

Études & documents

n° 120

Mars

2015

Prix des logements : Quels facteurs expliquent leur disparité au sein et entre les aires urbaines ?

ÉCONOMIE ET ÉVALUATION



Collection « Études et documents » du Service de l'Économie, de l'Évaluation et de l'Intégration du Développement Durable (SEEIDD) du Commissariat Général au Développement Durable (CGDD)

Titre du document : Prix des logements : Quels facteurs expliquent leur disparité au sein et entre les aires urbaines françaises ?

Directeur de la publication : Xavier **Bonnet**

Auteur(s) : Bruno **Vermont**

Date de publication : Mars 2015

Ce document n'engage que ses auteurs et non les institutions auxquelles ils appartiennent.
L'objet de cette diffusion est de stimuler le débat et d'appeler des commentaires et des critiques.

Sommaire

Résumé.....	3
Introduction.....	5
1 Prix immobiliers et prix implicites des caractéristiques des biens.....	6
1.1 Quelques fondements théoriques.....	6
1.2 Les fonctions de prix hédoniques.....	6
1.3 Choix de la spécification de la fonction de prix hédoniques.....	8
2 Les données utilisées.....	10
2.1 Les données notariales.....	10
2.2 Géolocalisation des données notariales et distance au centre des aires urbaines.....	12
2.3 Les données complémentaires rassemblées.....	15
3 Résultats d'estimations sur les caractéristiques intrinsèques des logements.....	16
4 Prix immobiliers et localisation.....	19
4.1 Prix immobiliers et distance au centre des aires urbaines.....	19
4.2 Prix immobiliers et taille des aires urbaines.....	21
4.3 Estimations sur l'ensemble de la France métropolitaine : effet de la taille de l'aire urbaine et de la distance au centre de l'aire urbaine.....	23
4.4 Estimations par tranche d'aires urbaines avec variables de localisation.....	25
4.5 Bilan : prix hédoniques de la surface habitable et de l'éloignement du centre de l'aire urbaine.....	32
5 Analyse comparative des prix de l'immobilier par aire urbaine.....	35
5.1 Estimations des prix par aires urbaines.....	35
5.2 Quels déterminants socio-démographiques et économiques expliquent les différences de prix de l'immobilier entre les aires urbaines ?.....	37
6 Analyse de l'évolution des prix.....	45
6.1 Analyse descriptive des évolutions de prix.....	45
6.2 Analyse économétrique de l'évolution des prix.....	48
Conclusion.....	50
Annexes	51
1. Choix de la forme fonctionnelle des modèles hédoniques.....	51
2. Régressions en coupe complémentaires.....	53
3. Régressions complémentaires sur les évolutions de prix.....	54
Bibliographie.....	56

Résumé

Le prix des logements a augmenté de 60 % hors inflation entre 2000 et 2012 en France métropolitaine, avec une augmentation différenciée selon les localisations qui a renforcé les disparités entre zones urbaines.

Le marché du logement français est caractérisé par une forte hétérogénéité qui se décline à la fois dans la diversité des types de biens échangés et dans la variabilité spatiale des prix, autant entre grandes aires urbaines qu'au sein d'une même aire urbaine. L'existence de ces différentes dimensions affectant les prix de l'immobilier complexifie l'analyse des marchés de l'immobilier, notamment les comparaisons entre zones géographiques et l'étude de l'évolution temporelle des prix.

Dans ce contexte, cette étude analyse les disparités entre aires urbaines et leur évolution récente.

Le recours à la méthode des prix hédoniques pour différencier le prix de chaque composante du logement (caractéristiques propres mais aussi localisation) permet de comparer les niveaux de prix par aire urbaine toutes choses égales par ailleurs.

Les résultats de l'étude fournissent des valeurs chiffrées précisément et donnent ainsi des informations sur les facteurs qui déterminent la disparité des prix.

Principaux résultats

L'étude analyse tout d'abord les déterminants des prix des logements anciens liés aux caractéristiques des logements (surface, période de construction, éléments de confort). Le prix de vente des logements est par exemple plus élevé dans les logements ayant certains éléments de confort supplémentaires ou dans les logements de construction la plus récente. Il est également en moyenne plus élevé dans les maisons individuelles que dans les appartements.

Une part importante de la variabilité des prix est due à la localisation du logement au sein de l'aire urbaine. Dans la plupart des grandes aires urbaines, les prix sont décroissants depuis la ville-centre vers la périphérie.

Une part importante de la variabilité des prix est enfin due aux caractéristiques propres de chaque aire urbaine. La taille de l'aire urbaine a notamment un effet positif sur les niveaux de prix moyens, mais elle ne suffit pourtant pas à expliquer les différences de niveaux de prix observés.

Afin d'approfondir ces questions, nous réalisons une analyse comparative des niveaux de prix par aire urbaine toutes choses égales par ailleurs, c'est-à-dire pour un appartement de référence dont les caractéristiques sont contrôlées. Cette analyse quantitative montre que les prix sont d'autant plus élevés que l'aire urbaine est peuplée et qu'elle a connu une croissance démographique importante dans les années 2000. D'autres aires urbaines présentent des prix élevés du fait de leur forte attractivité touristique, malgré une dynamique démographique modérée. La richesse de l'aire urbaine, qui témoigne de la qualité des services offerts par la ville, est fortement corrélée avec les niveaux de prix. Les villes les moins chères sont quant à elles caractérisées par une part plus importante de jeunes, de ménages ayant emménagé depuis de nombreuses années et un taux de vacance élevé des logements.

L'analyse de l'évolution des prix par aire urbaine montre que les écarts entre zones urbaines s'accroissent sur la période 2000-2012. Les niveaux de prix ont progressé dans la quasi-totalité des aires urbaines mais à des rythmes variés. Ces différences d'évolution entre les aires urbaines s'expliquent en grande partie par des évolutions diverses de certains facteurs économiques, démographiques et par l'évolution du parc et de l'offre en logements. Les évolutions de prix les plus importantes se concentrent dans les villes où les prix étaient déjà élevés au début des années 2000 et dans celles qui sont dynamiques sur le plan démographique et économique. La croissance de l'offre en logements permet de contenir la hausse des prix dans certaines aires urbaines.

Introduction

Le marché du logement français est caractérisé par une forte hétérogénéité qui se décline à la fois dans la diversité des types de biens échangés et dans la variabilité spatiale des prix tant entre grandes zones urbaines qu'au sein d'une même zone urbaine. Les prix des logements observés dans les transactions immobilières reflètent non seulement les caractéristiques physiques des biens mais aussi celles liées à leur localisation et à leur proximité avec un ensemble d'aménités. La hausse soutenue des prix de l'immobilier dans les années 2000 a renforcé ces disparités entre zones urbaines [Juillard et Alexandre, 2012]. L'existence des différentes dimensions (spatiale, temporelle et qualité des biens) affectant les prix de l'immobilier complexifie l'analyse des marchés de l'immobilier, notamment les comparaisons entre zones géographiques et l'étude de l'évolution temporelle des prix. La variabilité spatiale des prix de l'immobilier est pourtant au centre des choix de localisation résidentielle des ménages et l'évolution récente des prix pose un certain nombre de questions notamment sur la capacité des ménages à accéder à la propriété et sur l'impact des politiques publiques portant sur le logement.

Dans une étude récente, Combes et al. (2010), montrent que les prix fonciers sont fortement liés aux caractéristiques démographiques de l'aire urbaine dans laquelle le terrain se situe, à la localisation du terrain au sein de l'aire urbaine et à ces caractéristiques physiques. Nous chercherons dans cette étude à déterminer dans quelle mesure les prix des logements reflètent ces trois composantes en insistant sur l'effet de la localisation.

Les données notariales sur les transactions immobilières sur la France entière constituent une source de données intéressante pour une telle analyse des prix immobiliers du fait de leur richesse sur le plan des caractéristiques des biens qu'elles décrivent, sur la géolocalisation précise des logements et sur le nombre de transactions qu'elles représentent sur l'ensemble du territoire français.

La méthodologie utilisée dans cette étude repose sur des techniques qui sont aujourd'hui d'un usage relativement courant dans la littérature économique et notamment la méthode des prix hédoniques. L'utilisation de cette méthode a l'avantage d'avoir des bases théoriques solidement ancrées et de permettre les comparaisons avec les études déjà réalisées. D'autre part, elle permet de s'affranchir des effets de qualité des biens présents dans les données brutes et de comparer les prix immobiliers entre zones géographiques toutes choses égales par ailleurs.

L'étude est organisée comme suit. La première partie revient sur les fondements de la théorie hédonique et sur les fonctions de prix hédoniques estimées. Les données utilisées sont ensuite présentées puis une première application de la méthode hédonique analyse l'impact des caractéristiques des biens sur les prix sur l'ensemble des aires urbaines françaises et par tranche d'aires urbaines (TAU). L'étude se poursuit ensuite par l'analyse de l'impact de la localisation des logements par rapport au centre des aires urbaines et compare cet effet entre aires urbaines de taille différente. La partie suivante compare les prix entre les aires urbaines pour un bien de caractéristiques données et cherche à déterminer quels facteurs expliquent la variabilité des prix observés. La dernière partie analyse l'évolution temporelle des prix du logement et cherche à expliquer les différences observées dans les variations de prix entre les aires urbaines sur la période 2000-2012.

1 Prix immobiliers et prix implicites des caractéristiques des biens

1.1 Quelques fondements théoriques

L'étude des prix des caractéristiques intrinsèques des biens immobiliers a reçu un traitement abondant dans la littérature. Ces analyses partent du constat que les prix des caractéristiques des logements (surface habitable, confort, qualité, localisation...) ne sont pas observables directement sur le marché mais se reflètent dans le prix du logement. Ces éléments font référence à la notion de «marchés implicites» qui désignent *la production, l'échange et la consommation de biens qui sont échangés sous forme de «paniers» de caractéristiques* [Sheppard, 1999]. Le logement est ainsi vu comme un bien composite constitué d'un ensemble d'éléments vendus simultanément. Les prix et les transactions observés sur le «marché explicite» de l'immobilier représentent le prix de ces paniers. On peut cependant considérer que ce marché est constitué d'un ensemble de marchés implicites pour chacun des attributs du logement.

Plusieurs approches de ces marchés implicites existent, mais la grande majorité des travaux de la littérature utilisent une approche hédonique basée sur le travail fondateur de Rosen (1974). L'hypothèse fondamentale de cette approche est que les logements sont des biens hétérogènes (donc pas caractérisés par un prix unique) constitués d'un agrégat de caractéristiques qui elles sont homogènes (ou ont au moins un prix implicite unique). Ces caractéristiques vendues en bloc sont ensuite transformées en utilité par les consommateurs (elles servent en fait d'inputs à la fonction d'utilité des consommateurs).

Les estimations de fonction de prix hédoniques sont utilisées pour deux catégories d'analyse :

- L'analyse de l'évolution (temporelle) des prix en tenant compte des changements de qualité des biens d'une période à l'autre (ex : indice des prix Notaires-INSEE des logements anciens).
- L'analyse de la demande des consommateurs pour les attributs des logements. L'estimation de la fonction de prix hédoniques ne donne directement aucune information sur le comportement des consommateurs mais les prix implicites calculés peuvent être utilisés pour estimer un système de demande où les attributs du logement (surface, accessibilité, aménités) sont alors considérés comme des biens (ex : Cavailhès (2005)).

1.2 Les fonctions de prix hédoniques

Rappelons le cadre théorique de l'estimation des fonctions de prix hédoniques avant de déduire ce que l'on peut espérer en retirer sur le comportement des consommateurs. On se place dans le cadre d'un ménage j de caractéristiques α qui maximise l'utilité $U(H, Z, \alpha)$ qu'il retire de la consommation d'un bien composite Z et d'un logement H composé d'un ensemble de caractéristiques $x_k \quad k \in [1, \dots, n]$ sous la contrainte budgétaire $P(H) + p_z Z = W$ où $P(H)$ est le prix total de H , p_z le prix de Z (numéraire) et W le revenu. Il est à noter qu'en général la fonction de prix hédoniques est non-linéaire, c'est-à-dire que le prix du logement n'est pas linéaire par rapport aux quantités de ces attributs x_k . Ceci a plusieurs implications sur la méthode d'estimation de la fonction de prix hédoniques et des paramètres de la demande pour les caractéristiques des biens.

$$\begin{aligned} & \underset{x_k, Z}{\text{Max}} U(H(x_k), Z, \alpha) \\ & \text{s.c } P(H) + Z \leq W \end{aligned}$$

En posant λ le multiplicateur de Lagrange associé à la contrainte budgétaire, les conditions du premier ordre de cette maximisation donnent :

$$\frac{\partial U}{\partial Z} = \lambda$$

$$\frac{\partial U}{\partial x_k} = \lambda \frac{\partial P(H)}{\partial x_k} \quad \forall k$$

En notant p_k le prix hédonique de la caractéristique x_k , on obtient que p_k égalise le taux marginal de substitution de x_k et Z .

$$\frac{\partial P(H)}{\partial x_k} = p_k = \frac{U_{x_k}}{U_Z}$$

À noter que si la fonction $P(H)$ est non linéaire par rapport à x_k , p_k sera fonction de x_k , ce qui a pour incidence que le prix implicite d'un ménage pour une caractéristique varie selon la quantité demandée. $P(H)$ est exogène pour le ménage et il ne peut pas influencer sur ce prix. Le ménage peut cependant choisir pour un même prix total du logement, différentes combinaisons des caractéristiques des biens et donc différents prix implicites pour ces caractéristiques.

On définit la fonction d'enchère (« *bid-rent function* ») du ménage comme le montant qu'un ménage est prêt à payer pour différentes combinaisons des caractéristiques du logement à utilité u et revenu W donnés. est définie implicitement par.

$$\theta(H, u, W, \alpha) = \underbrace{\text{Max}}_{(x_k, Z)} P(H(x_k))$$

$$\text{s.c } U(H(x_k), Z, \alpha) \geq u$$

$P(H)$ étant fixé par le marché, c'est le prix minimal que le ménage doit payer. L'utilité est donc maximisée lorsque :

$$\theta(H(x^*), u^*, W, \alpha) = P(H(x^*))$$

En dérivant cette équation par rapport à x_k et en reprenant les conditions du premier ordre du programme du ménage, on obtient :

$$\frac{\partial \theta}{\partial x_k}(x^*) = \frac{U_{x_k}}{U_Z} = p_k(x^*) \quad \forall k$$

Ces équations montrent que sous les hypothèses ci-dessus, l'estimation des prix hédoniques pour chacune des caractéristiques du logement donne des informations sur le consentement à payer des ménages pour les attributs du logement au voisinage du couple prix total du logement-quantité demandée de l'attribut observé (c'est à dire à l'équilibre de marché si on suppose que les données observées le sont). Chaque observation du prix des logements permet donc d'obtenir des informations sur la fonction d'enchère des ménages au voisinage de ce prix. On peut donc espérer avec suffisamment d'observations inférer sur le comportement des ménages en retrouvant l'ensemble de la fonction d'enchère à partir des paramètres estimés d'une fonction de prix hédoniques. On se retrouve dans le cas de l'analyse d'un marché classique où on aurait suffisamment d'observations du couple prix de l'attribut/demande de l'attribut. En pratique, Sheppard (1999) et Follain et Jimenez (1985) montrent que l'application d'une telle méthode pose un certain nombre de problèmes appliqués. En particulier, de telles estimations nécessitent d'avoir suffisamment d'information sur les ménages (revenus et taille notamment). Dans cette étude, nous nous limitons à l'estimation de la seule fonction de prix hédoniques étant donné le peu d'informations dont nous disposons sur les

ménages. Nous utilisons ensuite cette fonction de prix pour comparer les niveaux de prix des différentes aires urbaines françaises et leur évolution. L'analyse du comportement des ménages pourra faire l'objet d'une étude future.

1.3 Choix de la spécification de la fonction de prix hédoniques

Le choix de la spécification de la fonction de prix hédoniques a reçu beaucoup d'attention de la part des économètres. L'estimation de telles fonctions pose un certain nombre de difficultés qui ont donné lieu à une large variété d'approches :

- choix de la forme fonctionnelle (linéaire, log, transformation de box-cox),
- choix des variables explicatives à inclure et multi-colinéarité,
- manque de variabilité dans les observations (préférences homogènes des ménages et limites de la technologie de construction de logement),
- endogénéité de certaines variables explicatives,
- auto-corrélation spatiale (non-indépendance des observations spatialement proches les unes avec les autres et donc biais potentiel dans l'estimation),
- hétéroscédasticité potentielle.

Sheppard (1999) passe en revue ces problèmes et met en avant le fait qu'il n'y a pas de « meilleure spécification » mais que la bonne spécification du modèle dépend à la fois des données disponibles et de l'objectif de l'analyse. Ainsi, si l'objectif est simplement d'estimer et de prédire le prix total du logement, une spécification qui a un fort pouvoir de prédiction est appropriée (ex : spécification flexible type box-cox). En revanche, si on souhaite estimer les prix implicites des attributs du logement, une spécification qui a un pouvoir de prédiction plus faible mais dont les paramètres estimés sont plus stables, peut être mieux adaptée.

Il insiste ensuite sur le fait que le choix des variables à inclure est tout aussi important que le choix de la spécification (notamment l'omission de variables de localisation, et de variables prenant en compte les aménités locales et/ou environnementales peut entraîner des biais importants). Il insiste enfin sur le fait que la prise en compte de la structure spatiale des observations peut conduire à des résultats plus fiables.

Une des hypothèses de base de la théorie est que le prix $P(H)$ est une fonction non-linéaire des quantités des caractéristiques. On évitera donc a priori la spécification linéaire-linéaire (lin-lin). Les spécifications log-linéaire et semi-log sont très simples à utiliser mais potentiellement biaisées [Cavailhès, 2005]. Enfin, la transformation box-cox, soit de la variable dépendante (le prix), soit de certaines variables explicatives, soit des deux est d'un usage relativement commun aujourd'hui dans la littérature.

Pour toutes ces spécifications, on peut résumer le modèle estimé par :

$$\tilde{P}_i = \beta_0 + \sum_k \beta_k \tilde{x}_{ik} + \sum_l \beta_l x_{il} + \epsilon_i$$

où P_i est le prix total de i , les β_k et β_l sont les paramètres estimés du modèle, le tilde désigne les variables transformées (log, quadratique, box-cox), x_{ik} sont les attributs du logement i que l'on transforme (typiquement les variables continues comme la surface habitable, la distance à un centre urbain...) et x_{il} les autres attributs non transformés (typiquement les variables discrétisées comme la présence ou non d'un élément de confort, l'état du logement, la période de construction...).

Ainsi, si g est la fonction qui transforme P_i en \hat{P}_i et g_k celle transformant x_{ik} en \tilde{x}_{ik} , on peut exprimer le prix hédonique estimé des différents attributs transformés par :

$$\begin{aligned}\frac{\partial \tilde{P}_i}{\partial x_k} &= \hat{\beta}_k \frac{\partial \tilde{x}_{ik}}{\partial x_{ik}} && \forall k \\ \frac{\partial \hat{P}_i}{\partial x_k} \frac{\partial g(\hat{P}_i)}{\partial P_i} &= \hat{\beta}_k \frac{\partial g_k(x_{ik})}{\partial x_{ik}} && \forall k \\ \hat{p}_{ik} &= \hat{\beta}_k \frac{\frac{\partial g_k(x_{ik})}{\partial x_{ik}}}{\frac{\partial g(\hat{P}_i)}{\partial P_i}} && \forall k\end{aligned}$$

où $\hat{\beta}_k$ représente la valeur estimée du paramètre et \hat{P}_i la valeur prédite du prix total du logement. De même, pour les attributs non transformés, on a :

$$\hat{p}_{il} = \hat{\beta}_l \frac{1}{\frac{\partial g(\hat{P}_i)}{\partial P_i}} \quad \forall l$$

Le tableau 1 résume la valeur des prix hédoniques estimés selon différentes spécifications de la fonction de prix hédoniques (différentes transformations g et g_k des variables).

Tableau 1 : Prix hédoniques estimés selon la spécification de la fonction de prix hédoniques

spécification	g	g_k	\hat{p}_{ik}	\hat{p}_{il}
semi-log	$\ln(P_i)$	x_{ik}	$\hat{\beta}_k \hat{P}_i$	$\hat{\beta}_l \hat{P}_i$
log-linéaire	$\ln(P_i)$	$\ln(x_{ik})$	$\hat{\beta}_k \frac{\hat{P}_i}{x_{ik}}$	$\hat{\beta}_l \hat{P}_i$
box-cox-linéaire	$\frac{P_i^\lambda - 1}{\lambda}$	x_{ik}	$\frac{\hat{\beta}_k}{\hat{P}_i^{\lambda-1}}$	$\frac{\hat{\beta}_l}{\hat{P}_i^{\lambda-1}}$
box-cox-box-cox	$\frac{P_i^\lambda - 1}{\lambda}$	$\frac{x_{ik}^{y_k} - 1}{y_k}$	$\hat{\beta}_k \frac{x_{ik}^{y_k - 1}}{\hat{P}_i^{\lambda-1}}$	$\frac{\hat{\beta}_l}{\hat{P}_i^{\lambda-1}}$

Il est important de noter que les prix implicites calculés ont des caractéristiques différentes pour les variables transformées et non-transformées. En effet, pour les variables transformées, le prix implicite fait intervenir la quantité achetée de la caractéristique $(\hat{\beta}_k \frac{\hat{P}_i}{x_{ik}}, \frac{x_{ik}^{y_k} - 1}{y_k})$. Le prix implicite varie donc en même temps que la quantité achetée.

