

# Chapitre 5

## Pouvoir de négociation sur le marché des appartements en Corse : une analyse en classes latentes

Steven B. Caudill, Claudio Detotto et Dominique Prunetti

1

### 5.1 Introduction

Le marché du logement corse présente plusieurs caractéristiques distinctives. Celles-ci s'expliquent tout d'abord par la topographie de l'île qui entraîne une segmentation du marché du logement. La Corse peut en effet être décrite comme une chaîne de montagnes entourée d'un anneau de plages de sable. Ces plages attirent l'intérêt de ceux qui, à l'intérieur et à l'extérieur de la Corse, recherchent des propriétés de vacances et de bord de mer. L'intérieur, d'une grande beauté naturelle, est un lieu propice aux randonnées et autres activités d'éco-tourisme.

Ce marché du logement corse est très vraisemblablement caractérisé par des asymétries d'information dans au moins deux dimensions. Tout d'abord, et de manière naturelle, les asymétries d'information existent et augmentent avec la distance entre l'acteur du marché et le lieu de la propriété. Mais, par ailleurs, des asymétries existent également en raison de barrières linguistiques et culturelles, les transactions immobilières se faisant en français et étant soumises au système juridique français. Les participants au marché qui ne sont pas familiers avec la langue et les coutumes juridiques françaises sont désavantagés par rapport aux participants qui parlent couramment français.

---

DOI : <https://doi.org/10.58110/estate-ea97>

1. Une version antérieure de ce chapitre est parue sous forme d'article : Caudill, S., Detotto, C., Prunetti, D., 2020. « Bargaining power in apartment sales in Corsica : A latent class approach », *Urban Studies*, Vol 57, Issue 13, 2754-2772 (<https://doi.org/10.1177/004209801988>).

Ces caractéristiques font du marché du logement corse un champ d'étude intéressant à examiner et qui pose des défis empiriques particuliers.

Des études précédentes sur les marchés fonciers et immobiliers corses ont porté soit sur les déterminants du marché des terres agricoles (Tafani et al., 2012; Beaumais et al., 2021), soit l'analyse de la tendance des prix (Prunetti et al., 2015). Notre examen du marché des appartements s'appuie sur un ensemble de données unique contenant des informations sur plus de 8 000 transactions d'appartements en Corse pour une période s'étendant de 2006 à 2016. En plus des informations sur les caractéristiques des appartements, notre ensemble de données contient des informations sur les caractéristiques des acheteurs et des vendeurs. Suivant l'approche développée par Harding et al. (2003) (ci-après notée *HRS*), nous utilisons ces informations sur les caractéristiques des acheteurs et des vendeurs afin de déterminer le pouvoir de négociation relatif des deux groupes. La détermination du pouvoir de négociation relatif constitue une question importante car, comme le notent Harding, Rosenthal et Sirmans, lorsque les marchés deviennent de plus en plus étroits, comme c'est le cas pour le logement, les prix sont influencés non seulement par les caractéristiques du produit mais aussi par les caractéristiques des acheteurs et des vendeurs et leur capacité de négociation. En outre, les acheteurs sont souvent intéressés par un petit groupe de propriétés similaires, de sorte qu'il peut y avoir une segmentation sur ces marchés, et que cette segmentation peut s'étendre au pouvoir de négociation des acheteurs et des vendeurs. Nous examinons cette possibilité en proposant une extension de l'approche de *HRS* consistant à estimer un modèle de classe latente permettant, tout d'abord, de mettre en évidence l'existence de segments du marché du logement et de les identifier et, ensuite, de mesurer le pouvoir de négociation des participants au marché sur ces segments.

En utilisant l'approche par classes latentes, nous parvenons à démontrer l'existence de deux segments. Dans le premier segment, nous identifions un pouvoir de négociation accru pour les résidents de la commune dans laquelle se situe le bien échangé et un faible pouvoir de négociation pour les étrangers. Dans le second segment, nous constatons une augmentation du pouvoir de négociation des résidents de la même commune, des autres résidents corses et des participants au marché étrangers. Les résultats de notre régression auxiliaire, ainsi que l'examen des moyennes pondérées des variables indépendantes, indiquent que les appartements du segment 1 sont plus susceptibles d'être ruraux et plus éloignés de la côte ainsi que de l'une des deux principales villes de l'île. D'autre part, les résidents communaux, les résidents corses et les acteurs du marché étrangers ont un fort pouvoir de négociation dans le segment 2.

## 5.2 Revue de la littérature

Les modèles de régression hédoniques sont utilisés depuis de nombreuses années dans les études de l'impact des caractéristiques des logements sur les prix de vente. Ces prix hédoniques sont souvent interprétés comme des prix d'équilibre du marché concurrentiel, mais, récemment, de nombreux auteurs ont reconnu que concernant les

marchés du logement, en particulier, cette interprétation pourrait être erronée. La raison réside dans le fait que les marchés du logement ne sont pas un lieu d'échange de produits homogènes. En effet, tant les propriétés échangées que les participants sur ce marché sont fortement hétérogènes, ce qui donne lieu à des marchés étroits. Lorsque les marchés sont étroits, il est probable que différents acheteurs puissent payer des prix sensiblement différents pour des propriétés très similaires. Cela est dû à la présence d'une asymétrie d'information, le plus souvent associée à des coûts de recherche différentiels pour les acheteurs et les vendeurs et à des différences en termes de préférences entre les participants au marché. La dispersion des prix observée peut également être due à certaines caractéristiques distinctives de la propriété échangée qui conduisent à un « amoindrissement » de l'intérêt des acheteurs. De nombreuses études dans le domaine du logement et, plus généralement, de l'immobilier, ont examiné les causes et les conséquences de cette dispersion des prix. Dans cette section, nous passons en revue les recherches empiriques sur les facteurs liés à la dispersion des prix sur les marchés du logement, de l'immobilier et sur d'autres marchés.

### 5.2.1 Caractéristiques des acheteurs

Turnbull et Sirmans (1993) suggèrent que les différences de prix pour des propriétés similaires peuvent être dues à des niveaux d'information différents, éventuellement imputables à des différences dans les coûts de recherche. Leurs résultats empiriques ne révèlent aucune différence systématique de prix entre les différents types d'acheteurs. Song (1995) constate que les résultats de la négociation sont affectés par le prix demandé par le vendeur et que les acheteurs d'une première maison ne négocient pas moins que les acheteurs d'une deuxième maison. Harding et al. (2003) examinent le pouvoir de négociation relatif des acheteurs et des vendeurs dans un modèle de régression hédonique. Les auteurs de cet article constatent que la richesse, le sexe et la présence d'enfants d'âge scolaire expliquent une partie des différences en terme de pouvoir de négociation. Brasington et Sarama Jr (2008) constatent que le type d'acte de propriété influe sur le prix de vente, les actes de propriété plus complexes entraînant des prix plus bas. Ihlanfeldt et Mayock (2012) constatent que les acheteurs paient des prix différents pour des maisons presque identiques et suggèrent que cette observation peut être imputable à des différences en terme de pouvoir de négociation relatif des acheteurs. Ils testent l'hypothèse selon laquelle le pouvoir de négociation relatif des acheteurs venant de l'extérieur du marché local est relativement faible en raison de coûts de recherche élevés et valident cette hypothèse à travers un examen des transactions de maisons individuelles en Floride. En utilisant des données portant sur la Chine, Zhou et al. (2015) constatent également que les acheteurs extérieurs au marché local paient des prix plus élevés et observent une stabilisation des prix. Chandra et al. (2017) examinent les différences de prix des voitures neuves et constatent que les différences de caractéristiques démographiques expliquent au moins 20 % de la variation observée des prix. Bayer et al. (2017) examinent les prix relatifs payés par les minorités par rapport aux caucasiens pour un logement similaire et constatent que les acheteurs noirs

et hispaniques paient environ deux pour cent de plus pour l'achat de leur maison. Holmes et Xie (2018) constatent que les acheteurs provenant hors de l'État paient 20% de plus que les acheteurs locaux. Considérées dans leur ensemble, ces études mettent en évidence une variation considérable des prix sur les marchés étroits. Cette différence est généralement imputable à des asymétries d'information liées à la distance ou à des différences dans les préférences des acheteurs.

### 5.2.2 Caractéristiques des vendeurs

Cottleer et al. (2008) constatent que les vendeurs de terres agricoles disposent d'un pouvoir de marché car les acheteurs s'efforcent d'acheter des parcelles de terre adjacentes. Hardin III et al. (2009) évaluent l'impact du régime de propriété et de la structure de gestion sur la performance des biens immobiliers et constatent que les propriétés multifamiliales gérées par des sociétés d'investissement immobilier<sup>2</sup> génèrent des loyers effectifs plus élevés que les propriétés qui ne sont pas gérées par de telles structures. Ces résultats impliquent que la structure de propriété peut avoir un impact sur la performance des biens immobiliers. En utilisant des données de ventes répétées, Anglin et Wiebe (2013) démontrent de manière empirique qu'un seul petit vendeur peut influencer le prix de vente de sa maison et que l'effet est plus important que prévu. Larsen et Coleman (2014) examinent le pouvoir de négociation des seniors sur le marché des maisons individuelles. Les résultats qu'ils obtiennent indiquent que les seniors ne subissent pas un différentiel de prix en cas d'achat, mais qu'ils vendent leur maison à un prix inférieur de 6% en moyenne à celui des autres vendeurs. Il est ainsi clair que les caractéristiques des vendeurs peuvent influencer sur les prix sur les marchés étroits.

