête phare de l'Insee en raison de son grand échantillon (40 000 logements) et de son ancienneté (1955). C'est la source statistique la plus importante pour décrire le parc de logements et ses conditions d'occupation par les ménages. Elle permet également de mesurer le coût du logement que supportent les différents types de ménages. Cependant, elle n'est pas une référence concernant le prix du foncier ou des loyers.

L'association Notaires-Insee pour le logement ancien

L'Insee, en partenariat avec l'association des Notaires de France, construit un indice sur les prix des logements anciens dit « Notaires-Insee » (1994) qu'il met à disposition chaque trimestre sur son site internet. Cet indice est construit selon une méthodologie hédonique uniquement à partir de données notariales.

En effet, depuis quelques années, les notaires collectent les multiples informations présentes sur les actes de vente pour en constituer une base de données : BIEN à Paris et PERVAL en province. Ces bases sont d'une grande richesse puisqu'elles enregistrent les transactions pour tous les types de biens (appartement, maison mais également immeubles, terrains, garages et autres locaux) et contiennent un grand nombre d'informations : caractéristiques du bien immobilier mais également de ses propriétaires.

Néanmoins, cette base étant constituée au bon vouloir des notaires, le taux de couverture des bases notariales n'était que de 66% en 2003 pour l'ensemble de la France (83% pour l'Île-de-France et 59% pour la province). De plus, les données ne sont pas directement disponibles et représentent un coût non négligeable pour quiconque voudrait les exploiter.

Les enquêtes du ministère de l'Écologie pour le logement neuf

Deux enquêtes majeures sont effectuées par le ministère de l'écologie, du développement durable et de l'énergie : l'enquête sur le prix des terrains à bâtir (EPTB) et l'enquête sur la commercialisation des logements neufs (ECLN). Ces enquêtes nécessitent l'utilisation de la base Sit@del qui recense les permis de construire.

La première enquête est mensuelle et son exploitation est annuelle. Elle se penche sur les terrains à bâtir destinés à la construction d'une maison individuelle et permet de suivre l'évolution du coût du foncier. On s'intéresse ici au type d'acquisition (prix, surface, avancement des travaux, caractéristiques de la maison) mais également aux caractéristiques du pétitionnaire (âge, PCS) ainsi qu'à la part qu'occupe le prix du terrain dans le coût total d'acquisition d'une maison.

La seconde enquête est trimestrielle. Elle se restreint à suivre la commercialisation des logements neufs destinés à la vente aux particuliers. Elle permet alors de connaître le prix au mètre carré des habitations.

le CGDD pour des enquêtes plus thématiques

L'enquête Performance de l'Habitat, Équipements, Besoins et USages de l'énergie (Phébus) est une enquête nouvelle et ponctuelle. Elle contient un volet « CLODE » sur les caractéristiques des résidences principales de France métropolitaine (taille, type de chauffage, liste des travaux, des équipements énergivores, consommation d'énergie), de ses occupants et des dépenses. Parmi les enquêtes concernées par ce volet, certains sont concernés par un second volet « DPE » qui consiste à faire effectuer par des diagnostiqueurs qualifiés et certifiés un diagnostic de performance énergétique (DPE) officiel visant à mesurer la performance énergétique du logement.

Les sources administratives

La source administrative sur le logement la plus mobilisée est Filocom, gérée par le SOeS. Ce fichier est fondé sur le rapprochement de fichiers fiscaux : taxe d'habitation, impôt sur le revenu, fichier foncier des propriétés bâties et fichier des propriétaires. Il permet donc d'avoir une richesse d'informations sur les caractéristiques du logement et des ménages y résidant. Malheureusement, ce fichier ne comporte pas d'information sur le prix du foncier ou le montant des loyers. Néanmoins, ce fichier est amené à être remplacé vers 2018 par le projet Insee du répertoire statistique du logement (RSL) qui utilisera, en plus des sources déjà présentes dans Filocom, d'autres sources administratives : répertoire des immeubles localisés et éventuellement base des demandes de valeurs foncières (DVF) pour avoir accès au prix du logement.

Le locatif entre Insee, OLAP et SOeS

L'enquête trimestrielle Loyers et charges de l'Insee permet le suivi des loyers à la fois neufs et anciens grâce à un panel effectué sur cinq trimestres consécutifs. Les évolutions sont donc mesurées à qualité constante des logements. Grâce aux résultats de cette enquête, l'Insee a la possibilité de réaliser une analyse conjoncturelle du secteur locatif (à la fois dans le secteur libre et social).