2 Les données utilisées

2.1 Les données notariales

Les données notariales (PERVAL pour la province et BIEN pour l'Île-de-France) rassemblent un certain nombre d'informations sur les transactions de maisons et d'appartements enregistrées par les notaires principalement dans l'ancien. Pour chaque transaction, le prix du bien, sa date de mutation, sa localisation, les caractéristiques de la mutation, les caractéristiques du bien ainsi que des éléments sur les caractéristiques de l'acheteur et du vendeur sont renseignés. Si les données communes à tous les types de biens sont plutôt bien renseignées, à savoir le prix, la date et les caractéristiques de l'acheteur et du vendeur, il n'en est pas de même pour les caractéristiques des biens qui présentent un taux de renseignement parfois très bas pour certaines variables.

L'information concernant les caractéristiques des acheteurs est relativement réduite, car elle se limite à l'âge, la profession, la nationalité, le sexe et le statut matrimonial. Les bases notariales ne contiennent pas d'informations sur la taille et le revenu du ménage auquel l'acheteur appartient. On peut difficilement faire le lien avec d'autres informations sur les ménages à part en associant à l'acheteur le revenu local médian de la commune (au mieux par catégorie socio-professionnelle). Ceci rend difficile de mener une analyse sur le comportement des ménages à partir de ces bases étant donné que les choix résidentiels des ménages sont fortement liés à leur taille et leur revenu. C'est pourquoi cette étude se limite à une analyse des caractéristiques des marchés immobiliers et des facteurs qui expliquent la variabilité des prix observés.

Enfin, les bases PERVAL et BIEN sont renseignées sur la base du volontariat et ne représentent pas l'ensemble des transactions annuelles. Le taux de couverture des transactions varie selon la zone (de moins de 30 % à plus de 80 % selon les départements en 2003 selon l'INSEE). En soi, un faible taux de couverture n'est pas un problème pour les estimations si le nombre d'observations est suffisant et que les biens enregistrés sont représentatifs de l'ensemble des transactions. Néanmoins, il peut exister un biais si les transactions concernant un certain type de bien, de client ou de marché ne sont jamais renseignées.

Les données brutes des bases notariales PERVAL ont été traitées en partie sur la base du site métier de l'entrepôt de données SIDAUH de la DGALN¹. Certaines données ont été retirées de l'échantillon notamment lorsque la date de mutation, la surface habitable ou le prix du bien n'étaient pas renseignés. D'autres filtres ont été imposés de manière à exclure certains biens atypiques ou non vendus aux conditions normales du marché :

- Nature de la mutation : vente de gré à gré ou non renseignée
- Acheteurs et vendeurs : Acquisitions effectuées par des marchands de biens exclues, acheteurs et vendeurs du secteur public et social exclus
- Type de propriété : bien en pleine propriété ou non renseignée
- Biens à usage d'habitation uniquement
- Occupation : bien libre de toute occupation ou non renseignée
- Part sociale : bien vendu sans part sociale ou non renseignée
- Appartements et maisons atypiques exclus (moulins, anciennes gares, loges de gardiens, greniers...)
- Viager : vente en viager exclue ou non renseignée
- Surface habitable : entre 10 et 200 m² pour les appartements, entre 20 et 400 m² pour les maisons
- Prix hors taxes supérieur à 1 500 euros.

1 <http://sidauh.metier.i2/documentation-r3.html>

Des traitements du même type ont été appliqués aux bases BIEN pour les années de 2000 à 2010. Le tableau ci-dessous reporte pour les bases PERVAL et BIEN le nombre d'observations dans les données brutes, le nombre d'observations supprimées, car le prix ou la surface habitable n'étaient pas renseignés et le nombre d'observations supprimées par les autres filtres.

Tableau 2 : Traitement des données PERVAL

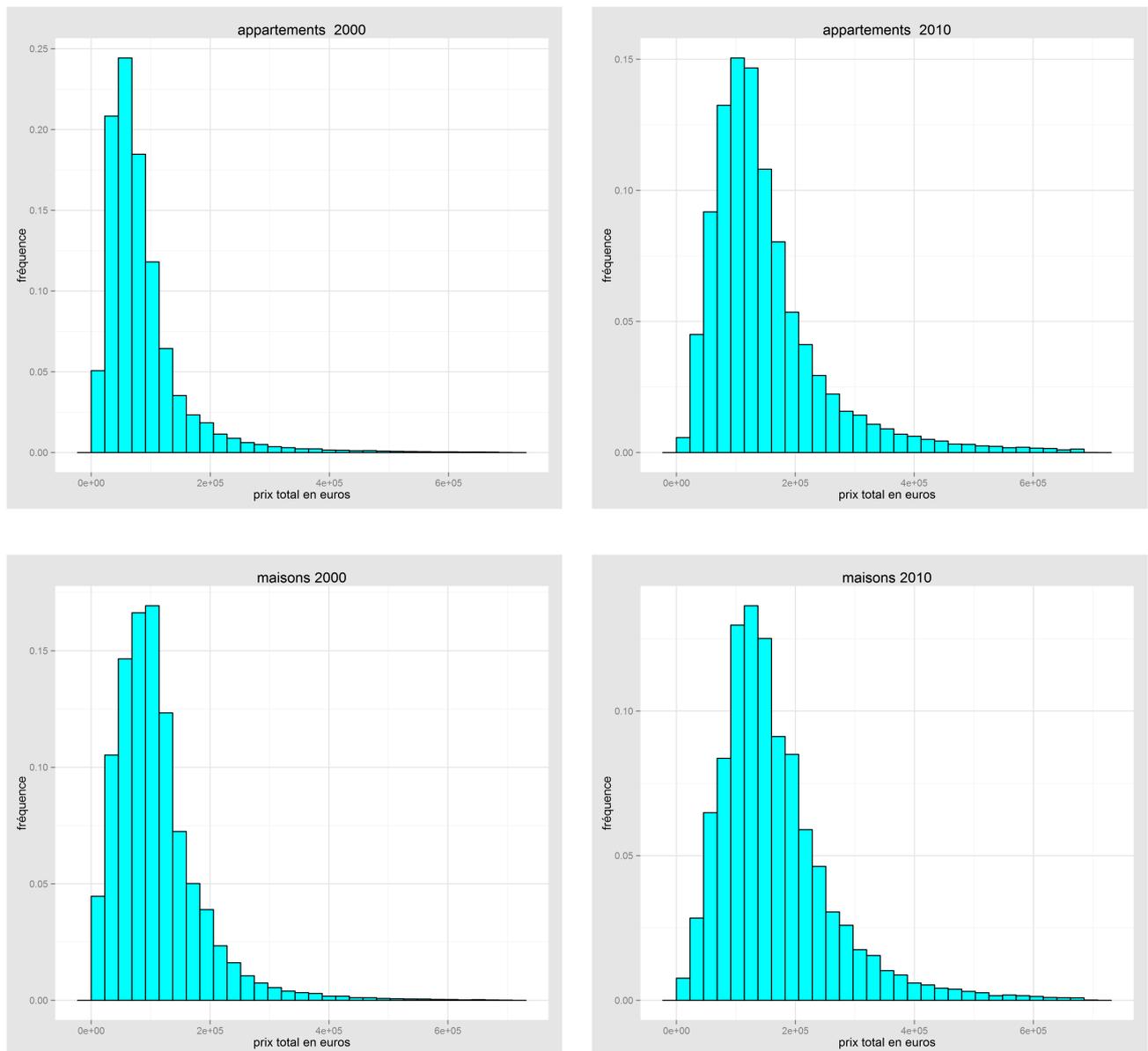
Traitement	Appartements	Maisons
Nombre d'observations avant traitement	1 307 395	1 594 657
Surface, prix non renseignés	170 142	690 738
Autres filtres	115 613	92 594

Les deux tableaux ci-dessous présentent quelques statistiques descriptives sur les échantillons retenus pour les bases PERVAL et BIEN après traitement. Les graphiques suivants montrent la distribution du prix hors taxes dans les deux bases et par type de bien pour 2000 et 2012. En 2010, la médiane du prix au m² sur l'ensemble de la France sur les biens conservés dans la base de données est de 2005 euros alors qu'elle est environ deux fois plus élevée sur l'aire urbaine de Paris. Sur les deux bases de données, les appartements ont un prix médian au m² plus élevés que les maisons. Le nombre de transactions par type de biens sur la base PERVAL est relativement équilibré entre appartements et maisons alors que les maisons sont largement minoritaires dans la base BIEN. De 2000 à 2010, les prix médians ont fortement augmenté (+75 % environ) et ont pratiquement doublé à Paris. Les distributions des prix par type de bien révèlent une grande hétérogénéité sur l'ensemble de l'échantillon. On peut également remarquer un certain étalement de la distribution des prix vers la droite entre 2000 et 2010 témoignant d'une hausse des écarts de prix entre biens. Il convient cependant d'être prudent à la lecture de ces statistiques descriptives, car les données notariales ne sont pas exhaustives et ne représentent qu'un échantillon des biens échangés chaque année. D'autre part, il peut exister des biais liés à la structure et à la qualité des biens échangés.

Tableau 3 : Statistiques descriptives des bases de données après traitement, prix en euros 2000

Année	2000	2002	2004	2006	2008	2010	2012
N. obs total	338 050	327 301	375 257	341 282	339 984	390 158	227 105
N. obs appartements	201 020	202 249	236 296	213 394	217 240	251 868	124 360
N. obs maisons	137 030	125 052	138 961	127 888	122 744	138 290	102 745
prix m ² médian total	1 143	1 307	1 612	1 882	1 990	2 005	1 651
prix m ² médian appartements	1 294	1 470	1 850	2 166	2 330	2 379	1 990
prix m ² médian maisons	953	1 052	1 234	1 442	1 469	1 436	1 330
prix m ² moyen total	1 308	1 495	1 869	2 144	2 359	2 386	1 847
prix m ² moyen appartements	1 501	1 711	2 166	2 476	2 754	2 813	2 162
prix m ² moyen maisons	1 025	1 146	1 364	1 591	1 658	1 608	1 465
N. obs médian par aire urbaine	33	29	36	39	36	36	34

Source : PERVAL et BIEN

Figure 1 : Distributions des prix hors taxes des appartements et des maisons en 2000 et 2012 (euros 2000)

Source : PERVAL et BIEN

2.2 Géolocalisation des données notariales et distance au centre des aires urbaines

Les bases notariales contiennent un certain nombre d'éléments permettant de localiser de manière plus ou moins précise les biens échangés. Les observations PERVAL et BIEN sont tout d'abord identifiées par le numéro de commune INSEE dans laquelle le bien est situé. Dans les données PERVAL récentes, on dispose de coordonnées pour un certain nombre d'observation. Ces coordonnées sont renseignées soit au numéro de voie soit à la voie, et ont été obtenues par une méthode de géolocalisation à l'adresse. Enfin, les données sont identifiées par leur code de section cadastrale et leur numéro de parcelle. Il est donc possible d'associer des coordonnées géographiques aux logements en retrouvant les coordonnées des parcelles cadastrales correspondantes.

Les deux méthodes de localisation des biens, par leur adresse ou par le cadastre, présentent chacune des avantages et des inconvénients. La géolocalisation à l'adresse a l'avantage d'être moins sensible aux changements temporels, les adresses évoluant moins régulièrement que le cadastre. Les sections cadastrales peuvent en effet disparaître ou être scindées ou regroupées au cours du temps. La géolocalisation à l'adresse nécessite néanmoins des opérations relativement fastidieuses pour formater les adresses présentes dans les données brutes et effectuer des requêtes sur une application de géolocalisation pour récupérer les coordonnées géographiques correspondantes. Comme le montre le tableau 4, dans les données brutes, seules 40 % des observations sont localisées au numéro de voie. D'autre part, les logements de la base BIEN ne sont pas géolocalisés. Nous n'utilisons donc pas ces informations incomplètes.

Dans le cadre de cette étude et dans la perspective de futures études nécessitant une localisation précise des logements, nous avons donc choisi de géolocaliser à nouveau les logements en appariant les données notariales avec la BD PARCELLAIRE de l'IGN à l'aide des codes cadastraux présents dans les données brutes. La BD PARCELLAIRE recense l'ensemble des parcelles cadastrales, leurs identifiants cadastraux (numéro de parcelle et section cadastrale) et leurs coordonnées géographiques pour une année. Les codes de section cadastrale des données notariales ont été traités de manière à retrouver des codes dont le format correspond à celui de la BD parcellaire. Nous apparions ensuite les bases notariales avec la BD PARCELLAIRE. La clé d'appariement est composée par les variables suivantes :

- le numéro INSEE de la commune (5 caractères numériques)
- le préfixe de la section cadastrale (3 caractères numériques) :
 - le plus souvent 000
 - dans les cas de fusion de communes (nécessairement au sein d'un même département), le préfixe est le code commune (3 caractères numériques) de la commune fusionnée.
- le suffixe de la section cadastrale (2 caractères alphanumériques)
- le numéro de la parcelle cadastrale (4 caractères compris entre '0000' et '9999')

Sur l'ensemble des données notariales, moins de 1 % des codes de section sont irrécupérables. En appariant avec la BD PARCELLAIRE, on parvient à associer une parcelle à environ 94 % des observations. Le tableau 4 montre la part de l'échantillon géolocalisé à la parcelle par année. On remarque que plus on remonte dans le temps, plus cette part diminue. Ceci est lié aux problèmes inhérents à cette méthode de géolocalisation à l'aide d'un parcellaire récent (2012) car certaines parcelles ont pu disparaître au cours du temps. Néanmoins, le taux de géolocalisation est très élevé et bien supérieur au taux de géolocalisation des données brutes. On attribue ensuite aux logements non géolocalisés par cette méthode, les coordonnées du centroïde de la commune dans laquelle ils sont situés.

Grâce à cette géolocalisation, il est donc possible d'attribuer une localisation fine aux logements, de calculer des distances précises au centre de l'aire urbaine dans laquelle ils sont situés, mais aussi à d'autres aménités (ex : infrastructures de transport, services, nature, littoral...). Enfin, on peut envisager d'utiliser des méthodes d'économétrie spatiale. Nous nous concentrons dans cette étude uniquement sur le paramètre de distance au centre de l'aire urbaine qui reste un bon indicateur de l'éloignement entre le logement et le centre même lorsque celui-ci est localisé à la commune.

Tableau 4 : Part de l'échantillon géolocalisé par niveau de géolocalisation

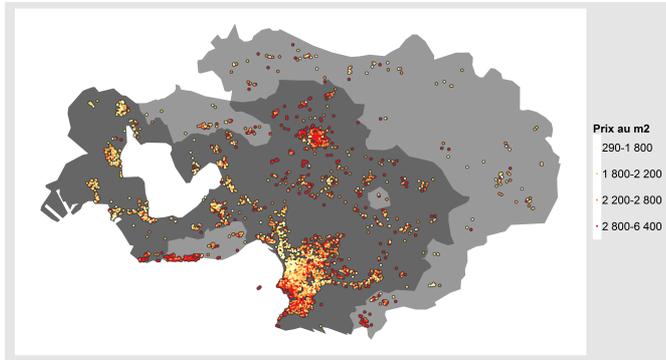
Niveau de géolocalisation	Année	Nombre d'observations	Part de l'échantillon	Niveau de géolocalisation	Année	Nombre d'observations	Part de l'échantillon
commune	2000	118 655	0.35	voie	2000	66 442	0.20
commune	2002	119 150	0.36	voie	2002	64 467	0.20
commune	2004	146 129	0.39	voie	2004	73 637	0.20
commune	2006	101 900	0.30	voie	2006	80 243	0.24
commune	2008	125 699	0.37	voie	2008	73 138	0.22
commune	2010	173 581	0.44	voie	2010	71 476	0.18
commune	2012	33 504	0.15	voie	2012	32 180	0.14
numéro de voie	2000	152 953	0.45	parcelle	2000	304 274	0.90
numéro de voie	2002	143 684	0.44	parcelle	2002	305 920	0.93
numéro de voie	2004	155 491	0.41	parcelle	2004	358 779	0.96
numéro de voie	2006	159 139	0.47	parcelle	2006	331 245	0.97
numéro de voie	2008	141 147	0.42	parcelle	2008	333 511	0.98
numéro de voie	2010	145 101	0.37	parcelle	2010	378 045	0.97
numéro de voie	2012	161 421	0.71	parcelle	2012	213 873	0.94

Pour calculer une distance au centre de l'aire urbaine, il a fallu définir un pôle de l'aire urbaine. Lorsque l'aire urbaine avait une commune centre explicite (ex : aire urbaine de Lyon), le centroïde de la commune centre a été choisi comme centre de l'aire urbaine. Lorsque la commune centre n'était pas définie explicitement, le centroïde du polygone représentant l'ensemble de l'aire urbaine a été choisi. On calcule ensuite les distances à vol d'oiseau entre les logements de la base de données et le centroïde de l'aire urbaine à laquelle ils appartiennent. Pour les biens situés dans des communes multipolarisées, on calcule la distance au centre de l'AU la plus proche du logement. Pour les biens non géolocalisés à la parcelle, on calcule la distance entre le centroïde de la commune à laquelle le bien appartient et le centre de l'AU correspondante. Pour Paris, Lyon et Marseille, on utilise la distance entre le centroïde de la ville-centre (ensemble des arrondissements) et le logement si celui-ci est géolocalisé, ou la distance entre le centroïde de la ville-centre et le centroïde de l'arrondissement où le logement est situé dans le cas contraire. Enfin, une distance forfaitaire a été appliquée aux logements situés dans la ville-centre et localisés à la commune ($d = \frac{2}{3} \sqrt{\frac{S_{commune}}{\pi}}$ où $S_{commune}$ est la superficie de la ville-centre et d la distance au centre attribuée à ces logements).

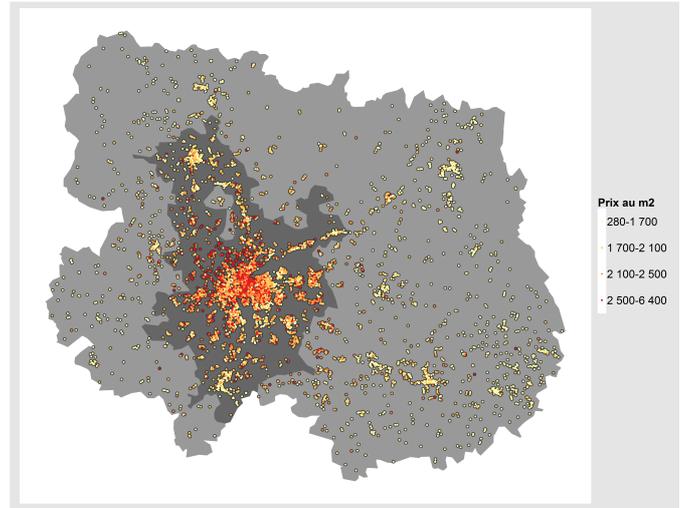
Les quatre cartes suivantes représentent le prix au m² en 2010 des logements géolocalisés sur les aires urbaines de Marseille, Lyon, Bordeaux et Toulouse.

Figure 2 : Prix au m² des logements en 2010 sur quatre aires urbaines

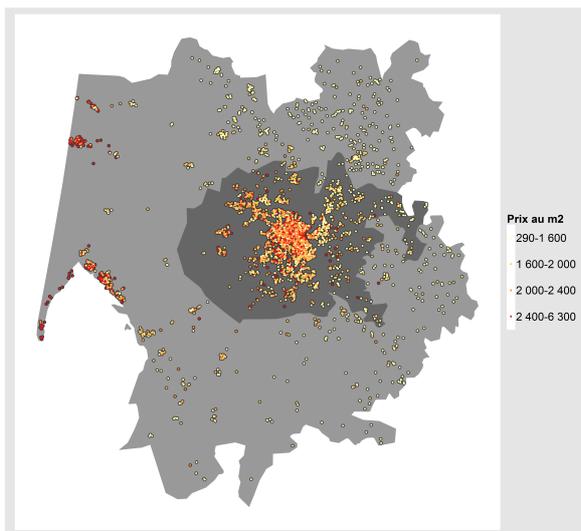
Aire urbaine de Marseille-Aix-en-Provence



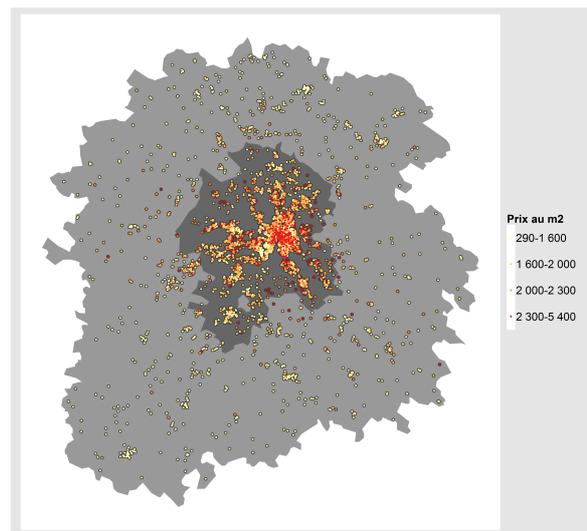
Aire urbaine de Lyon



Aire urbaine de Bordeaux



Aire urbaine de Toulouse



2.3 Les données complémentaires rassemblées

Un ensemble de données complémentaires a été rassemblé à l'échelle communale. Elles sont principalement issues de l'INSEE mais certaines proviennent d'autres sources comme les infocentres SIDAUH et GéoKit3 du MEDDE. Certaines de ces données ont été agrégées à partir des données communales de manière à constituer une base de données à l'échelle des aires urbaines.