### 5.2.3 Autres résultats empiriques

Pope (2008) suggère que l'asymétrie d'information entre acheteurs et vendeurs peut affecter les prix hédoniques. Il confirme cette idée en examinant les informations fournies par les vendeurs sur les zones inondables. Il constate une baisse de 4% des prix de vente après la divulgation de ces informations. Colwell et Munneke (2006) procèdent à une légère modification du modèle *HRS* et observent des différences systématiques dans le pouvoir de négociation et la catégorie de biens pour certains groupes d'acheteurs. Pour compléter cette revue de littérature, l'approche *HRS* étant la base de notre analyse, elle est examinée de manière séparée et en détail dans la section suivante.

## 5.3 Le modèle de négociation de *HRS*

*HRS* ont modifié le modèle de régression hédonique standard afin d'incorporer des effets de négociation concernant les acheteurs et les vendeurs. Leur cadre général est le

---

2. *Real Estate Investment Trusts* (REITS).

suivant :

$$P_i = sC_i + B_i \quad (5.1)$$

Où  $P_i$  représente le prix du bien foncier ou immobilier échangé,  $C_i$  les caractéristiques du bien,  $s$  les prix hédoniques qui sont influencés par les acheteurs et les vendeurs et  $B$  l'impact de la négociation sur la fonction de prix hédonique.

Afin de rendre le modèle opérationnel, c'est-à-dire d'estimer l'effet du marchandage,  $HRS$  intègrent les caractéristiques de l'acheteur et du vendeur dans la relation hédonique.  $HRS$  représentent le marchandage,  $B$ , comme une fonction des caractéristiques de l'acheteur et du vendeur,  $D^{sell}$  et  $D^{buy}$ , les impacts marginaux de ces caractéristiques sur le marchandage, donnés par  $b^{sell}$  et  $b^{buy}$ , et un terme d'erreur capturant les différences idiosyncratiques du pouvoir de marchandage entre les acheteurs et les vendeurs  $e_B$  :

$$B = b^{sell} D^{sell} + b^{buy} D^{buy} + e_B \quad (5.2)$$

En substituant (5.2) dans (5.1), on obtient le modèle de régression hédonique suivant qui incorpore l'impact des caractéristiques de l'acheteur et du vendeur sur la négociation :

$$P = sC + b^{sell} D^{sell} + b^{buy} D^{buy} + e_B \quad (5.3)$$

$HRS$  supposent que les caractéristiques du logement,  $C$ , sont connues de tous les acteurs du marché mais ne sont que partiellement observées par l'observateur. Ainsi, les caractéristiques des logements sont divisées en deux groupes. Le premier groupe,  $C_1$ , correspond aux caractéristiques qui sont accessibles pour l'observateur et le second groupe,  $C_2$ , correspond à celles non observées par l'observateur. De plus,  $HRS$  supposent que ces caractéristiques inconnues  $C_2$  sont évaluées par les acheteurs et les vendeurs, ce qui entraîne que le groupe  $C_2$  est corrélé avec  $D^{sell}$  et  $D^{buy}$  selon la relation suivante :

$$s_2 C_2 = d^{sell} D^{sell} + d^{buy} D^{buy} + e_D \quad (5.4)$$

Où  $s_2$  est le vecteur des prix implicites sur  $C_2$ . Ne pas disposer d'informations sur ces caractéristiques omises conduit à un problème de variables omises qui biaise les estimations des coefficients décrivant le pouvoir de négociation dans l'équation (5.3) ci-dessus.  $HRS$  affirment que surmonter ce problème de variables omises est la clé pour mesurer les effets de la négociation. Leur solution consiste à substituer l'équation (5.4) dans l'équation (5.3) afin d'obtenir la relation suivante :

$$P = s_1 C_1 + (b^{sell} + d^{sell}) D^{sell} + (b^{buy} + d^{buy}) D^{buy} + \epsilon \quad (5.5)$$

L'ensemble des termes de l'équation (5.5) ont été précédemment définis à l'exception de  $s_1$ , vecteur des prix implicites sur  $C_1$ , et du terme d'erreur aléatoire  $\epsilon = e_B + e_D$ . Si les expressions entre parenthèses de l'équation (5.5) sont écrites de la manière suivante, un problème d'identification apparaît de manière évidente :

$$\Omega^{sell} = b^{sell} + d^{sell} \quad (5.6)$$

$$\Omega^{buy} = b^{buy} + d^{buy} \quad (5.7)$$

En effet, les coefficients des caractéristiques individuelles ne sont pas identifiés, à moins que certaines restrictions portant sur les paramètres soient imposées dans l'estimation.

Afin de résoudre ce problème, *HRS* font deux hypothèses qui entraînent des restrictions de paramètres sur le modèle. La première hypothèse, qu'ils appellent *pouvoir de négociation symétrique*, implique que  $b \equiv b^{sell} = -b^{buy}$ . Comme son nom l'indique, cette contrainte implique que les acheteurs et les vendeurs ont un pouvoir de négociation égal. La deuxième contrainte qu'ils imposent à l'estimation est appelée *demande symétrique*. Elle implique que  $d \equiv d^{sell} = d^{buy}$  et signifie que les caractéristiques du vendeur et de l'acheteur ont le même impact sur la valeur des caractéristiques inconnues du bien.

En imposant ces contraintes, *HRS* obtiennent une équation qui est facilement estimable par la méthode des moindres carrés ordinaires (MCO) :

$$P = s_1 C_1 + b(D^{sell} - D^{buy}) + d(D^{sell} + D^{buy}) + \epsilon \quad (5.8)$$

Dans ce cadre,  $b$  fournit une mesure directe de l'effet des caractéristiques individuelles du vendeur et de l'acheteur sur le pouvoir de négociation<sup>3</sup>.

Dans ce chapitre, nous allons présenter une extension du modèle *HRS* que nous avons développée afin de pouvoir utiliser cette approche en intégrant une approche en classes latentes ou en mélange fini. Les développements de Dempster et al. (1977) ont fourni la base de l'estimation par maximum de vraisemblance de ces modèles et, grâce aux améliorations en termes de vitesse et de capacité des ordinateurs, les mélanges finis de modèles de régression sont désormais largement utilisés<sup>4</sup>. Cependant, ils n'ont pas été utilisés de manière intensive dans la littérature portant sur le logement et l'immobilier. En fait, une recherche sur la littérature du domaine ne nous a permis d'identifier que trois applications. Ainsi, Belasco et al. (2012) ont utilisé un mélange fini de modèles de régression pour localiser des sous-marchés du logement inobservables. Les travaux de Lu et al. (2015) ont utilisé un mélange fini de modèles de régression pour identifier les sous-marchés latents des quartiers à croissance intelligente. Ils déterminent quatre classes distinctes mais inobservables d'individus sur le marché du logement. Enfin, plus récemment, l'article d'Ibraimovic et Hess (2018) a utilisé un modèle de mélange fini pour identifier trois classes latentes démontrant l'hétérogénéité des choix résidentiels pour les personnes vivant à Lugano, en Suisse. Cet article enrichit cette littérature en étendant le modèle de *HRS* à un modèle de classes latentes ou de mélange fini. L'utilisation d'un mélange fini ou d'une approche par classes latentes va, dans le cadre de l'approche développée dans ce chapitre, nous permettre de déterminer s'il existe en Corse des sous-populations identifiables caractérisées par différents modèles de régression hédonique,

3. Dans la littérature utilisant les caractéristiques de l'acheteur et du vendeur, les études de Colwell et Munneke (2006) et de Steegmans et Hassink (2017) utilisent le cadre fourni par *HRS* et sont donc en mesure de séparer les effets. Les études incluant les caractéristiques de l'acheteur et/ou du vendeur mais n'utilisant pas le cadre fourni par *HRS* incluent Song (1995), Ihlanfeldt et Mayock (2012), Daneshvary et Claretie (2013), Larsen et Coleman (2014), Carrillo et al. (2015), Bø (2018), Hayunga (2018) et Holmes et Xie (2018).

4. Une application précoce en Economie est constituée par le travail de Beard et al. (1991). Ces auteurs ont estimé un modèle de mélange fini de fonctions de coût statistiques qui diffèrent en raison de différences inobservables en termes de technologie.

tout en nous permettant, éventuellement, d'identifier des modèles et des degrés différents de pouvoir de négociation<sup>5</sup>.

