

Annexe 1 : La transmission des actes de ventes aux bases notariales et l'estimation des taux de couverture

Les mutations de logements, dans leur quasi-totalité⁵⁷, donnent lieu à la rédaction d'un acte authentique par un notaire. Ce document est transmis à l'administration fiscale (Conservation des hypothèques, puis services du cadastre et des domaines), pour l'enregistrement, le paiement des droits de mutation dans la plupart des cas, et celui de la TVA si la mutation porte sur un logement neuf. Il existe donc bien une source exhaustive. Mais, d'une part, il s'agit d'un document papier et d'autre part, ce document ne se présente pas sous la forme administrative classique d'un formulaire : c'est un texte entièrement rédigé. Par conséquent, pour constituer un fichier informatique exploitable, c'est-à-dire une base de données, il faut passer par le codage de l'information dans une grille de saisie, opération lourde et coûteuse compte tenu de la taille du document (généralement plusieurs pages) et du volume (environ 800 000 transactions de logements anciens par an sur l'ensemble de la France).

La transmission des informations sur les actes de vente et leur saisie a été entreprise par les notaires eux-mêmes, d'abord par la Chambre interdépartementale des notaires de Paris (CINP) à la fin des années 1970 (et surtout depuis 1990), puis par le Conseil supérieur du notariat, qui a créé à cette fin la société anonyme Perval en 1993. Ces deux organismes reçoivent les informations des études notariales et alimentent chacun une base de données, le premier couvrant l'Île-de-France, le second la province et les DOM. C'est ce qui a permis le développement, toujours à l'initiative des notaires et avec la collaboration de l'Insee, des indices de prix présentés ici. Notons toutefois qu'un indice existait déjà : il avait vu le jour en 1983 et concernait les ventes d'appartements anciens à Paris. Cet indice était calculé à partir d'une moyenne des prix pondérés par le parc de logements au dernier recensement de la population dans 72 strates. Il a été rénové et repose désormais sur la même méthodologie que les autres indices Notaires-Insee.

Jusqu'en 2010 les informations étaient transmises sur papier⁵⁸. Depuis cette date, une partie des actes sont transmis sous forme numérique. La télétransmission est encore en cours de mise en place dans les offices notariaux, tant pour l'alimentation des Conservations des hypothèques (Projet Télé@ctes) que des bases notariales (projet de télétransmission aux bases notariales). Ces projets de longue haleine ont nécessité la modification de l'ensemble des logiciels de rédaction d'actes (LRA) et de l'organisation des offices notariaux afin que toutes les informations nécessaires soient saisies dans les LRA. Fin 2012, environ 50% des actes étaient télétransmis pour les besoins des indices.

Malheureusement, les études notariales ne transmettent pas tous leurs actes, et toutes les études ne transmettent pas. Le taux de couverture des bases notariales est variable d'un département à l'autre. Il est important de le connaître puisqu'il intervient dans le calcul de la pondération des indices.

Taux de couverture

On appelle taux de couverture des bases notariales le rapport entre le montant des transactions figurant dans les bases et le montant total des transactions⁵⁹. Le taux est estimé en comparant le montant des transactions figurant dans les bases notariales avec l'assiette des droits de mutation perçus par la Direction générale des finances publiques (DGFIP). La DGFIP fournissant les montants de droits de mutation perçus dans chaque département, le taux de couverture peut être calculé par département.

En revanche, il n'est pas possible de le faire séparément pour les logements anciens, car ils ne sont plus soumis à des régimes de droits de mutation différents. En effet, avant 1999, il existait un régime de droit de mutation spécifique à l'immobilier d'habitation ancien, dont le produit était enregistré dans la comptabilité de la DGFIP. Il suffisait de diviser ce produit par le taux des droits pour obtenir l'assiette. Depuis les réformes de 1998 et 1999, les transactions de logements anciens sont incluses dans l'assiette du régime de droit commun. Il existe aujourd'hui trois régimes de droits de mutation immobiliers :

- Le régime « de droit commun »

⁵⁷ A l'exception de certaines transactions réalisées par l'administration, ou des transferts de propriété lors des rachats d'entreprises.

⁵⁸ Les études des notaires font une photocopie de l'acte (ou plus souvent de l'extrait d'acte) et elles joignent à ce document un papillon autocollant comportant des informations qui ne sont pas forcément présentes dans l'extrait d'acte.

⁵⁹ Les taux calculés à partir des montants sont supérieurs aux taux calculés à partir du nombre de transactions, car les départements où le taux de couverture est le plus élevé sont en général ceux où le montant moyen des transactions est le plus élevé, notamment ceux de l'Île-de-France.

Ce régime s'applique aux mutations à titre exclusivement onéreux d'immeubles anciens et de terrains non soumis à la taxe sur la valeur ajoutée. Depuis le 1^{er} janvier 2011, il inclut un taux de 3,80%⁶⁰ revenant au département et un taux de 1,20% revenant à la commune, soit en tout 5,00%. Auparavant, il incluait un taux de 3,60% revenant au département, un taux de 1,20% revenant à la commune et, sur les dernières années de la période, un taux de 0,20% revenant à l'État qui n'est désormais plus perçu.

- Le régime « dérogatoire »

Ce régime s'applique :

- aux mutations à titre onéreux d'immeubles neufs et assimilés, de biens achetés en vue d'une revente rapide (« marchands de biens ») et de terrains soumis à la taxe sur la valeur ajoutée,
- à certaines mutations d'immeubles à titre gratuit (donations) ; les mutations immobilières à titre gratuit représentent environ 40% de l'assiette de ce régime en moyenne sur l'ensemble de la France.

Depuis le 1^{er} janvier 2011, le régime dérogatoire inclut un taux revenant au département, de 0,70% sur les mutations à titre onéreux et 0,60% sur les mutations à titre gratuit. Auparavant, ce régime incluait un taux de 0,60% revenant au département, et, sur les dernières années de la période et pour les mutations à titre onéreux, un taux de 0,10% revenant à l'État qui n'est désormais plus perçu.

- L'exonération des droits de mutation

Ce régime s'applique principalement aux acquisitions de l'État, des collectivités locales et de certains établissements publics.

Depuis 1999, les transactions de logements anciens ne faisant plus l'objet d'une ligne spécifique dans la comptabilité de la DGFIP, il est devenu nécessaire d'en estimer le montant. Les transactions de logements anciens représentaient, fin 1999, environ 80% de l'assiette du régime de droit commun. Les contrôles de cohérence effectués depuis indiquent que ce poids élevé a peu varié par la suite⁶¹.

Si le taux de couverture des bases notariales était le même pour l'immobilier d'habitation ancien et pour l'ensemble des biens relevant du régime de droit commun, la difficulté serait résolue : le poids de l'immobilier d'habitation ancien dans l'ensemble du régime de droit commun serait le même que l'on considère l'ensemble des mutations ou les seules mutations présentes dans les bases notariales. Cependant, pour les années antérieures à 2000, pour lesquelles les différents types de biens pouvaient être distingués au sein du produit des droits de mutation, le taux de couverture des bases notariales était légèrement plus élevé (d'environ 2%) pour l'immobilier d'habitation ancien que pour l'ensemble des mutations soumises au régime de droit commun actuel. Ce différentiel variait notablement d'un département à l'autre.

L'hypothèse consiste à approcher le taux de couverture sur les seuls logements anciens par le taux de couverture pour l'ensemble du régime de droit commun. L'erreur ainsi faite peut être élevée pour le taux de couverture sur un département donné, mais cela n'a pas d'effet sur la valeur de l'indice Notaires-Insee sur ce département. L'erreur possible sur la valeur des indices très agrégés (France, Île-de-France et province) est également faible, les erreurs sur les pondérations sur l'ensemble du territoire se compensant mutuellement. C'est à un niveau supra-départemental peu agrégé (par exemple, une région constituée de deux départements) que l'effet sur la valeur des indices est potentiellement le plus élevé, mais les tests effectués ont montré qu'il reste très faible.

⁶⁰ A la date de la rédaction était annoncé un projet de relèvement de ce taux jusqu'à 4,50% (soit une augmentation de 0,7%) à compter du 1^{er} mars 2014 dans les départements qui le souhaitent.

⁶¹ Il s'agit des contrôles effectués par le Conseil scientifique des indices Notaires-Insee lors du calcul du nombre de transactions de logements anciens.

Tableau A1.1 - Taux de couverture des bases notariales (régime de droit commun de droits de mutation)

Année	Île-de-France	Province	France
2000	74,0%	58,2%	62,7%
2001	78,4%	59,0%	64,5%
2002	77,4%	53,6%	60,7%
2003	80,3%	55,3%	62,8%
2004	81,7%	54,1%	62,3%
2005	81,5%	55,6%	63,3%
2006	80,3%	56,5%	63,6%
2007	83,0%	57,6%	64,8%
2008	80,4%	58,6%	64,8%
2009	76,6%	56,0%	62,2%
2010	80,0%	56,5%	63,1%
2011	70,1 %	51,0 %	56,8 %
2012	73,4%	51,9%	58,4%

Source : DGFIP et Notaires de France - Bases Perval et BIEN

Lecture : Les taux de couverture sont relativement stables dans le temps. Cependant ils ont fléchi nettement en 2011. Ceci s'explique par la tendance suivie par le nombre de transactions, qui ont repris après avoir fortement diminué en 2008-2009. La remontée n'a en effet pas pu être accompagnée par le dispositif de codification mis en place.

Annexe 2 : Le découpage géographique

Les indices élémentaires de prix des logements anciens sont calculés à un niveau géographique détaillé, appelé strate, puis agrégés sous forme de somme pondérée par la structure du parc de référence⁶². On découpe donc le territoire à un niveau le plus fin possible pour obtenir un maillage de calcul. Une strate doit respecter deux contraintes pour être réputée d'une qualité suffisante :

- Elle doit contenir un nombre minimum de transactions, égal à 110 pour un trimestre sur la période de l'étude.
- A l'intérieur de la strate, les prix doivent être homogènes.

La stratification a été faite selon deux méthodes différentes en province et en Île-de-France, compte tenu des contextes différents des marchés immobiliers.

- En province, il peut exister des marchés locaux très cloisonnés tels les zones touristiques (stations balnéaires, stations de ski). La démarche est basée sur un critère de proximité géographique en partant de la commune.
- En Île-de-France, on a d'abord défini des zones de prix : deux pour le marché des appartements (Paris composant une zone et le reste de l'Île-de-France l'autre zone), une pour les maisons (du fait du trop faible effectif des maisons à Paris). Puis les communes ont été regroupées sur des critères d'homogénéité des niveaux de prix. Les communes appartenant à une même strate ne sont pas toujours contiguës.

Dans les deux cas, on cherche à avoir des prix homogènes à l'intérieur d'une strate. Cette homogénéité s'apprécie en niveau de prix : prix de vente pour les maisons, prix de vente au m² pour les appartements. On vérifie de plus a posteriori que les strates sont hétérogènes entre elles.

En province, les travaux conduisent à un nombre supérieur de strates dans la nouvelle version des indices :

- 97 strates pour les appartements, contre 88 en 2002,
- 174 strates pour les maisons en 2008 contre 146 en 2002.

L'augmentation du nombre de strates provient aussi bien de l'amélioration de l'alimentation de la base que de l'augmentation du nombre de transactions total.

La liste des communes composant les strates est donnée exhaustivement dans le classeur Excel joint.

En Île-de-France, les travaux ont conduit au contraire à réduire le nombre de strates :

- 5 strates pour les appartements de Paris, contre 18 en 2002,
- 10 strates pour les appartements pour l'Île-de-France hors Paris, contre 37 en 2002,
- 7 strates pour les maisons de l'Île-de-France comme en 2002.

La liste des communes composant les strates est donnée exhaustivement le classeur Excel joint.

La stratification en province

Les transactions analysées couvrent la période du 1^{er} trimestre 1998 au 3^{ème} trimestre 2007, soit 39 trimestres. On commence la stratification à un niveau géographique fin, c'est-à-dire la commune. On retient comme strate toute commune qui dépasse le seuil minimum fixé à 110 transactions par trimestre. Ceci n'est réalisé dans la pratique que pour les plus grandes communes. Si ce n'est pas le cas, on agrège les communes dans une même strate lorsqu'elles présentent des niveaux de prix homogènes. On distingue de plus deux types de communes particulières : les stations de ski et les stations balnéaires.

La province est ainsi découpée en strates de niveaux géographiques et de types de marché immobilier différents. Les niveaux de la segmentation peuvent être :

⁶² Voir définitions du chapitre 3, page 17 et suivantes.

- la commune : c'est le découpage le plus fin⁶³. Ce découpage constitue la première étape du processus. Le nombre de ventes trimestrielles est calculé pour toutes les communes de France ; on conserve celles qui dépassent le seuil minimum de 110 transactions par trimestre.
- l'unité urbaine : on isole les strates par unités urbaines et on retire les communes qui ont déjà été sélectionnées en tant que strate. Il peut donc y avoir des strates formées par une banlieue seule ou bien par l'agglomération tout entière quand une grande ville (ou commune) n'était pas suffisamment importante pour constituer une strate à elle seule.
- les stations de sport d'hiver alpines : elles sont regroupées en strates par massif montagneux voire par département lorsque cela est possible. Cela concerne la Savoie, la Haute-Savoie et l'Isère, vu l'importance des domaines skiables. On crée trois strates de communes pour les appartements, selon des massifs montagneux donnés :
 - o Strate 01 : Tarentaise
 - o Strate 02 : Maurienne, Val d'Arly, Beaufortin, les Bauges, Chablais, Giffre, Les Bornes, et Chartreuse,
 - o Strate 03 : Mont-Blanc et Aravis.

En Isère, on crée une classe particulière pour les communes ayant sur leur territoire une station de ski (soit la strate 04 : Isère). Pour les autres massifs montagneux (Alpes du sud, Pyrénées, Vosges et Massif central), il n'a pas été possible de constituer des strates particulières en raison d'un volume de ventes trop faible.

- les communes littorales : les strates sont obtenues par regroupement géographique de communes de même département ou région. Elles se distinguent des communes proches par un prix moyen des logements plus élevé que dans les communes limitrophes (l'écart est fixé à 15 ou 20 % selon les cas).

Les communes sont regroupées sur la base de la proximité géographique, à partir d'unités de territoire national : la commune, l'unité urbaine, un regroupement de communes littorales, un regroupement de station de ski, jusqu'à un département voire une région.

Quelques strates, qui passent sous le seuil de 110 transactions au cours de quelques trimestres seulement, peuvent être sélectionnées si l'évolution du nombre de mutations y est favorable. Ainsi, pour les appartements, la strate d'Aix-en-Provence est supprimée car le nombre de mutations s'est fortement réduit. Il est passé régulièrement en-dessous du seuil limite à partir de l'année 2003. Au contraire, à Caen le nombre de mutations croît sur la période et se maintient au-dessus du seuil à partir de l'année 1998. Cette strate a donc été créée. (*figures A2.1 et A2.2*).

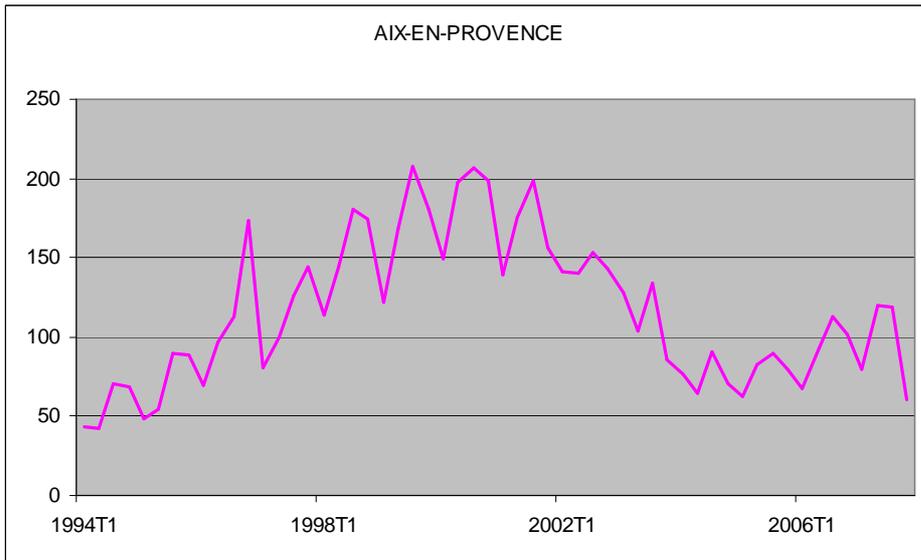
On traite successivement les unités urbaines, les communes littorales et les communes appartenant aux massifs montagneux les plus touristiques, et les communes rurales. Ces dernières n'ont pas un nombre de ventes suffisant par trimestre pour être sélectionnées en tant que telles et n'ont pas de caractéristique particulière non plus, qui aurait pu les faire ressortir. Dans ce cas, on répartit les communes en classes d'après les quantiles de revenu moyen de l'année 2006 par habitant, chaque classe constituant une strate. Ce travail est réalisé à partir d'un processus descendant en regardant d'abord au niveau des zones économiques et d'aménagement du territoire (ZEAT). Chaque strate doit respecter les deux critères suivants :

- un volume de transactions supérieur à 8 000 ou 9 000 mutations entre 1998 et 2007,
- un minimum de 110 transactions par trimestre entre 1998 et 2007.

Si les seuils limites pour une ZEAT sont largement dépassés, on descend au niveau régional. Il en va de même pour une région donnée, on est amené parfois à descendre à un niveau encore plus fin, le département. On regroupe dans d'autres cas deux départements d'une même région pour former une strate (*tableau A2.1* ; cas du regroupement des départements des Ardennes et de la Marne par exemple).

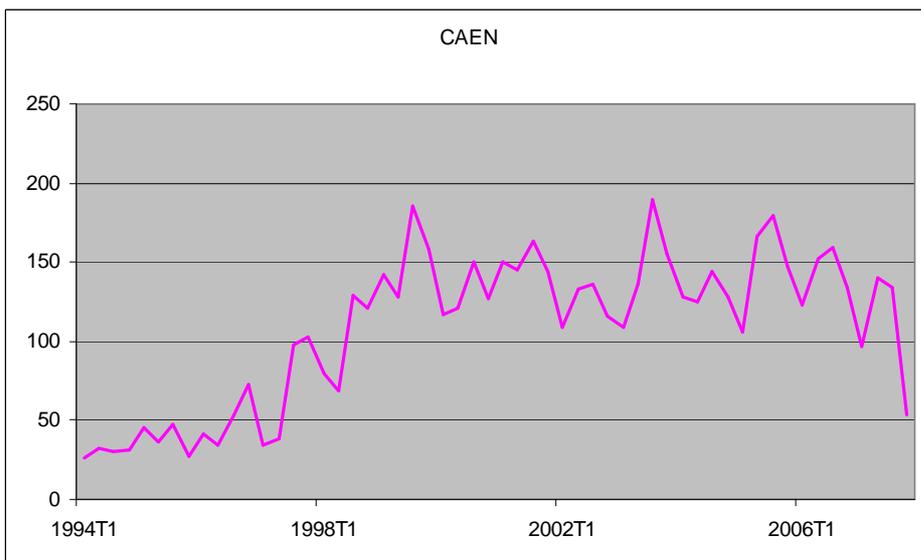
⁶³ Il n'y a pas de strate pour les arrondissements. Les arrondissements de Lyon et Marseille sont traités via des indicatrices dans les régressions.