- À l'échelle communale :
 - Démographie : population, densité de population, structure de la population par CSP et tranche d'âge au recensement de 1999 et aux recensements annuels de 2006 à 2010. Part de la population scolarisée et part de la population sortie du système scolaire par type de diplôme obtenu.
 - Emploi : population active, emploi au lieu de travail, taux de chômage pour 1999 et 2010 (de 2000 à 2012 à l'échelle des zones d'emplois).
 - Parc de logements : nombre de logements, nombre de logements par type, statut d'occupation des logements aux années de recensement.

- Tourisme : part de résidences secondaires, nombre d'hôtels et de chambres d'hôtel, nombre de campings et d'emplacements de campings, recensement des communes littorales d'une mer ou d'un lac.
- Ménages : Revenus fiscaux des ménages moyens par unité de consommation de 2001 à 2011, nombre de ménages et ancienneté d'emménagement aux années de recensement.
- Typologie des communes au sein des aires urbaines (commune appartenant à un pôle urbain, à la couronne du pôle, à l'espace périurbain...)
- Ces données communales ou par zone d'emploi ont été utilisées pour créer une base de données à l'échelle des aires urbaines :
 - Population, densité de population et structure de la population pour les années de recensement, superficie de l'AU
 - Emplois et taux de chômage (moyenne pondérée par la population active des communes en 1999)
 - Revenus par UC moyens sur l'aire urbaine, nombre de ménages et ancienneté d'emménagement
 - Hôtels, campings, part de résidence secondaires
 - Parc de logement et son évolution
 - Typologie des aires urbaines (tranches d'aire urbaine, grande aire urbaine / moyen pôle / petit pôle, pôle urbain / périurbain).

3 Résultats d'estimations sur les caractéristiques intrinsèques des logements

Les estimations suivantes ont été réalisées sur les bases PERVAL et BIEN, en empilant les données des deux bases pour les maisons et les appartements. L'avantage d'estimer le modèle sur l'ensemble des biens est de pouvoir avoir une représentation globale du marché du logement à l'échelle de l'aire urbaine. En effet, en omettant les maisons individuelles, on aurait clairement une sur-représentation des communes proches du centre rendant difficile l'inclusion de variables de localisation comme la distance au centre de l'AU. L'inconvénient est de ne pas pouvoir inclure certaines caractéristiques spécifiques à un type de bien (ex : surface du terrain pour les maisons, présence d'un ascenseur pour les appartements). À titre de comparaison, deux estimations par type de bien sont également présentées dans cette partie.

Les estimations ont été conduites sur l'ensemble des aires urbaines de France métropolitaine (PERVAL et BIEN), par tranche d'aires urbaines en province (PERVAL) et sur l'aire urbaine de Paris (BIEN). La forme fonctionnelle retenue est la forme logarithmique pour le prix total du bien et la surface, linéaire pour les autres caractéristiques discrétisées des logements. Le tableau 25 en annexes présente des estimations réalisées avec différentes formes fonctionnelles et justifie ce choix. Les tranches d'aires urbaines de l'INSEE ont été regroupées pour définir 5 classes : les AU de moins de 50 000 habitants, de 50 à 200 000 habitants, de 200 000 à 500 000 habitants, de plus de 500 000 habitants, ainsi qu'un dernier ensemble regroupant les communes multipolarisées et les communes rurales.

Le modèle estimé ici est donc :

$$\log P_{it} = \beta_0 + \beta_t + \beta_s \log s_{it} + \sum_l \beta_l x_{itl} + \epsilon_{it}$$

où P_{it} est le prix total hors taxe du logement en euros 2000, s_{it} est la surface habitable en m^2 , les β_t sont des indicatrices temporelles pour chaque année de l'échantillon et les x_{itl} sont un ensemble de variables polytomiques qui représentent certaines caractéristiques du logement. Ces estimations permettent de tester quelles caractéristiques des biens ont un effet significatif sur les prix et serviront de référence lorsque d'autres variables comme des variables de localisation seront introduites. De cette manière, il est possible d'analyser l'effet de la localisation sur le prix des biens à caractéristiques physiques données des logements.

Tableau 5 : Paramètres estimés, effets des caractéristiques des biens

Variable	France	France appart	France maison	Hors AU	AU inf50k	AU 50_200k	AU 200_500k	AU sup500k	AU Paris
Constante	8.3***	8.1***	8.3***	8.5***	8.7***	8.2***	7.8***	7.9***	7.9***
log (surface habitable)	0.73***	0.75***	0.62***	0.66***	0.6***	0.69***	0.83***	0.79***	0.92***
nombre de sdb manquant	-0.16***	-0.044***	-0.35***	-0.42***	-0.29***	-0.21***	-0.17***	-0.14***	-0.062***
nombre sdb >1	0.33***	0.36***	0.2***	0.28***	0.31***	0.22***	0.19***	0.24***	0.18***
présence d'un parking	0.083***	0.084***	0.12***	0.13***	0.15***	0.14***	0.11***	0.12***	-0.1***
période de construction manquante	-0.086***	-0.063***	-0.1***	-0.27***	-0.2***	-0.17***	-0.17***	-0.013***	-0.065***
construction avant 1913	-0.1***	0.33***	-0.52***	-0.51***	-0.56***	-0.4***	-0.3***	-0.26***	0.2***
construction 1914-1948	-0.12***	0.094***	-0.37***	-0.5***	-0.49***	-0.37***	-0.26***	-0.24***	0.041***
construction 1948-1969	-0.12***	-0.09***	-0.22***	-0.32***	-0.33***	-0.31***	-0.23***	-0.15***	-0.081***
construction 1970-1980	-0.09***	-0.12***	-0.12***	-0.17***	-0.19***	-0.21***	-0.17***	-0.11***	-0.13***
construction 1981-1991	0.0037**	-0.0061***	-0.035***	-0.06***	-0.054***	-0.073***	-0.026***	0.0089***	-0.079***
construction neuve	0.0014	0.038***	-0.029***	0.019***	-0.01**	0.022***	0.073***	-0.0066***	0.11***
maison	-0.35***			-0.22***	-0.089***	-0.04***	-0.097***	-0.093***	-0.25***
absence d'ascenseur		-0.073***							
2e étage		0.032***							
3e étage		0.073***							
4e étage ou plus		0.16***							
rez-de-chaussée		-0.013***							
log(surface du terrain)			0.043***						
présence d'une piscine			0.34***						
maison de plus d'1 étage			0.29***						
maison de plain-pied			-0.044***						
année 2002	0.12***	0.13***	0.1***	0.1***	0.1***	0.099***	0.082***	0.14***	0.11***
année 2004	0.32***	0.36***	0.27***	0.28***	0.28***	0.25***	0.27***	0.37***	0.34***
année 2006	0.48***	0.52***	0.44***	0.51***	0.47***	0.45***	0.47***	0.58***	0.57***
année 2008	0.55***	0.6***	0.46***	0.52***	0.48***	0.47***	0.46***	0.6***	0.63***
année 2010	0.55***	0.63***	0.41***	0.47***	0.46***	0.43***	0.46***	0.58***	0.64***
année 2012				0.45***	0.43***	0.42***	0.45***	0.59***	
R2 ajusté	0.45	0.50	0.52	0.51	0.49	0.53	0.64	0.62	0.62
Nombre d'observations	2 107 344	1 319 303	776 621	239 653	228 746	295 996	347 165	685 407	536 757

Les résultats pour l'ensemble de la France (colonne 1) montrent que la plupart des caractéristiques introduites ont un effet significatif sur le prix total du logement. Les caractéristiques des biens expliquent en moyenne 45 % de la variabilité des prix des logements. L'élasticité estimée du prix total des logements par rapport à leur surface est de 0,73 (élasticité du prix au m² des logements par rapport à leur surface de -0,27). L'absence de salle de bain a bien un effet négatif significatif sur le prix du bien alors que la présence de plus d'une salle de bain qui est un élément de confort supplémentaire augmente significativement le prix du logement. La présence d'un parking ou d'un garage a un effet positif sur le prix. Les logements dont la période de construction est la plus ancienne ont globalement un prix significativement plus faible que les autres (La référence étant les logements construits après 1992 non neufs). Ces résultats sont relativement cohérents avec différentes études utilisant des méthodes de prix hédoniques sur la France (ex: [Fauvet, 2008], [Cavailhès, 2005]). Le résultat sur le prix plus faible des maisons est cependant contraire à celui reporté dans ces deux études. Ce paramètre peut cependant être biaisé du fait que la localisation n'est pas contrôlée pour ces premières estimations.

Les deux colonnes suivantes montrent les résultats d'estimations sur la France entière mais par type de bien. Des variables spécifiques aux maisons et aux appartements sont introduites. Les résultats sont similaires pour les variables « présence d'une salle de bain » ou « d'un parking ». Le paramètre associé à la surface habitable estimé est plus faible pour les maisons (0,61) que pour les appartements (0,75) soulignant le fait que le prix des appartements augmente plus sensiblement avec la surface habitable du bien que celui des maisons. La surface du terrain a un effet positif et significatif sur le prix des maisons. Pour les deux types de biens, la présence d'un élément de confort en plus (ascenseur, piscine) a un effet positif et significatif sur le prix du bien. Au-delà d'un étage, le prix des maisons est significativement plus élevé tandis que les maisons de plain-pied ont un prix significativement plus faible. Dans le cas des appartements, le prix des appartements situés à un étage supérieur au premier est significativement plus élevé alors que les appartements situés au rez-de-chaussée sont moins valorisés. La variable « année de construction » a un signe ambigu et surprenant pour les appartements. On peut supposer que les appartements parisiens, qui représentent une part importante de l'échantillon dans ces estimations, tirent vers le haut le paramètre associé aux logements dont l'époque de construction est ancienne.

Les colonnes suivantes présentent les résultats d'estimations par tranches d'aires urbaines et sur l'aire urbaine de Paris. Ils vont globalement dans le même sens que ceux sur l'ensemble de la France. Les signes et la significativité des paramètres estimés sont semblables bien que leurs effets varient en magnitude selon la tranche d'aire urbaine (TAU) considérée. On note notamment la variabilité de l'élasticité du prix total du logement par rapport à sa surface qui semble plus importante dans les aires urbaines les plus peuplées (0,79 pour les aires urbaines de plus de 500 000 habitants et 0,92 pour l'aire urbaine de Paris, 0,6 pour les aires urbaines de moins de 50 000 habitants). Les caractéristiques des biens expliquent entre 49 % et 64 % de la variabilité des prix selon la TAU. Les résultats d'estimations sur l'aire urbaine de Paris montrent certaines particularités. L'élasticité du prix total des logements par rapport à leur surface est très élevée. De manière inattendue, le paramètre pour la présence d'un parking est négatif. Enfin, les logements construits avant 1992 semblent avoir un prix significativement plus bas que les autres à l'exception de ceux construits avant 1948. On peut ici penser que les immeubles anciens de Paris (haussmanniens notamment) sont valorisés par les acheteurs ce qui tire le paramètre associé aux logements construits avant 1918 vers le haut. Ceci explique également les résultats sur la période de construction dans les estimations sur la France entière. Contrairement à Paris, dans les aires urbaines de Province, les logements plus anciens sont moins valorisés que les logements plus récents.

Ces résultats suggèrent des différences de prix implicites (et donc potentiellement de marchés) pour certaines caractéristiques selon l'aire urbaine considérée. La variabilité des paramètres estimés pour certaines caractéristiques d'une tranche d'aire urbaine à l'autre justifie également le fait d'estimer des fonctions de prix hédoniques sur des sous-échantillons d'aires urbaines plutôt qu'une estimation sur l'ensemble des données. Un test de Chow permet de tester l'existence ou non de ces ruptures entre tailles d'aires urbaines. Le test comparant le modèle sur l'ensemble de la France et les modèles par tranche d'aires urbaines montre des différences significatives entre les coefficients estimés.

Enfin, les indicatrices temporelles sont toutes significatives et montrent une croissance des prix sur la France entière et sur chaque tranche d'aires urbaines notamment entre 2000 et 2006. Après 2006, l'évolution est variable d'une tranche d'aires urbaines à l'autre.

4 Prix immobiliers et localisation

4.1 Prix immobiliers et distance au centre des aires urbaines

Dans les modèles standards d'économie urbaine, la distance au centre du bassin d'emploi influe sur le prix du foncier. L'arbitrage des ménages entre la proximité au centre-ville ou au centre d'emploi et la surface du logement qu'ils décident d'acheter va se refléter dans le prix unitaire du sol. Les observations du prix du foncier et de l'immobilier devraient refléter en partie cet arbitrage. Cette vision simplifiée de la ville présente cependant l'avantage de pouvoir être confrontée facilement aux données et d'être compatible avec une fonction de prix hédoniques dans laquelle la distance au centre de l'aire urbaine capture en partie les effets sur les prix de l'accessibilité des logements par rapport au centre-ville ainsi que certaines aménités qui évoluent selon la distance au centre.

Les figures 4 et 3 représentent graphiquement le prix unitaire des logements en fonction de la distance au centre de l'aire urbaine pour l'ensemble de la France et pour Paris, et les figures 2 et 5 représentent le prix au m² des biens sur une carte de différentes aires urbaines.

Sur l'Île-de-France en 2010, on observe bien une tendance générale négative du gradient de prix en fonction de la distance au centre de Paris. Cette tendance se retrouve lorsqu'on différencie par type de bien. Les cartes par type de bien permettent également de se rendre compte de la sous-représentation des communes éloignées du centre si on se limite aux données concernant les appartements. On observe enfin que certaines zones relativement éloignées du centre présentent un prix unitaire élevé comme le littoral à Marseille et à Bordeaux. Néanmoins, on observe bien sur l'Île-de-France comme sur les aires urbaines de Lyon ou de Toulouse (cf. figure 2) des prix unitaires moins élevés lorsque l'on s'éloigne de la ville-centre.

Sur l'ensemble des aires urbaines (figure 4), l'effet de la distance au centre sur les prix est moins clair notamment à cause de certaines aires urbaines qui ne semblent pas suivre le schéma classique de la ville monocentrique. On observe néanmoins une tendance générale à la baisse des prix en fonction de l'éloignement au centre. Le fait que l'on n'observe pas un effet net de la distance au centre n'est pas ici étonnant car d'autres sources de variabilité liées à d'autres paramètres comme la taille de l'aire urbaine ou les différentes caractéristiques des biens existent dans les données. Enfin, d'autres aménités de localisation (positives ou négatives) ont de l'influence sur le prix unitaire et certaines aires urbaines sont polycentriques et composées de plusieurs bassins d'emploi influençant les prix. Une illustration de ces effets peut être observée sur la carte de l'aire urbaine de Marseille (figure 2) où on observe notamment que les logements en bordure du littoral et au centre d'Aix-en-Provence ont un prix au m² plus élevés que ceux situés au centre de Marseille.

Figure 3 : Prix au m² des appartements et des maisons en fonction de la distance au centre de Paris, 2010, Île-de-France

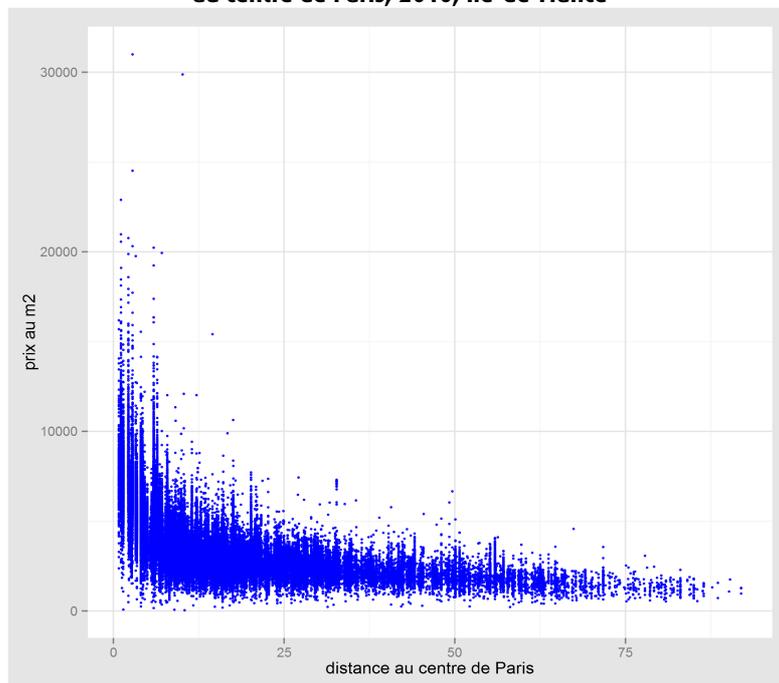
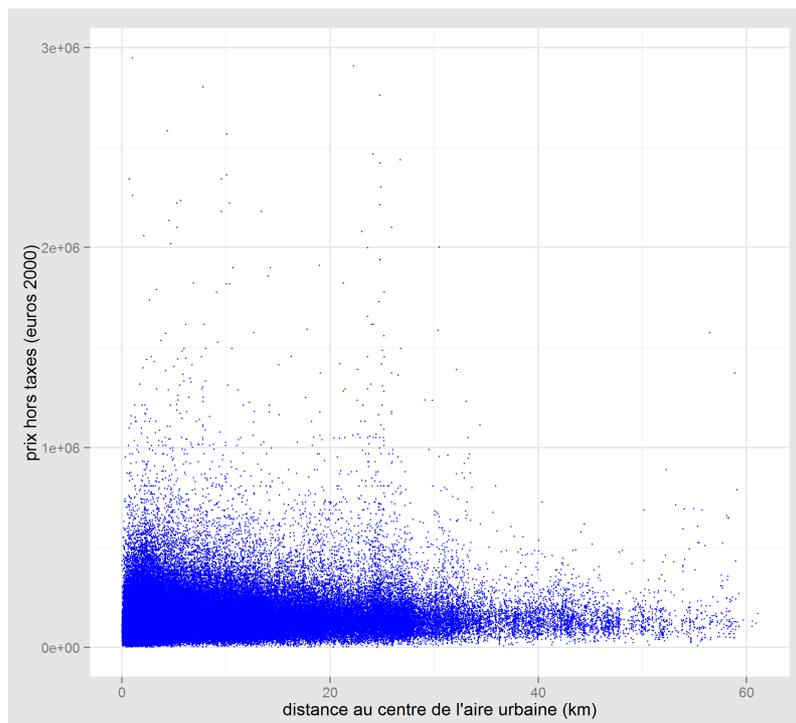


Figure 4 : Prix au m² des appartements et des maisons en fonction de la distance au centre des aires urbaines, 2010, France entière



Source : bases BIEN et PERVAL

Figure 5 : Prix au m² par type de logement, 2010, Île-de-France

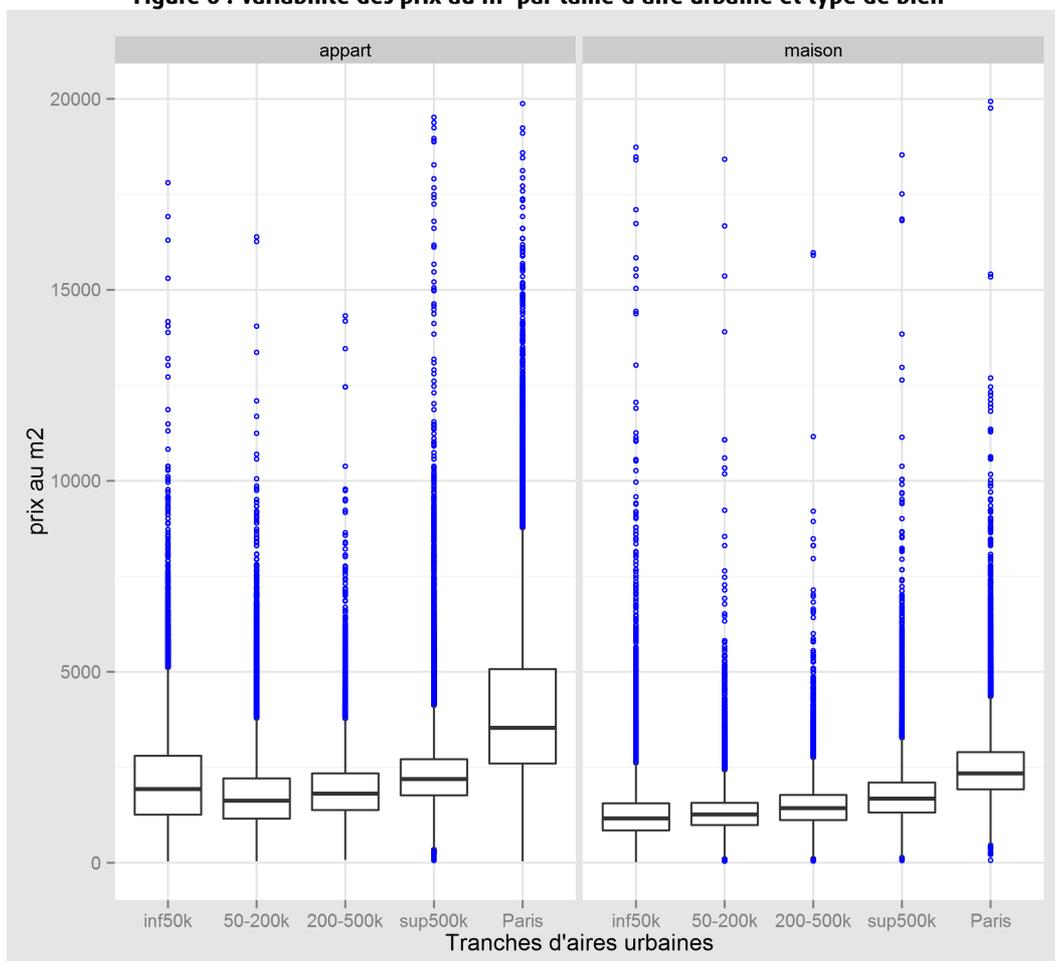
Source : base BIEN

4.2 Prix immobiliers et taille des aires urbaines

L'échantillon de données dont on dispose couvre une large variété de zones urbaines qui ne forment pas un marché unique ce qui peut être à l'origine d'une partie de l'hétérogénéité observée dans les prix immobiliers. Combes et al. (2011) montrent par exemple que les prix fonciers sont d'autant plus importants que la taille de l'aire urbaine est importante. La taille des aires urbaines et de manière générale leurs caractéristiques démographiques et économiques sont susceptibles d'expliquer la part de variabilité observée dans les prix des biens qui ne s'explique pas par des différences de caractéristiques physiques des biens ou de localisation au sein d'une ville.