## 5.4 Approche en classes latentes

Dans cette section, nous présentons une description de l'approche par classes latentes. Pour commencer, supposons que nous ayons un échantillon aléatoire  $y_1, \dots, y_n$  d'une variable aléatoire avec la fonction de densité suivante :

$$f(y_i|\theta_1, \dots, \theta_K, \lambda_1, \dots, \lambda_{K-1}) = \sum_{k=1}^K \lambda_k f_k(y_i|\theta_k) \quad (5.9)$$

$$0 \leq \lambda_k \leq 1 \text{ et } \lambda_k = 1 - \sum_{k=1}^{K-1} \lambda_k \quad (5.10)$$

Dans laquelle  $K$  est le nombre de composants de la fonction de densité<sup>6</sup>. La densité mélange,  $f$ , est donc une combinaison linéaire pondérée des fonctions de densité des composants  $K$ , la somme des poids de chaque composant  $\lambda_k$ , étant égaux à un. Le nombre de composants des fonctions de densité,  $K$ , doit être choisi avant l'estimation, bien que des modèles basés sur différentes valeurs de  $K$  puissent être estimés et que des critères d'information puissent être utilisés pour déterminer la meilleure valeur de  $K$  et donc le modèle le plus approprié. L'hypothèse habituelle est que les fonctions de densité des composantes sont de la même famille (suivent toutes une loi normale, par exemple), mais ce n'est pas impératif. Pour opérationnaliser le modèle, il convient de sélectionner une forme fonctionnelle pour les fonctions de densité des composants,  $f_k$ , ainsi que le nombre de fonctions de densité des composants dans le mélange,  $K$ . Pour démontrer la mise en œuvre et l'intérêt de l'estimation d'un modèle mélange, commençons par le cas habituel de l'estimation d'un modèle de régression unique issu des informations d'un échantillon. Considérons le modèle suivant :

$$y_i = X_i\beta + \epsilon_i \quad (5.11)$$

Dans lequel  $y_i$  correspond à la variable dépendante,  $X_i$  à un vecteur de variables indépendantes et  $\beta$  à un vecteur de paramètres à estimer. Supposons également que

---

5. Nos résultats d'estimation des classes latentes sont obtenus à l'aide de la procédure PROC FMM du progiciel statistique SAS. Nous supposons un mélange de régression à deux composantes avec des erreurs normales pour chacune d'elles indépendantes les unes des autres. L'hypothèse de normalité fait des MCO à régime unique un cas particulier. Lors de l'estimation, nous obtenons deux ensembles complets de coefficients de régression, deux estimations de la variance résiduelle et une estimation du poids de mélange qui indique quelle fraction des observations de l'échantillon est associée à chaque régime. Après l'estimation, nous disposons de la probabilité postérieure que chaque observation soit associée à chaque régime. Les poids de mélange sont les moyennes de ces probabilités postérieures.

6. Cette discussion est basée sur les présentations sur le sujet de Frühwirth-Schnatter (2006) et Quant (1988).

les termes d'erreur aléatoires sont normalement distribués avec une moyenne nulle et une variance  $\sigma^2$ . Avec l'hypothèse de normalité, le modèle est facilement estimé par la méthode du maximum de vraisemblance. L'hypothèse implicite lors de l'estimation de ce modèle est que la population dans son ensemble est caractérisée par une seule relation. Cette hypothèse n'est pas toujours correcte, mais elle est relativement facile à vérifier.

Supposons, à présent, que nous considérons que notre population est caractérisée par deux sous-populations distinctes mais inobservables, chacune représentée par des modèles de régression différents. Autrement dit, au lieu d'un échantillon provenant d'une seule population, notre ensemble de données est obtenu à partir de deux sous-populations représentées chacune par des modèles de régression différents s'écrivant :

$$y_i = X_i\beta_1 + \epsilon_{i1} \text{ avec une probabilité } \lambda \quad (5.12)$$

$$y_i = X_i\beta_2 + \epsilon_{i2} \text{ avec une probabilité } 1 - \lambda \quad (5.13)$$

Le premier modèle de régression représente une part  $\lambda$  des données et le second modèle de régression représente la part restante  $1 - \lambda$ .

Il y a de nombreuses occasions pour lesquelles les populations sont caractérisées par plus d'une relation de régression. Par exemple, si les groupes ou sous-populations sont identifiés dans l'échantillon par une variable muette indiquant, par exemple, *homme* ou *femme*, l'échantillon peut être séparé par sexe, puis des modèles de régression peuvent être estimés pour chaque groupe. Un grand avantage de l'estimation d'un modèle de mélange fini réside dans le fait que la procédure d'estimation sépare de manière probabiliste les observations en groupes. En d'autres termes, la procédure détermine le meilleur regroupement et la variable muette (le sexe dans l'exemple ci-dessus) n'est pas nécessaire.

Un mélange de régressions à deux composants est basé sur l'idée que notre échantillon contient des observations provenant de deux sous-populations distinctes qui sont caractérisées par des paramètres de régression différents : soit  $\beta_1 \neq \beta_2$ , soit  $\sigma_1 \neq \sigma_2$ , soit les deux. Les critères d'information nous permettent de déterminer si deux modèles de régression sont validés ou non par les données. Les fonctions de densité de nos deux modèles de régression avec erreurs normales sont données par les équations suivantes :

$$f_1(y_i) = \frac{1}{\sqrt{(2\pi\sigma_1^2)}} \exp \left\{ \frac{-(y_i - X_i\beta_1)^2}{2\sigma_1^2} \right\} \quad (5.14)$$

$$f_2(y_i) = \frac{1}{\sqrt{(2\pi\sigma_2^2)}} \exp \left\{ \frac{-(y_i - X_i\beta_2)^2}{2\sigma_2^2} \right\} \quad (5.15)$$

Avec l'ajout du poids de mélange  $\lambda$ , la fonction de densité pour un mélange à deux composants du modèle de régression normale est donnée par :

$$f(y_i|X_i\beta_1, X_i\beta_2, \sigma_1, \sigma_2, \lambda) = \lambda f_1(y_i|X_i\beta_1, \sigma_1) + (1 - \lambda)f_2(y_i|X_i\beta_2, \sigma_2) \quad (5.16)$$

Nous pouvons maximiser la fonction de vraisemblance basée sur la densité du mélange de l'équation (5.16) pour les valeurs de  $\beta_1$ ,  $\beta_2$ ,  $\sigma_1$ ,  $\sigma_2$ , et  $\lambda$ . La procédure donnera des estimations du maximum de vraisemblance pour tous les paramètres, y compris  $\lambda$ , qui est la fraction des observations associées à chaque sous-population.

Après l'estimation, nous pouvons utiliser les estimations des paramètres pour calculer les probabilités futures que chaque observation appartienne à chaque groupe. Cette probabilité est donnée par :

$$Pr(\text{Groupe } k|y_i) = \frac{\lambda_k f_k(y_i|\theta_k)}{\sum_{k=1}^K \lambda_k f_k(y_i|\theta_k)} \quad (5.17)$$

Ces estimations de probabilité dépendent des paramètres non connus du modèle et ne sont donc connues qu'après l'estimation du modèle. Une fois ces probabilités obtenues, elles peuvent être utilisées comme variable dépendante dans un modèle de régression auxiliaire pour étudier les facteurs corrélés avec la probabilité d'appartenance à un groupe.

Dans la section suivante, nous décrivons notre base de données ainsi que les définitions des variables utilisées dans nos modèles de régression du pouvoir de négociation en Corse.

## 5.5 La base de données et les variables

### 5.5.1 La base de données

Notre base de données est constituée de tous les actes de vente d'appartements en Corse entre 2006 et 2016. La principale source de ces informations est la base de données PERVAL<sup>7</sup>. Cette base de données, produite par la Chambre des Notaires, contient des informations détaillées sur les transactions immobilières en France, notamment le prix d'achat, la date de signature du contrat et plusieurs caractéristiques du bien. Toutes les observations sont géo-référencées et des informations sur les caractéristiques de l'acheteur et du vendeur sont disponibles, notamment la nationalité de l'acheteur et du vendeur ainsi que le lieu de résidence enregistré. Les coordonnées spatiales des propriétés nous permettent d'estimer les temps de trajet entre chaque appartement et diverses commodités du voisinage, comme les services (médecins, pharmacies et écoles) et les biens environnementaux (vue sur la mer, plages et centre-ville). Notre base de données finale comprend des informations sur 8259 ventes d'appartements.