Figure A2.1 - Nombre de mutations trimestriel à Aix-en-Provence (appartements)



Source : base Perval

Figure A2.2 - Nombre de mutations trimestriel à Caen (appartements)



Source : base Perval

Tableau A2.1 - Exemple d'éclatement de zones géographiques selon la variable « revenu » (maisons)

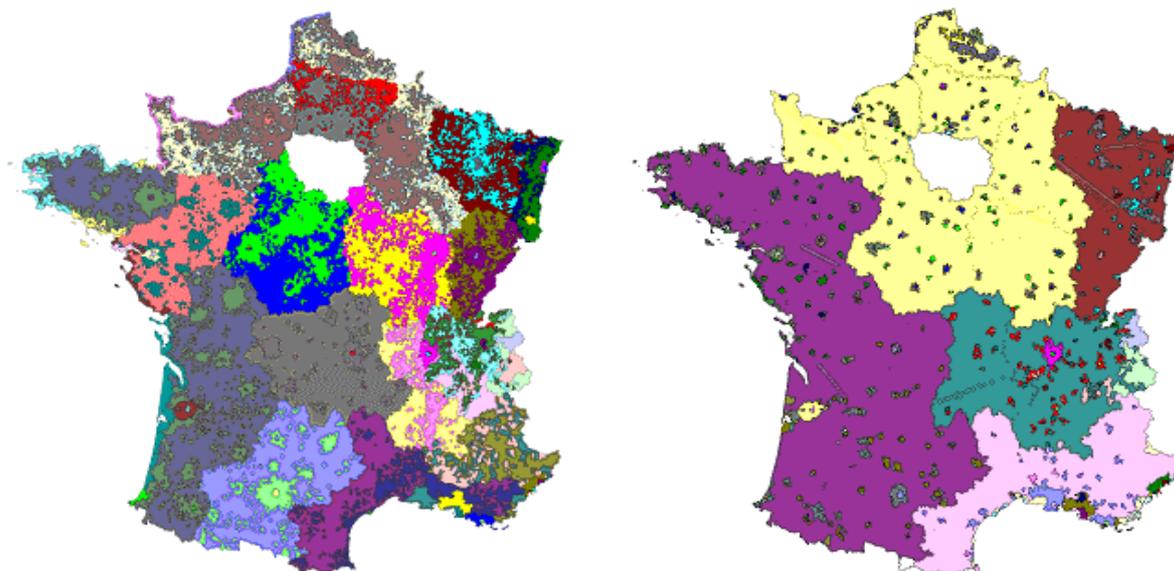
Code de la strate	Département	Région	Borne inférieure de revenu pour la strate (en €)	Borne supérieure de revenu pour la strate (en €)	Nombre total de trimestres où le seuil de 110 logements est dépassé	Nombre total de ventes de 1998 à 2007
99210	Regroupement des départements Ardennes et Marne	Champagne Ardennes	7 824	14 480	39	9 564
99211			14 480	18 821	39	11 535
99212			18 821	55 946	39	10 678
99213	Regroupement des départements Aube et Haute-Marne		8 321	15 106	39	8 921
99214			15 106	55 835	39	8 640
99220	Somme		Picardie	9 342	13 543	39
99221		13 543		15 638	39	9 373
99222		15 638		40 192	39	9 159
99223	Aisne	9 998		14 889	39	11 616
99224		14 889		17 456	39	11 741
99225		17 456		20 190	39	11 635
99226		20 190		72 021	39	12 147
99227	Oise	8 299		13 431	39	8 717
99228		13 431		15 505	39	8 868
99229		15 505		44 817	39	9 118

Source : base Perval

Figure A2.3 - Strates pour les appartements de la province

Nouvelle base (version 3 des indices)

Ancienne base (version 2 des indices)



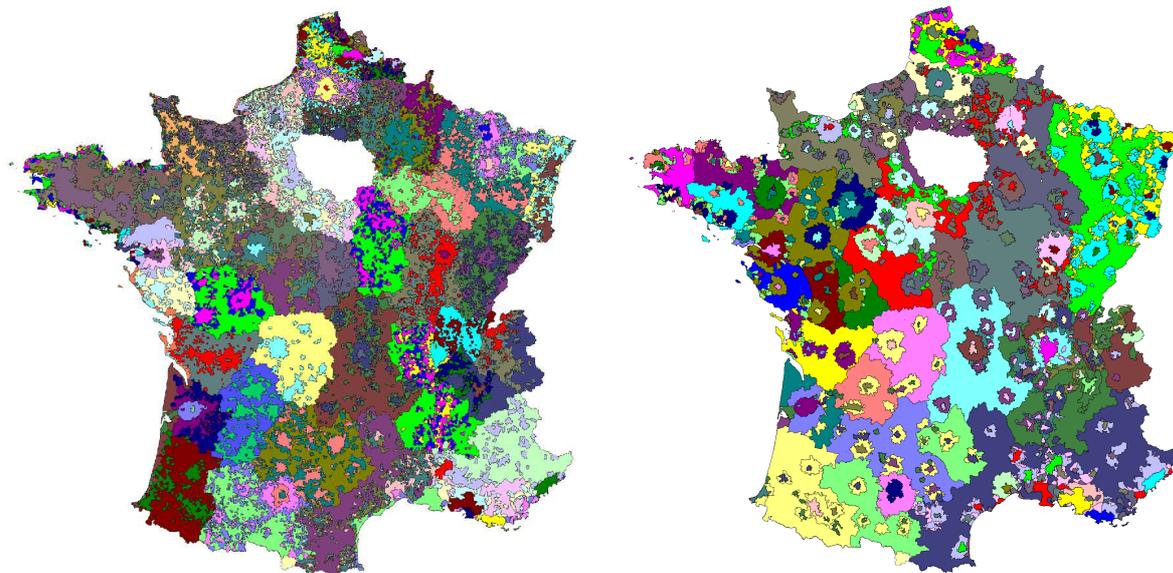
Source : base Perval

Lecture : on distingue une strate donnée grâce à sa couleur. Par exemple, en version 2, la grande tache jaune autour de l'Île-de-France correspond à la strate des communes rurales du Nord-Ouest de la France. Sur le graphique, vu le nombre de strates important, certaines strates ont la même couleur. Elles se différencient par le fait qu'elles ne sont pas contiguës.

Figure A2.4 - Strates pour les maisons de la province

Nouvelle base (version 3 des indices)

Ancienne base (version 2 des indices)



Source : base Perval
Lecture : cf. ci-dessus.

La stratification en Île-de-France

En Île-de-France, les transactions analysées sont celles des années 2003 à 2007, soit 20 trimestres.

La première étape consiste à synthétiser l'information grâce à une analyse factorielle des correspondances (AFC). La stratification proprement dite est effectuée dans une seconde étape grâce à une classification ascendante hiérarchique (CAH).

- Première étape : synthétiser l'information

La CAH (classification ascendante hiérarchique) n'admettant que des variables qualitatives, on discrétise les variables continues. On recherche le meilleur résumé possible de l'information en réduisant le nombre d'axes d'analyse. Chaque modalité de variable peut être vue comme un axe d'analyse. On vérifie que la contribution propre de chaque variable n'est pas trop dominante par rapport à la contribution moyenne sur les premiers axes. La méthode s'appuie, pour établir les coupures, sur des indicateurs statistiques comme les centiles, la moyenne et les écarts-types. Ensuite, un regroupement des modalités peu fréquentes est réalisé. Si les effectifs de la modalité ne représentent pas 5 % du total, elle est regroupée avec une autre.

- Deuxième étape : constituer les strates

Avant de commencer la stratification, on calcule des statistiques au niveau des quartiers (pour Paris) ou des cantons (pour le reste de l'Île-de-France), pour le croisement des variables suivantes : la surface habitable du logement (en tranches), la surface du terrain pour les maisons (en tranches également), le nombre de salles de bain, le nombre de pièces, l'époque de construction, la présence d'ascenseurs pour les appartements, l'étage, la présence de caves, le nombre de garages et la présence de balcons ou de terrasses. Paris, où les prix moyens sont plus élevés, est traité à part. Le prix au m² pour les appartements et le prix de vente pour les maisons sont introduites en variables supplémentaires dans notre analyse. Elles serviront ainsi à caractériser les classes obtenues.

La classification se fait pas à pas, en agrégeant les observations qui « se ressemblent » le plus. Pour cela, on utilise la distance entre deux observations ou deux classes pour regrouper les plus proches. A fur et à mesure de l'agrégation, les classes constituées comportent de plus en plus d'observations.

- Exemple : la classification réalisée pour les appartements de Paris

Dans un premier temps, il faut déterminer le nombre de classes. *L'arbre hiérarchique (figure A2.5)* permet de déterminer visuellement le nombre de classes optimal. Ce graphique renseigne sur l'ordre de formation des classes et mesure également la valeur de la distance d'agrégation des classes. Un grand

« saut » dans le graphique correspond à une forte hausse de la valeur de cette distance. En pratique, on choisit la coupure dans la partie haute du diagramme, c'est là qu'on a des classes construites avec des distances d'agrégation encore faibles (donc des classes assez homogènes) et juste avant un grand saut (en évitant de prendre trop haut dans l'arbre les classes trop hétérogènes). Pour notre cas, le grand saut correspond à une coupure en dessous d'une coupure de 5 classes.

Un second diagramme (*figure A2.6*) permet de connaître pour chaque classe, les modalités de variables et les individus (les quartiers regroupés par arrondissement) qui contribuent le plus à la détermination de la classe. Le diagramme permet de caractériser les cinq classes obtenues en termes de poids dans le total des transactions réalisées à Paris, de modalités de variables bien représentées dans la classe et de répartition géographique des logements. Le croisement de ces trois critères donne le profil de chaque classe.

Profil de la classe 1

Cette classe représente 20 % des transactions de Paris. Les modalités des variables les mieux représentées sont :

- prix par m² élevé,
- époque de construction ancienne,
- appartement sans cave,
- appartement sans ascenseur,
- appartement sans garage.

La répartition géographique des logements de cette classe s'étend à :

- 100 % des logements des 1^{er}, 2^{ème}, 3^{ème}, 4^{ème} et 6^{ème} arrondissements,
- 100 % des logements du 9^{ème} arrondissement,
- 72 % des logements du 10^{ème} arrondissement,
- 70 % des logements du 5^{ème} arrondissement,
- 50 % des logements du 7^{ème} arrondissement.
- 20% du 12^{ème}

Profil de la classe 2

Cette classe représente 34 % des transactions de Paris. Les modalités des variables les mieux représentées sont :

- prix par m² moyen,
- époque de construction ancienne,
- petite surface entre 20 et 40 m²,
- appartement sans ascenseur,
- appartement sans salle de bain,
- appartement sans garage.

La répartition géographique des logements de cette classe s'étend à :

- 100 % des logements des 11^{ème} et 18^{ème} arrondissements,
- 63 % des logements du 17^{ème} arrondissement,
- 50 % des logements du 20^{ème} arrondissement,
- 39 % des logements du 14^{ème} arrondissement
- 30 % des logements du 5^{ème} arrondissement,
- 28% du 10^{ème} et du 13^{ème}.

Profil de la classe 3

Cette classe couvre 24 % des transactions de Paris. Les modalités des variables les mieux représentées dans cette classe sont :

- prix par m² plutôt élevé,
- époque de construction plutôt récente,
- surface moyenne,
- appartement avec garage,
- appartement avec cave
- appartement avec ascenseur.

La répartition géographique des logements de cette classe s'étend à :

- 100 % des logements du 15^{ème} arrondissement,
- 80% des logements du 12^{ème} arrondissement,
- 47 % des logements du 14^{ème} arrondissement,
- 30% des logements du 19^{ème} arrondissement,
- 27 % des logements du 20^{ème} arrondissement.
- 12% du 17^{ème}.

Profil de la classe 4

Cette classe représente 10 % des transactions de Paris. Les modalités des variables les mieux représentées sont :

- prix par m² bas,
- époque de construction récente,
- appartement avec garage,
- appartement avec cave.

La répartition géographique des logements de cette classe s'étend à :

- 72% des logements du 13^{ème} arrondissement,
- 70% des logements du 19^{ème} arrondissement,
- 23% des logements du 20^{ème} arrondissement,
- 15% des logements du 14^{ème} arrondissement.

Profil de la classe 5

Cette classe représente 12 % des transactions de Paris. Les modalités des variables les mieux représentées sont :

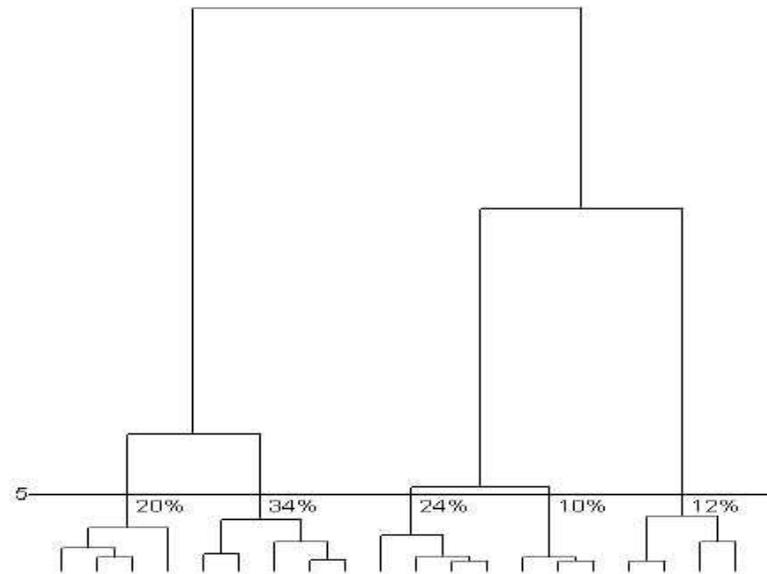
- Prix par m² élevé,
- époque de construction des années 1950 à 1970,
- appartement de grande surface,
- appartement avec 2 salles de bain ou plus,
- appartement avec ascenseur.

La répartition géographique des logements de cette classe s'étend à :

- 100 % des logements des 8^{ème} et 16^{ème} arrondissements,
- 50 % des logements du 7^{ème} arrondissement,
- 15 % des logements du 17^{ème} arrondissement.

La figure A2.7 permet de visualiser la répartition spatiale des cinq classes obtenues pour les appartements de Paris.

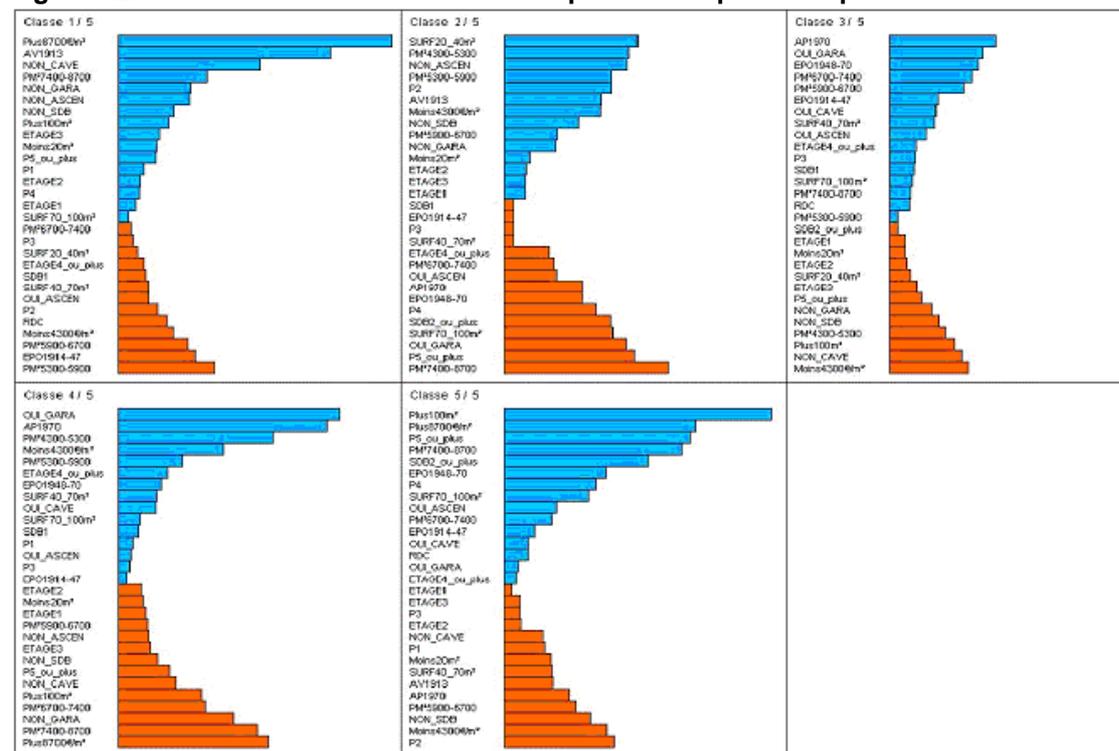
Figure A2.5 - Arbre hiérarchique pour les appartements de Paris



Source : Base BIEN

Lecture : la ligne droite définissant une coupe à 5 classes correspond à un choix optimal en termes de stratégie.

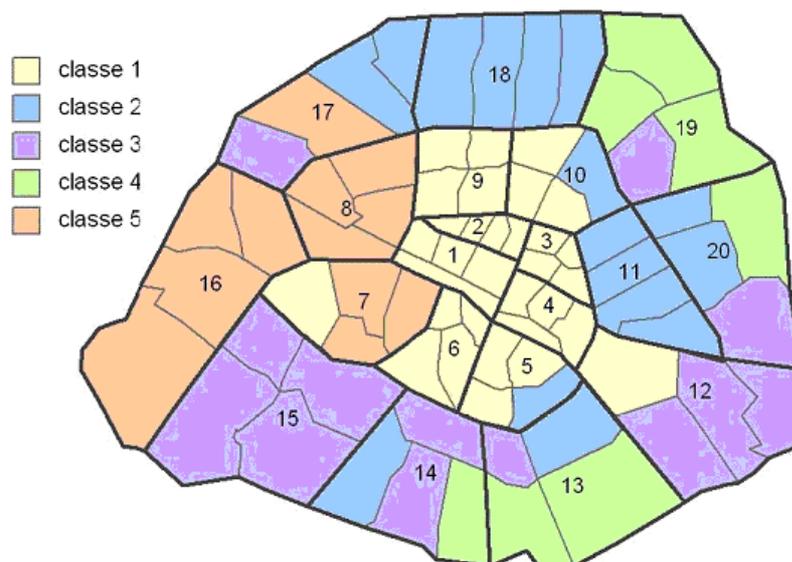
Figure A2.6 - Recherche des variables surreprésentées pour chaque classe



Source : Base BIEN

Lecture : Les bâtons bleus correspondent aux variables qui sont surreprésentées dans la classe.

Figure A2.7 - Découpage en classes pour les appartements de Paris



Source : Base BIEN

Les tests d'homogénéité des strates

Une fois le découpage en strates réalisé, on vérifie l'homogénéité des prix à l'intérieur d'une strate, et l'hétérogénéité inter-strate. Nous allons prendre comme exemple les travaux réalisés par PNS pour les appartements de Paris. Ceux de Min.not sont similaires.

Le tableau A2.2 permet de faire un premier examen visuel sur les moyennes de prix selon les strates pour les appartements de Paris. Ces moyennes, sur la période 2003 - 2007, varient de 4 514 € à 6 549 € le mètre carré. Les intervalles de confiance sont le plus souvent disjoints, ce qui semble indiquer que les prix sont différents d'une strate à l'autre.

Tableau A2.2 - Résultats statistiques relatifs aux prix au m² dans les strates de Paris pour les appartements

Strate	Nombre de transactions	Moyenne	Écart-type	Erreur standard	Intervalle de confiance à 95% pour la moyenne		Minimum	Maximum
					Borne inférieure	Borne supérieure		
7501	24 724	6 190,61	2 037,04	12,96	6 165,21	6 216,00	322,58	24 851,37
7502	40 714	4 919,52	1 323,92	6,56	4 906,66	4 932,38	148,24	18 152,17
7503	28 771	5 453,07	1 361,13	8,02	5 437,34	5 468,80	435,15	19 402,71
7504	12 488	4 514,04	1 208,99	10,82	4 492,83	4 535,24	431,97	17 784,15
7505	16 339	6 549,16	1 874,81	14,67	6 520,41	6 577,90	877,37	22 000,00
Total	123 036	5 474,97	1 707,11	4,87	5 465,43	5 484,51	148,24	24 851,37

Source : PNS

La méthode de Levene permet de vérifier que n échantillons extraits d'une même population ont une variance égale (tableau A2.3). On réalise enfin un test de Fisher (tableau A2.4). Une analyse de la variance à un facteur permet également de rechercher si la zone géographique (facteur) a une influence sur le prix du bien. L'hypothèse d'homogénéité en termes de prix des strates peut être refusée pour un niveau de risque $\alpha = 0,05$. Il existe donc des variances différentes et des effets moyens différents. Mais on ne sait pas si les cinq strates sont différentes ou bien si une partie seulement des strates ne sont pas homogènes entre elles. Pour résoudre ce problème, dans un deuxième temps, un test post-hoc (tableau A2.5) est réalisé pour voir si toutes les strates sont différentes deux à deux. Dans notre exemple, ce test compare les cinq strates parisiennes. Pour chaque strate, le niveau de significativité est inférieur au seuil $\alpha = 0,05$. Des différences de prix sont donc constatées entre deux

strates pour les appartements de Paris. Des résultats équivalents sont trouvés pour les autres strates constituées en Île-de-France et en province.

Tableau A2.3 - Test d'homogénéité des variances

Statistique de Levene W	Degrés de liberté 1	Degrés de liberté 2	Pvalue du test
1 276,53	4	123 031	0,000

Source : Base BIEN

H0 : il n'y a pas de différence de variance entre les strates

H1 : il existe au moins une variance différente dans une strate

Si l'hypothèse H0 n'est pas confirmée, cela indique que le résultat n'est pas le fruit d'un tirage simple provenant d'une même population. Dans ce cas, les variances des groupes ne sont plus égales entre elles au risque α .

Au risque α égal à 0,05, on rejette l'hypothèse H0 ; les variances ne sont pas égales.