Le graphique 6 montre la variabilité des prix unitaires en fonction de la taille des AU et du type de bien. En moyenne, les prix unitaires des maisons et des appartements augmentent avec la taille de l’AU avec une exception concernant les appartements dans les AU de moins de 50 000 habitants qui ont globalement des prix unitaires élevés. Ce graphique montre également que la dispersion des prix unitaires au sein d’une même tranche d’AU est grande et que la taille de l’AU n’explique qu’en partie cette variabilité.

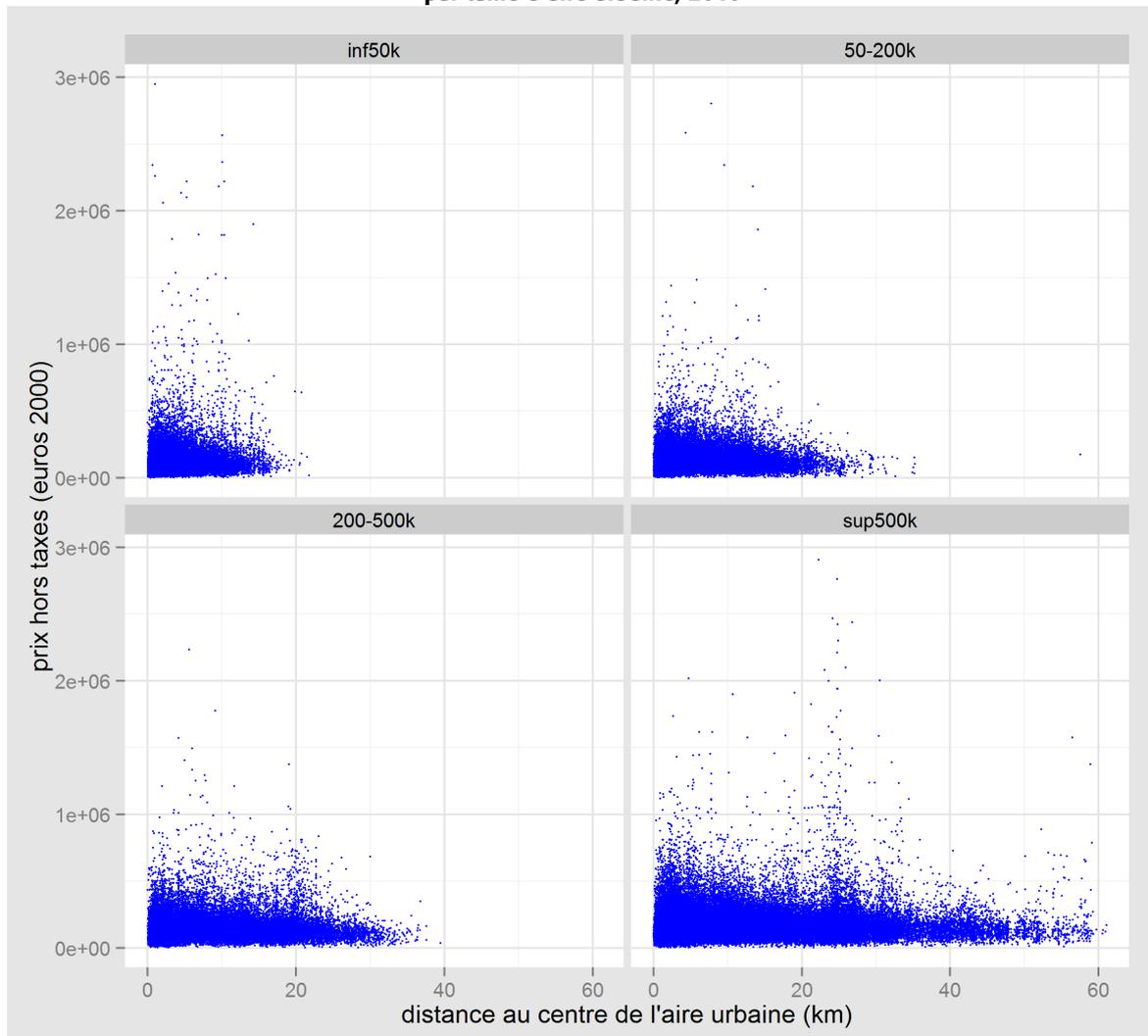
Figure 6 : Variabilité des prix au m² par taille d’aire urbaine et type de bien



Note de lecture : les « boîtes à moustaches » représentées ici donnent des indications sur la distribution et de la dispersion des prix au m² des logements selon le type de bien et la taille de l’aire urbaine. Pour chacune, la barre centrale représente la médiane du prix au m², la boîte en elle-même encadre le premier quartile (Q1) et le troisième quartile (Q3), les « moustaches » en noir correspondent à 1,5 fois l’écart inter-quartile (Q3-Q1). Les points représentés en bleu correspondent aux observations situées en dehors de cet intervalle.

La figure 7 illustre l’effet conjoint de la taille de l’AU et de la distance au centre. On observe que les limites des aires urbaines (en tout cas dans les observations des bases notariales) varient selon la taille de l’aire urbaine et que la décroissance des prix est différente selon la taille de l’AU. Ceci explique que l’on n’observe pas un effet net de la distance au centre sur l’ensemble des données empilées (figure 4). Enfin, on observe qu’à une même distance du centre de l’AU, la variabilité des prix reste importante. La distance au centre, de même que la taille de l’aire urbaine explique donc seulement une partie de la variabilité observée sur les prix de l’immobilier. La taille de l’AU semble jouer à la fois sur les prix moyens dans l’AU et aussi sur le gradient de prix au sein de l’AU. On testera ces deux effets dans les estimations.

Figure 7 : Prix total des appartements et des maisons en fonction de la distance au centre de l'aire urbaine par taille d'aire urbaine, 2010



Source : base PERVAL

4.3 Estimations sur l'ensemble de la France métropolitaine : effet de la taille de l'aire urbaine et de la distance au centre de l'aire urbaine

Les estimations présentées ici étudient l'influence de différentes variables de localisation sur la variabilité des prix entre les aires urbaines et au sein d'une aire urbaine donnée. Les différences de prix entre aires urbaines sont capturées soit par l'introduction d'indicatrices par aire urbaine soit par l'introduction de variables explicatives continues comme la population, la densité ou la croissance démographique de l'aire urbaine. La variabilité des prix au sein d'une aire urbaine est capturée par la distance entre le bien et le centre de l'aire urbaine dans laquelle il se situe. Le modèle estimé est ici :

$$\ln(P_{ij}) = \beta_0 + \alpha_j + \beta_s \ln(s_{ij}) + \beta_{jd} d_{ij} + \sum_l \beta_l x_{ijl} + \epsilon_{ij}$$

où α_j est une indicatrice pour l'aire urbaine j , d_{ij} est alternativement la distance en km entre la commune où est situé le bien i et le centre de l'aire urbaine j quand celui-ci est localisé à la commune ou la distance entre les coordonnées du bien dans la base de données et le centre de l'aire urbaine quand celui-ci est localisé plus précisément, β_{jd} est le paramètre

estimé pour l'effet de la distance sur le prix des logements, x_{ijl} sont les caractéristiques des biens.

L'effet des indicatrices par aire urbaine est d'abord testé en l'absence des autres caractéristiques des biens à l'exception de la surface habitable (colonne 1), puis en ajoutant la distance au centre de l'aire urbaine (colonne 2) et les autres caractéristiques des biens (colonne 3). On reporte seulement une partie de la distribution des paramètres estimés pour les indicatrices par aire urbaine (moyenne, premier et dernier décile sur les 314 paramètres estimés). Enfin, dans les colonnes 4 et 5, on remplace les indicatrices par aire urbaine par des variables continues calculées à l'échelle des aires urbaines.

On restreint ici l'analyse à un sous-échantillon de 314 aires urbaines pour lesquelles on dispose d'au moins 25 observations de transactions. Les estimations sont réalisées pour chaque année de l'échantillon de données même si on ne présente ici que les résultats pour l'année 2010. Les résultats d'estimations sont d'abord présentés pour l'ensemble de la France métropolitaine et par tranche d'aires urbaines.

Tableau 6 : Résultats d'estimations sur la France métropolitaine, 2010

Variable	(1)	(2)	(3)	(4)	(5)
Constante	9.02***	8.96***	9.28***	6.99***	6.49***
log(surface habitable)	0.778***	0.844***	0.785***	0.694***	0.705***
nombre de sdb manquant			-0.0354***	-0.0372***	-0.0455***
nombre sdb >1			0.138***	0.246***	0.241***
présence d'un parking			0.0495***	0.04***	0.0423***
période de construction manquante			-0.125***	-0.0784***	-0.068***
construction avant 1913			-0.0336***	-0.0315***	-0.00461
construction 1914-1948			-0.151***	-0.165***	-0.149***
construction 1948-1969			-0.177***	-0.171***	-0.15***
construction 1970-1980			-0.155***	-0.131***	-0.12***
construction 1981-1991			-0.0605***	-0.0144***	-0.0158***
maison			0.0549***	-0.0827***	-0.0695***
construction neuve			0.072***	0.0569***	0.0521***
log(population de l'AU en 2010)				0.152***	0.105***
log(densité de population de l'AU en 2010)					0.166***
log(Population en 1999/Population en 2010)					1.40***
distance au centre de l'AU		-0.0144***	-0.0162***	-0.011***	-0.0115***
moyenne des effets fixes	8.18	7.96	8.24		
premier décile des effets fixes	7.8	7.56	7.86		
dernier décile des effets fixes	8.63	8.48	8.74		
R2 ajusté	0.63	0.68	0.71	0.55	0.58
Nombre d'observations	342 889	342 889	342 889	342 889	342 889
Nombre d'AU	314	314	314	314	314
Nombre de communes	11 516	11 516	11 516	11 516	11 516

Les indicatrices par aire urbaine sont quasiment toutes significatives au seuil de 1 % à l'exception de celles pour 3 aires urbaines. Elles expliquent une part assez importante de la variabilité des prix totaux (environ 21 %, la surface habitable expliquant 28 % seule). Ces résultats montrent qu'une part importante de la variabilité observée dans les transactions est due à des caractéristiques propres à chaque aire urbaine. Le modèle le plus complet (caractéristiques des logements, distance au centre et indicatrices par aire urbaine) explique plus de 70 % de la variabilité des prix.

L'effet de la distance au centre de l'aire urbaine sur le prix total et sur le prix unitaire des logements est négatif et très significatif dans chacune des estimations. L'ajout de ce paramètre fait monter sensiblement le R^2 d'environ 5 %. En moyenne, pour une aire urbaine donnée et un bien de caractéristiques données, l'éloignement du centre de l'aire urbaine d'1 km fait baisser le prix d'environ 1 %. Il convient d'interpréter ce paramètre avec précaution car il semble assez sensible aux variables explicatives introduites (la distance au centre peut-être corrélée avec beaucoup de variables explicatives: année de construction, présence d'un parking ou d'un garage, type de bien ...). Il peut également exister des variables omises qui biaisent ce paramètre surtout sur ces estimations où l'ensemble des données pour la France entière est utilisé.

La significativité et le signe des coefficients associés aux caractéristiques des biens sont peu modifiés par l'introduction de ces variables de localisation et vont dans le même sens que les estimations incluant seulement les caractéristiques des biens à l'exception du paramètre associé aux maisons qui est cette fois-ci positif lorsque la localisation est contrôlée (colonne 3) ce qui est conforme avec la majorité des études hédoniques.

Les colonnes 4 et 5 étudient l'influence de variables explicatives à l'échelle des aires urbaines. Elles montrent que les prix sont d'autant plus élevés que la taille, la densité et la croissance de la population des aires urbaines sont importantes. Les R^2 pour ces deux estimations sont sensiblement plus faibles que ceux des estimations avec des indicatrices par aire urbaine signalant que ces trois variables capturent seulement une partie de la variabilité des prix entre AU. Nous étudierons plus précisément dans la partie 6 quels facteurs (en plus de la taille de l'aire urbaine) permettent d'expliquer les différences de prix moyens entre AU.

4.4 Estimations par tranche d'aires urbaines avec variables de localisation

Pour clarifier l'effet de certaines variables de localisation comme la distance au centre de l'aire urbaine et la taille de l'aire urbaine, des estimations sont effectuées en séparant l'échantillon de données par tranche d'aires urbaines. Ceci permet notamment d'estimer des valeurs du paramètre associé à la distance au centre de l'AU sur des aires urbaines de taille (superficie) équivalente et avec un nombre d'observations par aire urbaine plus équilibré. D'autre part, Cavailhès (2005) montre que des fonctions de prix hédoniques estimées sur des sous-échantillons par tranche d'aires urbaines capturent une part de l'hétérogénéité des prix plus importante que des estimations à l'échelle nationale. Le test de Chow comparant le modèle sur l'ensemble de la France et les modèles par tranche d'aires urbaines montre des différences significatives entre les coefficients estimés pour chaque année de l'échantillon. Enfin, les tests statistiques usuels montrent que l'hétéroscédasticité est réduite dans ces estimations par rapport à celle sur l'échantillon national.

Le modèle estimé est identique à celui estimé sur l'ensemble de la France. Les tableaux suivants comparent les résultats de plusieurs modèles par tranche d'AU. On teste tout d'abord l'arbitrage surface habitable-distance au centre (colonne 1) puis l'effet des caractéristiques propres des aires urbaines à travers les indicatrices par aire urbaine (colonne 2) puis ces deux effets conjointement (colonne 3). On ajoute ensuite l'ensemble des caractéristiques des logements testées précédemment (colonne 4). Dans la colonne 5, on autorise le paramètre associé à la distance au centre à varier entre les aires urbaines pour tester la présence de différents types de gradients de prix entre les AU. Pour les indicatrices et les gradients de prix par aire urbaine, on reporte seulement la moyenne, le premier et le dernier décile de la distribution des paramètres.

Les résultats montrent qu'à surface habitable du logement et aire urbaine donnée, l'effet de la distance sur le prix total est négatif et significatif pour toutes les tranches d'aires urbaines sauf celles de moins de 50 000 habitants. La proximité au centre

de l'AU dans la majeure partie de ces petites aires urbaines ne semble pas être valorisée. Dans le reste des aires urbaines, le fait de s'éloigner d'1 km du centre fait baisser les prix de 0,5 à 0,8 % en Province et de 2 % à Paris.

L'ensemble surface habitable, distance au centre et indicatrices par aire urbaine explique une part importante de la variance du prix total (entre 62 et 68 % selon la TAU), le reste des caractéristiques expliquant 3 à 7 % de la variabilité restante. Les paramètres associés aux caractéristiques restent très significatifs et vont globalement dans le même sens que dans les régressions incluant seulement les caractéristiques du bien. Le paramètre associé aux maisons devient positif une fois la localisation contrôlée ce qui suggère que les maisons ont globalement un prix total plus élevé que les appartements à localisation donnée comme le rapportent la plupart des études hédoniques sur la France.

Les indicatrices par aire urbaine expliquent une part importante de la variance suggérant des différences significatives de prix entre les aires urbaines d'une même tranche. Le fait d'autoriser le paramètre associé à la distance au centre à varier entre les AU (colonne 5) améliore légèrement le R^2 et permet de constater que l'éloignement du centre de l'AU n'a pas un effet négatif sur le prix des logements dans toutes les aires urbaines (dernier décile de la distribution des paramètres de distance positif).

Les gradients de prix varient donc d'une aire urbaine à l'autre indiquant une valorisation plus ou moins forte du centre urbain selon la ville. Une analyse plus fine permettrait d'identifier les AU pour lesquelles le paramètre est positif et de faire des hypothèses sur l'origine de cette valeur (polycentrisme, aménités négatives du centre urbain,...).

Au final, l'ajout de ces quelques variables de localisation améliore sensiblement le R^2 pour toutes les tranches d'aires urbaines par rapport aux modèles ne considérant que les caractéristiques des biens. L'ensemble caractéristiques des logements-variables de localisation explique entre 72 et 78 % de la variabilité des prix. Ces valeurs sont proches de celles obtenues par Fauvet (2008) qui utilise les bases notariales pour estimer des fonctions de prix similaires.