---

7. Source des données au lien suivant : <https://www.perval.fr>.

### 5.5.2 Les variables

Notre variable dépendante est le logarithme du prix de vente de l'appartement. Nos variables indépendantes comprennent l'*Intermédiation* qui est une variable muette pour laquelle une valeur de un indique la présence d'un intermédiaire dans la transaction. *Hypothèque* est une variable muette indiquant la présence d'une hypothèque sur la propriété. *Meublé* est une variable fictive indiquant si le logement est meublé. *Viager* indique la présence d'un prêt viager hypothécaire. *Pièces* est le nombre de pièces dans l'appartement. *Salle de bain* est le nombre de salles de bain dans l'appartement. *Etages* est l'étage sur lequel l'appartement est situé. *Neuf* indique une nouvelle construction. *Pré* indique une vente avant la construction. *Jardin* est une variable muette égale à un si l'appartement dispose d'un jardin privé. *Taille* indique le nombre de mètres carrés de surface utile de l'appartement. *Piscine* est une variable muette indiquant si l'appartement dispose d'une piscine. *Ville* est une variable muette indiquant que l'appartement est situé dans l'une des deux principales villes corses : Ajaccio ou Bastia. *Plage* est le temps d'accès à la plage en minutes. *Médecin* est le temps d'accès au médecin le plus proche en minutes. *Pharmacie* est le temps d'accès à la pharmacie la plus proche en minutes. *École primaire* est le temps d'accès à l'école primaire la plus proche en minutes. *Centre-ville* est le temps d'accès au centre ville en minutes. *Ville principale* est le temps d'accès à la ville principale la plus proche en minutes, à savoir Ajaccio, Bastia, Calvi, Corte, Ghisonaccia, L'Île-Rousse, Porto-Vecchio ou Propriano.

*Indice de vue sur mer* est un indice construit qui évalue la qualité de la vue sur la mer depuis l'appartement, calculé à l'aide d'ArcGIS (Nagy, 1994; O'Sullivan et Turner, 2001). En explorant le voisinage de l'appartement à l'aide des géo-coordonnées, nous pouvons déterminer le ratio de la surface de mer visible dans un rayon de 20 km autour de l'appartement. Ce rapport est compris entre 0 et 100, 0 indiquant une absence de vue sur la mer et 100 une vue complète.

### 5.5.3 Variables construites d'*HRS*

Afin d'évaluer l'impact des caractéristiques individuelles des acheteurs et des vendeurs sur les prix de vente, nous avons construit les sommes et les différences des caractéristiques des acheteurs et des vendeurs de la même manière que *HRS*. Deux des variables sont basées sur la localisation et limitées aux citoyens français. La première variable indique si l'acheteur ou le vendeur résident dans la même commune que le bien (*Communes*) et sont de nationalité française et la seconde indique si l'acheteur ou le vendeur résident dans une autre commune de Corse (*Corse*) et sont de nationalité française. Une troisième variable muette a la valeur un si l'acheteur ou le vendeur n'est pas de nationalité française (*Etranger*). Nous pensons que les résidents seront plus familiers avec la Corse que les non-résidents et que les citoyens de pays étrangers seront moins familiers avec la langue et la structure juridique que les citoyens français. Enfin, nous construisons des variables muettes pour indiquer si les acheteurs ou les vendeurs font partie d'un groupe (*Co-propriétaire*).

Ces variables muettes sont utilisées pour construire quatre nouvelles variables de « somme » et quatre nouvelles variables de « différence ». Les variables de somme prennent les valeurs 0 si la caractéristique n'indique ni l'acheteur ni le vendeur, 1 si la caractéristique indique soit l'acheteur soit le vendeur ou 2 si elle indique l'acheteur et le vendeur. Harding et al. (2003) suggèrent que les coefficients de ces variables mesurent les effets sur la demande des caractéristiques des acheteurs et des vendeurs.

Nos quatre variables muettes sont également utilisées pour construire quatre nouvelles variables de différence. À l'instar de *HRS*, nous adoptons la convention de calcul des différences vendeur-acheteur suivante. Comme ces caractéristiques sont des variables muettes, les différences peuvent prendre les valeurs -1, 0 ou 1. Une valeur de -1 indique que la caractéristique est présente chez l'acheteur mais pas chez le vendeur. Une valeur de 0 indique que la caractéristique est présente à la fois chez l'acheteur et le vendeur ou absente chez les deux, ce qui signifie essentiellement que l'acheteur et le vendeur sont identiques. Une valeur de 1 indique que la caractéristique est présente chez le vendeur mais absente chez l'acheteur.

Par exemple, considérons la paire de variables muettes indiquant si le vendeur est *Corse* ou si l'acheteur est *Corse*. Pour les variables de somme construite, il existe quatre possibilités. Une transaction entre un vendeur *Corse* et un acheteur *Corse* rapporte une somme de 2. Une transaction entre un vendeur *Corse* et un acheteur non corse ou un vendeur non corse et un acheteur *Corse* donne une valeur de 1 (soit  $1 + 0$  ou  $0 + 1$ ). Une transaction entre un vendeur non corse et un acheteur non corse donne une valeur de 0. Les variables de différence associées à ces transactions sont, respectivement,  $1 - 1 = 0$ ,  $1 - 0 = 1$  ou  $0 - 1 = -1$ , et  $0 - 0 = 0$ . Toutes les variables de somme et de différence sont construites de manière similaire.

*HRS* utilisent les coefficients des variables de différence pour indiquer le pouvoir de négociation des personnes présentant la caractéristique (ou son absence). Un coefficient négatif associé à cette variable de différence, compte tenu de notre construction vendeur-acheteur, signifie que, quel que soit le bien, les vendeurs présentant cette caractéristique sont associés à des prix inférieurs à la moyenne, c'est-à-dire qu'ils ne sont pas de bons négociateurs. Le signe négatif implique également que les acheteurs présentant la caractéristique paient, en moyenne, des prix plus élevés que la moyenne, ce qui indique également qu'ils sont de mauvais négociateurs. Ainsi, le coefficient négatif indique que les acheteurs et les vendeurs présentant cette caractéristique ont un faible pouvoir de négociation.

En revanche, si le coefficient d'une variable de différence est positif, le contraire est vrai. Là encore, étant donnée notre convention vendeur-acheteur, un coefficient positif signifie que, quelle que soit la propriété, les vendeurs possédant la caractéristique vendent à des prix supérieurs à la moyenne et les acheteurs possédant la caractéristique achètent à des prix inférieurs à la moyenne. Dans ce cas, il y a un grand pouvoir de négociation associé à la caractéristique vendeur-acheteur.

Les statistiques récapitulatives de toutes les variables utilisées dans le modèle figurent dans la colonne 2 du tableau 5.2. Toutes les variables monétaires utilisées dans cette étude ont été déflatées à l'aide de l'indice annuel des prix français. Comme le

montre le tableau, le prix réel moyen de l'appartement vendu au cours de cette période est légèrement inférieur à 175 000 euros. Le tableau 5.2 montre qu'un peu moins de 30% de tous les logements ont été achetés avec un prêt hypothécaire et que le nombre moyen de pièces par appartement est légèrement inférieur à trois. Un peu moins de 5% des logements ont un sous-sol et la taille moyenne du logement est d'un peu plus de 65 mètres carrés.

## 5.6 Le modèle conceptuel

Nous utilisons cette base de données très détaillée pour estimer un modèle de régression hédonique log-linéaire. Nous utilisons le logarithme du prix de vente comme variable dépendante et incluons plusieurs variables indépendantes que nous avons décrites précédemment.

Notre modèle conceptuel de base est présenté ci-dessous :

$$\begin{aligned}
 \ln P = & \beta_0 + \beta_1 \textit{Intermédiation} + \beta_2 \textit{Hypothèque} + \beta_3 \textit{Meublé} + \beta_4 \textit{Viager} + \beta_5 \textit{Pièces} \\
 & + \beta_6 \textit{Salledebain} + \beta_7 \textit{Etages} + \beta_8 \textit{Neuf} + \beta_9 \textit{Pré} + \beta_{10} \textit{SousSol} + \beta_{11} \textit{Jardin} \\
 & + \beta_{12} \textit{Taille} + \beta_{13} \textit{Plage} + \beta_{14} \textit{Ville} + \beta_{15} \textit{Médecin} + \beta_{16} \textit{Pharmacie} \\
 & + \beta_{17} \textit{Ecoleprimaire} + \beta_{18} \textit{CentreVille} + \beta_{19} \textit{Villeprincipale} + \beta_{20} \textit{Indicedevuesurmer} \\
 & + \beta_{21} \textit{DiffCommune} + \beta_{22} \textit{DiffCopropriétaire} + \beta_{23} \textit{DiffCorse} \\
 & + \beta_{24} \textit{DiffEtranger} + \beta_{25} \textit{SommeCommune} + \beta_{26} \textit{SommeCopropriétaire} \\
 & + \beta_{27} \textit{SommeCorse} + \beta_{28} \textit{SommeEtranger} + y_r_i + \epsilon
 \end{aligned}
 \tag{5.18}$$

Nous nous attendons à ce que les coefficients *Intermédiation*, *Hypothèque*<sup>8</sup>, *Meublé*, *Viager*, *Pièces*, *Salle de Bain*, *Etage*, *Neuf*, *Pré*, *Sous-sol*, *Jardin*, *Taille*, *Vue sur la mer*, et *Ville* soient positifs. En nous basant sur Davidoff et Welke (2017), nous anticipons que le coefficient *Viager* sera négatif<sup>9</sup>. Nous nous attendons également à ce que les coefficients de toutes les variables de temps de déplacement, à savoir *Plage*, *Médecin*, *Pharmacie*, *Primaire*, *Centre-ville* et *Ville principale*, soient négatifs, bien que la proximité du bruit et des foules puisse entraîner une baisse des prix des maisons dans certains de ces cas. Nous incluons également un ensemble de variables indicatrices de l'année afin de saisir les effets éventuels de l'évolution des conditions du marché.