Une autre entité s'occupant des questions du locatif est l'Observatoire des loyers. Il collecte de l'information sur le niveau et l'évolution des loyers du parc locatif libre de l'agglomération parisienne et de onze agglomérations de province et en diffuse notamment quelques données agrégées. La collecte est multisource : les données proviennent à la fois d'agences, de notaires, de bailleurs, de caisses d'allocations familiales, etc.

Concernant le secteur social, le ministère de l'Écologie utilise un répertoire des logements locatifs fourni par les bailleurs sociaux (RPLS). L'Insee effectue également une enquête sur les bailleurs sociaux.

ANNEXE B

La méthode des prix hédoniques

La méthode des prix hédoniques (Rosen, 1974) part de l'hypothèse de modélisation suivante : un logement j est un bien hétérogène (prix unitaire non unique) composé d'un ensemble de K caractéristiques $Z_j = (z_{1j}, \dots, z_{Kj})$: un panier de biens homogènes vendus en bloc. Ces caractéristiques peuvent être intrinsèques au logement (surface, étage, confort...) mais également extrinsèques (localisation et caractéristiques des territoires qui l'accueillent).

La méthode hédonique a alors pour objectif de déceler les prix marginaux implicites p_k de ces caractéristiques à partir des prix globaux P_j des logements observés sur le marché explicite de l'immobilier. La théorie économique donne à l'équilibre :

$$p_k = \frac{\partial P(Z)}{\partial z_k}$$

Autrement dit, chaque prix marginal implicite p_k est égal au consentement marginal à payer de l'acheteur pour la caractéristique z_k .

Les prix marginaux p_k sont calculés grâce à l'estimation d'une fonction de prix hédoniques f en régressant le prix de vente du logement sur ses caractéristiques :

$$P_j = P(Z_j) = \hat{f}(z_{1j}, \dots, z_{Kj}, \varepsilon_j)$$

où :

- f est la fonction hédonique que l'on cherche à approcher afin de connaître les prix marginaux, \hat{f} est donc une approximation de cette fonction.
- ε_j sont les erreurs de mesures (qui suivent par exemple une loi gaussienne). Elles servent à capturer les effets des caractéristiques des biens omises dans la modélisation.

La régression hédonique prendra donc la forme suivante :

$$\tilde{P}_j = \beta_0 + \sum_k \beta_k \tilde{z}_{kj} + \sum_l \beta_l z_{lj} + \varepsilon_j$$

avec :

- \tilde{P}_j est une transformation du prix du logement (les types de transformation possibles sont indiqués dans le tableau B.1) ;
- \tilde{z}_{kj} sont certaines variables explicatives transformées (variables continues comme la surface) ;
- z_{lj} sont des variables explicatives non transformées (indicatrices d'appartenance à une AU, de non-viabilisation du terrain...) ;
- β_k et β_l sont des paramètres à estimer.

La théorie suppose que f n'est pas une fonction linéaire des z_k . C'est pourquoi nous éviterons la spécification linéaire-linéaire. Nous avons choisi d'utiliser dans notre mémoire les spécifications semi et double logarithmiques.¹

1. Une autre transformation, la transformation Box-Cox $\frac{X^\lambda - 1}{\lambda}$, est souvent utilisée dans la littérature. Comme avec le logarithme, il est possible d'effectuer cette transformation sur les variables explicatives et la variable à expliquer. Néanmoins, les coefficients sont plus complexes à interpréter qu'avec une spécification logarithmique.

Transformation	\tilde{P}_j	\tilde{z}_{kj}
Linéaire-linéaire	P_j	z_{kj}
Semi-logarithmique	$\ln(P_j)$	z_{kj}
Double-logarithmique	$\ln(P_j)$	$\ln(z_{kj})$

TABLEAU B.1— Formes fonctionnelles des variables dans les modèles hédoniques

Ces régressions hédoniques doivent être réalisées en gardant à l'esprit que la plupart des caractéristiques des logements sont endogènes à leurs prix. Autrement dit, les prix influencent ces caractéristiques de la même manière que ces caractéristiques influencent les prix. Ce fait biaise les estimations par moindres carrés ordinaires : il faut alors recourir à des méthodes de variables instrumentales.