Tableau A2.4 - Analyse de variance pour les strates des appartements dans Paris

	Somme des carrés	Degrés de liberté	Moyenne des carrés	Statistique de Fisher	Niveau de significativité du test de Fisher F (Pvalue)
Variance inter-strate	55 621 427 170,7	4	13 905 356 792,5	5 647,5	>0,000
Variance intra-strate	302 929 525 100,8	123 031	2 462 221,1		
Total	358 550 952 270,9	123 035			

Source : Base BIEN

H0 : il n'y a pas de différence de prix entre les moyennes des strates

H1 : il existe une différence de prix entre les moyennes des strates

Tableau A2.5 - Test post-hoc pour les strates des appartements dans Paris

(I) Strate de référence	(J) Comparée avec les strates	Différence des moyennes (I-J)	Erreur standard	Signification	Intervalle de confiance à 95%	
					Borne supérieure	Borne inférieure
7501	7502	1 271,08	14,52	0,03	1 227,42	1 314,75
	7503	737,54	15,24	0,03	690,34	784,74
	7504	1 676,57	16,88	0,02	626,24	1 726,90
	7505	-358,55	19,57	0,02	-417,15	-299,95
7502	7501	-1 271,08	14,52	0,03	-1 314,75	-1 227,42
	7503	-533,55	10,37	0,03	-566,18	-500,92
	7504	405,49	12,65	0,00	370,30	440,67
	7505	-1 629,63	16,07	0,00	-1 674,40	-1 584,86
7503	7501	-737,54	15,24	0,03	-784,74	-690,34
	7502	533,55	10,37	0,03	500,92	566,18
	7504	939,04	13,47	0,00	901,05	977,02
	7505	-1 096,08	16,72	0,00	-1 143,17	-1 049,00
7504	7501	-1676,57	16,88	0,02	-1 726,90	-1 626,24
	7502	-405,49	12,65	0,00	-440,67	-370,30
	7503	-939,04	13,47	0,00	-977,02	-901,05
	7505	-2035,12	18,23	0,01	-2 086,88	-1 983,36
7505	7501	358,55	19,57	0,02	299,95	417,15
	7502	1629,63	16,07	0,00	1 584,86	1 674,40
	7503	1096,08	16,72	0,00	1 049,00	1 143,17
	7504	2035,12	18,23	0,01	1 983,36	2 086,88

Source : Base BIEN

Description des strates

Les strates sont des zones géographiques à l'intérieur desquelles les prix sont homogènes (cf. chapitre 3). Leur découpage communal est précisé dans un classeur Excel, qui peut être obtenu sur demande auprès des notaires.

Strates de province

Pour la province, la construction des strates est fondée sur un critère de proximité géographique en partant de la commune. On trouvera dans le tableau A2.6, la répartition des transactions du parc de référence de 2007-2008 en province, selon la tranche d'unités urbaines 1999.

Tableau A2.6 - Province : répartition des transactions (appartements et maisons) selon la tranche d'unité urbaine et le regroupement de commune (%)

Tranche d'unité urbaine 1999	Rural	UU < 5 000 hab.*	5 000-10 000 hab.	10 000-20 000 hab.	20 000-50 000 hab.	50 000-100 000 hab.	100 000-200 000 hab.	Plus de 200 000 hab.	Total
Appartements									
Rural	6%	0%	0%	0%	0%	0%	0%	0%	6%
Ville centre	0%	2%	3%	5%	6%	9%	8%	33%	67%
Banlieue	0%	0%	0%	1%	1%	2%	3%	20%	27%
Total	6%	2%	3%	6%	8%	11%	11%	53%	100%
Maisons									
Rural	39%	0%	0%	0%	0%	0%	0%	0%	39%
Ville -centre	0%	8%	6%	5%	5%	4%	2%	5%	35%
Banlieue	0%	1%	1%	2%	2%	3%	3%	14%	26%
Total	39%	8%	7%	7%	7%	8%	5%	20%	100%

* Hab. = habitants

Parc de référence 2007-2008

Strates pour les indices des appartements de province

L'ensemble du territoire a été découpé en 97 strates pour les appartements anciens : 33 strates sont constituées d'une seule ville (*tableau A2.7*), ce qui permet de calculer un indice pour chacune de ces 33 villes. Concernant les autres strates (*tableaux A2.8 et suivants*) : 12 sont des banlieues de ces villes, 5 correspondent à la totalité d'une unité urbaine, 14 sont composées de communes atypiques (littorales et stations de sports d'hiver) ; les 33 strates restantes sont des regroupements de villes en fonction de leur région et de leur revenu moyen par habitant (voir les définitions ci-après).

Un certain nombre de séries d'indices n'ayant pas l'appellation Notaires-Insee sont diffusées par les notaires. Il s'agit essentiellement des séries régionales, départementales ainsi que celles des 33 communes du *tableau A2.7*.

Les supports de diffusion sont les sites internet immoprix.com et perval.fr, les conférences de presse organisées par les chambres des notaires sur les prix de l'immobilier ainsi que les rapports d'expertise réalisés par les notaires.

La liste des indices diffusables fait l'objet d'une vérification tous les deux ans environ. Une série d'indice est diffusable lorsque le nombre de transactions par trimestre est supérieur ou égal à 110 sur plusieurs trimestres consécutifs.

Tableau A2.7 - Liste des 33 villes de province faisant l'objet d'un indice pour les appartements

Commune	Population*	Nombre de logements*	Nombre de transactions du parc de référence**
Antibes	76 994	27 135	1 871
Agde	22 487	3 862	1 719
Annecy	50 115	20 372	1 020
Besançon	117 599	36 396	1 974
Bordeaux	235 891	78 724	2 539
Brest	142 097	40 430	2 411
Caen	109 899	33 206	942
Cannes	72 939	26 181	2 836
Le Cannet	40 940	15 169	1 205
Clermont-Ferrand	139 006	42 322	1 349
Dijon	151 576	52 136	2 628
Grenoble	156 659	61 561	2 212
Le Havre	178 769	32 486	1 304
Lille	225 784	61 103	2 328
Lyon	474 946	188 776	6 808
Marseille	851 420	244 867	6 160
Menton	28 833	9 775	1 133
Metz	122 838	35 076	1 566
Montpellier	252 998	84 784	2 399
Nancy	106 361	40 915	1 756
Nantes	283 288	82 438	4 123
Nice	439 553	132 175	8 541
Orléans	113257	28 652	1 593
Pau	84 036	27 637	1 209
Reims	181 468	33 262	1 913
Rennes	206 655	65 019	2 972
Rouen	109 425	38 595	1 448
Saint-Étienne	172 696	53 944	1 743
Strasbourg	272 116	84 070	2 929
Toulon	166 733	50 808	1 906
Toulouse	439 553	152 830	4 674
Tours	135 480	37 007	1 380
Villeurbanne	141 106	47 334	1 963

* Source : recensement de la population de 2008

** Parc 2007-2008

A l'exception de Lyon et de Marseille, les indices associés n'ont pas l'appellation Notaires-Insee.

Tableau A2.8 - Liste des 17 strates (banlieue ou unité urbaine de province) faisant l'objet d'indices pour les appartements

Strate	Nombre de transactions du parc de référence*
Banlieue de Lille (partie française)	1 903
Banlieue de Annecy	1 111
Banlieue de Bordeaux	1 861
Banlieue de Grenoble	2 340
Banlieue de Nancy	1 458
Banlieue de Nantes	1 099
Banlieue de Rouen	936
Banlieue de Strasbourg (partie française)	1 547
Banlieue de Nice	4 941
Banlieue de Toulon	3 807
Banlieue de Lyon	4 689
Banlieue de Marseille-Aix-en-Provence	2 108
Unité urbaine de Chambéry	1 148
Unité urbaine de Mulhouse	1 602
Unité urbaine de Saint-Nazaire	848
Unité urbaine de Genève-Annemasse (partie française)	1 944
Unité urbaine de Bayonne	1 953

* Parc 2007-2008

Douze strates sont constituées de la banlieue d'une unité urbaine et cinq sont composées de l'ensemble de l'unité urbaine. Les indices associés n'ont pas l'appellation Notaires-Insee et ne sont pas publiés.

Tableau A2.9 - Liste des 14 strates de montagne et du littoral faisant l'objet d'indices pour les appartements

Strate	Nombre de transactions du parc de référence*
Littoral Grand Nord des régions Nord-Pas-de-Calais et Picardie	2 013
Littoral des régions Haute et Basse Normandie	2 228
Littoral Breton des départements Ille-et-Vilaine et Morbihan	2 062
Littoral Breton des départements Finistère et Côte d'Armor	2 262
Littoral de la région Pays de la Loire	1 140
Littoral Sud-ouest des régions Poitou-Charentes et Aquitaine	2 584
Littoral Languedoc-Roussillon Sud	2 457
Littoral Languedoc-Roussillon Nord	2 038
Littoral PACA Ouest	1 551
Littoral PACA Est	3 058
Stations de ski de Tarentaise	1 871
Stations de ski de Maurienne, Val d'Arly, Beaufortin, les Bauges, Chablais, Giffre, Les Bornes et Chartreuse	1 775
Stations de ski de Mont-Blanc et Aravis	1 136
Stations de ski de l'Isère	1 198

* Parc 2007-2008

Quatorze strates sont composées de communes « atypiques » : communes littorales, stations de sport d'hiver, qu'il a été nécessaire d'isoler pour obtenir des strates cohérentes dans les régressions. Les indices associés n'ont pas l'appellation Notaires-Insee et ne sont pas publiés.

Tableau A2.10 - Autres strates faisant l'objet d'indices pour les appartements, en fonction du revenu moyen 2006

Strate	Nombre de transactions du parc de référence selon la tranche de revenu*		
	Revenu faible	Revenu moyen	Revenu élevé
Autres communes d'Alsace	2 052		1 594
Autres communes de Bourgogne	1 703		1 444
Autres communes de Franche-Comté	1 657		1 512
Autres communes de Lorraine	2 669		2 505
Autres communes de l'Ouest sauf Pays de la Loire	2 580		1 812
Autres communes de PACA Nord	1 775	1 667	1 454
Autres communes de Picardie	1 867		1 419
Autres communes du Centre	2 526		1 956
Autres communes du Limousin et d'Auvergne	1 763		1 503
Autres communes du Midi-Pyrénées	2 076		1 596
Autres communes du Nord sauf Picardie	1 908	2 118	1 518
Autres communes du Pays de la Loire	1 936		1 689
Autres communes de Rhône-Alpes-Est	2 051	1 900	2 498
Autres communes de Rhône-Alpes-Ouest	2 404		2 198
Autres communes du Sud-est (Languedoc-Roussillon et PACA)	2 016		3 236

* Parc 2007-2008

Les communes des 33 strates ci-dessus n'ont pas un nombre de ventes suffisant par trimestre pour être sélectionnées en tant que telles, et n'ont pas de caractéristiques particulières non plus, qui auraient pu les faire ressortir. Dans ce cas, on les a réparties en classes, en fonction du revenu moyen de l'année 2006 par habitant. Les indices associés n'ont pas l'appellation Notaires-Insee et ne sont pas publiés.

Strates pour les indices des maisons de province

Trente-sept strates concernent des villes, des banlieues ou la totalité de l'unité urbaine (*tableau A2.11*) ; les cent-trente-sept autres sont des regroupements de villes en fonction de leur région et de leur revenu moyen par habitant observé en 2006 (*tableau A2.12*).

Tableau A2.11 - Liste des 37 villes, banlieues et unités urbaines faisant l'objet d'un indice pour les maisons

Commune	Nombre de transactions du parc de référence**
Amiens	970
Bordeaux	1 192
Le Havre	868
Le Mans	1 394
Lille	931
Marseille	826
Nantes	1 146
Toulouse	744
Tourcoing	1 060
Banlieue de Bordeaux	3 631
Banlieue de Lille (partie française)	6 772
Banlieue de Marseille-Aix-en-Provence	1 990
Banlieue de Nantes	2 640
Banlieue de Toulouse	2 222
Unité Urbaine d'Avignon	949
Unité Urbaine de Béthune	1 417
Unité Urbaine de Brest	1 227
Unité Urbaine de Calais	1 129
Unité Urbaine de Clermont-Ferrand	929
Unité Urbaine de Dijon	907
Unité Urbaine de Douai-Lens	3 298
Unité Urbaine de Dunkerque	970
Unité Urbaine de Grenoble	800
Unité Urbaine de Lyon	3 811
Unité Urbaine de Maubeuge (partie française)	903
Unité Urbaine de Metz	1 547
Unité Urbaine de Montpellier	815
Unité Urbaine de Nancy	2 258
Unité Urbaine de Nice	3 241
Unité Urbaine de Orléans	1 966
Unité Urbaine de Reims	906
Unité Urbaine de Rennes	1 047
Unité Urbaine de Rouen	2 708
Unité Urbaine de Saint-Nazaire	770
Unité Urbaine de Toulon	2 089
Unité Urbaine de Tours	1 709
Unité Urbaine de Valenciennes (partie française)	1 991

* Parc 2007-2008

Hormis celui de Lille, les indices associés n'ont pas l'appellation Notaires-Insee et ne sont pas publiés.

Tableau A2.12 - Autres strates pour les maisons en province

Strate	Nombre de transactions du parc de référence**			
	Revenu			
	Très faible	Faible	Moyen	Élevé
Stations de Ski d'Isère, Savoie et Haute Savoie				1 597
Littoral de Gironde, des Landes et des Pyrénées-Atlantiques		1 573		1 127
Littoral de Loire-Atlantique, Vendée et Charente-Maritime		2 186		1 762
Littoral d'Ille-et-Vilaine, et du Morbihan	1 128	985	958	1 044
Littoral du Finistère et de Côte d'Armor	1 434	1 263	1 139	1 186
Littoral Du Nord-Pas-de-Calais et des deux Normandie	1 244	1 385	1 359	1 148
Littoral Méditerranéen		1 789	1 490	1 583
Autres communes d'Alsace		1 567	1 685	1 686
Autres communes d'Aquitaine du Nord-est		2 020		1 852
Autres communes d'Aquitaine du Sud-ouest		2 179		2 301
Autres communes d'Auvergne		1 409	1 779	1 985
Autres communes de Basse-Normandie	2 320	2 220	2 219	2 134
Autres communes de Bourgogne Est (21,71)		2 287	2 254	2 184
Autres communes de Bourgogne Ouest (89,58)		1 453	1 620	1 967
Autres communes de Bretagne du Nord		2 640	2 455	2 309
Autres communes de Bretagne du Sud		2 371		2 451
Autres communes de Champagne-Ardenne Nord		1 951	1 996	1 940
Autres communes de Champagne-Ardenne Sud		1 836		1 703
Autres communes de Franche-Comté		2 308	2 148	2 134
Autres communes de Gironde		1 908		2 003
Autres communes de la Somme		1 611	1 633	1 511
Autres communes de l'Ain		2 117		2 236
Autres communes de l'Aisne	1 760	1 976	1 723	1 934
Autres communes de l'Eure		1 999	1 853	1 870
Autres communes de l'Isère		1 889		1 973
Autres communes de Loire-Atlantique		2 354		2 141
Autres communes de l'Oise		1 468	1 455	1 450
Autres communes de Lorraine Est (54,57)		2 429	2 408	2 548
Autres communes de Lorraine Ouest (55,88)		2 279		2 126
Autres communes de Midi-Pyrénées du Nord		1 272	1 144	1 537
Autres communes de Midi-Pyrénées du Sud		1 500	1 479	1 624
Autres communes de PACA du Nord		1 708		1 717
Autres communes de PACA du Sud		1 398	1 570	1 187
Autres communes de Poitou-Charentes Nord		1 703	1 792	1 621
Autres communes de Poitou-Charentes Sud		3 159		3 103
Autres communes de Savoie et de Haute-Savoie		1 532	1 525	1 425
Autres communes de Seine-Maritime		1 670	1 917	1 844
Autres communes de Vendée		1 703		1 609
Autres communes d'Eure-et-Loir		1 733	1 713	1 870
Autres communes d'Indre et d'Indre-et-Loire		1 314	1 442	1 536
Autres communes du Languedoc-Roussillon	1 569	1 611	1 917	1 763
Autres communes du Limousin		1 953		2 034
Autres communes du Loir et Cher		1 879		1 785
Autres communes du Loiret		1 493	1 403	1 574
Autres communes du Cher		1 362		1 283
Autres communes du Nord	1 608	1 604	1 580	1 542
Autres communes du Pas-de-Calais		1 625	1 660	1 768
Autres communes du Pays de la Loire Nord	2 904	2 731	3 012	2 942
Autres communes du Rhône-Alpes Ouest	1 883	2 031	1 936	2 017

* Parc 2007-2008

Les indices associés n'ont pas l'appellation Notaires-Insee et ne sont pas publiés.

Strates d'Île-de-France

En Île-de-France, les strates ont été déterminées par une analyse en composantes principales du prix moyen par commune. Il en résulte que les strates ne sont pas construites selon un critère de proximité géographique et qu'elles ne peuvent être identifiées par un libellé en clair comme en province.

On trouvera ci-dessous la répartition départementale du nombre de transactions par strate (transactions du parc de référence 2007 -2008) et par type de logement.

Tableau A2.13 - Île-de-France : nombre de strates par département (hors Paris) et effectifs correspondants du parc de référence* ; appartements

Appartements*								
Département	77	78	91	92	93	94	95	Total
Strate								
1	2 089	434	2 086	-	-	126	986	5 721
2	2 229	158	73	-	-	162	314	2 936
3	1 217	475	1 119	-	1 204	-	1 546	5 561
4	523	1 532	1 728	-	2 348	2 863	1 895	10 889
5	1 612	3 628	3 927	-	846	831	2 021	12 865
6	-	1 596	-	5 308	4 319	6 102	2 082	19 407
7	-	2 172	-	1 527	-	-	-	3 699
8	271	3 636	1 781	3 147	-	31	326	9 192
9	-	-	-	8 528	-	3 238	-	11 766
10	-	-	149	7 294	6 150	4 649	-	18 242
Total	7 941	13 631	10 863	25 804	14 867	18 002	9 170	100 278

* Parc 2007-2008

Tableau A2.14 - Île-de-France : nombre de strates par département (Paris inclus) et effectifs correspondants du parc de référence* ; maisons

Maisons*									
Strate	75	77	78	91	92	93	94	95	Total
1	133	234	1 321	259	2 248	-	1 163	1 081	6 439
2	-	-	-	-	690	709	379	-	1 778
3	-	-	984	1 082	109	4 988	2 040	2 623	11 826
4	-	1 603	466	1 959	-	770	829	-	5 627
5	-	2 304	956	828	-	-	484	3 193	7 765
6	-	2 867	3 918	3 009	49	-	332	154	10 329
7	-	6 801	2 110	1 817	-	-	-	1 300	12 028
Total	133	13 809	9 755	8 954	3 096	6 467	5 227	8 351	55 792

Lecture : Pour les maisons, la strate 1 inclut Paris.

* Parc 2007-2008

Strates pour les appartements d'Île-de-France

Tableau A2.15 - Liste des cinq strates pour les indices des appartements à Paris

Strate	Quartiers	Nombre de transactions du parc de référence*
7501	1 à 17, 19 à 24, 28, 33 à 39 et 48	9 346
7502	18, 40 à 44, 49, 56, 67 à 72, 77, 79	15 139
7503	45 à 47, 52, 53, 55, 57 à 60, 65, 76, 80	10 689
7504	50, 51, 54, 73 à 75, 78	4 635
7505	25 à 27, 29 à 32, 61 à 64, 66	6 002
Total	80 quartiers	45 811 transactions

* Parc de référence 2007-2008

Cinq strates ont été constituées à partir des 80 quartiers de Paris.

Tableau A2.16 - Liste des dix strates pour les indices des appartements en Île-de-France hors de Paris

Strate	Nombre de communes	Nombre de transactions du parc de référence*
1	790	5 721
2	87	2 936
3	23	5 561
4	41	10 889
5	184	12 865
6	51	19 407
7	3	3 699
8	61	9 192
9	13	11 766
10	27	18 242
Total	1280 communes	100 278 transactions

* Parc de référence 2007-2008

Dix strates pour les appartements ont été constituées à partir des 1 280 communes de l'Île-de-France hors Paris. Ces indices n'ont pas l'appellation Notaires-Insee et ne sont pas publiés.

Strates pour les maisons d'Île-de-France

Tableau A2.17 - Liste des sept strates pour les indices des maisons en Île-de-France

Strate	Nombre de communes	Nombre de transactions du parc de référence
1	64	6 439
2	23	1 778
3	81	11 826
4	56	5 627
5	98	7 765
6	286	10 329
7	673	12 028
Total	1281 communes	55 792 transactions

* Parc de référence 2007-2008

Sept strates ont été constituées pour les maisons de l'Île-de-France, à partir des 1 281 communes d'Île-de-France. Paris est inclus dans la strate 1 car le volume de ventes de maisons y est faible. Ces indices n'ont pas l'appellation Notaires-Insee et ne sont pas publiés.

Annexe 3 : Mise à jour de la spécification des modèles

La spécification des modèles est revue à chaque renouvellement de base. L'essentiel des modifications de la version 3 par rapport à la version 2 ont porté sur les modalités de prise en compte de la surface du terrain et de la surface habitable pour les maisons ; on a recherché également une harmonisation la plus complète possible des modèles d'Île-de-France et de province.