8. Comme indiqué dans Jin et Rafferty (2018), le taux hypothécaire affecte l'offre et la demande de logements. En utilisant un modèle de prix hédonique dynamique avec un ensemble de données de panel mensuel, les auteurs montrent que les taux hypothécaires sont positivement associés à la croissance des prix des logements. Jauregui et al. (2017) confirment ces résultats en utilisant une vaste base de données portant sur 9716 transactions immobilières dans le comté de Franklin, en Ohio. Ils mesurent la décote attendue des acheteurs pour les achats de logements au comptant.

9. Comme le soulignent Davidoff et Welke (2017), « le prêt viager hypothécaire réduit le gain net à la vente de la maison, car il y a plus de richesse en l'absence de vente et moins de richesse après une vente. » Cela explique le signe négatif du coefficient *Viager*.

## 5.7 Résultats de l'estimation

### 5.7.1 Résultats de l'estimation par la méthode des moindres carrés ordinaires (MCO)

Les résultats de l'estimation des MCO sont donnés dans la colonne 3 du tableau 5.2. Le  $R_2$  du modèle est de 0,63, ce qui est impressionnant étant donné le grand nombre de transactions de notre échantillon. Sur les 20 coefficients du modèle, 18 sont statistiquement significatifs au niveau  $\alpha = 0,05$  voire à un niveau inférieur. Les signes de tous les coefficients sont conformes à nos attentes, à l'exception du coefficient positif et statistiquement significatif de la variable *Ecole primaire*. Ce résultat peut être dû aux externalités dues au bruit et à la congestion du trafic générées par la présence d'écoles primaires. Compte tenu du grand nombre d'observations et du grand nombre de variables indépendantes, ce modèle est particulièrement performant.

Pour évaluer les résultats, nous examinons les estimations des variables construites de la somme et de la différence. Tous les coefficients de différence construits sont statistiquement significatifs *a minima* au niveau  $\alpha = 0,10$ . Les résultats indiquent un pouvoir de négociation plus important pour les acheteurs et les vendeurs appartenant à la même commune et lorsque les acheteurs et les vendeurs sont des résidents corses. Les résultats indiquent également un pouvoir de négociation important pour les citoyens étrangers et une marge de négociation moindre si la transaction immobilière implique un groupe d'acteurs du marché. Ces résultats sont cohérents avec l'idée que les individus vivant à proximité de la propriété disposent de meilleures informations que ceux qui sont plus éloignés.

En ce qui concerne le résultat des *Etrangers* présents sur le marché corse, il est possible qu'ils soient de riches investisseurs extérieurs (acheteurs), ou des parents de résidents corses vivant à l'étranger et ayant hérité de propriétés corses (vendeurs). Le manque de pouvoir de négociation des groupes d'acheteurs et de vendeurs est cohérent avec la présence de coûts de transaction et de coordination plus élevés dans ces situations. Les résultats contiennent également trois effets de demande statistiquement significatifs. Les prix et la demande sont plus faibles pour les locaux, les *Corses* et les *Etrangers* par rapport aux citoyens français non résidents (catégorie omise).

Ces résultats des MCO reposent sur l'hypothèse que le marché du logement n'est pas segmenté. Dans la section suivante, nous examinons la validité de cette hypothèse en estimant un modèle de classe latente ou de mélange fini.

### 5.7.2 Résultats de l'estimation par classe latente

Les résultats de l'estimation du modèle de classe latente sont donnés dans les colonnes quatre et six du tableau 5.2. Un modèle à deux régimes est favorisé par rapport à un modèle à un seul régime en utilisant le critère d'information bayésien ou BIC. Nous constatons que le régime 1 caractérise environ 18% des observations et que le régime 2 caractérise environ 82% des observations. Les résultats de l'estimation pour le régime

arbitrairement appelé régime 1 sont donnés dans la colonne 4 du tableau. Sur les 20 variables indépendantes associées aux caractéristiques des logements, seules trois ont des coefficients qui ne sont pas statistiquement différents de zéro à tous les niveaux habituels.

Les résultats de l'estimation du régime 1 indiquent que deux des coefficients de différence sont statistiquement différents de zéro au moins au niveau  $\alpha = 0,05$ . Ces résultats indiquent un fort pouvoir de négociation pour les résidents locaux qui vivent dans la même commune et un pouvoir de négociation réduit pour les *Etrangers*. Ces résultats sont cohérents, respectivement, avec un meilleur accès à l'information par les locaux et des déficiences en matière d'information linguistique et juridique pour les *Etrangers*. Par ailleurs, il existe deux effets de demande statistiquement significatifs dans le régime 1 qui indiquent une baisse des prix et de la demande par les résidents locaux et une baisse de la demande par les autres Corses.

La colonne 6 du tableau 5.2 contient les résultats de l'estimation pour le régime 2 qui représente environ 82% des données. Sur les vingt coefficients, cinq ne sont pas significativement différents de zéro (*Intermédiation*, *Etage*, *Pré*, *Médecin*, et *Pharmacie*) et deux, à savoir *Médecin* et *Ville*, ont un signe contraire à notre intuition mais seul le coefficient *Ville* est statistiquement significatif. De même que précédemment, le coefficient *Ecole Primaire* est statistiquement significatif mais négatif, ce qui peut s'expliquer par la présence d'externalités de localisation.

Nous examinons ensuite les coefficients des variables somme et différence et les implications du pouvoir de négociation et de la demande pour nos résultats du régime 2. Les coefficients de différence pour les variables *DiffCommune*, *DiffCorse*, *DiffEtrangers* sont positifs et statistiquement différents de zéro au niveau  $\alpha = 0,01$ . Les résultats indiquent un fort pouvoir de négociation pour ces trois catégories. Ces résultats sont similaires à ceux de nos MCO et sont dus aux asymétries d'information dans les deux premiers cas et peut-être à la richesse et/ou à la présence de propriétés héritées dans le dernier cas. Trois des coefficients de somme sont statistiquement significatifs. Les résultats indiquent une demande plus faible de la part des résidents de la même commune, des autres résidents corses et des étrangers. Par la suite, nous examinerons de plus près la nature de ces deux régimes en estimant les facteurs corrélés à l'appartenance à un état. Pour ce faire, nous estimerons des modèles de régression en utilisant la probabilité postérieure d'appartenance au régime 1 comme variable dépendante<sup>10</sup>.

### 5.7.3 Régressions auxiliaires

Comme mentionné précédemment, l'un des résultats de l'estimation du modèle de classe latente est une estimation de la probabilité postérieure que chaque observation soit associée à chaque état. Autrement dit, la procédure fournit une estimation de la probabilité qu'une transaction soit associée au régime 1 ou au régime 2. Nous utilisons cette probabilité comme variable dépendante dans une régression auxiliaire pour estimer

---

10. Nous n'avons pas besoin d'examiner les deux régimes car tout facteur augmentant la probabilité d'association au régime 1 doit diminuer la probabilité d'association au régime 2 et inversement.

l'impact de chaque variable indépendante sur la probabilité d'association au régime 1. Ces résultats sont donnés dans la colonne 2 du tableau 5.2. L'effet le plus important est associé à la présence d'un prêt viager hypothécaire qui augmente la probabilité d'être associé au régime 2 d'environ 0,18. En dehors de cet effet, les deux variables ayant le plus d'impact sur la base de la forte valeur de leurs coefficients sont *Meublé* et *Neuf*. Les deux coefficients sont statistiquement significatifs et négatifs. Combinés, ces résultats indiquent qu'un appartement neuf et meublé a une probabilité accrue d'être associé au régime 2 d'environ 0,10.

Pour ce modèle de régression auxiliaire, nous avons également inclus le groupe de huit variables fictives acheteur-vendeur utilisées pour construire nos variables de somme et de différence. Le tableau indique que quatre des coefficients associés à ces variables fictives sont statistiquement significatifs. Les résultats montrent que les vendeurs de la même commune, les autres vendeurs corses et les vendeurs étrangers sont plus étroitement associés au régime 2, tout comme les acheteurs copropriétaires.