Tableau 7 : Résultats d'estimations sur les aires urbaines de moins de 50 000 habitants, 2010

Variable	(1)	(2)	(3)	(4)	(5)
log(surface habitable)	0.556***	0.866***	0.862***	0.78***	0.809***
nombre de sdb manquant				-0.106***	-0.104***
nombre sdb >1				0.118***	0.0914***
présence d'un parking				0.106***	0.116***
période de construction manquante				-0.115***	-0.145***
construction avant 1913				-0.305***	-0.316***
construction 1914-1948				-0.27***	-0.286***
construction 1948-1969				-0.173***	-0.199***
construction 1970-1980				-0.127***	-0.153***
construction 1981-1991				-0.0295***	-0.0531***
construction neuve				0.0606***	0.0326***
maison				0.0604***	0.0666***
Constante	9.14***	8.3***	8.25***	8.52***	8.44***
moyenne des effets fixes		7.79	7.77	8.15	8.05
premier décile des effets fixes		7.36	7.34	7.76	7.66
dernier décile des effets fixes		8.37	8.35	8.66	8.54
distance au centre de l'AU	0.0256***		0.00918***	0.00829***	
moyenne des paramètres de distance					0.00173
premier décile des paramètres de distance					-0.0449
dernier décile des paramètres de distance					0.0595
R2 ajusté	0.28	0.65	0.66	0.70	0.72
Nombre d'observations	24 883	24 883	24 883	24 883	24 883
Nombre d'AU	173	173	173	173	173
Nombre de communes	1 500	1 500	1 500	1 500	1 500

Tableau 8 : Résultats d'estimations sur les aires urbaines de 50 000 à 200 000 habitants, 2010

Variable	(1)	(2)	(3)	(4)	(5)
log(surface habitable)	0.652***	0.814***	0.819***	0.736***	0.754***
nombre de sdb manquant				-0.0614***	-0.0699***
nombre sdb >1				0.123***	0.105***
présence d'un parking				0.115***	0.111***
période de construction manquante				-0.178***	-0.182***
construction avant 1913				-0.278***	-0.272***
construction 1914-1948				-0.279***	-0.277***
construction 1948-1969				-0.238***	-0.236***
construction 1970-1980				-0.164***	-0.166***
construction 1981-1991				-0.0824***	-0.0888***
construction neuve				0.0653***	0.0707***
maison				0.0793***	0.0936***
Constante	8.74***	8.37***	8.36***	8.74***	8.85***
moyenne des effets fixes		7.97	7.95	8.35	8.27
premier décile des effets fixes		7.67	7.66	8.09	8.01
dernier décile des effets fixes		8.31	8.31	8.66	8.61
distance au centre de l'AU	0.00401***		-0.00253***	-0.00492***	
moyenne des paramètres de distance					-0.00533
premier décile des paramètres de distance					-0.0204
dernier décile des paramètres de distance					0.0134
R2 ajusté	0.35	0.62	0.62	0.69	0.71
Nombre d'observations	45 395	45 395	45 395	45 395	45 395
Nombre d'AU	95	95	95	95	95
Nombre de communes	3 348	3 348	3 348	3 348	3 348

Tableau 9 : Résultats d'estimations sur les aires urbaines de 200 000 à 500 000 habitants, 2010

Variable	(1)	(2)	(3)	(4)	(5)
log(surface habitable)	0.772***	0.819***	0.841***	0.786***	0.799***
nombre de sdb manquant				-0.049***	-0.0501***
nombre sdb >1				0.114***	0.1***
présence d'un parking				0.0848***	0.0832***
période de construction manquante				-0.126***	-0.131***
construction avant 1913				-0.206***	-0.19***
construction 1914-1948				-0.206***	-0.206***
construction 1948-1969				-0.19***	-0.193***
construction 1970-1980				-0.152***	-0.151***
construction 1981-1991				-0.0531***	-0.06***
construction neuve				0.0758***	0.0726***
maison				0.0495***	0.069***
Constante	8.37***	8.14***	8.09***	8.35***	8.36***
moyenne des effets fixes		8.11	8.07	8.33	8.26
premier décile des effets fixes		7.85	7.8	8.1	8
dernier décile des effets fixes		8.47	8.44	8.68	8.49
distance au centre de l'AU	-0.000865***		-0.00686***	-0.00821***	
moyenne des paramètres de distance					-0.00784
premier décile des paramètres de distance					-0.0202
dernier décile des paramètres de distance					0.00446
R2 ajusté	0.50	0.67	0.68	0.72	0.74
Nombre d'observations	55 355	55 355	55 355	55 355	55 355
Nombre d'AU	29	29	29	29	29
Nombre de communes	2 833	2 833	2 833	2 833	2 833

Tableau 10 : Résultats d'estimations sur les aires urbaines de plus de 500 000 habitants, 2010

Variable	(1)	(2)	(3)	(4)	(5)
log(surface habitable)	0.723***	0.797***	0.821***	0.768***	0.782***
nombre de sdb manquant				-0.0379***	-0.0434***
nombre sdb >1				0.164***	0.148***
présence d'un parking				0.107***	0.099***
période de construction manquante				-0.0953***	-0.1***
construction avant 1913				-0.199***	-0.186***
construction 1914-1948				-0.205***	-0.203***
construction 1948-1969				-0.165***	-0.172***
construction 1970-1980				-0.121***	-0.128***
construction 1981-1991				-0.021***	-0.0383***
construction neuve				0.0446***	0.046***
maison				0.0432***	0.0821***
Constante	8.79***	8.48***	8.44***	8.67***	8.69***
moyenne des effets fixes		8.38	8.34	8.56	8.5
premier décile des effets fixes		8.11	8.05	8.29	8.2
dernier décile des effets fixes		8.6	8.56	8.81	8.66
distance au centre de l'AU	-0.00181***		-0.00532***	-0.00708***	
moyenne des paramètres de distance					-0.00753
premier décile des paramètres de distance					-0.0182
dernier décile des paramètres de distance					0.00341
R2 ajusté	0.49	0.62	0.63	0.68	0.71
Nombre d'observations	108 195	108 195	108 195	108 195	108 195
Nombre d'AU	16	16	16	16	16
Nombre de communes	2 661	2 661	2 661	2 661	2 661

Tableau 11 : Résultats d'estimations sur l'Île-de-France, 2010

Variable	(1)	(2)	(3)
log(surface habitable)	0.867***	0.865***	0.866***
nombre de sdb manquant		-0.0271***	-0.0338***
nombre sdb >1		0.139***	0.125***
présence d'un parking		-0.0453***	-0.0189***
période de construction manquante		-0.0561***	-0.0618***
construction avant 1913		0.142***	0.0976***
construction 1914-1948		-0.0126***	-0.0335***
construction 1948-1969		-0.112***	-0.118***
construction 1970-1980		-0.142***	-0.144***
construction 1981-1991		-0.0732***	-0.0672***
construction neuve		0.113***	0.117***
maison		-0.0168***	-0.0119***
Constante	8.98***	9.02***	9.75***
moyenne des effets fixes			9.47
premier décile des effets fixes			9.11
dernier décile des effets fixes			9.88
distance au centre de l'AU	-0.022***	-0.0211***	-0.0279***
moyenne des paramètres de distance			
premier décile des paramètres de distance			
dernier décile des paramètres de distance			
R2 ajusté	0.65	0.68	0.71
Nombre d'observations	109 381	109 381	109 381
Nombre de communes	1 174	1 174	1 174

Les résultats d'estimations similaires sur l'Île-de-France montrent que la distance au centre a également un effet négatif et significatif sur le prix des logements. La surface habitable et la distance au centre de Paris expliquent à elles seules une part très importante de la variabilité des prix, le reste des caractéristiques n'expliquant qu'environ 3 % de la variabilité restante. La valeur du paramètre associé à la distance au centre est élevée en comparaison avec les estimations sur les autres aires urbaines signifiant une forte valorisation du centre de Paris. L'ajout d'indicatrices par unité urbaine² dans la dernière colonne améliore légèrement le R² sans modifier grandement les autres paramètres estimés.

2 La notion d'unité urbaine repose sur la continuité du bâti et le nombre d'habitants. On appelle unité urbaine une commune ou un ensemble de communes présentant une zone de bâti continu (pas de coupure de plus de 200 mètres entre deux constructions) qui compte au moins 2 000 habitants. L'aire urbaine de Paris est composée de plusieurs unités urbaines : l'unité urbaine de Paris qui constitue le pôle urbain de l'aire urbaine et des pôles secondaires. Dans les estimations sur l'Île-de-France, nous ajoutons des indicatrices par unité urbaine pour tester si ces différents pôles présentent des niveaux de prix significativement différents.

4.5 Bilan : prix hédoniques de la surface habitable et de l'éloignement du centre de l'aire urbaine

Les deux tableaux suivants récapitulent les résultats des estimations par tranche d'aires urbaines et sur Paris en présentant les prix hédoniques estimés de la surface habitable et de la distance au centre. Les prix hédoniques sont estimés sur chaque observation de l'échantillon en utilisant les formules du tableau 1 et les moyennes de ces prix estimés sont ensuite calculées sur l'ensemble de l'échantillon, par tranche d'aires urbaines et selon la taille du logement et la distance du logement au centre de l'AU.

Tableau 12 : Prix hédoniques (moyens) de la surface estimés selon la surface habitable et par tranche d'aires urbaines, en euros par m², 2010

Tranche d'aires urbaines	Toutes surfaces	10-50 m ²	50-75 m ²	75-100 m ²	100-200 m ²	200-400 m ²
< à 50k habitants	1 417	1 943	1 319	1 106	1 035	981
50 à 200k habitants	1 130	1 509	1 159	1 006	927	801
200 à 500k habitants	1 340	1 589	1 371	1 234	1 161	1 024
> 500k habitants	1 605	1 930	1 619	1 438	1 329	1 156
Paris	3 040	3 547	2 879	2 658	2 561	2 017
Ensemble des AU	1 706	2 103	1 669	1 488	1 403	1 196

Le prix d'un m² de surface habitable est estimé sur l'ensemble de l'échantillon à 1 706 euros en 2010 (soit 68 euros par an et par m² si on fixe arbitrairement une durée de remboursement du bien à 25 ans). Le prix du m² est croissant avec la taille des aires urbaines pour les AU de plus de 50 000 habitants et il atteint 3 040 euros à Paris ce qui représente une valeur supérieure de 78 % à la moyenne sur l'ensemble de l'échantillon. Parmi les explications possibles pour expliquer ces différences on peut citer des différences de prix fonciers entre AU (Combes et al., 2012), des coûts de construction plus élevés dans les AU les plus denses ou encore la présence d'aménités positives plus importantes dans les villes densément peuplées (Cavailhès, 2005). Le prix unitaire dans les AU les moins peuplées est en moyenne supérieur à celui des AU de 50 à 200 000 habitants et ceux pour toutes les tailles de logement.

Le prix unitaire de la surface diminue avec la taille du logement pour toutes les tailles d'aires urbaines. Les petites surfaces inférieures à 50 m² ont un prix unitaire de 2 103 euros sur l'ensemble des AU soit environ 23 % supérieur à la moyenne sur l'ensemble des observations. Ces résultats sont cohérents avec la littérature et peuvent s'expliquer en partie par l'existence de coûts fixes dans la construction des logements qui augmentent le coût de construction unitaire des petits logements.

La figure 8 résume l'effet conjoint de la taille des aires urbaines et de la taille des logements sur les prix hédoniques de la surface habitable et met en avant la variabilité du consentement à payer des ménages pour un m² de surface habitable supplémentaire.

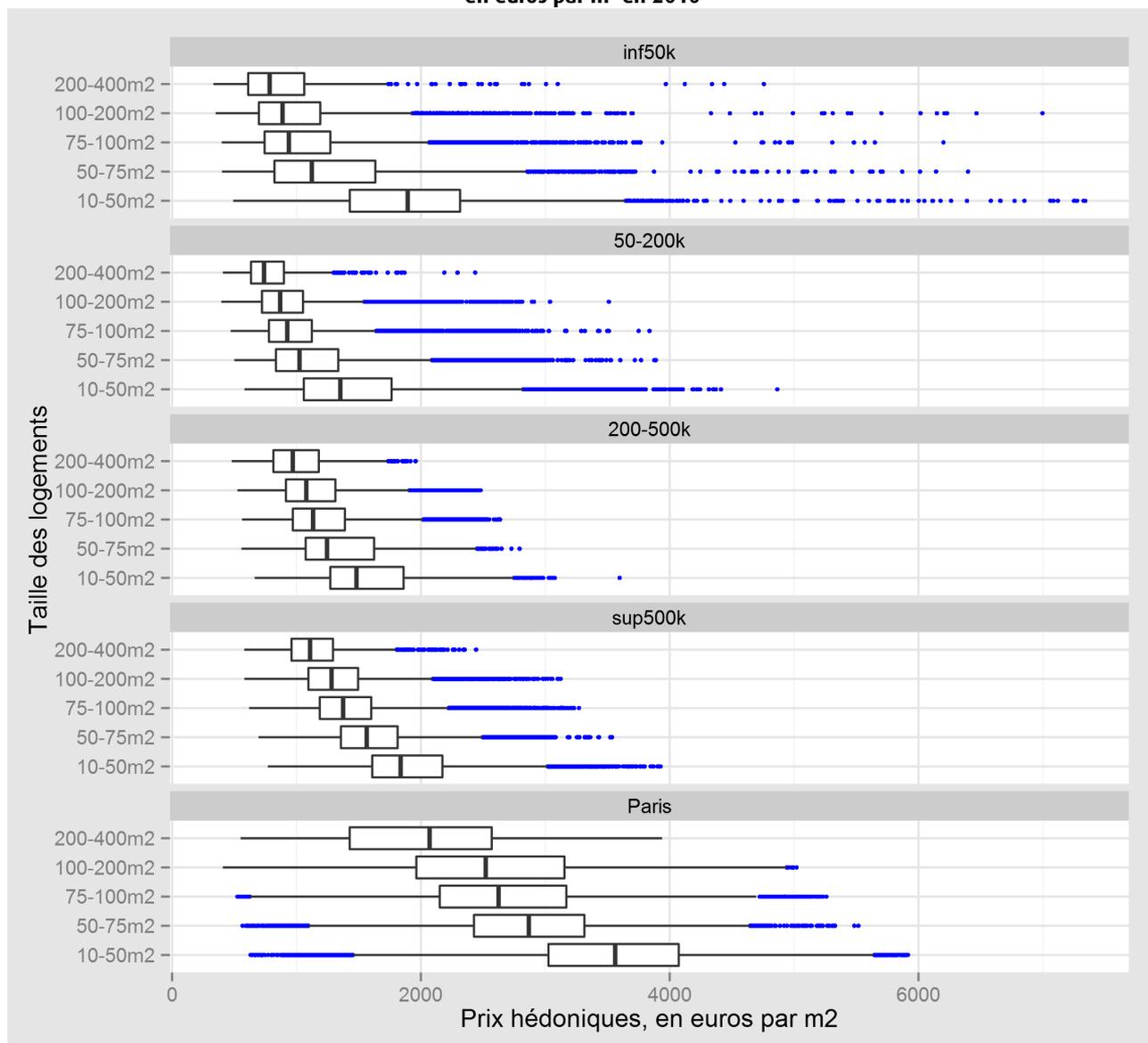
Tableau 13 : Prix hédoniques (moyens) de la distance au centre estimés selon la distance et par tranche d'aires urbaines, en euros par km, 2010

Tranche d'aires urbaines (en milliers d'habitants)	Toutes distances	0-2.5km	2.5-5km	5-7.5km	7.5-10km	10-20km	20-40km	> 40km
< à 50	936	868	972	1 064	970	990	1 059	
50 à 200	-617	-537	-660	-724	-714	-664	-668	
200 à 500	-1 058	-924	-1 016	-1 232	-1 254	-1 181	-1 183	-796
> 500	-1 040	-909	-997	-1 073	-1 129	-1 138	-1 103	-1 033
Paris	-4 593	-5 374	-4 974	-4 876	-4 889	-4 662	-3 979	-2 730
Ensemble des AU	-1 274	-1 375	-1 335	-1 368	-1 403	-1 331	-1 175	

Le prix hédonique de la distance au centre de l'AU qui représente le coût de l'éloignement du centre urbain (centre d'emploi) ou encore le coût (en valeur absolue) que les ménages sont prêts à supporter pour se rapprocher d'une unité de distance du centre est estimé à 1274 euros par km sur l'ensemble de l'échantillon (environ 50 euros par an et par km sur 25 ans). Ce coût est plus faible dans les AU les moins peuplées (de -617 euros par km dans les AU de 50 à 200 000 habitants à -4 593 euros pour Paris). Ce résultat est en contradiction avec Cavailhès (2005) qui obtient des coûts de l'éloignement au centre moins élevés dans les AU les plus denses et justifie ce résultat en attribuant cela à une efficacité plus grande des réseaux de transports dans les grandes agglomérations. On peut néanmoins facilement justifier le résultat inverse en rappelant que les grandes agglomérations concentrent beaucoup d'aménités positives dont il est coûteux de s'éloigner. Les résultats concernant les paramètres de distance sont cependant encore fragiles car comme l'évoque Cavailhès (2005) dans son article, la distance au centre est potentiellement corrélée avec certaines variables omises (comme les aménités positives et négatives présentant une structure spatiale par exemple). Il est donc difficile d'estimer précisément ce paramètre sans contrôler pour ces variables omises, soit en les incluant dans l'estimation sous forme de variables explicatives, soit en utilisant des méthodes plus complexes comme l'économétrie spatiale.

La forme de la courbe du prix hédonique de la distance en fonction de l'éloignement au centre est variable selon les aires urbaines. Ainsi, le coût de l'éloignement au centre est très élevé sur les premiers kilomètres et semble diminuer progressivement avec la distance dans l'AU de Paris suggérant une forte valorisation du centre de Paris par les ménages. Dans les AU de 200 à 500 000 habitants, il semble augmenter sur les premiers kilomètres puis diminuer ensuite (courbe en cloche). Les différences de prix sont cependant relativement faibles entre le centre et la périphérie des AU (sauf à Paris) et la fragilité des estimations impose d'être prudent en interprétant ces résultats. Pour approfondir cet aspect, il conviendrait de tester différentes formes fonctionnelles pour les gradients de prix et d'utiliser des méthodes permettant de prendre en compte l'auto-corrélation spatiale des données. Ce type d'analyse dépasse cependant le cadre de cette étude.

Figure 8 : Variabilité des prix unitaires de la surface estimés selon la taille de l'aire urbaine et la taille des logements, en euros par m² en 2010



Note de lecture : les « boîtes à moustaches » représentées ici donnent des indications sur la distribution et de la dispersion des prix hédoniques de la surface habitable selon la taille du logement et la taille de l'aire urbaine. Elles montrent que, pour chaque taille d'aire urbaine, la médiane du prix unitaire est plus élevée pour les petites surfaces. Elles montrent aussi que, à taille de logement donnée, le prix unitaire est plus élevé dans les aires urbaines les plus peuplées, l'ensemble des boîtes à moustache se décalant vers la droite lorsque la taille des aires urbaines augmente.

5 Analyse comparative des prix de l'immobilier par aire urbaine

5.1 Estimations des prix par aires urbaines

Les estimations hédoniques réalisées ci-dessus ont permis de constater que l'hétérogénéité observée dans les prix des biens était due à la fois à des différences de caractéristiques des biens, de localisation au sein d'une zone urbaine (et donc de proximité à un ensemble d'aménités) et à des différences entre aires urbaines. L'existence de ces 3 sources de variabilité dans les prix rend complexe la comparaison des niveaux de prix entre zones urbaines toutes choses égales par ailleurs.

De manière à comparer les niveaux de prix entre aires urbaines, on utilise les régressions hédoniques précédentes pour calculer le prix d'un bien de référence par aire urbaine pour chaque année de l'échantillon et s'affranchir des effets de qualité présents dans les données brutes. On peut ensuite étudier l'influence de facteurs démographiques et économiques sur les prix et leur évolution.

Lorsqu'on utilise les régressions hédoniques qui incluent seulement les caractéristiques intrinsèques des biens, on obtient le prix moyen d'un bien de référence par AU. Si on introduit une distance au centre de l'AU, on obtient le prix d'un bien de référence au centre de l'AU (à distance nulle du centre).

Le modèle utilisé pour le calcul de ces prix est similaire à celui de la partie 4) :

$$\ln(P_{ij}) = \beta_0 + \alpha_j + \beta_s \ln(s_{ij}) + \beta_{jd} d_{ij} + \sum_l \beta_l x_{ijl} + \epsilon_{ij}$$

Sachant que les caractéristiques x_{ijl} sont toutes discrétisées, $\hat{\beta}_0$ estime le prix moyen d'un bien pour le niveau de x_{ijl} choisi comme référence. Le prix moyen d'un bien de référence estimé à l'échelle d'une aire urbaine j , pour une surface de référence \bar{s} au centre de l'aire urbaine ($d_{ij}=0$) s'exprime donc :

$$\hat{P}_j = e^{(\hat{\beta}_0 + \hat{\alpha}_j + \hat{\beta}_s \ln(\bar{s}))} = \bar{s}^{\hat{\beta}_s} e^{(\hat{\beta}_0 + \hat{\alpha}_j)}$$

Le prix de ce bien de référence est calculé pour chaque année de l'échantillon et pour chaque AU. L'avantage d'estimer le prix pour chaque année de l'échantillon est de pouvoir obtenir des évolutions différenciées de prix entre les différentes aires urbaines ce qui ne pourrait pas être obtenu avec un modèle estimé sur toutes les années de l'échantillon et avec des indicatrices temporelles (toutes les aires urbaines auraient alors la même évolution).

On retient les aires urbaines pour lesquelles on dispose d'au moins 25 observations (314 aires urbaines au total). Le tableau 14 montre les prix moyens par tranche d'aires urbaines et sur l'ensemble des aires urbaines qui ont été estimés en utilisant un modèle hédonique incluant caractéristiques des biens et distance au centre et en prenant $\bar{s} = 65 m^2$ (valeur proche de la surface moyenne des biens sur l'échantillon). Le bien de référence est ici un appartement non neuf de $65 m^2$, avec une salle de bain, sans parking, construit après 1992. Le tableau 15 reporte le coefficient de variation des prix estimés pour le bien de référence par tranche d'aires urbaines et par année.