Modification du traitement de la surface des maisons

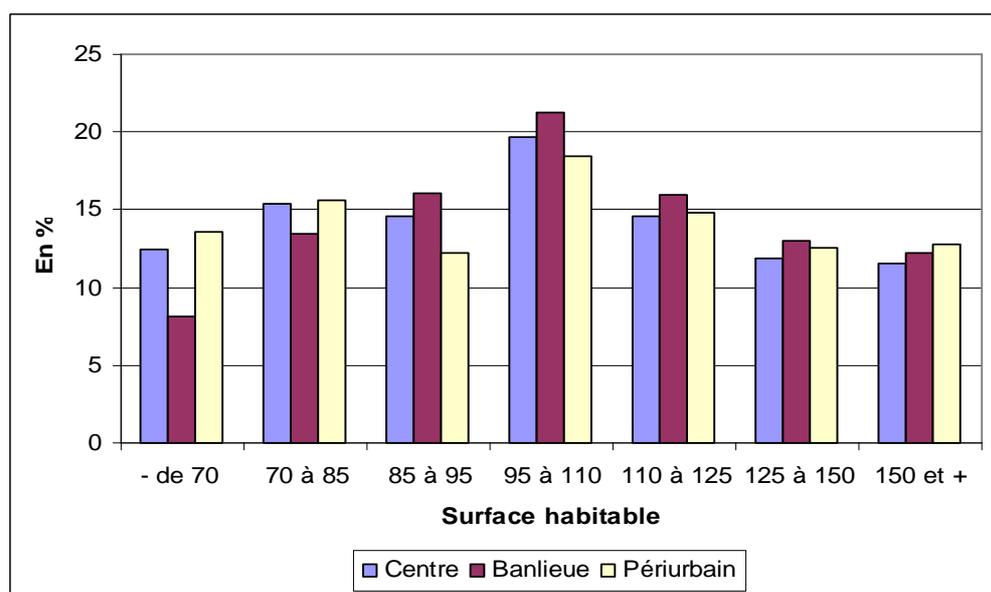
Analyse des variables de surface

Les prix constatés sont utilisés pour estimer la valeur d'un bien de référence, qui lui-même sert à estimer la valeur du parc de référence au temps t . Dans la version précédente des indices, la valeur de référence des variables continues n'avait pas été fixée, si bien qu'elle se trouvait mise à zéro par défaut. L'estimation des prix relatifs n'en est pas altérée. Cependant, il est plus correct de choisir comme référence un bien dont la surface de terrain et la surface habitable ne seraient pas nulles.

La question s'est posée de savoir s'il fallait établir des valeurs de référence différentes entre les communes centres, de banlieue et rurales. On a donc comparé les distributions respectives selon le type de commune. Au final, les mêmes valeurs ont été prises en Île-de-France et en province.

Pour la province, la classe modale de la surface habitable du logement est identique (figure A3.1) quel que soit le type de commune [classe 95 à 110 m²]. Il n'y a pas de différences notables entre les trois séries, on peut noter que les communes de banlieue sont moins représentées dans les classes de petites surfaces (en dessous de 85 m²).

Figure A3.1 - Répartition des transactions selon la surface habitable de la maison, en province



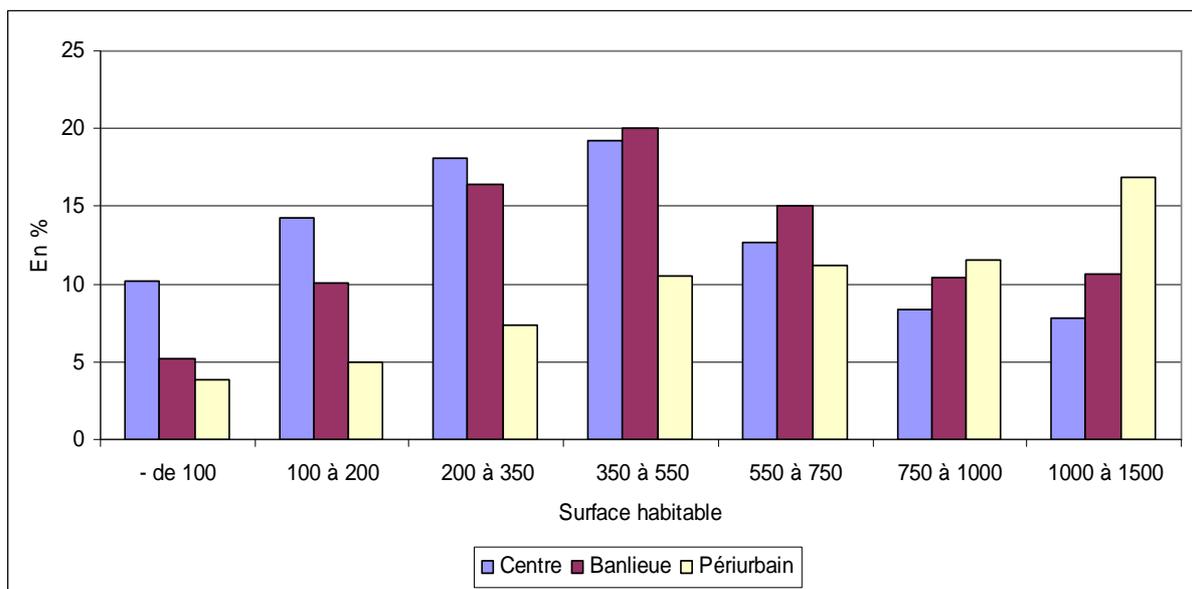
Source : Base Perval, transactions observées entre 2004 et 2008

Lecture : Les maisons de 95 à 110 m² représentent plus de 20 % de l'ensemble des maisons situées en banlieue.

Les différences sont plus marquées quant à la surface du terrain (figure A3.2). Les communes périurbaines ont des terrains plus grands : 50 % des maisons des communes périurbaines ont des terrains de plus de 1000 m² contre seulement 23 % pour les communes de banlieue et 17 % pour les communes centres. Ceci se traduit dans la valeur des classes modales des séries : les communes périurbaines ont pour valeur un intervalle de 1000 à 1500 m² contre une classe d'intervalle plus petit, de 350 à 550 m², pour les deux autres types de communes.

Cependant, prendre des modalités de référence différentes selon le type de commune n'a pas été possible dans tous les cas. En effet, certaines strates comportent à la fois des communes centres et des communes périurbaines.

Figure A3.2 - Répartition des transactions selon la surface du terrain de la maison, en province



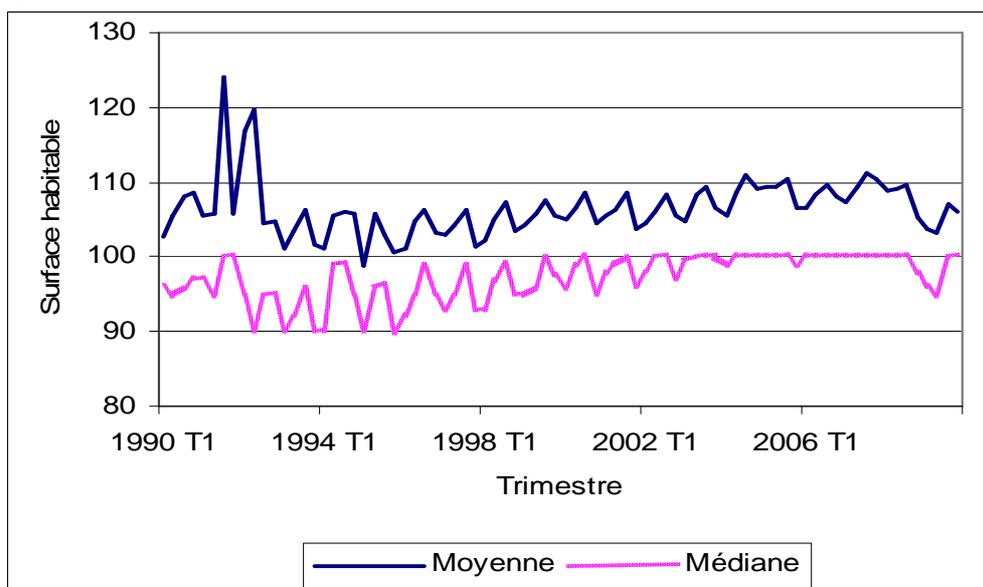
Source : Base Perval, transactions observées entre 2004 et 2008

Lecture : Les maisons ayant un terrain de 350 à 550 m² représentent près de 20 % de l'ensemble des maisons situées en banlieue.

Concernant les évolutions dans le temps, si l'on prend l'exemple de l'Île-de-France, la moyenne trimestrielle de la surface habitable a un profil irrégulier sur la période du 1^{er} trimestre 1990 au 4^{ème} trimestre 2009. La médiane, moins sensible aux valeurs extrêmes, est plus stable surtout à partir de 2003. Une saisonnalité notable est constatée (figure A3.4).

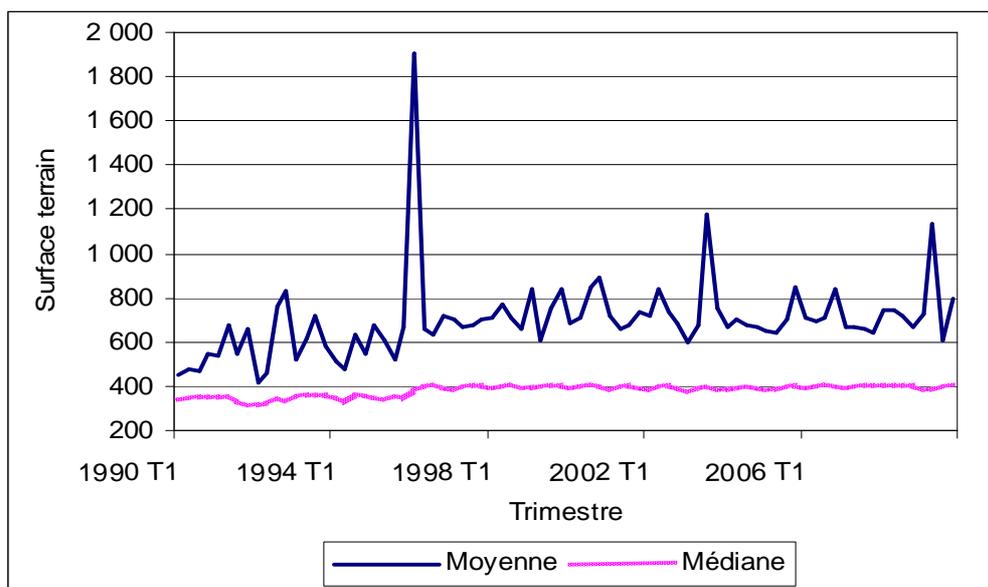
Concernant la surface du terrain, la médiane fluctue peu alors que la moyenne est sujette à une plus grande volatilité (figure A3.5).

Figure A3.3 - Surface habitable médiane et moyenne des maisons, en Île-de-France



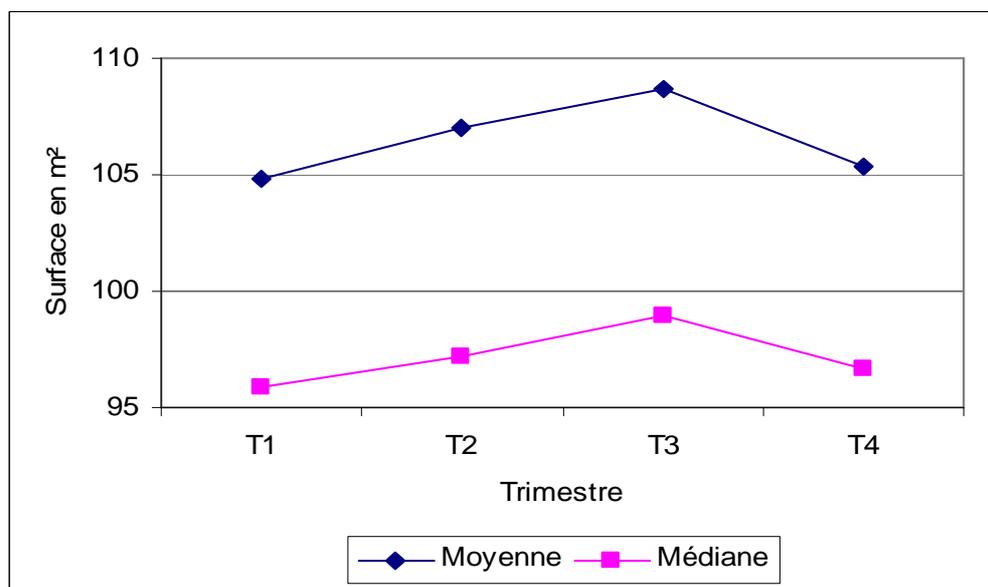
Source : Base BIEN, transactions observées entre le 1^{er} trimestre 1990 et le 4^{ème} trimestre 2009

Figure A3.4 - Surface médiane et moyenne du terrain des maisons en Île-de-France



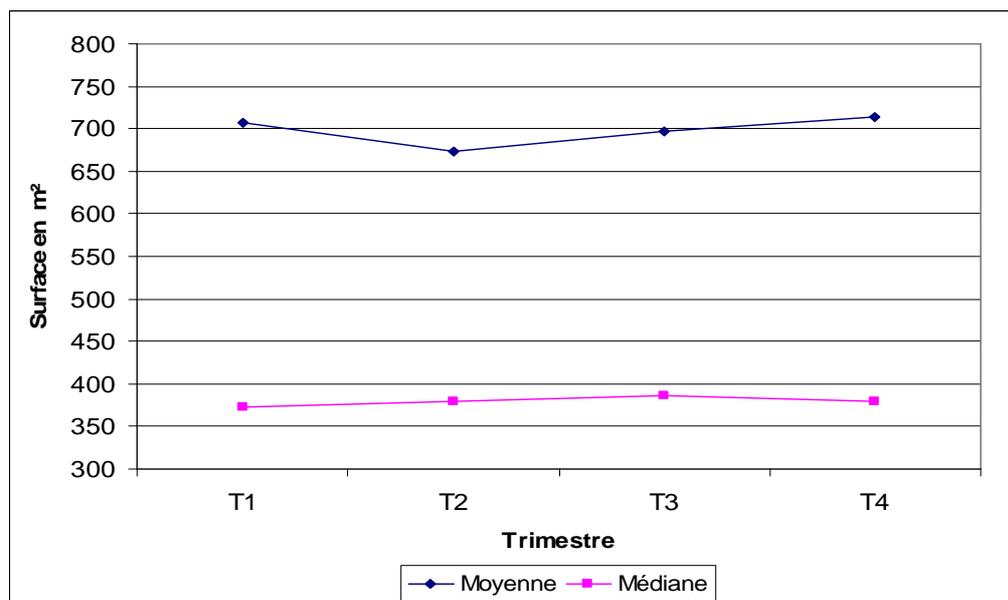
Source : Base BIEN, transactions observées entre le 1^{er} trimestre 1990 et le 4^{ème} trimestre 2009
 Remarque : l'évolution des séries est affectée par l'ajout de la grande couronne à partir de 1996.

Figure A3.5 - Surface habitable médiane et moyenne des maisons selon le trimestre, en Île-de-France



Source : Base BIEN, transactions observées par trimestre (entre le 1^{er} trimestre 1990 et le 4^{ème} trimestre 2009)

Figure A3.6 - Surface médiane et moyenne du terrain des maisons selon le trimestre en Île-de-France



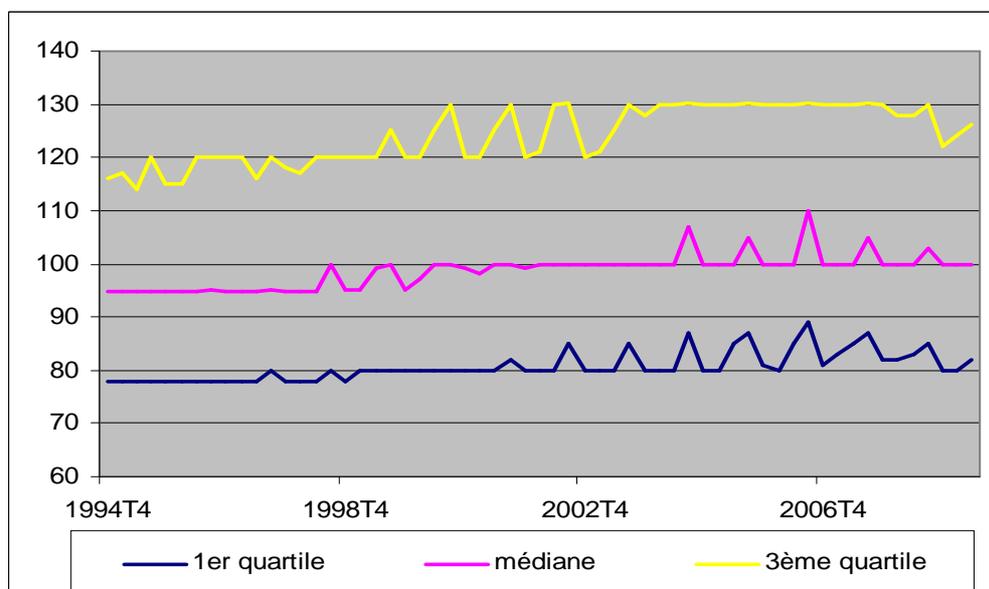
Source : Base BIEN, transactions observées par trimestre (entre le 1^{er} trimestre 1990 et le 4^{ème} trimestre 2009)

Choix des valeurs de référence

La volatilité de la surface du terrain est moindre que celle de la surface habitable. Il y a peu de saisonnalité de la médiane.

Les figures A3.7 et A3.8 décrivent pour la province l'évolution de la médiane et des deux déciles extrêmes pour les deux variables.

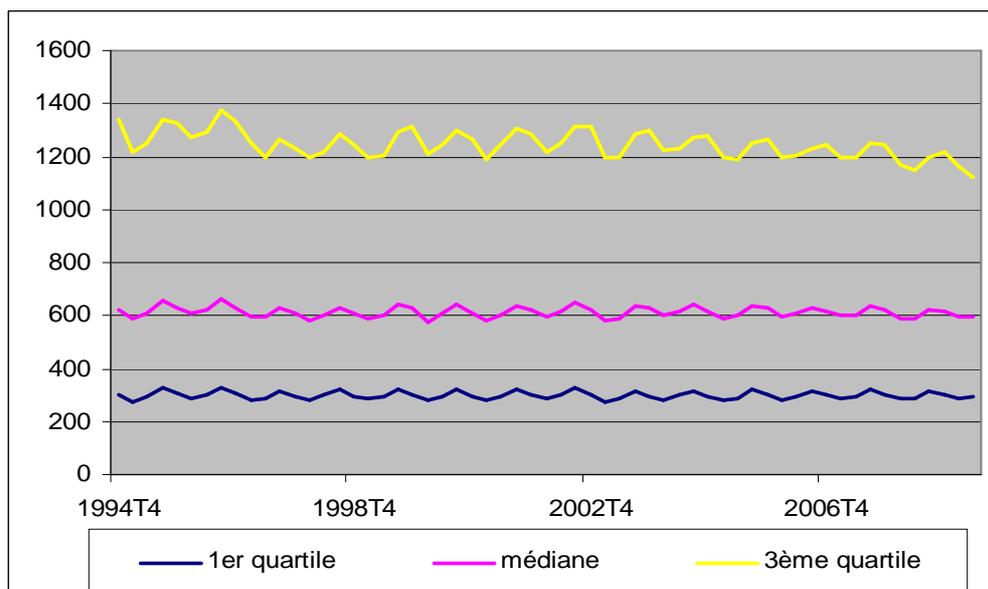
Figure A3.7 - Surface habitable des maisons de province selon trois quantiles



Source : Base Perval, transaction observées du 4^{ème} trimestre 1994 au 2^{ème} trimestre 2009

Lecture : la médiane de la surface habitable des maisons est relativement stable au cours des dix dernières années.

Figure A3.8 - Surface des terrains des maisons de province selon trois quantiles



Source : base Perval, transactions observées du 4^{ème} trimestre 1994 au 2^{ème} trimestre 2009

Prise en compte de la surface dans les modèles

La maison de référence retenue mesure 100 m² de surface habitable, avec une surface de terrain de 610 m². Et ces valeurs sont les mêmes en Île-de-France et en province.

Une autre question est de savoir s'il faut introduire une transformation des variables de surface, par exemple en prenant leur logarithme ou en les décomposant en tranches. Le choix dépend de l'idée que l'on se fait du lien entre le prix du bien et la surface habitable (ou la surface du terrain). Si on pense que le lien est linéaire, on opte pour les logarithmes. L'hypothèse faite est que l'élasticité-prix est constante. Un modèle log-log a l'avantage de rendre possible la lecture directe des élasticités. Si on pense au contraire que le lien entre la surface du terrain ou la surface habitable n'est pas linéaire, une solution est de découper les variables en classe. En outre, la discrétisation permet de mieux traiter la discontinuité de la surface habitable. En effet, les surfaces habitables des maisons sont souvent arrondies (à 5 ou 10 mètres carrés près)⁶⁴.

Dans la version précédente des indices, la surface du terrain et la surface habitable étaient exprimées en niveau brut et non en logarithme (on faisait l'hypothèse que le prix des maisons était une fonction exponentielle des coefficients $\beta_{k,\tau}$ du modèle hédonique au lieu d'être une fonction linéaire). On avait donc tendance à donner, dans la détermination du prix des maisons, une influence relative trop faible à ces deux variables par rapport aux autres variables. Ceci se traduisait par un coefficient correcteur proche de zéro. Cela avait pour conséquence d'accroître la saisonnalité de la série des indices de prix de façon artificielle.