Dans la colonne 3 du tableau 5.2, nous estimons le modèle de régression auxiliaire sans inclure aucune information sur les caractéristiques de l'acheteur ou du vendeur. Les résultats de l'estimation pour les variables restantes sont presque identiques aux résultats précédents, nous ne les discuterons donc pas davantage.

Nous pouvons obtenir plus d'informations sur la nature des régimes en examinant les moyennes des variables. Nous utilisons les probabilités postérieures comme poids dans le calcul des moyennes pondérées pour chacun des deux groupes. Ces moyennes sont données dans les colonnes quatre et six du tableau 5.1. Ces résultats reflètent, dans une large mesure, nos conclusions concernant les modèles de régression auxiliaires. D'après les moyennes, les appartements associés au régime 1 ont tendance à être plus spacieux, ils sont moins susceptibles d'être neufs, plus susceptibles d'avoir un jardin, et ont généralement des temps de trajet plus longs pour se rendre à l'une des commodités évoquées. Ils se situent, a priori, hors de la ville et pourraient être qualifiés de maisons de campagne.

Avec les résultats de la régression auxiliaire, une représentation du marché des appartements du régime 1 se dégage. Ces appartements semblent être plus ruraux et plus éloignés de la côte. La probabilité est relativement élevée qu'ils soient vendus plus fréquemment par des étrangers. Pour les propriétés situées en dehors des deux grandes villes, il est raisonnable d'observer un pouvoir de négociation plus élevé pour les locaux et les autres Corses en raison de leur meilleure connaissance du marché. Un pouvoir de négociation moindre pour le groupe *Etrangers* n'est pas non plus surprenant car ce groupe doit surmonter des barrières sociales, juridiques et éventuellement linguistiques pour effectuer des transactions. À l'inverse, les appartements du régime 2 sont plus petits, plus récents et plus étroitement associés aux deux grandes villes de Corse. Là encore, les acteurs du marché qui sont de la même commune ou d'autres communes corses font preuve d'un fort pouvoir de négociation, tandis que les étrangers ont également un pouvoir de négociation. Cela pourrait résulter d'un effet de richesse non observé sur le marché de l'investissement ou des propriétés de vacances ou, peut-être, s'agit-il de locaux étrangers.

Bon nombre des différences que nous constatons entre les différents types d'acheteurs et de vendeurs peuvent s'expliquer par les asymétries d'information qui, comme nous l'avons noté précédemment, peuvent être dues à la distance géographique ou à la distance linguistique et juridique. Les individus rationnels continueront à chercher une meilleure offre jusqu'à ce que le coût marginal de la recherche soit égal à son bénéfice marginal. Nous pourrions nous attendre à ce que les participants au marché locaux, corses et étrangers affichent des coûts de recherche différents, en raison de différences dans la proximité du bien ou dans la maîtrise de la langue française. En d'autres termes, le coût de la collecte d'informations (recherche d'une meilleure offre) augmente à mesure que l'agent est plus éloigné et moins familier avec la langue et le système juridique français. En suivant ce raisonnement, les vendeurs et les acheteurs non locaux ont tendance à arrêter la recherche et/ou à mettre fin à la négociation du prix plus tôt que les participants locaux.

Enfin, nous démontrons à quel point certains de ces marchés sont restreints en examinant les fréquences de diverses combinaisons de transactions présentées dans le tableau 5.3. Le tableau donne les fréquences des combinaisons de transactions pour chacune de nos quatre variables de différence. Ces fréquences sont fournies pour l'échantillon complet et pour les régimes 1 et 2, pondérées à nouveau par les probabilités postérieures. Il convient de noter que ces informations ne peuvent être fournies pour les régimes 1 et 2 qu'après l'estimation du modèle de classe latente, car il s'agit de fréquences pondérées, les poids étant les probabilités postérieures.

Le tableau 5.3 présente les fréquences par caractéristique pour chacune des trois valeurs de différence importantes :  $-1$ ,  $0$ , et  $+1$ . Comme le montre la colonne 2 du tableau 5.3, la majorité des transactions ont lieu entre des acheteurs et des vendeurs ayant la même caractéristique. L'examen des fréquences des valeurs différentes de zéro est instructif. Environ 44% des transactions impliquant des résidents d'une commune impliquent des partenaires qui n'appartiennent pas à cette commune. Pour les copropriétaires, le nombre de transactions n'impliquant pas d'autres copropriétaires est d'environ 20%. Pour les résidents corses, la fraction de ces transactions « mixtes » est d'environ 37%, et pour les étrangers, la part des transactions « mixtes » est d'environ 8%.

Les colonnes 3 et 4 du tableau 5.3 présentent les fréquences pondérées des combinaisons de transactions pour les régimes 1 et 2. Nous ne discutons pas de ces fréquences en détail ici car elles sont remarquablement similaires aux fréquences caractérisant l'échantillon complet. Autrement dit, l'étroitesse du marché ne diffère pas d'un régime à l'autre.

## 5.8 Conclusion

Bien que la segmentation des marchés du logement et les différences de pouvoir de négociation des acheteurs et des vendeurs dans ces segments aient été théorisées depuis longtemps, cette étude est la première à mesurer empiriquement ces différences. Pour ce faire, nous combinons deux nouvelles méthodes économétriques, à savoir l'estimation

d'un modèle de régression hédonique à classe latente et la modification du modèle pour inclure les variables du pouvoir de négociation de HRS. L'approche HRS permet d'estimer le pouvoir de négociation des acheteurs et des vendeurs et la méthode des classes latentes est utilisée pour identifier des segments de marché autrement inobservables. En combinant l'approche HRS avec l'identification du segment fournie par l'estimation de la classe latente, nous sommes en mesure de présenter la première preuve empirique de différences dans le pouvoir de négociation au niveau du segment de marché.

En utilisant la Corse comme étude de cas, nos résultats indiquent que le marché du logement corse a deux régimes distincts et que le pouvoir de négociation des acheteurs et des vendeurs n'est pas le même dans ces deux segments. En particulier, nos résultats suggèrent que les résidents locaux d'une commune ont un fort pouvoir de négociation dans un régime et un faible pouvoir de négociation avec des participants étrangers. Dans l'autre état, les individus d'une commune, les autres résidents corses et les étrangers ont un fort pouvoir de négociation. Nos régressions auxiliaires indiquent que les appartements associés au régime 1 ont tendance à être plus spacieux, qu'ils sont moins susceptibles d'être neufs, qu'ils sont plus susceptibles d'avoir un jardin et que les temps de trajet pour se rendre à l'une des commodités évoquées sont généralement plus longs. Nous en concluons que les appartements du régime 1 sont plus susceptibles d'être situés en zone rurale et à une plus grande distance de la côte, ce qui place les étrangers dans une situation de proximité et éventuellement de désavantage linguistique. Les appartements du régime 2 sont plus étroitement associés aux régions côtières de l'île où les désavantages en matière d'information des étrangers peuvent être compensés par l'achat d'informations auprès d'experts locaux et peut-être par une richesse importante, car certaines de ces transactions concernent probablement des appartements de vacances.

Notre approche générale devrait être utile dans de nombreuses autres applications empiriques. Une fois que la méthode des classes latentes a identifié les segments de marché, de nombreuses questions empiriques intéressantes, telles que le pouvoir de négociation relatif des acheteurs et des vendeurs examiné ici, peuvent être étudiées au niveau du segment de marché.