Tableau 14 : Moyenne des prix estimés pour le bien de référence par tranche d'aires urbaines et par année en euros 2000

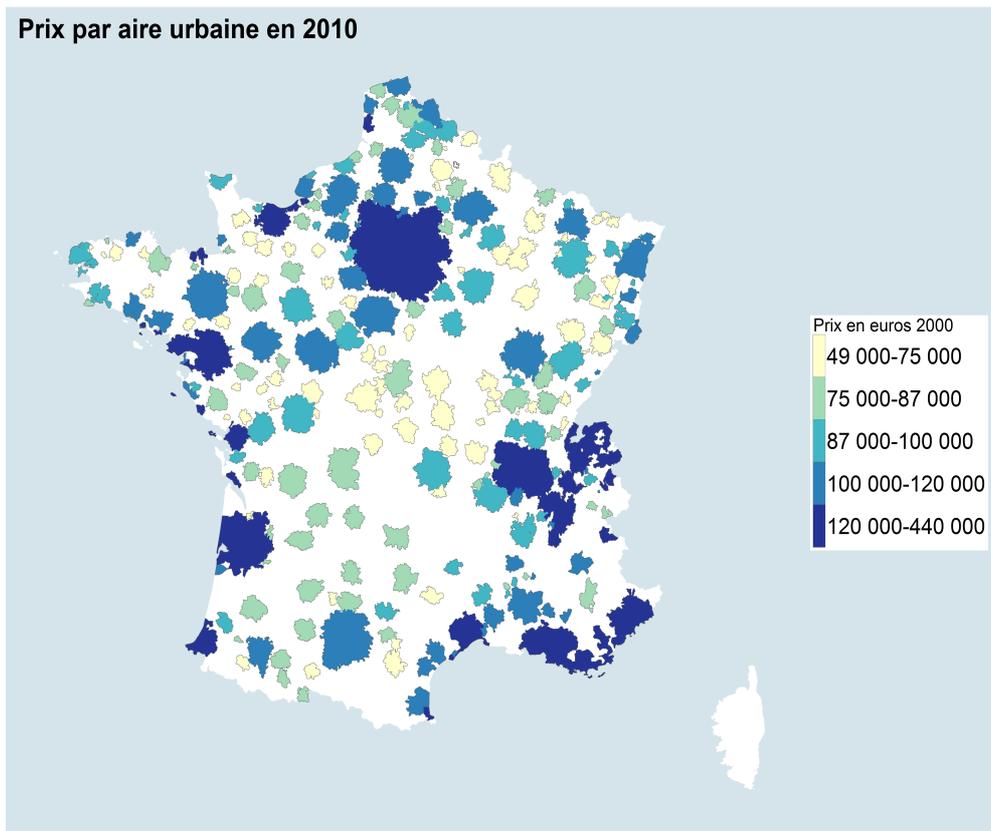
Tranche d'aires urbaines (en milliers d'habitants)	2000	2002	2004	2006	2008	2010	2012
< à 50	63 207	72 285	83 819	101 676	103 802	97 324	97 310
50 à 200	63 654	67 621	81 463	96 978	97 545	95 273	94 355
200 à 500	72 567	80 376	96 230	117 573	115 438	113 007	115 037
> 500	73 317	83 897	108 328	133 233	134 702	132 167	135 440
Paris	166 378	187 570	231 429	278 856	303 785	304 980	NA
Ensemble des AU	65 051	72 580	85 972	103 895	105 195	100 589	100 005

Tableau 15 : Coefficient de variation des prix estimés pour le bien de référence par tranche d'aires urbaines et par année en euros 2000

Tranche d'aires urbaines (en milliers d'habitants)	2000	2002	2004	2006	2008	2010	2012
< à 50	0.35	0.38	0.43	0.45	0.48	0.50	0.52
50 à 200	0.25	0.26	0.28	0.30	0.32	0.31	0.35
200 à 500	0.19	0.20	0.23	0.24	0.23	0.23	0.27
> 500	0.19	0.22	0.25	0.25	0.24	0.24	0.26
Ensemble des AU	0.31	0.34	0.38	0.40	0.42	0.43	0.45

La figure 9 montre la distribution spatiale des prix estimés en 2010. Les prix estimés sont en moyenne plus élevés dans les aires urbaines de taille plus importante. On note une forte variabilité des prix estimés sur l'ensemble des aires urbaines qui sont dans un rapport de 1 à 9 en 2010 entre l'aire urbaine la moins chère et l'aire urbaine la plus chère avec un coefficient de variation de 43 % en 2010. Les différences de prix au sein d'une même tranche d'aires urbaines peuvent également être relativement importantes suggérant que la taille de l'AU n'est pas le seul facteur qui influence les niveaux de prix. La carte montre que l'ensemble des aires urbaines dont les prix sont parmi les plus élevés sur l'ensemble de l'échantillon est principalement constitué par les aires urbaines les plus peuplées (Paris, Lyon, Marseille, Toulouse, Bordeaux, Nice...) et par les aires urbaines présentant une forte attractivité notamment touristique (aires urbaines proches du littoral).

Figure 9 : Distribution des prix estimés par aire urbaine en 2010



5.2 Quels déterminants socio-démographiques et économiques expliquent les différences de prix de l'immobilier entre les aires urbaines ?

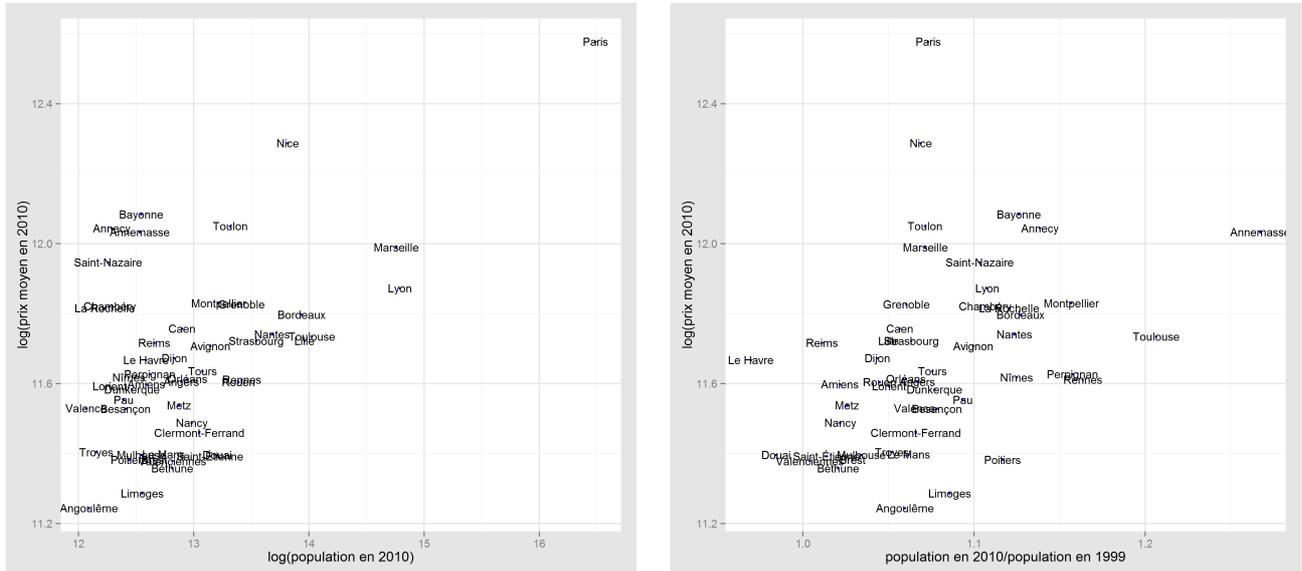
On cherche ici à déterminer quels facteurs permettent d'expliquer la variabilité observée des prix entre les aires urbaines. L'analyse proposée ici se concentre sur les prix estimés en coupe pour l'année 2010. La variabilité des prix entre aires urbaines reflète des différences d'attractivité que l'on cherche à expliquer par un ensemble de variables explicatives regroupées selon plusieurs thèmes :

- Population et structure de la population
- Revenus moyens
- Attractivité touristique
- Parc de logements
- Emploi

Pour ce faire, un ensemble de variables explicatives ont été compilées à l'échelle des aires urbaines. Ces différentes variables sont potentiellement corrélées les unes aux autres. C'est pourquoi on analyse dans un premier temps l'effet de ces variables groupe par groupe avant d'inclure simultanément certaines variables de chaque groupe dans les régressions finales.

5.2.1 Population et structure de la population

Figure 10: Prix de l'immobilier et population des aires urbaines dans les 50 plus grandes aires urbaines, 2010



Les figures 10 ci-dessus illustrent la corrélation entre d'une part les prix des logements par aire urbaine en 2010 et la population totale des aires urbaines en 2010 et d'autre part entre les prix en 2010 et la croissance de la population entre 1999 et 2010. Sur les 50 aires urbaines les plus peuplées, on observe une corrélation positive entre les niveaux de prix en 2010 et la population en 2010. De manière générale, la corrélation entre les prix et la population sur l'ensemble de l'échantillon des aires urbaines est positive et relativement faible (de l'ordre de 0,15) même si elle est statistiquement significative. De plus, une régression simple des prix sur la population montre que la population seule explique une part faible de la variabilité des prix entre aires urbaines (de l'ordre de 2 %). En revanche, lorsqu'on se limite aux 100 aires urbaines les plus peuplées, la corrélation entre les prix et la population est forte (0,60), statistiquement significative et la population seule explique 36 % de la variabilité des prix. Un certain nombre d'aires urbaines sortent de ce schéma de corrélation et notamment les aires urbaines côtières et touristiques qui présentent une population faible et des prix relativement élevés ce qui peut expliquer la corrélation faible entre population et prix sur l'ensemble de l'échantillon d'aires urbaines. Une corrélation entre les prix en 2010 et la croissance de la population des aires urbaines depuis 1999 existe également. Cette corrélation est stable, positive et significative lorsqu'on restreint l'échantillon aux AU les plus peuplées. De manière similaire, certaines aires urbaines sortent de ce schéma de corrélation comme Paris et Nice qui présentent des prix très élevés tout en ayant connu une croissance modérée de leur population.

Les régressions présentées dans les tableaux 16 et 17 étudient les liens entre les prix en 2010 et la population, la densité de population et des variables sur la structure de la population dans les aires urbaines comme la part de ménages avec une certaine ancienneté d'emménagement et la part de la population par tranche d'âge. On retrouve les observations précédentes sur la taille des aires urbaines qui a un effet positif et significatif sur les prix. La population a un effet bien plus marqué et explique une part bien plus importante de la variabilité des prix dans les 100 plus grandes aires urbaines que sur l'ensemble de l'échantillon. Similairement, les aires urbaines les plus densément peuplées présentent des prix significativement plus élevés. Les aires urbaines présentant une part importante de ménages ayant emménagé depuis moins de 20 ans ont des prix plus élevés. Des prix élevés étant un signe de l'attractivité de l'aire urbaine, il n'est pas étonnant de constater que les aires urbaines les plus chères sont celles dont la part de ménages ayant emménagé récemment est élevée. Les liens entre population par classe d'âge et prix montrent que les aires urbaines dont la part de la population de 15 à 29 ans est élevée ont des prix plus faibles par rapport à celles où la proportion de personnes de 30 à 44

et de 45 à 59 ans est élevée. Ce résultat peut s'expliquer par le fait que les jeunes (étudiants, revenus encore peu élevés) accèdent en général peu à la propriété ce qui tend à détendre le marché de l'immobilier. Il convient d'interpréter avec précaution ces résultats concernant la causalité entre le niveau des prix et la structure de la population. Une causalité inverse pourrait exister si on considère par exemple que les aires urbaines où les prix sont faibles sont privilégiées par les jeunes et les retraités disposant de ressources plus réduites.

Tableau 16 : Résultats d'estimations, population et structure de la population, ensemble des aires urbaines, 2010

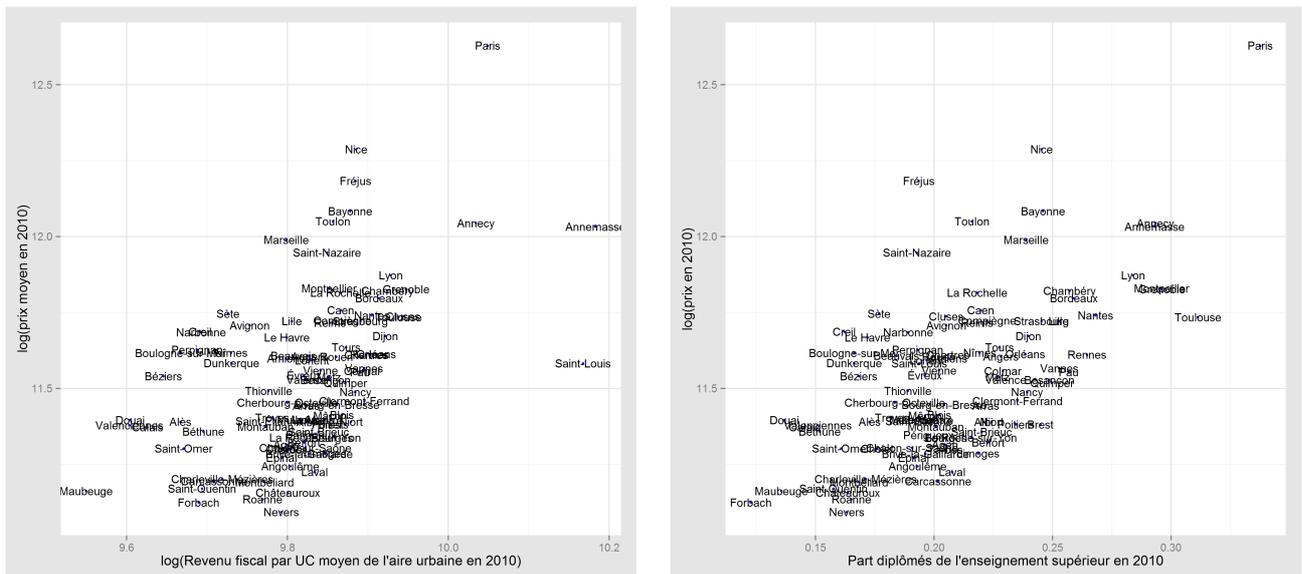
Variable	(1)	(2)	(3)	(4)
Constante	11***	10.7***	8.19***	10.2***
log(population de l'AU en 2010)	0.042***		0.0295***	0.0888***
log(densité de population de l'AU en 2010)		0.149***		
Part ménages ayant emménagé depuis - de 2 ans			1.63	
Part ménages ayant emménagé depuis 2 à 9 ans			2.48***	
Part ménages ayant emménagé depuis 10 à 19 ans			12.1***	
Part ménages ayant emménagé depuis 20 à 29 ans			-3.19*	
Part de 15-29 ans				-18.1***
Part de 30-44 ans				24.2***
Part de 45-59 ans				8.76***
Part de plus de 60 ans				-2.73***
R2 ajusté	0.023	0.083	0.5	0.58
Nombre d'observations	314	314	314	314

Tableau 17 : Résultats d'estimations, population et structure de la population, 100 plus grandes aires urbaines, 2010

Variable	(1)	(2)	(3)	(4)
Constante	9.31***	10.4***	8.83***	10.1***
log(population de l'AU en 2010)	0.181***		0.0915***	0.0981***
log(densité de population de l'AU en 2010)		0.21***		
Part ménages ayant emménagé depuis - de 2 ans			-0.119	
Part ménages ayant emménagé depuis 2 à 9 ans			1.85	
Part ménages ayant emménagé depuis 10 à 19 ans			11.4***	
Part ménages ayant emménagé depuis 20 à 29 ans			-9.2***	
Part de 15-29 ans				-20.2***
Part de 30-44 ans				22.5***
Part de 45-59 ans				7.24***
Part de plus de 60 ans				-2.19
R2 ajusté	0.36	0.21	0.74	0.57
Nombre d'observations	100	100	100	100

5.2.2 Revenus moyens

Figure 11 : Revenus moyens par UC et part de diplômés de l'enseignement supérieur dans les 100 plus grandes aires urbaines, 2010



D'après la figure 11 le prix des logements en 2010 est positivement corrélé avec le revenu fiscal moyen par unité de consommation et avec la part de diplômés de l'enseignement supérieur sur l'aire urbaine. Cette corrélation positive est assez forte à la fois sur l'ensemble de l'échantillon (0,56 pour le revenu) et sur les AU les plus peuplées. Ces deux variables sont fortement corrélées ensemble. La variable « revenu fiscal » est souvent prise comme proxy pour représenter la qualité des services publics offerts par les collectivités (étant donné le mode de financement d'un certain nombre de services par les impôts locaux). Ces deux variables capturent également les externalités positives de voisinage. Elles peuvent aussi capturer les différences de rentes foncières entre les AU. Comme le montrent les estimations ci-dessous, la variable « revenu par UC » explique plus de 30 % de la variabilité des prix sur l'ensemble de l'échantillon. Le paramètre estimé est positif et significatif dans tous les modèles suggérant que les aires urbaines les « plus riches », et donc les plus attractives sur le plan des services publics et des aménités de voisinage, présentent des prix plus élevés.

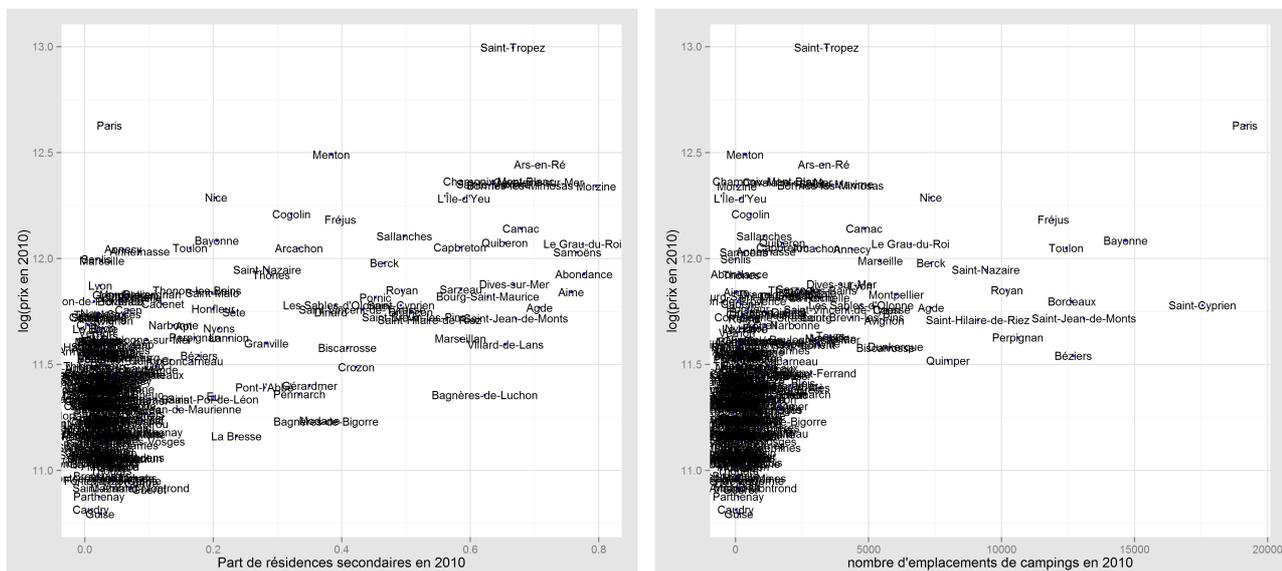
Les modèles du tableau 18 introduisent des indicateurs concernant la structure de la population par catégorie socioprofessionnelle. Seules certaines CSP sont introduites pour des raisons de multicolinéarité et la part de cadres qui est fortement corrélée au revenu par UC n'est pas introduite en même temps que ce dernier. Les estimations montrent que les proportions de cadres et de chefs d'entreprise ont un effet positif et significatif sur les prix alors que la proportion d'ouvriers et d'employés a un effet inverse. Ces résultats sont cohérents avec l'impact positif du revenu par UC sur les prix. La proportion d'agriculteurs sur l'aire urbaine a un effet négatif important sur les prix. On trouve ce résultat dans la littérature sur les prix fonciers qui suggère que le maintien de l'agriculture tend à modérer la hausse du prix foncier. Cet effet semble jouer ici également sur les prix de l'immobilier (cette variable peut aussi représenter le fait que l'aire urbaine est située à proximité de zones rurales et donc présente des prix plus faibles). L'introduction des variables « Revenu par unité de consommation » et « population » modifie peu ces effets et l'élasticité estimée du prix par rapport à la population est similaire aux estimations précédentes. L'élasticité du prix au revenu par UC est en revanche plus faible signalant la corrélation entre la proportion de la population par CSP et le revenu (la multicolinéarité a cependant été contrôlée et ne semble pas poser de problèmes).

Tableau 18 : Résultats d'estimations, Revenus par UC et structure de la population par CSP, ensemble des aires urbaines, 2010

Variable	(1)	(2)	(3)	(4)	(5)
Constante	-10.9***	10.7***	11.2***	0.617	-0.582
log(Revenu fiscal par UC moyen de l'AU en 2010)	2.28***			1.21***	1.18***
Part de diplômés de l'enseignement supérieur		4.18***			
Part d'agriculteurs			-16***	-15.8***	-15.8***
Part de cadres			5.79***		
Part de chef d'entreprises/artisans/commerçants			15.4***	11.9***	14.3***
Part d'employés/ouvriers			-1.22*	-2.66***	-1.28***
Part de retraités			-0.477	-1.6***	-0.599*
log(population de l'AU en 2010)					0.0626***
R2 ajusté	0.33	0.26	0.61	0.64	0.67
Nombre d'observations	314	314	314	314	314

5.2.3 Attractivité touristique et résidences secondaires

Figure 12: Prix de l'immobilier et attractivité touristique des aires urbaines, 2010



Les deux graphiques de la figure 12 illustrent la corrélation entre l'attractivité touristique d'une aire urbaine et le niveau de prix en 2010 à travers deux variables censées capturer cet effet : la part de résidences secondaires dans l'ensemble des logements en 2010 et le nombre d'emplacements de campings sur l'aire urbaine. On voit apparaître dans le quart Nord-est du premier graphique, un ensemble d'aires urbaines réputées pour leur attrait touristique et qui sont caractérisées par des prix élevés et une part importante de résidences secondaires. La corrélation positive entre la part de résidence secondaires et les prix sur l'ensemble de l'échantillon est forte et significative. Cette corrélation est moins marquée sur les 100 aires urbaines les plus

peuplées qui ont globalement des prix élevés avec des taux de résidences secondaires modérés, mais elle est revanche très forte sur les autres aires urbaines (0,75 environ). Paris fait ici à nouveau figure d'exception avec une part faible de résidences secondaires faible et des prix très élevés.