Par souci de simplicité pour cette révision, le choix s'est porté sur l'utilisation du logarithme. La mise en œuvre de tranches suppose en effet de nombreuses étapes, qui pourront être menées lors de la prochaine révision : trouver le nombre optimal de tranches (de 5 à 7 a priori⁶⁵) et déterminer leurs valeurs ; les différencier éventuellement selon les strates ; tester pour chaque strate l'hypothèse de stabilité des élasticités des prix par rapport aux surfaces.

Impact du changement de méthode sur les séries d'indices

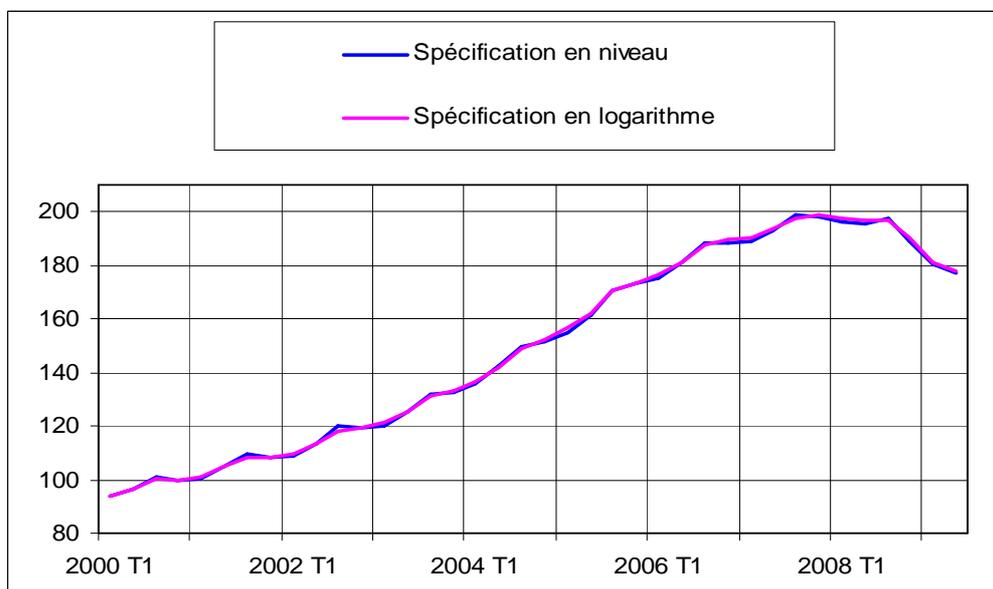
Nous avons également étudié l'impact de la nouvelle méthode sur la détermination des coefficients du vecteur des caractéristiques et sur les séries d'indices de prix. Nous présentons les résultats pour la province.

⁶⁴ Les surfaces des terrains sont quant à elles mieux mesurées car elles correspondent aux surfaces cadastrales renseignées dans les actes notariaux.

⁶⁵ Il ne faut pas faire trop de regroupements sinon on tend à revenir au cas continu. De plus cela accroît le nombre de coefficients à estimer.

L'ancienne méthode donne une série plus volatile que la nouvelle (figure A3.9). La saisonnalité des indices est réduite dans la nouvelle méthode. Les écarts d'indices sont assez faibles, de l'ordre de moins de 2 % en valeur absolue.

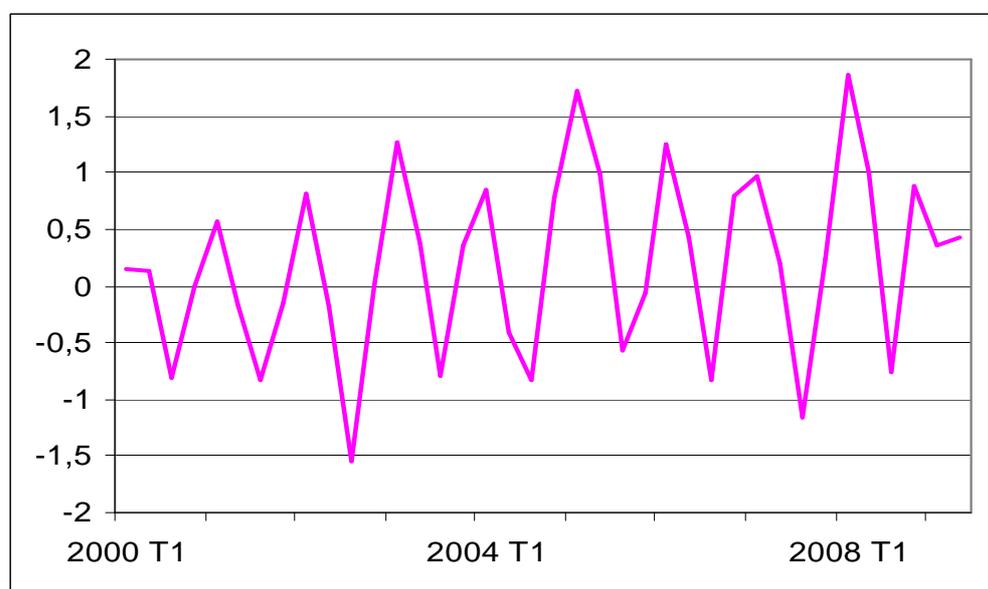
Figure A3.9 - Indice de prix des maisons en province



Source : base Perval, transactions observées entre le 1^{er} trimestre 2000 et le 3^{ème} trimestre 2009

Lecture : La série avec spécification des variables de surface en niveau a un profil saisonnier plus marqué que la série avec spécification des surfaces en logarithme.

Figure A3.10 - Écarts d'indices entre l'ancienne et la nouvelle méthode



Source : base Perval, transactions observées entre le 1^{er} trimestre 2000 et le 3^{ème} trimestre 2009

Lecture : l'écart correspond à l'indice « ancienne méthode - variables surface en niveau » moins l'indice « nouvelle méthode - variables surface en logarithme ».

Spécification des nouveaux modèles

Les variables expliquées ne sont pas modifiées : pour les maisons, le prix total est utilisé tandis que l'on prend le prix au m² pour les appartements. On retient pour la province les mêmes variables explicatives qu'en version 2.

Une seule évolution est à noter : le nombre de pièces et le quartier⁶⁶ étaient introduits séparément dans la version 2, tandis qu'en version 3, ils sont remplacés par leur croisement.

Les modèles ne peuvent être tout à fait identiques entre la province et l'Île-de-France, car certaines variables explicatives ne sont pas communes aux deux bases notariales (par exemple, la variable codifiant l'état général du bien n'est présente que dans la base Perval).

On trouvera dans les tableaux ci-dessous la liste des variables introduites dans les modèles. Toutes les variables qualitatives sont discrétisées.

Tableau A3.1 - Liste des variables explicatives contenues dans les modèles des appartements

Variables explicatives	Présence Modèle de Île-de-France	Présence Modèle de province
Année de la transaction	Oui	Oui
Mois de la transaction	Oui	Oui
Époque de construction	Oui	Oui
Nombre de salles de bains	Oui	Oui
Cave	Oui	Oui
Nombre de garages	Oui	Oui
Étage*	Oui	Oui
Ascenseur *	Oui	Oui
Croisement Ascenseur et nombre d'étage*	Oui	Oui
Croisement Nombre de pièces et quartier	Oui	Oui
Surface par pièce**	Oui	Oui
État du bien	Non	Oui
Terrasse ou balcon	Non	Oui

* Pour les appartements situés dans les étages les plus hauts (4ème étage et plus), on croise les variables « nombre d'étages » et « présence d'ascenseur ». En deçà, on estime que la variable « ascenseur » n'a pas d'effet sur les prix et on prend seulement la variable « étage ».

** En logarithme

Tableau A3.2 - Liste des variables explicatives contenues dans les modèles des maisons

Variables explicatives	Présence Modèle de Île-de-France	Présence Modèle de province
Surface habitable*	Oui	Oui
Surface du terrain*	Oui	Oui
Année de la transaction	Oui	Oui
Mois de la transaction	Oui	Oui
Époque de construction	Oui	Oui
Nombre de salle de bains	Oui	Oui
Sous-sol	Non	Oui
Nombre de garages	Oui	Oui
Nombre de niveaux	Oui	Oui
État du bien	Non	Oui
Cave	Oui	Non
Croisement Nombre de pièces et quartier	Oui	Oui

* En logarithme

⁶⁶ Au sens général de zone incluse dans une strate : il s'agit des quartiers dans les villes mais aussi de regroupements plus étendus en dehors des villes.

Annexe 4 : Tests de stabilité et durée de la période de référence

Les prix relatifs des caractéristiques des biens, obtenus à partir des modèles hédoniques, vont servir à calculer des indices de prix à caractéristiques constantes. Les modèles hédoniques sont estimés à partir des transactions d'une période donnée dite « période d'estimation »⁶⁷. Comme les transactions sont relativement rares mais que les prix relatifs des biens se modifient dans le temps, il faut retenir une période d'estimation assez longue pour disposer de suffisamment d'observations mais pas trop pour que les coefficients des modèles puissent être considérés comme constants. Des tests de stabilité visent à déterminer la durée optimale à retenir (en nombre d'années). Il s'agit de s'assurer que d'un point de vue statistique les coefficients $\hat{\beta}_k$ des caractéristiques peuvent être considérés comme constants sur la période d'estimation. L'effet temporel est par ailleurs capté par les indicatrices temporelles de l'équation économétrique.

Les tests de stabilité effectués dans la version 2

La durée de la période d'estimation en version 2 était de seize trimestres (années 1998 à 2001). On a testé la stabilité des coefficients $\hat{\beta}_k$ au cours de la période d'estimation en vérifiant que l'écart entre la valeur des biens ayant la caractéristique X_k , estimée par le modèle, et leur prix de vente réel (donc le résidu \mathcal{E}_i) satisfaisait les hypothèses stochastiques du modèle, et en particulier ne comportait pas de tendance déterministe oubliée.

Notons $\bar{\varepsilon}_{t,1}$ la moyenne des résidus mesurée sur l'ensemble des logements du trimestre t possédant la caractéristique X_1 . On construit la suite des résidus moyens $\bar{\varepsilon}_{1,1}$, $\bar{\varepsilon}_{2,1}$, ..., $\bar{\varepsilon}_{16,1}$ correspondant aux seize trimestres de la période d'estimation. L'observation de ces résidus, d'espérance mathématique nulle, en écart par rapport à 0, a permis de conclure à une stabilité globale des coefficients sur la période 1998 à 2001.

Les tests de stabilité effectués dans la version 3

La forte hausse des prix constatée dans les années 2000 puis son interruption en 2008 rendaient particulièrement cruciale cette question de la stabilité des modèles. Un test de stabilité sous la forme d'un test de Chow a été appliqué sur chacune des équations économétriques. En outre, pour s'assurer des conditions de validité du test de Chow, on a estimé dans une seconde étape les modèles à l'aide de l'estimateur des moindres carrés quasi-généralisés, pour se prémunir d'un éventuel problème de distribution induit par l'existence d'hétéroscédasticité.

Le test de Chow est réalisé à partir des sommes des carrés des résidus issues de la deuxième étape d'estimation. C'est un test d'analyse de la variance classique, qui prend la forme d'un test de Fisher. Le champ retenu est celui des observations du parc d'estimation. On répartit les observations en deux groupes d'années et deux sous-modèles : les années 2004-2005, « groupe des années de début de période » et les années 2006-2007, « groupe des années de fin de période ». Les tests ont établi que les coefficients estimés sur les deux périodes étaient significativement différents. Une estimation avec sélection de variables n'a pas permis d'améliorer les résultats.

Le tableau A4.1 présente à titre d'exemple les résultats de l'analyse pour les maisons et appartements de province. Il comporte le nombre et le poids des strates pour lesquelles l'hypothèse de stabilité des coefficients est rejetée. On aurait pu envisager de différencier la durée du parc de référence en fonction des strates selon les résultats du test : quatre ans pour les modèles acceptant l'hypothèse de stabilité et deux ans pour les autres par exemple. Cependant, pour des raisons pratiques - il faudrait notamment des méthodes différentes de chaînage d'indices de prix - on choisit de retenir une durée unique d'estimation du modèle pour toutes les strates. La proportion des strates rejetant l'hypothèse nulle croît naturellement avec la valeur du risque de première espèce. Pour le niveau de risque le plus souvent utilisé ($\alpha=0,05$), cette part est élevée dans le cas des maisons (46,0 %). Elle l'est encore plus dans le cas des appartements (pour $\alpha=0,01$, cette part atteint 80,4 % et passe à 85,6 % pour $\alpha=0,05$). Le tableau A4.2 indique que la correction de l'hétéroscédasticité ne change pas les résultats : l'hypothèse de stabilité des modèles ne peut être maintenue, aussi bien pour les maisons que pour les appartements, sur la période 2004-2007. Les prix relatifs des caractéristiques ne sont pas constants dans le temps. L'hypothèse de stabilité des modèles hédoniques sur une période de quatre ans ne peut pas être maintenue. Le même résultat est obtenu en Île-de-France.

⁶⁷ Voir définitions du chapitre 3.

Tableau A4.1 - Nombre de strates pour lesquelles l'hypothèse de stabilité est rejetée au seuil α^{68} ; estimateur des MCO

Selon le niveau de risque	Maisons		Appartements	
	Nombre de strates	En %	Nombre de strates	En %
$\alpha=0,001$	27	15,5	60	61,9
$\alpha=0,01$	52	29,9	78	80,4
$\alpha=0,05$	80	46,0	83	85,6

Source : Base Perval

Nombre de strates pour les maisons : 174, pour les appartements : 97

H0 : le modèle est stable dans le temps

H1 : le modèle n'est pas stable (un coefficient au moins est significativement modifié)

Tableau A4.2 - Nombre de strates pour lesquelles l'hypothèse de stabilité est rejetée au seuil α ; estimateur des MCQG

Selon le niveau de risque	Maisons		Appartements	
	Nombre de strates	En %	Nombre de strates	En %
$\alpha=0,001$	20	11,5	58	59,8
$\alpha=0,01$	44	25,3	72	74,2
$\alpha=0,05$	81	46,6	86	88,7

Source : Base Perval

Nombre de strates pour les maisons : 174, pour les appartements : 97

Une majorité des modèles hédoniques ne satisfont pas l'hypothèse de stabilité avec l'estimateur des MCO. Cependant, la fiabilité des tests peut être sujette à caution en présence d'hétéroscédasticité, auquel cas l'estimateur des MCO n'est pas efficace et les tests habituels ne fonctionnent plus. Des lectures graphiques reliant les carrés des résidus aux valeurs prédites des variables les plus influentes n'ont pas permis de déceler de variables associées à l'hétéroscédasticité. On a donc calculé l'estimateur des moindres carrés quasi-généralisés (MCQG), sous l'hypothèse d'une variance différente sur chaque sous période. Cette méthode est utilisée lorsque la forme de la matrice de la variance-covariance est inconnue. On applique cette correction lors de l'estimation de la fonction hédonique de la deuxième étape, c'est-à-dire après avoir retiré les points atypiques repérés en première étape. L'objectif est de trouver un bon estimateur des variances-covariances des aléas.

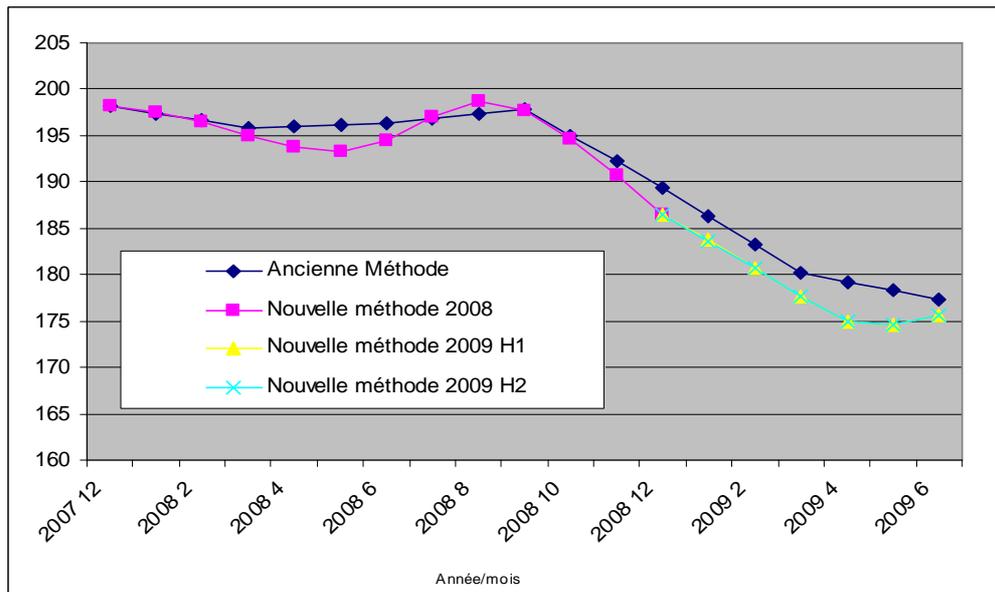
Choix des périodes de référence et de mise à jour du vecteur des prix relatifs des caractéristiques

Dans la version 3, la période d'estimation a donc été réduite à deux ans, contre quatre en version 2. Cette réduction est appliquée également au parc de référence. La définition des strates et la spécification des équations économétriques resteront inchangées sur la durée de vie de la version 3 mais les coefficients β_k seront mis à jour tous les deux ans. On notera qu'on aurait pu décider de les réactualiser tous les ans. Ces décisions reposent sur une comparaison pour les années 2008 et 2009 des indices de prix trimestriels calculés selon différentes méthodes relatives à la durée de la période d'estimation et la fréquence de réactualisation des coefficients (figure A4.1).

Les trois courbes ont approximativement le même profil d'évolution. Cependant, la série calculée avec un parc d'estimation de quatre ans (« ancienne méthode ») a un profil moins heurté que les deux autres séries, qui sont par ailleurs confondues en 2008 car elles sont calculées à partir du même parc d'estimation (celui des années 2006-2007). Les écarts entre la série calculée avec un parc d'estimation de quatre ans et celles reposant sur une durée de deux ans sont relativement importants : ils vont de 1,5 points à -4 points. La série « ancienne méthode » est toujours au-dessus des deux autres courbes sauf aux 2^{ème} et 3^{ème} trimestres 2008. Quant aux séries calculées avec un parc d'estimation de deux ans, elles diffèrent peu selon la méthode de calcul : les écarts sont nuls en 2008 et vont de -0,0012 à 0,1317 point en 2009.

⁶⁸ Le risque de première espèce est la probabilité de rejeter à tort l'hypothèse de stabilité des coefficients.

Figure A4.1 - Indice des maisons anciennes pour la province

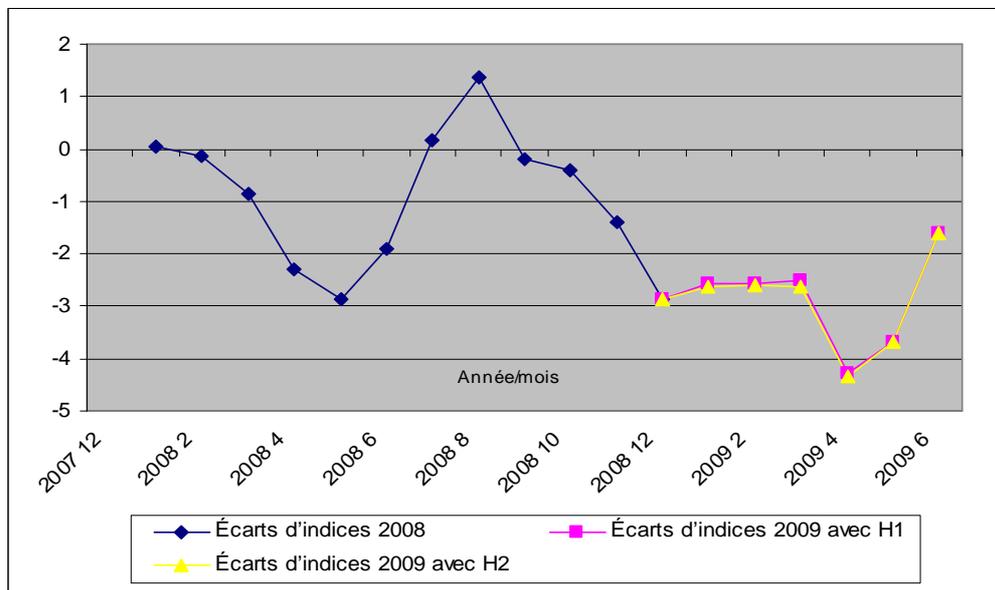


Source : Base Perval

Lecture :

- ancienne méthode = durée du parc d'estimation de 4 ans ;
- nouvelle méthode H1 : durée du parc d'estimation de 2 ans (sans réactualisation des β_k sur la période) ; les coefficients β_k des années 2008 et 2009 sont estimés sur les années 2006-2007 ;
- nouvelle méthode H2 : durée du parc d'estimation de 2 ans, avec réactualisation annuelle des β_k ; les coefficients β_k de l'année 2008 sont estimés sur 2006-2007 et ceux de l'année 2009, sur 2007-2008.