Variabes	Moyenne (Erreur standard)	Estimation MCO	Classe latente Régime 1	Régime 1 Moyennes et Erreurs standard	Classe latente Régime 2	Régime 2 Moyennes et Erreurs standard
Prix réel (ln)	11.889 (0.58)	—	—	11.758 (0.34)	—	11.918 (0.46)
Intermédiation	0.155 (0.36)	0.048*** (4.18)	0.186*** (3.79)	0.165 (0.15)	0.013 (1.29)	0.152 (0.32)
Hypothèque	0.297 (0.46)	0.008 (0.96)	0.102*** (2.69)	0.277 (0.19)	-0.013* (1.76)	0.301 (0.41)
Meublé	0.175 (0.38)	0.184*** (15.79)	0.403*** (7.91)	0.182 (0.16)	0.114*** (10.19)	0.173 (0.34)
Viager	0.002 (0.04)	-0.324*** (3.53)	-2.870*** (5.68)	0.000 (0.01)	-0.350*** (5.60)	0.002 (0.04)
Pièces	2.827 (1.22)	0.179*** (37.33)	0.211*** (15.48)	3.033 (0.68)	0.029*** (4.70)	2.782 (1.00)
Salle de bain	1.123 (0.42)	0.155*** (13.40)	0.210*** (5.98)	1.190 (0.24)	0.024* (1.93)	1.108 (0.33)
Etage	1.785 (1.78)	0.011*** (4.57)	0.033*** (3.13)	1.647 (0.75)	0.002 (1.06)	1.813 (1.60)
Neuf	0.514 (0.50)	0.214*** (15.05)	0.316*** (4.97)	0.425 (0.21)	0.182*** (14.91)	0.532 (0.45)
Pré	0.408 (0.49)	0.039*** (2.78)	0.210*** (3.30)	0.334 (0.20)	0.001 (0.09)	0.423 (0.44)
Sous-sol	0.242 (0.43)	0.047*** (4.92)	0.081** (2.11)	0.249 (0.18)	0.035*** (4.24)	0.239 (0.38)
Jardin	0.069 (0.25)	0.173*** (8.74)	0.263*** (3.61)	0.119 (0.13)	0.045** (2.15)	0.058 (0.21)
Taille	65.143 (40.87)	0.003*** (22.54)	0.0007*** (2.96)	72.634 (30.80)	0.012*** (45.14)	63.530 (26.70)
Plage	5.896 (7.42)	-0.013*** (22.93)	-0.022*** (7.37)	6.267 (3.42)	-0.010*** (20.40)	5.815 (6.59)
Médecin	3.570 (6.97)	-0.002 (0.96)	-0.008 (1.01)	5.270 (5.87)	0.003 (1.54)	3.205 (3.69)
Pharmacie	3.395 (7.00)	-0.011*** (4.89)	-0.014** (2.01)	5.177 (5.90)	-0.000 (0.02)	3.011 (3.69)
Ecole primaire	3.273 (6.41)	0.021*** (13.17)	0.030*** (5.50)	4.623 (5.70)	0.017*** (8.17)	2.983 (2.87)
Centre-ville	14.724 (15.12)	-0.008*** (20.26)	-0.006*** (4.52)	15.728 (6.72)	-0.009*** (22.20)	14.523 (13.54)
Ville principale	15.109 (13.95)	-0.003*** (7.49)	-0.003** (2.28)	17.059 (7.68)	-0.002*** (5.28)	14.703 (11.62)
Indice de vue sur mer	12.708 (11.47)	0.001*** (3.76)	0.000 (0.00)	12.004 (4.87)	0.001*** (3.90)	12.851 (10.37)
Ville	0.667 (0.47)	-0.025** (2.31)	0.004 (0.09)	0.619 (0.20)	-0.065*** (6.45)	0.676 (0.42)
Année 1	0.085 (0.28)	-0.120*** (6.11)	-0.188** (2.30)	0.076 (0.11)	-0.110*** (6.20)	0.086 (0.25)

Variabes	Moyenne (Erreur standard)	Estimation MCO	Classe latente Régime 1	Régime 1 Moyennes et Erreurs standard	Classe latente Régime 2	Régime 2 Moyennes et Erreurs standard
Année 2	0.092 (0.29)	0.026 (1.35)	-0.005 (0.68)	0.085 (0.11)	0.026 (1.50)	0.092 (0.26)
Année 3	0.104 (0.30)	0.021 (1.13)	-0.039 (0.49)	0.092 (0.12)	0.028 (1.64)	0.105 (0.27)
Année 4	0.093 (0.29)	0.009 (0.47)	0.015 (0.19)	0.085 (0.11)	0.007 (0.42)	0.094 (0.26)
Année 5	0.136 (0.34)	0.024 (1.33)	-0.073 (0.93)	0.111 (0.13)	0.019 (1.19)	0.140 (0.31)
Année 6	0.075 (0.26)	0.087*** (4.30)	0.044 (0.51)	0.066 (0.10)	0.060*** (3.33)	0.077 (0.24)
Année 7	0.075 (0.26)	0.095*** (4.69)	-0.009 (0.11)	0.078 (0.11)	0.094*** (5.14)	0.074 (0.23)
Année 8	0.093 (0.29)	0.023 (1.24)	-0.102 (1.27)	0.093 (0.12)	0.036** (2.03)	0.092 (0.26)
Année 9	0.069 (0.25)	0.037 (1.85)	-0.035 (0.45)	0.082 (0.11)	0.024 (1.29)	0.067 (0.22)
Année 10	0.038 (0.19)	0.026 (1.07)	-0.001 (0.02)	0.043 (0.08)	0.015 (0.69)	0.036 (0.17)
Année 11	0.054 (0.23)	0.012 (0.60)	-0.006 (0.08)	0.080 (0.11)	0.014 (0.63)	0.048 (0.19)
Diff Commune	0.202 (0.63)	0.084*** (9.72)	0.070** (1.97)	0.187 (0.27)	0.072*** (9.14)	0.205 (0.57)
Diff Copropriétaire	-0.030 (0.44)	-0.016* (1.80)	-0.055 (1.52)	-0.007 (0.19)	-0.005 (0.70)	-0.034 (0.39)
Diff Corse	0.008 (0.60)	0.060*** (6.61)	0.044 (1.52)	0.006 (0.25)	0.056*** (6.81)	0.008 (0.55)
Diff Etranger	-0.029 (0.27)	0.071*** (4.42)	-0.181*** (2.79)	-0.031 (0.12)	0.125*** (8.39)	-0.029 (0.24)
Somme Commune	0.794 (0.71)	-0.052*** (6.36)	-0.106*** (3.21)	0.784 (0.30)	-0.058*** (7.50)	0.797 (0.65)
Somme Copropriétaire	0.231 (0.46)	-0.010 (1.14)	-0.040 (1.11)	0.247 (0.20)	-0.011 (1.48)	0.228 (0.41)
Somme Corse	0.634 (0.70)	-0.085*** (10.24)	-0.090*** (2.67)	0.574 (0.29)	-0.090*** (11.54)	0.647 (0.63)
Somme Etranger	0.106 (0.35)	-0.147*** (11.30)	-0.084 (1.54)	0.120 (0.16)	-0.148*** (11.68)	0.103 (0.31)
Constante	—	11.086*** (460.32)	10.832*** (108.50)	—	11.188*** (456.09)	—
$R^2$	—	0.63	—	—	—	—
-2 log- vraisemblance	—	—	6252.00	—	3504.53	—
BIC	—	—	6621.78	—	4253.11	—
Poids de mélange	—	—	0.181*** (16.84)	—	0.819	—

TABLE 5.1 – Statistiques et résultats de l'estimation

Variabes	MCO avec probabilité postérieure du régime 1 comme variable dépendante	MCO avec probabilité postérieure du régime 2 comme variable dépendante
Constante	0.184*** (13.24)	0.169*** (13.27)
Intermédiation	-0.016** (2.46)	-0.015** (2.34)
Hypothèque	-0.011** (2.16)	-0.010** (2.10)
Meublé	-0.043*** (6.44)	-0.043*** (6.50)
Viager	-0.183*** (3.44)	-0.182*** (3.43)
Pièces	0.003 (1.12)	0.002 (1.04)
Salle de bain	0.023*** (3.52)	0.023*** (3.55)
Etage	-0.002 (1.42)	-0.001 (1.25)
Neuf	-0.051*** (6.23)	-0.053*** (6.61)
Pré	0.004 (0.61)	0.002 (0.26)
Sous-sol	-0.002 (0.50)	-0.001 (0.32)
Jardin	0.015 (1.36)	0.017 (1.53)
Taille	0.000*** (5.70)	0.000*** (5.74)
Plage	-0.000 (1.14)	-0.000 (1.60)
Médecin	0.001 (0.65)	0.001 (0.76)
Pharmacie	0.006*** (4.69)	0.006*** (4.69)
Ecole primaire	-0.002*** (3.15)	-0.002*** (3.15)
Centre-ville	-0.000 (1.47)	-0.000 (1.47)
Ville principale	0.000** (2.28)	0.000** (2.28)
Indice de vue sur mer	-0.000 (0.16)	-0.000 (0.47)
Ville	-0.019*** (3.04)	-0.020*** (3.45)
Année 1	-0.031*** (2.75)	-0.030*** (2.71)

Variabiles	MCO avec probabilité postérieure du régime 1 comme variable dépendante	MCO avec probabilité postérieure du régime 2 comme variable dépendante
Année 2	-0.022** (2.00)	-0.022** (1.97)
Année 3	-0.026** (2.40)	-0.026** (2.40)
Année 4	-0.014 (1.27)	-0.014 (1.29)
Année 5	-0.028*** (2.71)	-0.029*** (2.80)
Année 6	-0.021* (1.86)	-0.022* (1.89)
Année 7	0.002 (0.25)	0.001 (0.13)
Année 8	-0.007 (0.69)	-0.008 (0.75)
Année 9	-0.006 (0.58)	-0.006 (0.63)
Année 10	-0.010 (0.72)	-0.009 (0.71)
Année 11	0.051*** (4.16)	0.050*** (3.45)
Vendeur même commune	-0.021*** (2.82)	—
Vendeur Corse	-0.028*** (3.43)	—
Vendeur étranger	-0.043*** (3.16)	—
Vendeur Copropriétaire	0.002*** (0.29)	—
Acheteur même commune	0.006 (0.96)	—
Acheteur Corse	-0.002 (0.35)	—
Acheteur étranger	0.012 (1.19)	—
Acheteur Copropriétaire	-0.015** (2.22)	—
$R^2$	0.10	0.09