À taille d'aire urbaine donnée, l'attractivité touristique d'une aire urbaine explique une part importante de la variabilité des prix des logements (plus de 50 %). Les régressions du tableau 19 étudient l'effet des variables « part de résidences secondaires », « nombre de chambres d'hôtel » et « nombre d'emplacements de campings » sur l'ensemble des aires urbaines. Ces variables ont toutes un effet positif et significatif sur les prix suggérant un lien entre la cherté d'une aire urbaine et son attractivité touristique. La « part de résidences secondaires » explique une part importante de la variabilité des prix. L'effet de cette variable reste positif et significatif lorsque l'on ajoute les variables « population » et « revenu ». À noter que les trois variables « revenu », « population » et « résidences secondaires » à elles seules expliquent plus de 60 % de la variabilité des prix.

Tableau 19 : Résultats d'estimations, attractivité touristique des AU, ensemble des aires urbaines, 2010

Variable	(1)	(2)	(3)	(4)
Constante	11.3***	-2.36*	-8.02***	-9.81***
Part de résidences secondaires	1.11***	1.21***		
log(population de l'AU en 2010)		0.0972***	-0.0145	0.00808
log(Revenu fiscal par UC moyen de l'AU en 2010)		1.29***	1.99***	2.16***
Nombre d'emplacements de campings			5.19e-05***	
Nombre de chambres d'hôtel				3.47e-06***
R2 ajusté	0.37	0.66	0.49	0.35
Nombre d'observations	314	314	314	314

5.2.4 Parc de logements

Les régressions du tableau 20 étudient l'impact de variables décrivant le parc de logements et son occupation à l'échelle des aires urbaines. La part de logements vacants a un effet négatif sur les prix. Cette variable peut capturer différents effets notamment une faible attractivité de l'aire urbaine ou encore le faible dynamisme du marché de l'immobilier dans ces aires urbaines (difficultés pour vendre les logements). La part des résidences principales occupées par les propriétaires dans l'ensemble des logements a un fort pouvoir explicatif. Les aires urbaines dont la part de résidences principales occupées par les propriétaires est élevée présentent des prix significativement plus faibles. Ce résultat est difficile à interpréter et le sens de la causalité entre les niveaux de prix et la part de propriétaires n'est pas direct. D'autre part, la part de résidences principales occupées par les propriétaires est fortement corrélée négativement à la part de résidences secondaires. Il est donc possible que cette variable capture l'effet d'une attractivité touristique plus faible de certaines aires urbaines. Les aires urbaines ayant un parc HLM important ont des prix significativement plus faibles. Le pouvoir explicatif de la variable « Part de résidences principales HLM » est cependant faible. Lorsque que les variables sur la composition du parc de logement sont introduites simultanément, le signe des paramètres reste le même et le modèle explique 76 % de la variabilité des prix entre les aires urbaines.

Tableau 20 : Résultats d'estimations, parc de logement, ensemble des aires urbaines, 2010

Variable	(1)	(2)	(3)	(4)
Constante	-1.36	-4***	-7.83***	0.411
log(population de l'AU en 2010)	0.0219**	0.0803***	0.0394***	0.0738***
log(Revenu fiscal par UC moyen de l'AU en 2010)	1.34***	1.59***	1.94***	1.16***
Part de logements vacants	-7.56***			-4.12***
Part de résidences principales occupées par propriétaires		-2.13***		-1.65***
Part de résidences principales HLM			-1.21***	-0.355**
R2 ajusté	0.58	0.69	0.37	0.76
Nombre d'observations	314	314	314	314

5.2.5 Régressions sur l'ensemble des groupes de variables

Tableau 21 : Résultats d'estimations, modèles finaux, ensemble des aires urbaines, 2010

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Constante	-10.7***	-2.36*	0.411	1.62	3.57***	11.5***
log(population de l'AU en 2010)	0.0191	0.0972***	0.0738***	0.075***	0.0609***	0.0729***
log(Revenu fiscal par UC moyen de l'AU en 2010)	2.24***	1.29***	1.16***	0.939***	0.794***	
Part de résidences secondaires		1.21***		0.52***	0.361***	0.401***
Part de logements vacants			-4.12***	-3.54***	-3.4***	-4***
Part de résidences principales occupées par propriétaires			-1.65***			
Part de résidences principales HLM			-0.355**			
Part de 15-29 ans				-12.4***	-10.7***	-11.6***
Part de 30-44 ans				10.3***	10.5***	11.7***
Part de retraités				-0.615***	-0.53***	-0.662***
Part ménages ayant emménagé depuis plus de 30 ans					-2.24***	-2.54***
Taux de chômage en 2010						-0.0137***
R2 ajusté	0.34	0.66	0.76	0.75	0.80	0.77
Nombre d'observations	314	314	314	314	314	314

Tableau 22 : Résultats d'estimations, modèles finaux, 100 plus grandes aires urbaines, 2010

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Constante	0.432	1.22	2.2	6.46***	6.81***	11.3***
log(population de l'AU en 2010)	0.149***	0.161***	0.0694***	0.101***	0.0933***	0.0956***
log(Revenu fiscal par UC moyen de l'AU en 2010)	0.944***	0.837***	1.06***	0.447***	0.433***	
Part de résidences secondaires		2.21***		2.13***	1.89***	2.07***
Part de logements vacants			-5.58***	-3.37***	-3.33***	-3.17***
Part de résidences principales occupées par propriétaires			-2.85***			
Part de résidences principales HLM			-0.923***			
Part de 15-29 ans				-7.92**	-9.44**	-10.7***
Part de 30-44 ans				20.7***	18.5***	20.4***
Part de retraités				-3.31***	-2.81***	-3.37***
Part ménages ayant emménagé depuis plus de 30 ans					-0.97*	-1.04*
Taux de chômage en 2010						-0.0132**
R2 ajusté	0.47	0.73	0.78	0.84	0.84	0.83
Nombre d'observations	100	100	100	100	100	100

Tableau 23 : Résultats d'estimations, modèles finaux, aires urbaines hors les 100 plus grandes, 2010

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Constante	-15.7***	-5.02**	-1.01	0.302	1.76	11.6***
log(population de l'AU en 2010)	-0.0663***	0.043**	0.0521***	0.0535***	0.0456***	0.0578***
log(Revenu fiscal par UC moyen de l'AU en 2010)	2.84***	1.62***	1.32***	1.1***	0.995***	
Part de résidences secondaires		1.04***		0.438***	0.296**	0.34***
Part de logements vacants			-4.18***	-3.72***	-3.43***	-4.39***
Part de résidences principales occupées par propriétaires			-1.47***			
Part de résidences principales HLM			-0.22			
Part de 15-29 ans				-11.4***	-9.43***	-10.3***
Part de 30-44 ans				9.29***	9.88***	11***
Part de retraités				-0.602**	-0.529**	-0.649***
Part ménages ayant emménagé depuis plus de 30 ans					-2.17***	-2.41***
Taux de chômage en 2010						-0.0126*
R2 ajusté	0.40	0.66	0.75	0.75	0.80	0.77
Nombre d'observations	214	214	214	214	214	214

Dans les régressions présentées dans les tableaux 21 à 23, on introduit progressivement des variables de chaque groupe qui sont ressorties comme significatives dans les analyses précédentes et qui expliquaient une part importante de la variabilité des prix. Plusieurs spécifications sont testées et on exclut les variables qui entraînent des problèmes de multicollinéarité. La multicollinéarité a été testée en examinant les *variance inflation factors* de chaque variable qui mesurent l'augmentation de la variance d'un coefficient du fait de la colinéarité. Les modèles présentés ici ne posent pas de problèmes de multicollinéarité selon ce critère.

Sur l'ensemble des aires urbaines, les coefficients estimés présentent des signes cohérents avec les estimations par groupe de variables et ils sont presque tous significatifs à un seuil de 1 %. L'élasticité du prix par rapport à la population sur l'ensemble des aires est relativement faible (0.02 à 0.10) ce qui s'explique par le fait que la population influe surtout sur les prix dans les aires urbaines les plus peuplées (élasticité entre 0.09 et 0.16). Le revenu par UC a un impact positif et significatif sur les prix dans tous les modèles estimés avec une valeur de l'élasticité du prix au revenu moyen plus forte dans les aires urbaines les moins peuplées. La part de résidences secondaires a un effet positif et significatif sur les prix quel que soit le sous-échantillon considéré et est une des variables qui expliquent le plus la variabilité des prix entre aires urbaines. Parmi les variables concernant le parc de logement, la part de logements vacants dans le parc est celle qui semble avoir le plus d'effet sur les prix. Parmi les variables concernant la structure de la population, la part des 15-29 ans, la part de retraités et la part de ménages ayant emménagé depuis plus de 30 ans ont toutes un effet négatif sur les prix. La part de 30-44 ans a quant à elle un positif sur les prix. Enfin, le taux de chômage exerce un effet dépréciatif sur les prix.

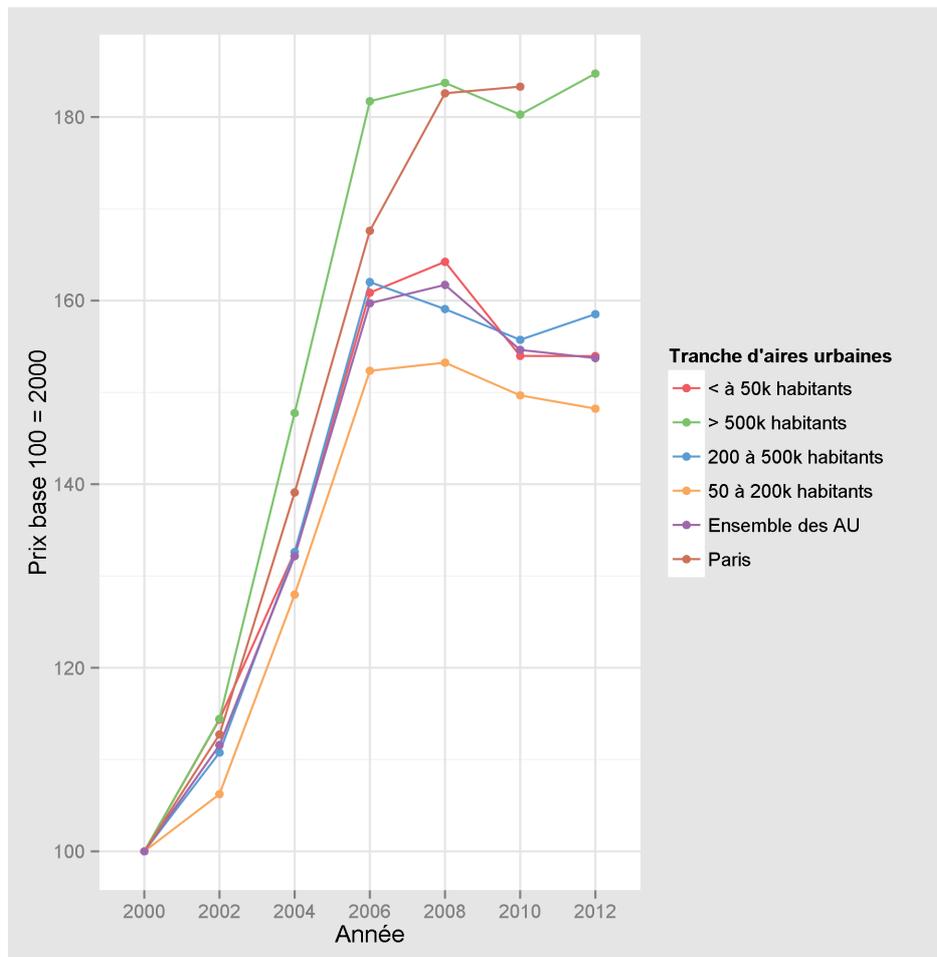
Ces résultats confirment ceux de l'approche par groupe de variables. Une part importante de la variabilité des prix entre les aires urbaines en 2010 s'explique par quelques variables socio-démographiques. Les R^2 des modèles incluant l'ensemble des variables sont élevés (entre 0,80 et 0,84). Parmi elles, la taille de l'aire urbaine, l'attractivité touristique et la richesse (qualité des services et voisinage) sont celles qui expliquent le plus les niveaux de prix élevés au sein des aires urbaines les plus attractives. Les villes les moins chères sont quant à elles caractérisées par une part plus importante de jeunes, de ménages ayant emménagé depuis de nombreuses années et un taux de vacance des logements élevé. Les mêmes régressions pour les années 2006 et 2008 présentées en annexes montrent des résultats similaires avec des signes et des valeurs de coefficients proches de ceux estimés pour 2010.

6 Analyse de l'évolution des prix

6.1 Analyse descriptive des évolutions de prix

L'évolution temporelle des prix au sein des aires urbaines montre une forte progression des prix entre 50 et 80 % de 2000 à 2012 (légère baisse entre 2008 et 2010) et ce sur l'ensemble des aires urbaines. Les coefficients de variations des prix entre aires urbaines sur la période (tableau 15) montrent que la variabilité des prix entre les aires urbaines s'accroît entre 2000 et 2012 suggérant que les phénomènes de « rattrapage » des prix immobiliers entre les aires urbaines sont minoritaires. Les écarts de prix entre les aires urbaines se creusent dans chacune des tranches d'aires urbaines.

Figure 13 : Evolution des prix de l'immobilier par tranche d'aires urbaines et sur Paris



La comparaison des cartes du niveau des prix en 2000 et de la croissance des prix entre 2000 et 2010 révèle qu’une grande partie des aires urbaines qui étaient caractérisées par des prix élevés en 2000 se classent également parmi celles qui ont subi une augmentation soutenue des prix. La plupart de ces aires urbaines sont aussi celles dont la population a augmenté le plus comme Toulouse ou Annemasse signalant que les prix déjà élevés en 1999 n’ont pas freiné l’attractivité de ces villes. Certaines aires urbaines comme Paris, Nice, Lille ou Valenciennes font exception avec une hausse modérée de leur population mais une forte augmentation des prix. À l’inverse, Rennes, Vannes et Montauban ont connu une hausse limitée des prix malgré une forte croissance de leur population.

Figure 14 : Prix immobiliers en 2000 et croissance des prix entre 2000 et 2010

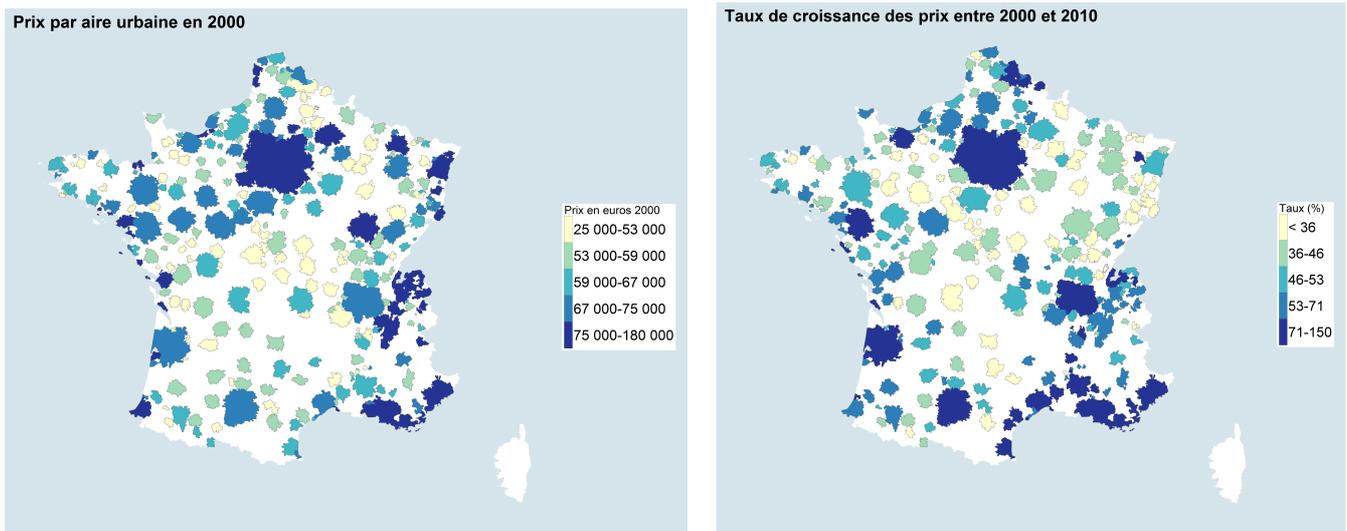
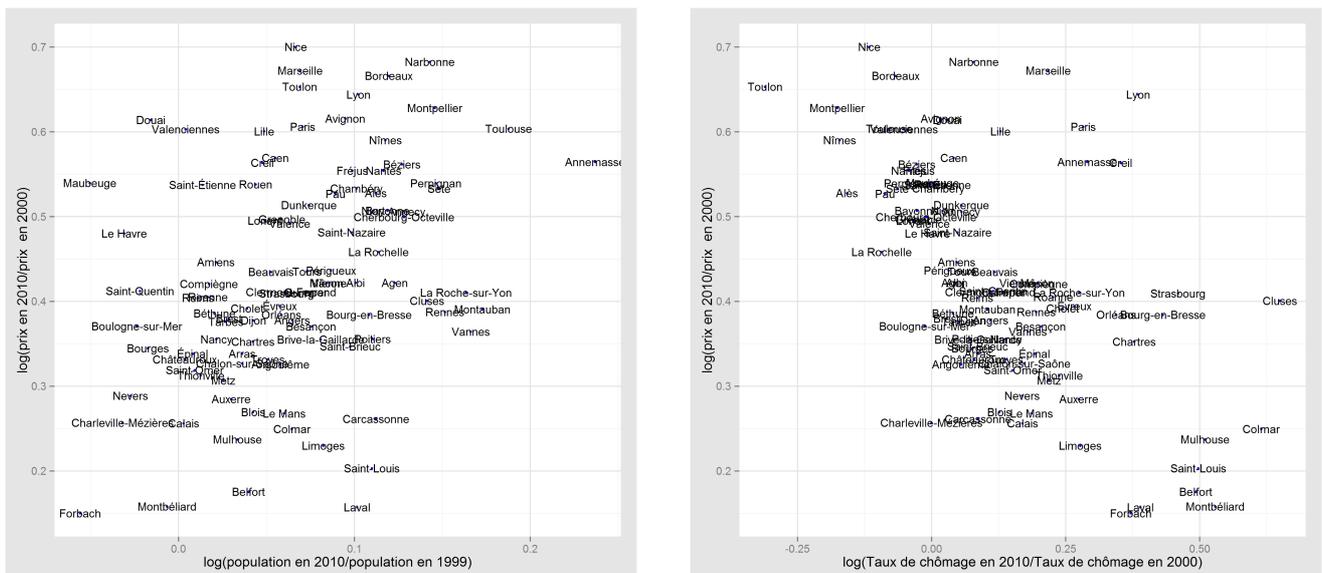


Figure 15 : Croissance des prix de l'immobilier et évolution de la population et du taux de chômage moyen dans les aires urbaines, 2010



D'autres facteurs semblent avoir également influencé les prix. La figure 15 montre par exemple que la croissance des prix de l'immobilier est négativement corrélée avec la croissance du taux de chômage suggérant que la hausse des prix de l'immobilier a été moindre dans les aires urbaines où le taux de chômage a augmenté de manière plus importante entre 2000 et 2010. Certaines villes font cependant exception, car elles ont connu une hausse importante des prix de l'immobilier et du taux de chômage simultanément (Paris, Lyon et Marseille notamment).

6.2 Analyse économétrique de l'évolution des prix

La hausse des prix de l'immobilier depuis 2000 a été généralisée à l'ensemble du territoire métropolitain. Les aires urbaines ont cependant connu des évolutions de prix différenciées et on cherche ici à identifier les facteurs pouvant expliquer l'évolution plus ou moins forte des prix dans certaines aires urbaines. L'analyse des niveaux de prix en coupe a montré que la population, la richesse et l'attractivité touristique des aires urbaines expliquaient une part importante des différences de prix entre grandes villes. On cherchera ici à déterminer si l'évolution temporelle de ces facteurs expliquent les différences d'évolution des prix entre aires urbaines. L'échantillon de données disponibles sur seulement 7 années ne permet pas de réaliser une analyse en économétrie des séries temporelles.

Nous nous limiterons ici à une analyse sur les variations de prix entre 2000 et les autres années de l'échantillon étant donné que l'on dispose de peu de variables explicatives qui varient à la fois dans le temps et entre aires urbaines (Revenu par UC et taux de chômage uniquement). La disponibilité future de données annuelles localisées sur l'offre de logement et la population pourrait permettre d'étendre cette analyse par une analyse en panel.