Figure A4.2 - Indice des maisons anciennes pour la province (écarts d'indice)



Source : Base Perval

Lecture : Les séries « nouvelles méthodes » sont identiques en 2008 (on utilise le même jeu de coefficients) et différentes en 2009 (on dispose de deux jeux différents).

Tableau A4.3 - Exemples de test de Chow pour quelques strates pour les maisons de province

code de la strate	Modèle total		Modèle 1 : 2004 et 2005		Modèle 2 : 2006 et 2007		Statistique du test	Pvalue
	degré de liberté	somme des carrés des résidus	degré de liberté	somme des carrés des résidus	degré de liberté	somme des carrés des résidus		
751	1 915	114,8	874	54,2	1 007	56,9	1,82	0,0028
752	2 690	142,4	1 325	76,3	1 326	61,7	2,23	0,0000
754	3 136	211,2	1 477	105,3	1 620	102,3	1,38	0,0581
755	7 182	412,1	3 456	204,1	3 677	202,2	2,09	0,0000
756	4 656	288,1	2 289	142,6	2 313	140,0	1,66	0,0018
757	8 938	428,0	4 574	225,8	4 320	198,4	1,80	0,0009
758	4 055	211,7	1 820	100,4	2 186	108,6	1,07	0,3401
6701	7 477	906,8	3 601	427,4	3 817	464,5	2,11	0,0000
13055	1 676	123,9	824	66,0	818	53,7	1,67	0,0092
21701	2 418	75,3	1 262	41,7	1 122	32,4	1,06	0,3742
29701	2 825	122,2	1 430	63,2	1 351	56,4	1,36	0,0573
31555	1 556	83,5	774	42,8	743	38,3	1,18	0,2137
31701	5 032	150,5	2 614	82,4	2 379	66,4	1,50	0,0245
33063	2 742	174,6	1 346	80,3	1 352	87,4	2,53	0,0000
33701	8 568	382,9	4 379	205,9	4 125	172,8	1,48	0,0080
34701	1 734	81,3	780	39,4	920	39,8	1,32	0,1061
35701	2 257	93,6	1 154	46,7	1 069	43,7	2,32	0,0000
37701	3 801	146,3	1 756	70,9	2 001	73,5	1,12	0,2656
38701	1 920	94,0	968	49,8	918	42,0	1,33	0,0977
44109	2 617	111,9	1 245	54,9	1 333	55,7	0,76	0,8619
44601	2 324	211,3	1 264	120,0	1 026	85,7	1,85	0,0021
44701	5 774	178,3	2 772	89,9	2 943	86,2	1,25	0,0982
45701	4 853	134,6	2 501	72,2	2 303	60,8	1,17	0,1984
51701	2 245	80,7	1 115	41,8	1 096	36,0	2,34	0,0000
54701	4 770	229,5	2 190	105,6	2 536	121,4	1,19	0,1790
59350	1 874	95,1	941	48,9	889	42,1	1,90	0,0004
59502	2 049	184,8	1 020	90,1	995	91,1	1,17	0,2332
59599	2 198	91,0	1 083	47,9	1 091	42,0	1,07	0,3675
59601	2 133	90,0	1 115	49,6	979	38,5	1,16	0,2291
59701	3 622	321,8	1 894	176,3	1 684	140,9	1,17	0,2019
59702	15 671	652,9	7 985	348,2	7 582	294,6	2,35	0,0000
62601	2 537	130,2	1 317	72,0	1 191	55,6	1,76	0,0072
63701	2 105	126,3	1 070	66,5	1 001	55,9	1,91	0,0013

Lecture : Dans la colonne Pvalue, les cases grisés indiquent les strates où l'on rejette l'hypothèse de stabilité $H_0 : \beta^1 = \beta^2 = \beta$, pour un niveau de risque $\alpha=0,05$.

Source : Base Perval

Tableau A4.4 - Exemples de test de Chow pour quelques strates après correction de l'hétéroscédasticité pour les maisons de province

Code de la strate	Modèle total		Modèle 1 : 2004 et 2005		Modèle 2 : 2006 et 2007		Statistique du test de Chow	Pvalue
	degré de liberté	somme des carrés des résidus	degré de liberté	somme des carrés des résidus	degré de liberté	somme des carrés des résidus		
751	1 921	462,5	865	209,0	1 014	238,3	1,52	0,0186
752	2 687	608,3	1 316	310,9	1 324	278,4	1,81	0,0007
754	3 150	812,7	1 471	393,5	1 632	402,7	1,37	0,0478
755	7 180	1 699,6	3 452	826,1	3 671	846,4	2,03	0,0000
756	4 653	1 143,1	2 281	560,9	2 310	557,9	1,61	0,0018
757	9 047	1 944,6	4 567	997,2	4 428	926,5	1,88	0,0001
758	4 064	919,9	1 814	415,3	2 193	490,9	1,06	0,3476
6701	7 474	2 607,6	3 588	1 235,4	3 819	1 340,0	1,38	0,0216
13055	1 700	465,3	817	229,1	841	218,7	1,55	0,0145
21701	2 408	422,3	1 252	227,8	1 114	185,3	1,25	0,1307
29701	2 882	593,9	1 423	293,7	1 407	284,9	1,45	0,0205
31555	1 634	419,7	810	212,2	792	198,1	1,15	0,2649
31701	5 028	865,1	2 607	461,2	2 374	392,3	1,44	0,0269
33063	2 734	693,5	1 339	332,5	1 343	335,6	1,95	0,0001
33701	8 561	1 782,7	4 367	926,2	4 122	834,3	1,49	0,0049
34701	1 734	384,0	774	176,4	918	197,7	1,07	0,3569
35701	2 270	462,2	1 151	231,6	1 077	215,3	1,81	0,0012
37701	3 798	737,0	1 750	347,8	1 996	379,3	0,97	0,5372
38701	1 914	431,9	960	222,5	912	198,4	1,16	0,2183
44109	2 611	538,1	1 239	262,0	1 325	267,4	0,90	0,6630
44601	2 326	698,5	1 257	383,6	1 027	294,5	1,64	0,0061
44701	5 775	1 006,5	2 771	498,0	2 937	495,4	1,12	0,2280
45701	4 850	800,1	2 497	422,9	2 296	366,6	1,13	0,2318
51701	2 238	418,5	1 106	213,5	1 090	192,0	1,68	0,0042
54701	4 789	1 044,2	2 185	477,8	2 552	553,2	1,17	0,1855
59350	1 868	417,4	930	208,7	886	189,9	1,65	0,0027
59502	2 039	610,0	1 014	302,4	983	293,6	1,11	0,2883
59599	2 190	458,6	1 074	228,2	1 084	223,7	1,00	0,4673
59601	2 126	437,4	1 107	232,5	972	193,7	1,16	0,2139
59701	3 616	1 074,0	1 885	571,3	1 679	487,5	0,98	0,5089
59702	15 683	3 148,4	7 976	1 634,3	7 595	1 467,6	2,08	0,0000
62601	2 541	576,5	1 311	304,4	1 193	259,6	1,51	0,0259
63701	2 097	505,9	1 061	260,8	994	230,9	1,41	0,0429

Lecture : Dans la colonne Pvalue, les cases grisées indiquent les strates où l'on rejette l'hypothèse de stabilité $H_0 : \beta^1 = \beta^2 = \beta$, pour un niveau de risque $\alpha=0,05$.

Source : Base Perval

Annexe 5 : Les méthodes de calcul alternatives

On a décrit dans ce volume la méthode de construction d'un indice de prix hédonique et présenté sa mise en œuvre pour obtenir les indices Notaires-Insee. Cette méthode présente plusieurs étapes :

- définition de strates, où les évolutions de prix sont supposées homogènes ;
- introduction de correctifs des effets qualité, strate par strate ;
- estimation des effets correctifs à partir d'un parc d'estimation ;
- calcul des évolutions de prix par strate à partir de l'ensemble des transactions ;
- calcul de l'indice en observant l'évolution de la valeur d'un parc de référence ;
- publication régulière d'indices et de sous-indices.

Une telle démarche se distingue par son caractère systématique et intégré. Nous discutons dans ce chapitre des méthodes alternatives employées, en France et surtout dans certains pays étrangers, pour déterminer si elles sont théoriquement supérieures, ou si elles sont préférées pour des raisons pratiques (contraintes budgétaires, disponibilité de données, contrainte plus ou moins forte de publication), ou du fait de comportements différents des ménages (une plus grande mobilité favorisant les ventes répétées). Finalement nous donnons dans le dernier paragraphe un descriptif des procédures suivies par d'autres organismes en France, et dans divers pays étrangers, en insistant sur les questions de collecte de données et sur les organismes gérant les calculs d'indices.

Dans l'éventail des méthodes mises en œuvre au niveau international pour établir des indices de prix des logements, on trouve des méthodes de calcul relativement simples et d'autres plus sophistiquées. Pour des raisons pratiques de coût ou pour des raisons juridiques, beaucoup de statistiques sur les prix des logements utilisent des prix moyens ou médians pour retracer l'évolution du marché des logements. On a expliqué pourquoi elles étaient biaisées (chapitre 2, paragraphe 2.1.2).

Les méthodes plus sophistiquées font en général appel à l'approche hédonique ou à l'approche par les ventes répétées. La première approche est de type économétrique ; elle s'appuie sur des régressions incorporant des effets qualité et un effet temporel, et elle assimile cet effet temporel à un pur effet prix à qualité constante. La deuxième approche vise à éliminer les effets qualité en ne conservant que les données sur les ventes successives, dites répétées. Nous allons maintenant examiner brièvement les principes et les hypothèses sur lesquels elles reposent et discuter certaines de leurs variantes.

Approche économétrique (hédonique) : interpréter les coefficients du temps dans une régression comme un pur effet de prix à qualité constante

Supposons qu'on dispose de plusieurs coupes instantanées d'échantillons de transactions immobilières, donnant une mesure du prix et des caractéristiques logement/immeuble/quartier permettant l'estimation d'un modèle hédonique. Le modèle définit le niveau de prix en fonction des caractéristiques du logement. Les coefficients de ces caractéristiques sont supposés stables dans le temps, et un effet temporel v_t est introduit dans le terme constant.

Le modèle s'écrit :

$$\log p_{i,t} = a + \sum_{k=1}^n \beta_k X_{k,i,t} + v_t + e_{i,t}$$

où les erreurs $e_{i,t}$ sont supposées indépendantes, de même loi, centrées. Notons que l'effet de strate n'est pas introduit, pour simplifier la discussion. De façon à rendre identifiable l'effet temporel, on suppose $v_{t_0} = 0$ pour une date donnée, qui constitue alors la période de base. L'évolution de l'indice entre t_0 et t est alors assimilée à la valeur v_t et celle de l'indice entre $t-1$ et t à $v_t - v_{t-1}$.

Les hypothèses sous-jacentes à cette spécification sont les suivantes :

- Les variables retenues pour caractériser un logement interviennent sous forme additive (après, rappelons-le, transformation éventuelle des caractéristiques initiales) ;

- Les prix relatifs β_k des caractéristiques X_k sont indépendants du temps, ce qui revient à supposer l'absence d'effet croisé de ces variables avec le temps ;
- Une fois corrigé des effets des caractéristiques, les variabilités de prix sont constantes (absence d'hétéroscédasticité des erreurs) ;
- Les transactions introduites comme observations sont représentatives de l'ensemble des biens sur lesquels on souhaite calculer l'indice, une fois corrigé des effets qualités.

Certains de ces aspects, comme la dépendance temporelle des paramètres β_k , sont classiquement étudiés dans la littérature. Si la constance de ces coefficients est habituellement rejetée sur longue période, ce qui explique la nécessité de révision régulière de l'indice, elle est habituellement acceptée sur des périodes plus petites de l'ordre de 4 à 5 ans. Il existe une façon simple de vérifier cette stabilité temporelle. En effet, si par exemple le coefficient de X_1 dépend de la date t , le modèle initial apparaît mal spécifié, car les variables croisées $1_\tau(t)X_{1,i,t}$, τ variant, ont été omises. Ces variables sont orthogonales les unes aux autres. Il suffit alors de reporter en fonction de la date τ , la corrélation empirique entre la variable omise $1_\tau(t)X_{1,i,t}$ et le résidu d'estimation $\hat{\epsilon}_{i,t}$. Si ces corrélations sont proches de zéro, le coefficient β_1 est considéré comme stable dans le temps. Sinon la forme de l'évolution de cette corrélation en fonction de τ donne de l'information sur le type d'évolution du coefficient.

Un exemple de cette méthode consiste en la méthode sur « périodes adjacentes ». On considère un modèle à indicatrices temporelles basé sur deux périodes consécutives. À chaque période t on réestime le modèle. L'indicatrice temporelle mesure l'évolution des prix à caractéristiques constantes entre les dates $t-1$ et t . L'indice entre deux dates est obtenu en chaînant les uns aux autres les indices des périodes comprises entre ces deux dates.

La représentativité des échantillons de transactions à chaque date t par rapport à ce qui est théoriquement souhaitable pour un indice est une question moins souvent discutée. Si, par exemple, l'échantillon de la date t comporte des transactions dont les prix sont systématiquement plus élevés que le prix théorique (c'est-à-dire le prix d'un parc de référence fixé et non de transactions qui varient d'une date à l'autre) sur la population, le coefficient temporel \hat{v}_t estimé à partir des données traduira à la fois le niveau de prix théorique et le biais dû à la non représentativité de l'échantillon, sans qu'il soit possible de facilement distinguer ces effets (problème d'identifiabilité). Deux approches peuvent cependant être suivies pour détecter des non-représentativités éventuelles.

- La première, suggérée par Griliches (1971, pp.7-8) dans un cadre différent (indice de prix des automobiles), consiste à distinguer dans les échantillons des dates $t-1$ et t par exemple, les ventes répétées. On peut alors calculer sur ces ventes répétées les moyennes des résidus à la date $t-1$ et à la date t et voir si elles sont proches de zéro. Cependant cette démarche vise plutôt à vérifier la représentativité du sous-échantillon de ventes répétées que celle des échantillons complets. Par ailleurs, elle est difficile à mettre en œuvre dans notre contexte, où le nombre de ventes répétées est faible.
- Une autre démarche repose sur l'évolution supposée de l'indice. Étant données des observations $\hat{v}_1, \dots, \hat{v}_T$ de l'indice entre 1 et T , on peut construire un modèle dynamique permettant de fournir un intervalle de prévision :

$$\left[\tilde{v}_{T+1}, \tilde{v}_{T+1} \right]$$

pour la valeur future v_{T+1} . Si l'estimation de la date $T+1$ n'est pas dans cet intervalle, on peut penser soit que l'échantillon de la date $T+1$ n'est pas représentatif, soit rechercher une cause structurelle à cette brusque modification des prix.

Méthode des ventes répétées

Dès 1943, Gaston Duon, travaillant au Service de la Statistique Nationale, prédécesseur de l'Insee, appliqua la méthode dite des ventes répétées pour reconstituer des indices du prix des logements à Paris de 1790 à 1944 (Duon, 1943 et 1946). Léo Grebler, qui apparemment ignorait les travaux de Gaston Duon, appliqua ensuite en

1956 cette méthode pour reconstituer un indice du prix des logements dans 22 villes des Etats-Unis⁶⁹. Tant Duon que Grebler étaient conscients que la qualité d'un bien donné change dans le temps sous les effets contraires de deux phénomènes, d'une part l'obsolescence générale du bâti et d'autre part les améliorations apportées au fil des ans. Ils ont donc tous deux produit deux séries d'indices, avant et après correction de ces effets. L'écart de croissance de ces deux séries est de l'ordre de 1% par an tant chez Duon que chez Grebler⁷⁰.

La méthode des ventes répétées a été ensuite appliquée, avec des moyens de calcul plus importants, par Bailey, Muth et Nourse (1963). Les auteurs ne posent pas directement la question du panier fixe de biens mais font deux remarques. D'une part, « la variation de qualité des logements vendus d'une période à l'autre fait que la moyenne des prix varie davantage que le prix de chaque bien pris individuellement » ; d'autre part un changement progressif de la qualité des logements échangés au cours du temps biaise l'évolution des prix moyens.

Devant la difficulté de spécifier un modèle hédonique avec indicatrice de temps, en l'absence fréquente de données sur les caractéristiques des logements, ils proposent d'utiliser le fait que certains logements connaissent plusieurs ventes successives. Ces données sur des ventes répétées permettraient de se passer de renseignements détaillés sur les caractéristiques des biens⁷¹.

Le manque de données sur les caractéristiques des biens et la difficulté technique de l'approche économétrique restent aujourd'hui les principales raisons invoquées en faveur d'une approche par les ventes répétées. L'approche est facile à expliquer lorsque les ventes répétées ont lieu aux dates $t-1$ et t . Dans ce cas on assimile l'évolution des prix à l'évolution moyenne constatée sur ces ventes répétées, selon un modèle du type :

$$\log p_{i,t} = \log p_{i,t-1} + b_t + u_{i,t-1,t}, i \in I_{t-1,t}$$

à une date t donnée. $I_{t-1,t}$ désigne l'ensemble des logements échangés à la fois en $t-1$ et t , et b_t l'évolution recherchée.

Cependant, même si la mobilité est grande, il y a peu de ventes répétées aussi rapprochées. L'approche est donc étendue pour prendre en compte des ventes répétées ayant lieu à deux dates t_1 et t_2 , $t_1 < t_2$, pouvant être plus éloignées. Le modèle sous-jacent devient alors :

$$\log p_{i,t_2} = \log p_{i,t_1} + \sum_{t=t_1+1}^{t_2} b_t + u_{i,t_1,t_2}, i \in I_{t_1,t_2}$$

pour t_1 et t_2 variant.

Il peut être réécrit en faisant intervenir les variables explicatives des dates $Z_{i,t} = 1$, si la date t figure dans la période entre les dates d'échange, $Z_{i,t} = 0$, sinon. Le modèle devient alors :

$$\log \frac{p_{i,t_2}}{p_{i,t_1}} = \sum_{t=T_1}^{T_2} b_t Z_{i,t} + u_{i,t_1,t_2}, i \in I_{t_1,t_2}$$

où $[T_1, T_2]$ donne l'intervalle de temps union de tous les intervalles $[t_1, t_2]$. Sous cette forme il s'agit d'un modèle linéaire dans les paramètres d'intérêt b_t , t variant, donnant les évolutions de prix aux diverses dates. Ce modèle est habituellement estimé par moindres carrés ordinaires.

⁶⁹ Les indices séculaires du prix des logements recourent généralement à la méthode des ventes répétées. C'est notamment par cette méthode que Piet Eichholtz a construit l'indice du Herengracht (Eichholtz, 1996). Font exception les indices de d'Avenel qui couvrent la période 1200-1800 et à ce titre constituent semble-t-il les indices immobiliers remontant le plus loin dans le temps, au prix de certaines incertitudes.

⁷⁰ Sur les indices séculaires du prix des logements, et notamment la comparaison des travaux de Duon et Grebler, cf. Comparing Four Secular Home Price Indices, J. Friggitt, juin 2008, http://www.cgedd.developpement-durable.gouv.fr/IMG/doc/house-price-index-Paris-and-others-secular_cle7fed11.doc.

Cependant, si l'intervalle de temps entre ventes répétées couvre plus de deux périodes, il y a des recouvrements de périodes correspondant aux divers logements, donc il peut y avoir des corrélations entre termes d'erreurs, qui devraient alors être prises en compte au niveau de la méthode d'estimation. Pour mettre en évidence ce problème éventuel et comprendre pourquoi les effets qualité ont disparu, il est intéressant de se placer dans le cadre du modèle hédonique vu au paragraphe précédent :

$$\log p_{i,t} = a + \sum \beta_k X_{k,i,t} + v_t + e_{i,t}$$

avec $e_{i,t} = u_{i,t} + \eta_i$,

où les termes d'erreur $u_{i,t}, \eta_i$ sont supposés indépendants entre eux, centrés. Si l'ensemble des ventes répétées ayant lieu en t_1, t_2 est représentatif de l'ensemble de tous les logements (une fois corrigé de l'effet caractéristique), on a :

$$\log \frac{p_{i,t_2}}{p_{i,t_1}} = v_{t_2} - v_{t_1} + u_{i,t_2} - u_{i,t_1}, i \in I_{t_1, t_2}$$

On vérifie alors que l'approche par ventes répétées avec estimation par moindres carrés est cohérente avec

$b_t = v_t - v_{t-1}$ et $\sum_{t_1+1}^{t_2} b_t = v_{t_2} - v_{t_1}$, puisque les termes d'erreurs $u_{i,t_1, t_2} = u_{i,t_2} - u_{i,t_1}$ sont bien indépendants, de même loi.

On peut finalement compliquer et affiner le modèle en prenant en compte la dépréciation naturelle des logements (nette des améliorations apportées), ou d'autres variables de qualité susceptibles d'évoluer entre deux ventes (par exemple un élément de confort qui a été ajouté au logement).