Ainsi, la méthode des prix hédoniques est celle que nous utilisons pour comparer les prix des terrains dans les différentes aires urbaines et en déduire leurs déterminants. En effet, l'utilisation de cette méthode repose sur des bases théoriques solides très souvent utilisées et permet de comparer cette étude sur le prix des terrains à d'autres études déjà réalisées.


```
# packages

library(sp) # importation et représentation de cartes
library(rgdal) # définitions des objets spatiaux
library(maptools) # définition de cartes
library(gdistance) # SIG : calcul de distance, calcul de géostatistique
library(raster) # SIG : calcul de distance, calcul de géostatistique
library(fields) # SIG : calcul de distance, calcul de géostatistique
library(RColorBrewer) # couleurs
library(classInt) # créer des intervalles pour les cartes (type Jenks)
library(gstat) # krigeage
library(spdep) # modèles d'économétrie spatiale
library(sphet) # modèles d'économétrie spatiale suite (SARAR)

#####
##### Cartographie #####
#####

carte <- readOGR(".", " Au2010")
vPal5 <- brewer.pal(n = 5, name = "Purples")
lJenks <- classIntervals(var = carte@data$prixm2 , n = 5, style = "jenks")
vJenks <- lJenks$brks
carte@data$legende <- as.character (cut(carte@data$prixm2 , breaks = vJenks ,
  labels = vPal5 , include.lowest = TRUE , right = FALSE))
carte@data$legende[is.na(carte@data$legende)]<-"#BDBDBD"
vLegendBoxJ5 <- as.character ( levels (cut( carte@data$prixm2 , breaks = vJenks ,
include.lowest = TRUE , right = FALSE )))
par(mar=c(0.5,0.5,0.5,0.5))
plot (carte , col = carte@data$legende , border = "black", cex=0.001)
legend ("bottomleft", legend = c("non diffusé",vLegendBoxJ5) , bty = "n",
fill = c("grey",vPal5) , cex = 0.8 , title = "Nombre d'habitants au m2")
title ( main = " Prix au m2 par aire urbaine (EPTB 2013)")
dev.off()

#####
##### krigeage #####
#####

# variogramme
vgm<-variogram(prixm2~1,data,cutoff=300000)
plot(vgm,pch=3,type="o")
```

```

# ajustement modèle
vgm.fit<-fit.variogram(vgm,vgm(psill=7000000,model="Exp",range=40000, nugget = 2500000))

# grille de krigage
x.range <- as.integer(c(-900000,1300000))
y.range <- as.integer(c(4870000,6800000))
grd <- expand.grid(x=seq(from=x.range[1], to=x.range[2], by=20000), y=seq(from=y.range[1],
to=y.range[2], by=500))
coordinates(grd) <- ~ x+y

# projection
prj <- CRS("+init=epsg:2154")
proj4string(grd) <- prj
grd <- grd[carte, ]
gridded(grd) <- TRUE

# krigage
krigeage<-krige(prixm2~1,data,grd,vgm.fit,nmax = 20) #peut mettre du temps à tourner

# couleur
library(RColorBrewer)
colors <- brewer.pal(9, "Purples")

#carte krigée
plot(spplot(krigeage[krigeage,"var1.pred"],main="Test", cuts=7, col.regions = colors))

#####
##### Autocorrélation spatiale #####
#####

# Matrices de voisinage
coor <- coor[, which(colnames(coor)%in%c("x_au", "y_au"))]
distance <- rdist(coor,coor)
diag(distance)<-0
distance[distance >=100000] <- 0
dist <- ( distance /100000* distance /100000)
dist2 <- (distance * distance)
dist.w<- mat2listw(dist, style ="W") #matrice de distance inverse (100km)
carte.knn <- knearneigh(coor,k =300, RANN=TRUE)
kn5.nb <- knn2nb(carte.knn)
kn5.w<- nb2listw ( kn5.nb, style ="W") #matrice des 10 plus proches voisins

# Indice de Moran et test
moran.test(col, kn10.w, na.action=na.omit, randomisation=TRUE, zero.policy=FALSE)
# (Ho) rejetée : il y a autocorrélation spatiale des résidus

# Diagramme de Moran
moran.plot(x=col , listw = kn10.w , xlab ="prix de référence au m2" ,
  main="Diagramme de Moran par AU (10 plus proches voisins)",
  ylab ="densité de prix dans le voisinage", labels =as.character(label),
  zero.policy=FALSE)