Les régressions du tableau 24 sont réalisées en utilisant comme variable dépendante le logarithme du rapport de prix de l'aire urbaine entre 2000 et 2010. Les variables explicatives correspondent au logarithme du rapport entre le niveau de la variable en 2000 (ou 1999 pour les données issues du recensement de la population) et son niveau en 2010. En procédant de cette manière on neutralise les variations de prix qui sont dues à des facteurs qui affectent l'ensemble des aires urbaines de manière homogène (déterminants macro-économiques, aides et conditions de financement appliquées à l'échelle nationale). On cherche donc bien ici les caractéristiques des aires urbaines qui expliquent les différences de variations de prix 2000-2010 entre les aires urbaines. Elle complète l'analyse sur le niveau des prix en étudiant notamment l'effet de l'évolution de la démographie, de l'emploi et de l'offre en logements sur les évolutions de prix. Le modèle estimé est :

$$\ln\left(\frac{P_{it_f}}{P_{it_d}}\right) = \beta_0 + \sum_k \beta_k \ln\left(\frac{x_{ikt_f}}{x_{ikt_d}}\right) + \epsilon_i$$

où t_f et t_d désignent respectivement l'année finale (2010) et l'année de départ (1999-2000) utilisées.

Les régressions du tableau 24 indiquent que la croissance du taux de chômage entre 2000 et 2010 a eu un impact négatif significatif sur l'évolution des prix. En moyenne, les aires urbaines qui ont connu une augmentation de leur taux de chômage 10 % plus importante que les autres ont subi une évolution des prix en moyenne de 2,4 % plus faible que les autres. Les variations de population et les variations du revenu moyen par UC ont quant à elles un effet positif significatif sur les évolutions de prix. Ces trois variables expliquent 35 % des différences de variations de prix entre aires urbaines de 2000 à 2010, le taux de chômage étant cette fois la variable qui explique la plus grande partie de cette variabilité (24 %). Les résultats sur les 100 plus grandes aires urbaines sont similaires et les coefficients estimés ont des valeurs proches (voir tableau 30 en annexes).

Les régressions 2, 3 et 4 du tableau 24 étudient l'impact de l'évolution du parc de logement et de son occupation sur l'évolution des prix. À revenu fixé, la croissance du parc de logements (nombre de logements et nombre de logement par habitant) a un effet négatif sur l'évolution des prix suggérant qu'une augmentation de l'offre en logement à revenu fixé des ménages limite l'augmentation des prix. L'élasticité du prix des logements par rapport à leur nombre est estimée à -0,4 sur l'ensemble des aires urbaines (-1,8 sur les 100 plus grandes aires urbaines). On identifie donc bien ici un effet négatif du développement de l'offre en logement sur les prix.

La croissance du nombre de logements vacants a bien un effet négatif sur les prix suggérant qu'une diminution de l'attractivité du parc de logements d'une aire urbaine modère également l'évolution des prix. Enfin, l'évolution du taux de résidences secondaires n'a pas d'effet sur l'évolution des prix. Ceci est probablement lié au fait qu'il y a eu peu d'évolution

du caractère « touristique » des aires urbaines sur la période 2000-2010.

La dernière régression du tableau a été réalisée en introduisant l'ensemble des variables précédentes et en sélectionnant les variables explicatives sur des critères d'information. La variation de la population et la variation du nombre de logements étant très corrélées, il est préférable d'utiliser le ratio nombre de logement par habitant dans cette estimation. Ces modèles montrent que les variables « Revenu par UC », « taux de chômage », « nombre de logement par habitant » et « part de logements vacants » ont un impact significatif sur l'évolution des prix. Elles expliquent 43 % de ces variations sur l'ensemble des aires urbaines et plus de 60 % sur les 100 plus grandes aires urbaines (voir tableau 30 en annexes). Ces R^2 moins élevés que dans l'analyse sur les niveaux de prix montrent que les évolutions de prix sont plus difficiles à expliquer que leur niveau. Cette analyse permet néanmoins de mettre en évidence certains des facteurs pouvant expliquer les évolutions différentes des prix d'une aire urbaine à l'autre.

Tableau 24 : Régressions sur les variations de prix entre 2000 et 2010, ensemble des aires urbaines

Variable	(1)	(2)	(3)	(4)	(5)
Constante	0.152**	0.173***	0.171***	0.24***	0.232***
log(revenu par UC 2010/revenu par UC 2000)	1.1***	1.17***	1.3***	1.05***	1.07***
log(population 2010/population 1999)	0.431***	0.621***			
log(taux de chômage 2010/taux de chômage 2000)	-0.237***	-0.227***	-0.228***	-0.196***	-0.199***
log(nombre de logements 2010/nombre de logements 1999)		-0.371**			
log(nombre de logements par habitant 2010/nombre de logements par habitant 1999)			-0.578***	-0.38***	-0.367***
log(Part de résidences secondaires/Part de résidences secondaires 1999)				0.017	
log(Part de logements vacants/Part de logements vacants 1999)				-0.151***	-0.154***
R2 ajusté	0.35	0.36	0.36	0.43	0.43
Nombre d'observations	314	314	314	314	314

Conclusion

Cette étude vise à déterminer et à quantifier les facteurs qui expliquent l'hétérogénéité observée sur les marchés du logement en France métropolitaine. L'analyse hédonique réalisée sur la majeure partie des aires urbaines métropolitaines montre que les caractéristiques des biens, leur localisation au sein de l'aire urbaine et les caractéristiques économiques et démographiques des aires urbaines expliquent une grande partie de cette hétérogénéité sur chaque année de l'échantillon 2000-2012 dont on dispose. À caractéristiques des biens données et pour une aire urbaine donnée, l'éloignement de la ville-centre a un effet négatif sur les prix dans la majeure partie des aires urbaines et notamment dans les plus grandes. Les prix hédoniques estimés pour la surface habitable et la distance au centre de l'aire urbaine montrent que la taille des aires urbaines semble jouer à la fois sur le consentement des ménages à payer pour un m² de surface habitable supplémentaire et sur leur consentement à s'éloigner de la ville-centre, les plus grandes aires urbaines étant caractérisées par des prix unitaires plus élevés et une forte valorisation de la ville-centre.

Les différences de prix entre aires urbaines sont étroitement liées aux caractéristiques propres de chaque ville comme leur population et sa structure, leur attractivité touristique et leur richesse qui témoigne de la qualité des services publics offerts par la ville. Les prix sont d'autant plus élevés que l'aire urbaine est peuplée et a connu une croissance démographique importante dans les années 2000. D'autres aires urbaines présentent des prix élevés du fait de leur forte attractivité touristique et de la proximité d'aménités positives (montagne, littoral) malgré une dynamique démographique modérée. Les villes les moins chères sont caractérisées par une part plus importante de jeunes, de ménages ayant emménagé depuis de nombreuses années et un taux de vacance des logements élevé.

Des différences importantes de prix existent entre aires urbaines et les écarts entre villes s'accroissent sur la période 2000-2012. Ces différences de variations de prix entre les aires urbaines s'expliquent en partie par des évolutions diverses de certains facteurs économiques, démographiques et par l'évolution du parc et de l'offre en logement. Les hausses de prix les plus importantes se concentrent dans les villes dont les prix étaient déjà élevés au début des années 2000 qui se classent toujours parmi les aires urbaines les plus attractives. D'autres aires urbaines ont connu une croissance soutenue des prix depuis 2000 du fait de leur dynamisme démographique. À l'inverse, la croissance du nombre de logement par habitant semble avoir limité la hausse des prix dans certaines aires urbaines.

Plusieurs voies d'amélioration de cette étude sont possibles. L'effet de la localisation des logements au sein d'une ville est pris en compte ici dans le cadre simplifié d'une aire urbaine monocentrique par l'intermédiaire de la distance au centre de l'aire urbaine. La prise en compte du polycentrisme des villes, soit par l'intermédiaire d'autres paramètres de distance (centres d'emploi secondaires, transports), soit par l'introduction d'indicateurs d'accessibilité pourrait permettre de renforcer l'analyse de l'impact de l'accessibilité des logements sur les prix.

La prise en compte des autres aménités qui influencent le prix des logements, sous forme de variables explicatives locales ou par l'utilisation de méthode d'économétrie spatiale, pourrait également rendre plus fiable l'estimation des paramètres liés à la localisation comme l'effet de la distance au pôle urbain. La géolocalisation des données notariales à la parcelle cadastrale et la disponibilité actuelle de nombreuses sources de données géolocalisées offrent de nombreuses perspectives sur ces aspects.

Enfin, l'étude proposée ici donne des indications sur le choix des ménages en termes de localisation à un instant donné ou sur une période donnée mais fournit peu d'informations sur le comportement des ménages face à un changement sur les marchés de l'immobilier. L'absence de données précises sur les caractéristiques des ménages dans les données rend difficile cette analyse qui est pourtant une extension possible du cadre hédonique utilisé ici. Une telle analyse nécessiterait cependant d'avoir accès à une source de données qui regroupent des informations précises à la fois sur le prix et les caractéristiques des biens mais aussi sur les ménages acquéreurs.

Annexes

1. Choix de la forme fonctionnelle des modèles hédoniques

Le tableau 25 présente des estimations réalisées avec différentes formes fonctionnelles sur les 100 plus grandes aires urbaines et avec des variables explicatives incluant caractéristiques des biens, indicatrices par aires urbaines et localisation. Le modèle 1 est celui utilisé dans la majeure partie de l'étude, le second introduit la distance au centre sous forme logarithmique, le modèle 3 utilise une forme linéaire pour la surface habitable et la distance, le modèle 4 introduit un polynôme de degré 4 de la surface habitable et le dernier utilise une transformation box-cox pour la variable dépendante à savoir le prix total du bien. Les résultats montrent que les R^2 sont équivalents pour toutes les formes fonctionnelles (moins élevé pour la forme linéaire) et que les valeurs et la significativité des paramètres autres que la distance au centre et la surface habitable sont peu modifiées (à l'exception du modèle box-cox, car la variable dépendante n'est plus la même dans ce cas). L'introduction d'une forme polynomiale pour la surface améliore sensiblement le R^2 par rapport au cas linéaire mais explique une part similaire de la variabilité par rapport à la forme logarithmique. Dans un souci de simplicité, nous préférons donc la forme fonctionnelle du modèle 1 (le 2 étant équivalent) d'autant plus que les paramètres estimés pour les transformations polynomiales et box cox sont moins faciles à interpréter.

Tableau 25 : Choix de la forme fonctionnelle, estimations sur les 100 plus grandes aires urbaines, 2010

variable	(1)	(2)	(3)	(4)	(5)
Constante	9.31***	9.4***	11.8***	11.2***	8.54***
nombre de sdb manquant	-0.0314***	-0.0248***	-0.0624***	-0.0384***	-0.0259***
nombre sdb >1	0.139***	0.139***	0.0616***	0.113***	0.106***
présence d'un parking	0.0516***	0.0622***	0.106***	0.0555***	0.0415***
période de construction manquante	-0.156***	-0.171***	-0.147***	-0.159***	-0.124***
construction avant 1913	-0.0408***	-0.0795***	-0.077***	-0.0493***	-0.0338***
construction 1914-1948	-0.173***	-0.18***	-0.181***	-0.176***	-0.137***
construction 1948-1969	-0.215***	-0.216***	-0.182***	-0.217***	-0.169***
construction 1970-1980	-0.199***	-0.204***	-0.183***	-0.205***	-0.157***
construction 1981-1991	-0.105***	-0.115***	-0.105***	-0.112***	-0.0831***
maison	0.0438***	0.041***	0.0489***	0.0195***	0.0352***
log(surface habitable)	0.786***	0.789***			0.619***
distance au centre de l'AU	-0.0162***		-0.0157***	-0.016***	-0.0127***
log(distance au centre de l'AU)		-0.148***			
surface habitable			0.0101***	0.0309***	
surface habitable ²				-0.00018***	
surface habitable ³				5.43e-07***	
surface habitable ⁴				-6.15e-10***	
moyenne des effets fixes	8.36	8.45	10.9	10.3	7.79
premier décile des effets fixes	8.03	8.09	10.5	9.92	7.53
dernier décile des effets fixes	8.7	8.79	11.2	10.6	8.06
R2 ajusté	0.71	0.70	0.66	0.71	0.71
Nombre d'observations	303 604	303 604	303 604	303 604	303 604
Nombre d'AU	100	100	100	100	100
Nombre de communes	8 877	8 877	8 877	8 877	8 877

2. Régressions en coupe complémentaires

Tableau 26 : Résultats d'estimations, modèles finaux, ensemble des aires urbaines, 2006

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Constante	-9.51***	-1.76	1.14	0.893	3.29**	10.9***
log(population de l'AU en 2006)	0.01	0.0868***	0.0713***	0.0709***	0.0599***	0.0758***
log(Revenu fiscal par UC moyen de l'AU en 2006)	2.15***	1.25***	1.1***	1.05***	0.882***	
Part de résidences secondaires		1.17***		0.998***	0.807***	0.845***
Part de logements vacants			-4.23***	-2.96***	-2.9***	-3.48***
Part de résidences principales occupées par propriétaires			-1.54***			
Part de résidences principales HLM			-0.716***			
Part de 15-29 ans				-0.65	-1.74*	-1.54
Part de 30-44 ans				0.124	0.298	3.24**
Part de retraités				-0.855	-0.959	-0.137
Part ménages ayant emménagé depuis plus de 30 ans					-2.42***	-2.73***
Taux de chômage en 2006						-0.00338
R2 ajusté	0.29	0.63	0.72	0.66	0.71	0.68
Nombre d'observations	314	314	314	314	314	314

Tableau 27 : Résultats d'estimations, modèles finaux, ensemble des aires urbaines, 2008

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Constante	-10.8***	-2.63*	0.392	0.236	2.61*	11.8***
log(population de l'AU en 2008)	0.00353	0.0832***	0.0637***	0.059***	0.0477***	0.0652***
log(Revenu fiscal par UC moyen de l'AU en 2008)	2.28***	1.34***	1.18***	1.16***	1***	
Part de résidences secondaires		1.21***		1.03***	0.824***	0.846***
Part de logements vacants			-4.45***	-3.24***	-3.27***	-4.07***
Part de résidences principales occupées par propriétaires			-1.58***			
Part de résidences principales HLM			-0.568***			
Part de 15-29 ans				-0.187	-1.38	-1.67
Part de 30-44 ans				-1.15	-1.39	1.04
Part de retraités				-1.12	-1.37**	-0.879
Part ménages ayant emménagé depuis plus de 30 ans					-2.26***	-2.62***
Taux de chômage en 2008						-0.00891
R2 ajusté	0.31	0.65	0.74	0.68	0.72	0.69
Nombre d'observations	314	314	314	314	314	314

3. Régressions complémentaires sur les évolutions de prix

Tableau 28 : Régressions sur les variations de prix entre 2000 et 2006, ensemble des aires urbaines

Variable	(1)	(2)	(3)	(4)	(5)
Constante	0.236***	0.248***	0.257***	0.301***	0.283***
log(revenu par UC 2006/revenu par UC 2000)	1.31***	1.37***	1.71***	1.38***	1.45***
log(population 2006/population 1999)	0.866***	1.04***			
log(taux de chômage 2006/taux de chômage 2000)	-0.192***	-0.188***	-0.198***	-0.175***	-0.177***
log(nombre de logements 2006/nombre de logements 1999)		-0.312			
log(nombre de logements par habitant 2006/nombre de logements par habitant 1999)			-0.912***	-0.746***	-0.706***
log(Part de résidences secondaires/Part de résidences secondaires 1999)				0.0424	
log(Part de logements vacants/Part de logements vacants 1999)				-0.099***	-0.108***
R2 ajusté	0.34	0.34	0.30	0.33	0.33
Nombre d'observations	314	314	314	314	314

Tableau 29 : Régressions sur les variations de prix entre 2000 et 2008, ensemble des aires urbaines

Variable	(1)	(2)	(3)	(4)	(5)
Constante	0.0951	0.117*	0.124*	0.201***	0.201***
log(revenu par UC 2008/revenu par UC 2000)	1.48***	1.54***	1.68***	1.4***	1.4***
log(population 2008/population 1999)	0.537***	0.772***			
log(taux de chômage 2008/taux de chômage 2000)	-0.204***	-0.195***	-0.205***	-0.172***	-0.172***
log(nombre de logements 2008/nombre de logements 1999)		-0.409**			
log(nombre de logements par habitant 2008/nombre de logements par habitant 1999)			-0.722***	-0.591***	-0.591***
log(Part de résidences secondaires/Part de résidences secondaires 1999)				0.0861**	0.0861**
log(Part de logements vacants/Part de logements vacants 1999)				-0.1***	-0.1***
R2 ajusté	0.32	0.33	0.32	0.37	0.37
Nombre d'observations	314	314	314	314	314

Tableau 30 : Régressions sur les variations de prix entre 2000 et 2010, 100 plus grandes aires urbaines

Variable	(1)	(2)	(3)	(4)	(5)
Constante	0.176	0.281**	0.284***	0.218**	0.226**
log(revenu par UC 2010/revenu par UC 2000)	1.05**	1.27***	1.24***	1.34***	1.32***
log(population 2010/population 1999)	0.514**	1.81***			
log(taux de chômage 2010/taux de chômage 2000)	-0.312***	-0.251***	-0.253***	-0.201***	-0.198***
log(nombre de logements 2010/nombre de logements 1999)		-1.84***			
log(nombre de logements par habitant 2010/nombre de logements par habitant 1999)			-1.82***	-0.933**	-0.939**
log(Part de résidences secondaires/Part de résidences secondaires 1999)				-0.0134	
log(Part de logements vacants/Part de logements vacants 1999)				-0.255***	-0.251***
R2 ajusté	0.42	0.54	0.55	0.60	0.61
Nombre d'observations	100	100	100	100	100

Bibliographie

Cavailhès Jean. Le prix des attributs du logement, 2005. *Economie et statistique*, 381 : 91-123.

Combes Pierre-Philippe, Duranton Gilles et Gobillon Laurent. Le prix des terrains en France : la localisation, encore la localisation, toujours la localisation, 2011. CGDD, *Le point sur* n°74.

Combes, Pierre-Philippe, Duranton, Gilles and Gobillon, Laurent. The costs of agglomeration: Land prices in French cities, 2012

Follain, James R and Jimenez, Emmanuel. Estimating the demand for housing characteristics: a survey and critique, 1985. *Regional science and urban economics*, 15 : 77-107.

Fauvet Laurent. Prix des logements anciens : les déterminants liés aux biens et la négociation, 2008. *Notes de synthèse du SESP* n°169.

Juillard Claire et Alexandre Hervé. La flambée immobilière dans les villes françaises, une question d'attractivité ?, 2012. *L'Observateur de l'immobilier du Crédit Foncier*, 84 : 34-47

Rosen, Sherwin. Hedonic prices and implicit markets: product differentiation in pure competition, 1974. *The journal of political economy*, 34-55

Sheppard, Stephen. Hedonic analysis of housing markets, 1999. *Handbook of regional and urban economics*, 3 :1595-1635

Commissariat général au développement durable

Service de l'économie, de l'évaluation et de l'intégration du développement durable

Tour Séquoia

92055 La Défense cedex

Tél : 01.40.81.21.22

Retrouvez cette publication sur le site :

<http://www.developpement-durable.gouv.fr/developpement-durable/>

Prix des logements : Quels facteurs expliquent leur disparité au sein et entre les aires urbaines françaises ?

Entre 2000 et 2012, le prix des logements a augmenté de 60 % hors inflation en France métropolitaine. Cette augmentation a été différenciée selon les localisations ce qui a renforcé les disparités entre zones urbaines.

Le marché du logement français est caractérisé par une forte hétérogénéité qui se décline à la fois dans la diversité des types de biens échangés et dans la variabilité spatiale des prix, autant entre grandes zones urbaines qu'au sein d'une même zone urbaine. L'existence de ces différentes dimensions affectant les prix de l'immobilier complexifie l'analyse des marchés de l'immobilier, notamment les comparaisons entre zones géographiques et l'étude de l'évolution temporelle des prix.

Dans ce contexte, cette étude analyse les disparités entre zones urbaines et leur évolution récente. Nous utilisons pour cela, la méthode des prix hédoniques pour différencier le prix de chaque composante du logement (caractéristiques propres mais aussi localisation) et comparer les niveaux de prix par aire urbaine toutes choses égales par ailleurs. Cette étude fournit une analyse quantitative sur les facteurs qui déterminent la disparité des prix entre aires urbaines mais aussi au sein d'une même aire urbaine. La géolocalisation précise des logements dans les données utilisées offrent des perspectives d'approfondissements sur l'effet de la localisation sur le prix des logements.

What determinants explain the variability in housing prices within and between French urban areas?

Housing prices increased by 60% between 2000 to 2012 in France. The magnitude of this price increase was different from one urban area to another which strengthened the disparities between urban areas.

The French housing market is highly heterogeneous. A large diversity of housing goods are traded and housing prices are characterized by an important spatial heterogeneity both between urban areas and inside each urban area. These various components affect housing prices simultaneously. This make difficult to analyse housing markets and to compare price levels between geographic areas.

In this study, we estimate hedonic price functions including both physical and locational attributes of housing goods in order to compare housing prices between urban areas. This study provides a quantitative analysis of determinants of the variability in prices between urban areas. Housing data used in this study are geolocated with a high degree of precision which may open opportunities for further studies of housing markets.



Dépôt légal : Mars 2015
ISSN : 2102 - 4723