En résumé, la méthode des ventes répétées suppose :

- que les prix relatifs des caractéristiques soient constants dans le temps. C'est la même hypothèse que dans les applications usuelles de la méthode hédonique. Mais elle a sans doute moins de chances d'être vérifiée sur la période relativement longue qui peut séparer deux ventes successives du même logement ;
- qu'il n'y ait pas de biais de sélection. Cependant les logements vendus fréquemment ne sont vraisemblablement représentatifs ni de l'ensemble des transactions, ni du parc de logements. Ce seraient par exemple des petits logements (premiers logements de jeunes couples), dont les prix évolueraient différemment de ceux des logements plus grands ; ou alors des logements revendus très vite présentant des caractéristiques inobservables, liées au vendeur, qui peuvent expliquer des plus-values importantes. Clapp et al. (1991) trouvent ainsi une différence dans l'évolution à court terme des indices de ventes répétées par rapport aux indices hédoniques, mais cette différence disparaît à long terme (3 ans), ce qui leur semble logique : si le marché fonctionne, il ne peut y avoir de déséquilibre à long terme des prix relatifs⁷². Vu les distorsions sur des périodes inférieures à trois ans, ils recommandent l'application des méthodes hédoniques. Case et al. (1997) proposent de corriger cet effet en incorporant l'information sur le lien entre taux d'appréciation d'un logement et fréquence de transaction ;
- que le logement soit bien le même. En réalité, le simple passage du temps fait qu'il se déprécie. En sens inverse, les réhabilitations, voire des modifications plus importantes (extensions, aménagements) sont fréquentes, qui font que le logement n'est plus le même. Ces deux derniers problèmes peuvent être pris en compte en combinant modèle hédonique et données sur ventes répétées ; ceci permet aussi de corriger un défaut de la méthode des ventes répétées : le fait qu'elle n'utilise que peu d'observations par rapport à l'ensemble des transactions, en effet les logements changent de main assez peu souvent, tout au moins en France ;
- que les termes d'erreurs $u_{i,t}$ sont bien indépendants, centrés, de même variance. Or cette hypothèse est vraisemblablement non satisfaite. Ces erreurs portent en effet sur des prix, qui en moyenne croissent avec le temps, d'où de l'hétéroscédasticité. Par ailleurs on s'attend à une corrélation plus forte entre le

⁷² Mark et Goldberg (1984) trouvent, eux, une différence persistante sur longue période.

prix de dates proches que de dates éloignées. Il apparaît ainsi important d'introduire une corrélation temporelle entre les erreurs $e_{i,t}$, ce qui modifie la méthode d'estimation des taux d'évolution $b_{i,t}$.

Notons que pour la France, la méthode des ventes répétées⁷³, pour les périodes et les zones qui se recourent, ne donne pas de résultats très différents de ceux des indices Notaires-Insee⁷⁴.

Indice de Laspeyres ou indice-chaîne

Une large part de la théorie classique des indices est consacrée au choix des pondérations. Trois choix sont en général possibles : elles peuvent être fixées une fois pour toutes égales à un ensemble de quantités échangées soit à une date initiale (indice de Laspeyres), soit à une date terminale (indice de Paasche), ou bien elles peuvent être modifiées à chaque date de calcul de l'indice (indice-chaîne).

Ces quantités échangées sont liées à la composition du parc par strate. La méthode des ventes répétées conduit naturellement à une optique d'indice-chaîne, où les poids affectés aux strates les plus fines varient dans le temps, selon le degré d'agrégation mis en œuvre lors du calcul. Cette agrégation est rendue nécessaire à un niveau plus ou moins fin selon le nombre des ventes utilisables dans le calcul. La structure des ventes répétées se modifiant dans le temps, la façon dont s'effectue le chaînage ne peut être définie a priori. Ce problème n'est pas évoqué par les utilisateurs ni les tenants de la méthode.

A contrario, l'approche hédonique permet de calculer des indices de type Laspeyres, Paasche ou chaîne puisqu'il s'agit du prix d'un bien de référence. Afin de faciliter les comparaisons de prix avec les autres types d'investissement, financiers par exemple, il semble cependant préférable de ne pas retenir un chaînage d'une période à l'autre, mais de maintenir un parc de référence sur une certaine durée. En effet, la pratique en matière de portefeuille financier est de suivre un portefeuille de composition fixe (portefeuille « cristallisé ») de façon à ne pas mêler les effets prix à ceux de la mise à jour de la structure des portefeuilles (stratégie d'investissement)⁷⁵.

Ceci semble d'autant plus justifié pour des indices de prix des logements que le logement est un bien durable, peu liquide, qui entraîne de forts coûts de transaction et pour lequel les préférences des investisseurs entre les différents types de biens évoluent lentement. L'indice des prix du logement est en cela différent de celui des prix à la consommation. La justification du chaînage pour les indices de prix à la consommation (ou celle de la révision annuelle des pondérations) réside dans le fait que l'on souhaite capter les évolutions progressives des préférences des consommateurs qui se révèlent par la part de budget allouée à chaque grande catégorie de biens. En matière de logement, ces adaptations sont lentes selon toute vraisemblance. Au niveau fin, les indices sont agrégés de manière multiplicative pour prendre en compte une possible substituabilité des biens. Au niveau supérieur d'agrégation, ils le sont de manière additive.

⁷³ Appliquée aux données notariales, sur longue période (comparaison de prix, sans économétrie). Voir Friggit J., "Comparing Four Secular Home Price Indices", CGEDD, June 2008, <http://www.cgedd.developpement-durable.gouv.fr/prix-immobilier-evolution-1200-a1048.html>.

⁷⁴ La méthode de ventes répétées donne des résultats médiocres sur le passé récent, à cause du faible nombre d'observations et du biais de sélection mentionné ci-dessus sur les durées de détention courtes (la plus-value sur les biens revendus rapidement est supérieure à la plus-value de l'indice). La comparaison a donc été effectuée après incorporation d'un coefficient correcteur calibré sur Paris.

⁷⁵ Les indices de prix des logements anciens sont des indices de Laspeyres chaînés. Par rapport à la version 2, la nouvelle méthode de calcul correspond à une succession plus rapprochée dans le temps de Laspeyres - tous les deux ans au lieu de tous les cinq ans (cf. chapitre 3).

Annexe 6 : Les indices de prix avancés

Usuellement, la vente d'un bien immobilier est précédée par la signature d'un avant-contrat⁷⁶. Il se passe en général environ trois mois entre la signature de l'avant-contrat et celle du contrat de vente. Le développement d'indices de prix avancés, calculés à partir des avant-contrats permettrait donc de disposer d'indices plus réactifs par rapport à l'état du marché.

Collecte des avant-contrats

La transmission des avant-contrats par les notaires aux gestionnaires des bases de données immobilières du notariat est une obligation qui découle de la loi de mars 2011. Toutefois, la collecte des avant-contrats a été mise en place par le notariat auprès des études depuis la mi-2009. Deux projets se sont succédé. Dans un premier temps, les notaires ont mis à la disposition des études de province, un portail sur lequel saisir les principales informations relatives aux avant-contrats. Ce système a permis la collecte d'un nombre important d'avant-contrats (*tableau A6.1*) mais il avait l'inconvénient de ne pas contenir toutes les variables nécessaires au calcul des indices.

Tableau A6.1 - Nombre de ventes et d'avant-contrats de logements anciens collectés en 2012

	Ventes	Avant-contrats
Île-de-France	95 927	15 526
Province	308 650	111 216

La décision a donc été prise de développer un module spécifique à la télétransmission des avant-contrats, dans les logiciels de rédaction d'actes utilisés dans les études. Il est constitué selon le même principe que le module de télétransmission des ventes, ce dernier étant amené à remplacer progressivement l'envoi de copies papier des actes par les études.

L'outil de télétransmission des avant-contrats a été proposé aux études de façon progressive à partir de 2010 (selon les SSII et les logiciels de rédaction d'acte). Il évolue depuis en fonction des demandes des gestionnaires de bases de données et des développements des SSII notariales.

A partir de la nouvelle version du logiciel, une procédure permettant de rapprocher l'avant-contrat de la vente a été mise en place courant 2013. Elle permettra de réaliser un certain nombre de calculs qui ne sont pas faits aujourd'hui par absence d'information : en particulier l'étude détaillée du délai entre l'avant-contrat et la vente, ou celle de la proportion des avant-contrats qui ne font pas finalement l'objet d'une vente.

Méthodologie et diffusion

L'historique et les volumes d'avant-contrats transmis ne permettent pas encore de développer un indice des prix reposant uniquement sur eux, ce qui pourrait constituer l'objectif à terme. Une méthode de calcul provisoire a été développée sur la base des indices actuels.

L'hypothèse de travail est que les modèles économétriques développés pour les ventes doivent également être valables pour les avant-contrats. Ainsi l'application de ces modèles aux données d'avant-contrats doit permettre de calculer des prix standardisés (« équivalents bien de référence »⁷⁷) pour les avant-contrats et de les comparer aux prix standardisés des ventes des trimestres précédents. Ceci permet de calculer des indicateurs avancés à condition que le volume d'avant-contrats reçus soit suffisant.

Des indicateurs avancés sur l'évolution des prix à Paris et dans les Hauts-de-Seine sont diffusés par la Chambre des Notaires de Paris dans le cadre de ses notes de conjoncture et de ses conférences de presse. En province, un indicateur général est donné lors des notes de conjoncture trimestrielles et des indications sur les évolutions de prix dans certaines grandes villes de province sont également diffusées lors de la conférence nationale annuelle du Conseil Supérieur du Notariat.

⁷⁶ L'ensemble des avant-contrats (compromis ou promesse de vente) sont collectés par les bases notariales.

⁷⁷ Cf. chapitre 3.

Annexe 7 : Convention avec PNS

Convention relative à la poursuite et au développement du partenariat entre l'Insee et PNS en matière d'indices de prix des logements en région Île-de-France,

entre :

- La **CHAMBRE INTERDEPARTEMENTALE DES NOTAIRES DE PARIS**, représentée par Maître BENASSE, notaire, son Président,
 - La **CHAMBRE DES NOTAIRES DE SEINE-ET-MARNE**, représentée par Maître HAUTEBAS, notaire, son Président,
 - La **CHAMBRE INTERDEPARTEMENTALE DES NOTAIRES DE VERSAILLES**, représentée par Maître SAVOURE, notaire, son Président,
 - La **CHAMBRE DES NOTAIRES DE L'ESSONNE**, représentée par Maître LEMOINE, notaire, son Président,
 - La **CHAMBRE DES NOTAIRES DES HAUTS-DE-SEINE**, représentée par Maître HERRNBERGER, notaire, son Président,
 - **L'ASSOCIATION DES NOTAIRES DU CHATELET (PARIS NOTAIRES SERVICES)**, association régie par la loi de 1901, représentée par Maître CAURO, notaire, son Président ;
- agissant conjointement et solidairement et désignées ci-après par les initiales PNS,

d'une part,

et :

L'INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES, désigné ci-après par l'acronyme Insee et représenté par son directeur général M. Jean-Luc TAVERNIER,

d'autre part,

Il est convenu et arrêté ce qui suit :

Préambule

PNS et l'Insee ont conclu une première convention en date du 6 décembre 1990, qui a mis en place un partenariat en vue du calcul et de la publication de l'indice de prix des appartements anciens vendus libres à Paris.

Une seconde convention, en date du 16 mai 2000 (convention 2000 00094), et un avenant en date du 26 novembre 2002 ont rénové le mode de calcul de cet indice et ont étendu le champ d'application de ce partenariat à d'autres indices de prix des logements anciens vendus libres relatifs à tout ou partie de la région Île-de-France.

Une troisième convention, en date du 8 décembre 2005 (convention 2005 00353), et un avenant en date du 26 novembre 2009 ont de nouveau étendu le champ d'application des indices et mis en place une diffusion et une publication concertées desdits indices.

Les conventions signées en 1990, 2000 et 2005 sont arrivées à expiration. Elles sont remplacées par la présente convention, qui vise à poursuivre et renforcer le partenariat entre PNS et l'Insee en matière d'indices de prix des logements. La présente convention prend acte des différentes avancées récentes dans ce partenariat.

Article 1er : Objet de la convention

La présente convention définit les conditions de participation de PNS et de l'Insee au calcul, à la validation, à la publication et à la diffusion d'indices de prix des logements anciens vendus libres. Ceux-ci sont calculés à partir des données de transactions immobilières contenues dans la base BIEN, appartenant à et gérée par PNS, et alimentée par les notaires d'Île-de-France.

PNS informe l'Insee que l'existence de la base BIEN en tant que traitement informatisé d'informations nominatives a été déclarée à la Commission nationale de l'informatique et des libertés, conformément aux dispositions de l'article 16 de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, modifiée par la loi 2004-801 du 6 août 2004.

Les indices couverts par la présente convention sont les suivants, étant entendu que les trimestres sont entendus au sens de trimestre civil :

- indices trimestriels des prix des logements (appartements et maisons) pour l'ensemble de l'Île-de-France,
- indices trimestriels des prix des logements (appartements et maisons) dans chacun des sept départements d'Île-de-France en dehors de Paris, pour l'ensemble des trois départements de la petite couronne, pour l'ensemble des quatre départements de la grande couronne et pour l'ensemble de l'Île-de-France hors Paris,
- indices trimestriels des prix des appartements pour l'ensemble de l'Île-de-France,
- indices trimestriels des prix des appartements dans chacun des huit départements de l'Île-de-France, pour l'ensemble des trois départements de la petite couronne, pour l'ensemble des quatre départements de la grande couronne et pour l'ensemble de l'Île-de-France hors Paris,
- indices trimestriels des prix maisons pour l'ensemble de l'Île-de-France,
- indices trimestriels des prix des maisons dans chacun des sept départements de l'Île-de-France en dehors de Paris, pour l'ensemble des trois départements de la petite couronne, pour l'ensemble des quatre départements de la grande couronne et pour l'ensemble de l'Île-de-France hors Paris.

Chacun de ces indices se décline en un indice provisoire, calculé environ un mois et demi après la fin du trimestre auquel il se rapporte, et un indice définitif, calculé environ quatre mois et demi après la fin du trimestre auquel il se rapporte.

De plus, chacun de ces indices se décline en un indice non corrigé des variations saisonnières et un indice corrigé des variations saisonnières

Dans toute la suite, les indices couverts par la présente convention sont dénommés génériquement « les indices ».

PNS informe l'Insee que des indices de prix des logements anciens mensuels sur trimestre glissant sont également calculés par PNS pour les trimestres ne correspondant pas à des trimestres civils. Ces indices mensuels sur trimestre glissant sont calculés selon la même méthode que les indices couverts par la présente convention et sont en continuité temporelle avec eux.

Article 2 : Appellation « Indices Notaires-Insee » et labellisation par l'Autorité de la statistique publique

Les indices bénéficient, sous réserve des dispositions de l'article 5 de la présente convention, de l'appellation « Indice Notaires-Insee ». A ce titre, l'Insee autorise PNS à utiliser cette appellation dans sa communication externe ainsi que dans ses relations avec les tiers.

PNS est informée que des indices analogues calculés pour la province par la société Min.not et le Conseil Supérieur du Notariat en partenariat avec l'Insee bénéficient également de l'appellation « Indice Notaires-Insee », ainsi que les indices relatifs à la France métropolitaine calculés par la société Min.not à partir des indices « Île-de-France » de PNS et « Province » de Min.not.

En outre, les indices d'Île-de-France ont fait l'objet d'une procédure de labellisation par l'Autorité de la statistique publique. Les indices Notaires-Insee d'Île-de-France ont obtenu cette labellisation.

Article 3 : Conseil scientifique

Le conseil scientifique des indices Notaires-Insee exerce un rôle de réflexion et de conseil vis-à-vis des parties de la présente convention. Ses interventions portent notamment (mais non exclusivement) sur le mode de calcul des indices et sur les modalités d'alimentation des bases notariales.

Il définit par exemple les taux de couverture en deçà desquels les indices ne sont pas validés.

Le Conseil scientifique comprend :

- i) un président, choisi d'un commun accord entre l'Insee, PNS, Min.not et le Conseil Supérieur du Notariat ;
- ii) deux représentants de l'Insee ;
- iii) deux représentants de PNS ;
- iv) un représentant de Min.not ;

v) un représentant du Conseil Supérieur du Notariat.

Le Conseil scientifique s'adjoit le concours des personnes qualifiées qu'il juge nécessaires à l'exercice de sa mission, pour la durée de la présente convention ou de manière plus ponctuelle.

Il se réunit, à la demande de son président, au moins une fois par trimestre. Ses réunions sont organisées par l'Insee et font l'objet de comptes rendus écrits. Chaque partie peut saisir le président d'une demande de réunion en tant que de besoin.

Le Conseil scientifique est commun aux deux conventions relatives au calcul, à la validation et à la diffusion des indices bénéficiant de l'appellation « Indice Notaires-Insee » :

- celle qui lie PNS et l'Insee ;
- celle qui lie la société Min.not, le Conseil Supérieur du Notariat et l'Insee.

Article 4 : Modifications de la présente convention

Toute modification apportée à la présente convention fera l'objet d'un avenant.

Article 5 : Engagements de l'Insee

L'Insee s'engage :

(i) à faire bénéficier PNS de son expertise et de ses conseils en matière statistique au profit du calcul des indices et à lui fournir les données en sa possession de nature à assurer la qualité des indices ;

(ii) à fournir à PNS chaque année au mois d'août pour ses séries les coefficients de correction des variations saisonnières nécessaires au calcul des indices corrigés des variations saisonnières ;

(iii) à faire figurer dans ses publications les indices couverts par la présente convention dont la diffusion lui paraîtra justifiée ;

L'Insee met en œuvre une méthode de suivi de la qualité des indices. Avant publication, les indices lui sont soumis pour approbation au moins une semaine avant. Il apporte sa réponse dans un délai maximal de deux jours ouvrés. Toute non-réponse au bout d'un délai de deux jours ouvrés vaut validation.

L'atteinte de taux de couverture minimaux est désormais une condition nécessaire pour la validation des indices. La méthodologie des indices est décrite en détail dans un numéro de la série Insee-Méthodes de l'Insee régulièrement mis à jour.

Article 6 : Engagements de PNS

PNS s'engage :

(i) à respecter les modalités de calcul des indices préconisées par le Conseil scientifique des indices Notaires-Insee,

(ii) à consacrer les moyens humains et matériels (en particulier informatiques) nécessaires à leur entretien et leur mise à jour et en particulier à adapter ces moyens à l'augmentation de la charge qui résultera de l'entrée en vigueur de l'obligation pour les notaires d'alimenter les bases servant au calcul des indices ; PNS mettra en place une évaluation régulière de ces besoins en moyens humains et financiers et des moyens effectivement consommés durant les derniers trimestres et prévus pour les trimestres suivants. Ces évaluations seront soumises à l'avis du Conseil scientifique des indices Notaires-Insee ;

(iii) à maintenir et améliorer le dispositif actuel de collecte des informations relatives aux ventes immobilières en Île-de-France : en particulier le taux de collecte et le délai d'intégration seront suivis via des tableaux de bord (nombre de mutations notariées et nombre de mutations intégrées par mois, stock en attente de codification) remis à l'Insee tous les trimestres. PNS s'engage à informer l'Insee avant toute modification du processus de constitution des bases de données notariales (définition des champs, modifications pouvant avoir un impact sur le taux de collecte et d'intégration, le taux de renseignement de certains champs, la valeur des indices) ;

(iv) à transmettre à l'Insee les informations définies en accord avec lui pour permettre la validation des indices. Celles-ci incluent notamment les indices détaillés, les niveaux de prix ainsi que les montants et les volumes de transactions réalisés au cours du trimestre ;

(v) à publier trimestriellement les indices ;

(vi) à fournir à l'Insee, dans le cadre de l'objet de la convention, toute information issue de ses bases utile à ses études statistiques internes, ainsi que toute série d'indices de prix calculés par elle-même pour ses propres besoins et qui seraient utiles à l'Institut à des fins d'études ;

(vii) à signaler dans ses publications toute modification de la méthode de calcul des indices mensuels sur trimestre glissant non couverts par la présente convention.

Article 7 : Publication et diffusion concertées des indices

La publication et la diffusion des indices, de la part de l'Insee et de PNS, sont soumises à embargo.

La date et l'heure de levée de l'embargo, c'est-à-dire la date et l'heure à compter desquelles l'Insee et PNS sont autorisés à publier et diffuser les indices, sont fixées selon les modalités ci-dessous :

- Les dates des publications de l'Insee et des conférences de presse de la Chambre interdépartementale des notaires de Paris sont proposées par le Conseil scientifique des indices Notaires-Insee et validées par l'Insee au moins un trimestre à l'avance. Le principe du calendrier est de les organiser dans la dernière semaine des mois de février, mai, août et novembre. Toutefois, la date de fin août sera systématiquement décalée d'environ deux semaines.
- L'heure de levée de l'embargo est fixée à 8 heures 45.