```

```
#####
#####Econométrie spatiale #####
#####

# modèles
model.ols<-lm(log(prixm2) ~ log(densite) + densiteevol + climat, table)
model.sar <- lagsarlm(log(prixm2) ~ log(densite) + densiteevol+ climat, data=table,
  kn10.w, tol.solve=1.0e-30)
model.sem <- errorsarlm(log(prixm2) ~ log(densite) + densiteevol+ climat,data=table,
  kn10.w, tol.solve=1.0e-30)
model.sdm<- lagsarlm(log(prixm2) ~ log(densite) + densiteevol+ climat, data=table,
  kn10.w,type="mixed" , tol.solve=1.0e-30)
model.sac <- sacsarlm(log(prixm2) ~ log(densite) + densiteevol+ climat, data=table
, kn10.w)
model.slx <- lmSLX(log(prixm2) ~ log(densite) + densiteevol+ climat, data=table
, kn10.w)
model.sarar <- spreg(log(prixm2) ~ climat, data=table, kn10.w, listw2=dist.w.,
  endog= ~ log(densite) + densiteevol, instruments= ~ partm20ans + evol_partm20ans,
  lag.instr = TRUE, model = "sarar", na.action = na.omit,type=c("Triangular"))

# exemples de comparaisons de modèles
test <- LR.sarlm(model.sdm,model.sem)
Hausman.test(model.sem, tol=1.0e-30)
AIC(model.sdm) < AIC(model.sar)
lagrange <-lm.LMtests(model.ols,kn10.w,test=c("LMerr","RLMerr","LMlag","RLMlag","SARMA"),
  zero.policy=FALSE)
```


Bibliographie

- [1] BAILLARGEON S., *Le krigeage : revue de la théorie et application à l'interpolation spatiale de données de précipitations* (2005).
- [2] CAVAILHÈS J., *Le prix des attributs du logement*, Économie et statistique (2005).
- [3] COMBES P-P. , DURANTON G., GOBILLON L., *The costs of agglomeration : Land prices in French cities* (2012).
- [4] DONZEL A., FRANÇOIS D., GENIAUX G. ET NAPOLEONE C., *Les déterminants socio-économiques des marchés fonciers* (2008).
- [5] DRUKKER P., EGGER P., PRUCHA I., *On Two-step Estimation of a Spatial Auto regressive Model with Autoregressive Disturbances and Endogenous Regressors*, Econometric Review (2013).
- [6] FINGLETON B., LE GALLO J., *Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances : Finite sample properties* (2008).
- [7] FLOCH J-M., LE SAOUT R., *Économétrie spatiale : une introduction pratique*, JMS 2015.
- [8] FRIGGIT J., CHAPPERT A. ET AL., *Rapport sur l'organisation du service statistique dans le domaine du logement* (Juin 2014).
- [9] HENNEQUI M., *Spatialisation des donnees de modelisation par Krigeage*, Methodology (2010).
- [10] JOUBERT M-P., VIGNOLLES B., *Marché du terrain et des maisons neuves : des liens forts mais des logiques spécifiques*, Document de travail (2015).
- [11] KELEJIAN H., PRUCHA I., *A generalized moments estimator for the autoregressive parameter in a spatial model*, International Economic Review (1999).
- [12] KELEJIAN H., PRUCHA I., *HAC Estimation in a Spatial Framework*, Journal of Econometrics (2007).
- [13] KELEJIAN H., PRUCHA I., *Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroscedastic Disturbances*, Journal of Econometrics (2010).
- [14] LE GALLO J., *Économétrie spatiale*, Juin 2000.
- [15] LESAGE J., PACE R., *Introduction to Spatial Econometrics*, 2009.
- [16] PIRAS G., *sphet : Spatial Models with Heteroskedastic Innovations in R*.
- [17] SHEPPARD S., *Hedonic analysis of housing markets*, *Handbook of Regional and Urban Economics* (1999).
- [18] TRAVERS M., APPÉRE G., LARUE S., *Évaluation des aménités urbaines par la méthode des prix hédoniques : une application au cas de la ville d'Angers* (2013).
- [19] VÉDRINE L., *Introduction à l'économétrie spatiale*, Cours Ensae (Avril 2014).
- [20] VERMONT B., *Prix des logements : Quels facteurs expliquent leur disparité au sein et entre les aires urbaines ?*, Études et documents, CGDD (Mars 2015).