**TABLE 5.2** – *Résultats des régressions auxiliaires*

	Echantillon complet	Régime 1 pondéré par la probabilité postérieure	Régime 2 pondéré par la probabilité postérieure
DiffCommune			
-1	980 (11.87)	188.27 (12.56)	791.73 (11.71)
0	4626 (56.01)	840.98 (56.11)	3785.02 (55.99)
1	2653 (32.12)	469.65 (31.33)	2183.36 (32.30)
DiffCoproriétaire			
-1	930 (11.26)	157.93 (10.54)	772.07 (11.42)
0	6647 (80.48)	1194.51 (70.69)	5452.49 (80.66)
1	682 (8.26)	146.46 (9.77)	151.03 (2.22)
DiffCorse			
-1	1503 (18.20)	256.58 (17.12)	1246.42 (18.44)
0	5186 (62.79)	975.41 (65.08)	4110.59 (62.29)
1	1570 (19.01)	266.91 (17.81)	1303.10 (19.27)
DiffEtranger			
-1	432 (5.23)	84.92 (5.67)	347.08 (5.13)
0	7639 (92.49)	1376.00 (91.87)	6262.00 (92.63)
1	188 (2.28)	36.97 (2.47)	151.03 (2.24)

**TABLE 5.3** – *Fréquences et pourcentage des variables de différence pour le pouvoir de négociation*

Pour les trois tableaux ci-dessus, nous pouvons noter que les chiffres entre parenthèses sont les valeurs absolues des t-ratios. \*\*\* Indique une significativité statistique au niveau  $\alpha = 0,01$ ; \*\* Indique une significativité statistique au niveau  $\alpha = 0,05$ ; \* Indique une significativité statistique au niveau  $\alpha = 0,10$ .

# Bibliographie

- Anglin, Paul et Robin Wiebe (2013). « Pricing in an illiquid real estate market ». In : *Journal of Real Estate Research* 35(1), p. 83-102.
- Bayer, Patrick, Marcus Casey, Fernando Ferreira et Robert McMillan (2017). « Racial and ethnic price differentials in the housing market ». In : *Journal of Urban Economics* 102, p. 91-105.
- Beard, T Randolph, Steven B Caudill et Daniel M Gropper (1991). « Finite mixture estimation of multiproduct cost functions ». In : *The review of economics and statistics*, p. 654-664.
- Beaumais, Olivier, Sauveur Giannoni et Caroline Tafani (2021). « Local buyers' exclusion through price on the farmland market: a potential outcome approach ». In : *Applied Economics* 53(19), p. 2206-2217.
- Belasco, Eric, Michael Farmer et Clifford Lipscomb (2012). « Using a finite mixture model of heterogeneous households to delineate housing submarkets ». In : *Journal of Real Estate Research* 34(4), p. 577-594.
- Bø, Erlend Eide (2018). « Housing match quality and demand: What can we learn from comparing buyer characteristics? » In : *Journal of Housing Economics* 41, p. 184-199.
- Brasington, David M et Robert F Sarama Jr (2008). « Deed types, house prices and mortgage interest rates ». In : *Real Estate Economics* 36(3), p. 587-610.
- Carrillo, Paul E, Eric R de Wit et William Larson (2015). « Can tightness in the housing market help predict subsequent home price appreciation? Evidence from the United States and the Netherlands ». In : *Real Estate Economics* 43(3), p. 609-651.
- Chandra, Ambarish, Sumeet Gulati et James M Sallee (2017). « Who loses when prices are negotiated? An analysis of the new car market ». In : *The Journal of Industrial Economics* 65(2), p. 235-274.
- Colwell, Peter F et Henry J Munneke (2006). « Bargaining strength and property class in office markets ». In : *The Journal of Real Estate Finance and Economics* 33(3), p. 197-213.
- Cotteleer, Geerte, Cornelis Gardebroek et Jan Luijt (2008). « Market power in a GIS-based hedonic price model of local farmland markets ». In : *Land Economics* 84(4), p. 573-592.

- Daneshvary, Nasser et Terrence M Clauretje (2013). « Agent change and seller bargaining power: a case of principal agent problem in the housing market ». In : *The Journal of Real Estate Finance and Economics* 47(3), p. 416-433.
- Davidoff, Thomas et Gerd M Welke (2017). « The role of appreciation and borrower characteristics in reverse mortgage terminations ». In : *Journal of Real Estate Research* 39(1), p. 99-126.
- Dempster, AP, NM Laird et DB Rubin (1977). « Maximum likelihood from incomplete data via the EM algorithm ». In : *Journal of the Royal Statistical Society* 39(1), p. 1-38.
- Frühwirth-Schnatter, Sylvia (2006). *Finite mixture and Markov switching models*. T. 425. Springer.
- Hardin III, William, Matthew Hill et James Hopper (2009). « Ownership structure, property performance, multifamily properties, and REITs ». In : *Journal of Real Estate Research* 31(3), p. 285-306.
- Harding, John P., Stuart S. Rosenthal et Clemon F. Sirmans (2003). « Estimating bargaining power in the market for existing homes ». In : *Review of Economics and Statistics* 85(1), p. 178-188.
- Hayunga, Darren K (2018). « Sales concessions in the US housing market ». In : *The Journal of Real Estate Finance and Economics* 56(1), p. 33-75.
- Holmes, Cynthia et Jia Xie (2018). « Distortions in real estate transactions with out-of-state participants ». In : *The Journal of Real Estate Finance and Economics* 57(4), p. 592-617.
- Ibraimovic, Tatjana et Stephane Hess (2018). « A latent class model of residential choice behaviour and ethnic segregation preferences ». In : *Housing Studies* 33(4), p. 544-564.
- Ihlanfeldt, Keith et Tom Mayock (2012). « Information, search, and house prices: Revisited ». In : *The Journal of Real Estate Finance and Economics* 44(1), p. 90-115.
- Jauregui, Andres, Alan Tidwell et Diane Hite (2017). « Sample selection approaches to estimating house price cash differentials ». In : *The Journal of Real Estate Finance and Economics* 54(1), p. 117-137.
- Jin, Jangik et Peter Rafferty (2018). « Externalities of auto traffic congestion growth: Evidence from the residential property values in the US Great Lakes megaregion ». In : *Journal of Transport Geography* 70, p. 131-140.
- Larsen, James E et Joseph W Coleman (2014). « Senior citizen's bargaining power in residential real estate markets ». In : *International Journal of Housing Markets and Analysis*.
- Lu, Zhongming, Frank Southworth, John Crittenden et Ellen Dunham-Jones (2015). « Market potential for smart growth neighbourhoods in the USA: A latent class analysis on heterogeneous preference and choice ». In : *Urban Studies* 52(16), p. 3001-3017.
- Nagy, George (1994). « Terrain visibility ». In : *Computers & Graphics* 18(6), p. 763-773.

- O'Sullivan, David et Alasdair Turner (2001). « Visibility graphs and landscape visibility analysis ». In : *International journal of geographical information science* 15(3), p. 221-237.
- Pope, Jaren C (2008). « Do seller disclosures affect property values? Buyer information and the hedonic model ». In : *Land Economics* 84(4), p. 551-572.
- Prunetti, Dominique, Xavier Pieri et Marie-Antoinette Maupertuis (2015). « Identification empirique d'un double marché résidentiel en Corse ». In : *Le double marché foncier et immobilier en Corse*. Sous la dir. de Guillaume Kessler et Caroline Tafani. Editions L'Harmattan, p. 103-129.
- Quant, Richard E (1988). *The econometrics of disequilibrium*. Blackwell.
- Song, Shunfeng (1995). « Determinants of bargaining outcome in single-family housing transactions: an empirical examination ». In : *Urban Studies* 32(3), p. 605-614.
- Stegmans, Joep et Wolter Hassink (2017). « Financial position and house price determination: An empirical study of income and wealth effects ». In : *Journal of Housing Economics* 36, p. 8-24.
- Tafani, Caroline, Marie-Antoinette Maupertuis et Xavier Pieri (2012). « Impact du tourisme sur le prix du foncier agricole littoral en Corse : la durabilité de l'agriculture en question ». In : *Aménagement durable des territoires méditerranéens*. Sous la dir. de B. Messini. PUAM, p. 73-92.
- Turnbull, Geoffrey K et Casey F Sirmans (1993). « Information, search, and house prices ». In : *Regional Science and Urban Economics* 23(4), p. 545-557.
- Zhou, Xiaorong, Karen Gibler et Velma Zahirovic-Herbert (2015). « Asymmetric buyer information influence on price in a homogeneous housing market ». In : *Urban Studies* 52(5), p. 891-905.