L'information relative aux indices est diffusée et publiée selon les modalités suivantes :

i) à l'attention des médias, par PNS :

Par le biais d'une conférence de presse, qui a lieu systématiquement le matin du jour de la levée de l'embargo, et dont l'horaire est fixé à huit heures quarante-cinq sauf exceptions. Une fois la conférence de presse tenue, la diffusion à l'attention des médias peut s'opérer par tout moyen approprié ;

Conformément aux recommandations de l'Autorité de la statistique publique, le dossier de presse remis aux journalistes ainsi que les documents présentés en séance devront présenter de façon clairement différenciée les indices qui font l'objet de l'appellation indices Notaires-Insee des autres informations.

ii) à l'attention des médias, par l'Insee :

Les informations sont communiquées aux agences de presse à huit heures et trente minutes et aux autres médias à huit heures et quarante-cinq minutes, le matin de la levée de l'embargo ;

iii) à l'attention du grand public, par PNS et / ou par l'Insee :

Les informations sont communiquées à partir de huit heures et quarante-cinq minutes, par tout moyen approprié.

Les dates d'arrêté de la base BIEN pour le calcul des indices sont fixées par PNS de façon à permettre la diffusion des indices à la date de levée de l'embargo, compte tenu des délais nécessaires pour procéder aux calculs, à leur vérification et leur validation, ainsi qu'à la préparation des publications.

Les coefficients ainsi que le zonage détaillé des régressions ne peuvent faire l'objet d'une communication publique qu'avec l'accord exprès de l'Insee et de PNS.

Les indices qui ne reçoivent pas l'approbation de l'Insee ne peuvent pas faire l'objet d'une publication avec la mention Indices Notaires INSEE

Article 8 : Conditions financières

Considérant que la constitution, le suivi et la publication régulière des indices est une tâche qui entre dans les missions respectives de l'ensemble des partenaires et que le travail de coopération prévu par la présente convention met à la charge de chacun d'entre eux des contreparties équilibrées, la présente convention est conclue à titre gratuit.

Article 9 : Durée

La présente convention entrera en vigueur dès qu'elle aura été signée par l'ensemble des partenaires. Elle est conclue pour une durée de cinq ans.

Article 10 : Conditions de dénonciation de la convention

La présente convention peut être dénoncée par l'une des parties au moyen d'une lettre recommandée avec avis de réception. Cette dénonciation prendra effet six mois après réception de cette lettre.

En cas de dénonciation, PNS et l'Insee renoncent pour l'avenir à l'utilisation de l'appellation « Indices Notaires-Insee ».

Toutefois les parties conviennent par avance de se concerter en vue de permettre la continuation des opérations financières engagées par des organismes financiers ou autres, dans le cadre des éventuels accords de licence d'exploitation de l'indice conclus par PNS.

Article 11 : Représentants de l'Insee et de PNS

Sont chargés de l'exécution des termes de la présente convention :

- i) pour l'Insee, le chef de la division Logement du département des Prix à la consommation, des ressources et des conditions de vie des ménages ;
- ii) pour PNS, le président en exercice de l'ASSOCIATION DES NOTAIRES DU CHATELET (PARIS NOTAIRES SERVICES).

Article 12 : Clauses exécutoires

La présente convention est dispensée du droit de timbre et de la formalité d'enregistrement.

L'entrée en vigueur de la présente convention a pour effet immédiat l'abrogation de la convention 2005 00353 et de son avenant de 2009.

Fait à Paris en sept exemplaires originaux, le 13 novembre 2013,

Pour PNS et les CHAMBRES DES NOTAIRES
D'ILE-DE-FRANCE

Signé : Maître Bénasse

Signé : Maître Hautebas

Signé : Maître Savoure

Signé : Maître Lemoine

Signé : Maître Herrnberger

Signé : Maître Cauro

Pour l'INSEE, le Directeur Général

Signé : Monsieur Jean-Luc Tavernier

Annexe 8 : Convention avec le Conseil supérieur du notariat et Min.not

Convention relative à la poursuite et au développement du partenariat entre l'Insee, le CSN et Min.not en matière d'indices de prix des logements en province,

entre :

- Le **CONSEIL SUPERIEUR DU NOTARIAT**, représenté par Me Jean Tarrade, notaire, son président,
 - La Société Min.not - **Marché Immobilier des Notaires**, SASU au capital de 2 450 000 €, immatriculée au R.C.S. d' Aix en Provence sous le numéro 381 000 611, représentée par M. Daniel BOUCHON, son Président,
- d'une part,

et :

L'INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES, désigné ci-après par l'acronyme Insee, et représenté par son directeur général M. Jean-Luc TAVERNIER,

d'autre part,

Il est convenu et arrêté ce qui suit :

Préambule

Le CSN, Perval et l'Insee ont conclu une première convention en date du 15 juin 1998, qui a mis en place un partenariat en vue du calcul et de la publication d'un indice de prix des appartements anciens vendus libres dans les villes-centres des agglomérations de 10.000 habitants ou plus, situées en France métropolitaine hors Île-de-France. Un avenant en date du 7 octobre 1999 a étendu le champ d'application de ce partenariat au calcul et à la diffusion d'un indice du prix des maisons anciennes dans l'ensemble des communes de province, ainsi qu'à un indice du prix des appartements anciens dans l'ensemble des communes de province des agglomérations de plus de 10.000 habitants (au lieu des seules villes-centres).

Le CSN, Perval et l'Insee ont conclu une deuxième convention en 2005 visant à poursuivre et renforcer le partenariat entre le CSN, Perval et l'Insee en matière d'indices de prix des logements, en étendant le champ d'application à des indices calculés au niveau des régions administratives et des grandes agglomérations, dès lors que la qualité statistique desdits indices serait reconnue comme satisfaisante par les parties. Cette deuxième convention a fait l'objet d'un avenant en 2010 visant d'une part à décrire la nature de l'information transmise par les notaires au Conseil scientifique des indices Notaires-Insee pour la validation trimestrielle des indices et d'autre part à préciser les conditions de diffusion des données et notamment à avancer les dates de publication des indices, pour tenir compte de l'amélioration de l'information recueillie par les bases notariales.

Les conventions signées en 1998 et 2005 sont arrivées à expiration. Elles sont remplacées par la présente convention, qui vise à poursuivre et renforcer le partenariat entre le CSN, Min.not et l'Insee en matière d'indices de prix des logements. La présente convention prend acte des différentes avancées récentes dans ce partenariat.

Article 1er : Objet de la convention

La présente convention définit les conditions de participation du CSN, de Min.not et de l'Insee au calcul, à la validation, à la publication et à la diffusion d'indices de prix des logements anciens vendus libres. Ceux-ci sont calculés à partir du fichier des transactions immobilières alimenté par les notaires de province, d'une part et à partir des indices de prix des logements anciens en Île-de-France calculés par Paris Notaires Service (PNS) d'autre part.

Min.not informe l'Insee que le traitement informatisé d'exploitation d'informations nominatives du fichier des transactions immobilières alimenté par les notaires de province et géré par l'ADSN est référencé dans la liste maintenue par le Correspondant Informatique et Libertés désigné par Min.not, conformément à la loi n°78-17 « Informatique et Libertés » du 6 janvier 1978.

Chacun des indices couverts par la présente convention se décline en un indice provisoire avancé, calculé environ un mois et demi après la fin du trimestre auquel il se rapporte, un indice provisoire, calculé environ trois mois après la fin du trimestre auquel il se rapporte, un indice semi-définitif, calculé environ quatre mois et demi

après la fin du trimestre auquel il se rapporte et un indice définitif, calculé environ six mois après la fin du trimestre auquel il se rapporte.

De plus, chacun de ces indices se décline en un indice non corrigé des variations saisonnières et un indice corrigé des variations saisonnières.

Les indices couverts par la présente convention sont les indices trimestriels des prix des logements anciens, bruts et corrigés des variations saisonnières :

- des appartements et maisons regroupés,
- des appartements,
- des maisons ;

pour les champs géographiques suivants :

- métropole,
- province,
- région Nord-Pas-de-Calais (indices provisoires, semi-définitifs et définitifs seulement),
- région Provence-Alpes-Côte-d'Azur (indices provisoires, semi-définitifs et définitifs seulement),
- région Rhône-Alpes (indices provisoires, semi-définitifs et définitifs seulement) ;

Les indices bruts et CVS suivants sont également publiés :

- maisons dans l'agglomération de Lille (indices provisoires, semi-définitifs et définitifs seulement),
- appartements :
 - o dans l'ensemble des agglomérations de plus de 10.000 habitants en province (indices provisoires, semi-définitifs et définitifs seulement),
 - o dans les villes-centres des agglomérations de plus de 10.000 habitants en province (indices provisoires, semi-définitifs et définitifs seulement),
 - o dans les banlieues des agglomérations de plus de 10.000 habitants en province (indices provisoires, semi-définitifs et définitifs seulement),
 - o dans les communes rurales et les communes des agglomérations de moins de 10.000 habitants en province (indices provisoires, semi-définitifs et définitifs seulement),
 - o dans l'agglomération de Lyon (indices provisoires, semi-définitifs et définitifs seulement),
 - o dans l'agglomération de Marseille (indices provisoires, semi-définitifs et définitifs seulement).

Min.not informe l'Insee que des indices de prix des logements anciens mensuels sur trimestre glissant sont également calculés par Min.not pour les trimestres ne correspondant pas à des trimestres civils. Ces indices mensuels sur trimestre glissant sont calculés selon la même méthode que les indices couverts par la présente convention et sont en continuité temporelle avec eux

Dans toute la suite, les indices couverts par la présente convention sont dénommés génériquement « les indices ».

Article 2 : Appellation « Indices Notaires-Insee »

Les indices bénéficient, sous réserve des dispositions de l'article 5 de la présente convention, de l'appellation « Indice Notaires-Insee » et le cas échéant d'une appellation personnalisée par région telle que « Indice Notaires-Insee Rhône-Alpes ». A ce titre, chacune des parties, le CSN, Min.not et l'Insee, est autorisée à utiliser ces appellations dans sa communication externe ainsi que dans ses relations avec les tiers.

Le CSN et Min.not sont informés que des indices analogues calculés pour l'Île-de-France par Paris Notaires Services (PNS) en partenariat avec l'Insee bénéficient également de l'appellation « Indice Notaires-Insee ».

Article 3 : Conseil scientifique

Le conseil scientifique des indices Notaires-Insee exerce un rôle de réflexion et de conseil vis-à-vis des parties de la présente convention. Ses interventions portent notamment (mais non exclusivement) sur le mode de calcul des indices et sur les modalités d'alimentation des bases notariales.

Il définit par exemple les taux de couverture en deçà desquels les indices ne sont pas validés.

Le Conseil scientifique comprend :

- i) un président, choisi d'un commun accord entre l'Insee, PNS, Min.not et le Conseil Supérieur du Notariat ;
- ii) deux représentants de l'Insee ;
- iii) deux représentants de PNS ;
- iv) un représentant de la société Min.not ;
- v) un représentant du Conseil Supérieur du Notariat.

Le Conseil scientifique s'adjoit le concours des personnes qualifiées qu'il juge nécessaires à l'exercice de sa mission, pour la durée de la présente convention ou de manière plus ponctuelle.

Il se réunit, à la demande de son président, au moins une fois par trimestre. Chaque partie peut saisir le président d'une demande de réunion en tant que de besoin. Ses réunions sont organisées par l'Insee et font l'objet de comptes rendus écrits.

Le Conseil scientifique est commun aux deux conventions relatives au calcul, à la validation et à la diffusion des indices bénéficiant de l'appellation « Indice Notaires-Insee » :

- celle qui lie PNS, les instances notariales des départements d'Île-de-France et l'Insee
- celle qui lie la société Min.not et le Conseil Supérieur du Notariat et l'Insee.

Article 4 : Modifications de la présente convention

Toute modification apportée à la présente convention fera l'objet d'un avenant.

Article 5 : Engagements de l'Insee

L'Insee s'engage :

- (i) à faire bénéficier Min.not de son expertise et de ses conseils en matière statistique au profit du calcul des indices, et à lui fournir les données en sa possession de nature à assurer la qualité des indices ;
- (ii) à fournir à Min.not chaque année au mois d'août les coefficients de correction des variations saisonnières nécessaires au calcul des indices corrigés des variations saisonnières relatifs à la province ;
- (iii) à faire figurer dans ses publications les indices couverts par la présente convention dont la diffusion répond aux critères de diffusion définis par le CSIN.

L'Insee met en œuvre une méthode de suivi de la qualité des indices. Avant publication, les indices lui sont soumis pour approbation au moins une semaine avant. Il apporte sa réponse dans un délai maximal de deux jours ouvrés. Toute non-réponse au bout d'un délai de deux jours ouvrés vaut validation.

L'atteinte de taux de couverture (méthode définie par le CSIN) minimaux est désormais une condition nécessaire pour la validation et la diffusion des indices. La méthodologie des indices est décrite en détail dans un numéro de la série Insee-Méthode de l'Insee régulièrement mis à jour.

Article 6 : Engagements du CSN et de Min.not

Min.not s'engage :

- (i) à respecter les modalités de calcul des indices préconisées par le conseil scientifique des indices Notaires-Insee,
- (ii) à consacrer les moyens humains et matériels (en particulier informatiques) nécessaires à leur entretien et leur mise à jour et en particulier à les adapter à l'augmentation de la charge qui résultera de l'entrée en vigueur de l'obligation pour les notaires d'alimenter les bases servant au calcul des indices ;

(iii) à suivre l'alimentation des données via des tableaux de bord (nombre de mutations notariées et nombre de mutations intégrées par trimestre, stock en attente de codification) remis à l'Insee tous les trimestres. Min.not s'engage à informer l'Insee de toute modification du processus de constitution des bases de données notariales (définition des champs, modifications pouvant avoir un impact sur le taux de collecte et d'intégration, le taux de renseignement de certains champs, la valeur des indices);

(iv) à transmettre à l'Insee les informations définies en accord avec lui pour permettre la validation des indices. Celles-ci incluent notamment les indices détaillés, les niveaux de prix ainsi que les montants et les volumes de transactions réalisés au cours du trimestre ;

(v) à publier trimestriellement les indices ;

(vi) à fournir à l'Insee, dans le cadre de l'objet de la convention, toute série d'indices de prix calculés par Min.not pour ses propres besoins et qui seraient utiles à l'Institut à des fins d'études ;

Le CSN s'engage :

(i) à maintenir et améliorer le dispositif actuel de collecte des informations relatives aux ventes immobilières en province : en particulier le taux de collecte et le délai d'intégration.

Article 6 bis – Protection des données à caractère personnel

Le CSN, Min.not et l'Insee, chacun en ce qui le concerne, s'engagent à respecter la réglementation relative à la protection des données à caractère personnel, dont la loi n°78-17 « Informatique et Libertés » du 6 janvier 1978 et ses décrets d'application.

Ainsi, ils s'engagent à prendre toutes précautions utiles afin de préserver la sécurité, la confidentialité des données à caractère personnel dont ils auraient connaissance dans le cadre de cette convention et notamment d'empêcher qu'elles ne soient déformées, endommagées ou communiquées à des personnes non expressément autorisées.

Il appartient au CSN, à Min.not et à l'Insee d'effectuer toute démarche administrative qui pourrait leur incomber au titre de la réglementation relative à la protection des données à caractère personnel.

A cet égard, il est précisé que le CSN et Min.not, ont chacun désigné un Correspondant Informatique et Libertés qui se charge de la tenue et de la mise à jour du registre des traitements et le tient à disposition de la Commission Nationale de l'Informatique et des Libertés.

Le CSN, Min.not et l'Insee se portent fort, au sens de l'article 1120 du Code civil, du respect par leurs préposés et/ou éventuels sous-traitants dûment autorisés, du présent article.

Il est convenu que la présente Convention pourra faire l'objet d'une communication à la Commission Nationale de l'Informatique et des Libertés.

Article 6 ter – confidentialité

Le CSN, Min.not et l'Insee sont tenus conventionnellement au secret le plus absolu sur toutes les informations issues des bases, documents ou tout élément notamment techniques, financiers ou organisationnels auxquels ils auraient accès dans le cadre de la présente Convention et/ou des conventions précédentes.

Chacune des parties s'engage à faire respecter cette obligation de confidentialité par l'ensemble de son personnel, tout préposé, tout contractant et tout sous-traitant.

La méconnaissance de cette obligation de confidentialité engagerait la responsabilité du CSN, Min.not ou de l'Insee et constituerait un des cas de résiliation anticipée dont la procédure est décrite à l'article 10 « Conditions de dénonciation de la convention ».

Article 7 : Publication et diffusion concertées des indices

La publication et la diffusion des indices de la part de l'Insee du CSN et de Min.not, sont soumises à embargo. La date et l'heure de levée de l'embargo, c'est-à-dire la date et l'heure à compter desquelles l'Insee, le CSN et Min.not sont autorisés à publier et diffuser les indices, sont fixées selon les modalités ci-dessous.

Les dates des publications de l'Insee sont proposées par le Conseil scientifique des indices Notaires-Insee et validées par l'Insee au moins un trimestre à l'avance

L'heure de levée de l'embargo est fixée à 8 heures 45.

L'information relative aux indices est diffusée et publiée selon les modalités suivantes :

i) à l'attention des médias, par le CSN et Min.not :

La diffusion à l'attention des médias peut s'opérer par le biais d'une conférence de presse. Si celle-ci a lieu, elle se tient le matin du jour de la levée de l'embargo à huit heures quarante-cinq sauf exceptions. Une fois la conférence de presse tenue, la diffusion à l'attention des médias peut s'opérer par tout moyen approprié.

Conformément aux recommandations de l'Autorité de la statistique publique, le dossier de presse remis aux journalistes ainsi que les documents présentés en séance devront présenter de façon clairement différenciée les indices qui font l'objet de l'appellation indices Notaires-Insee des autres informations.

ii) à l'attention des médias, par l'Insee :

Les informations sont communiquées aux agences de presse à huit heures et trente minutes et aux autres médias à huit heures et quarante-cinq minutes le matin de la levée de l'embargo ;

iii) à l'attention du grand public, par le CSN, Min.not et/ou par l'Insee :

Les informations sont communiquées à partir de huit heures et quarante-cinq minutes, par tout moyen approprié.

Les dates d'arrêt de la base pour le calcul des indices sont fixées par Min.not de façon à permettre la diffusion des indices à la date de levée de l'embargo, compte tenu des délais nécessaires pour procéder aux calculs, à leur vérification et leur validation, ainsi qu'à la préparation des publications.

Les coefficients ainsi que le zonage détaillé des régressions ne peuvent faire l'objet d'une communication publique qu'avec l'accord exprès de l'Insee, du CSN et de Min.not.

Les indices qui ne reçoivent pas l'approbation de l'Insee ne peuvent pas faire l'objet d'une publication sous l'appellation indices Notaires-Insee par l'Insee.

Article 8 : Conditions financières

Considérant que la constitution, le suivi et la publication régulière des indices est une tâche qui entre dans les missions respectives des parties et que le travail de coopération prévu par la présente convention met à la charge de chacune d'entre elles des contreparties équilibrées, la présente convention est conclue à titre gratuit.

Article 9 : Durée

La présente convention entrera en vigueur dès qu'elle aura été signée par les trois parties. Elle est conclue pour une durée de cinq ans.

Article 10 : Conditions de dénonciation de la convention

La présente convention peut être dénoncée par l'une des parties au moyen d'une lettre recommandée avec avis de réception. Cette dénonciation prendra effet six mois après réception de cette lettre.

En cas de dénonciation, le CSN, Min.not et l'Insee renoncent pour l'avenir à l'utilisation de l'appellation 'Indices Notaires-Insee'. Toutefois l'Insee et Min.not conviennent par avance de se concerter en vue de permettre la continuation des opérations financières engagées par des organismes financiers ou autres dans le cadre des éventuels accords de licence d'exploitation de l'indice conclus par Min.not et le CSN.

Article 11 : Représentants de l'Insee, du CSN et de Min.not

Sont chargés de l'exécution des termes de la présente convention :

i) pour l'Insee, le chef de la division Logement du département des Prix à la consommation, des ressources et des conditions de vie des ménages ;

ii) pour le CSN, son Président ;

iii) pour Min.not, son Président.

Article 12 : Clauses exécutoires

La présente convention est dispensée du droit de timbre et de la formalité d'enregistrement.

L'entrée en vigueur de la présente convention a pour effet immédiat l'abrogation de la convention 2005 00329 et de son avenant de 2010.

Fait à Paris, en trois exemplaires originaux,

Pour Min.not et le CSN

Signé :

Signé :

Pour l'INSEE, le Directeur Général

Signé : Monsieur Jean-Luc Tavernier

Annexe 9 : Lexique

Ancien

On utilise une définition fiscale du logement ancien. Une transaction est réputée porter sur un bien ancien s'il s'agit de la première vente plus de cinq ans après la date d'achèvement des travaux, ou s'il s'agit d'une seconde vente, quelle que soit la date d'achèvement des travaux. Il peut donc s'agir d'une première vente d'un bien déjà ancien (plus de cinq ans), ou d'une seconde vente d'un bien quasi-neuf. La distinction fiscale correspond à une taxation différente : 0,60% sur le neuf, 4,80% sur l'ancien.

Banlieue

Il s'agit des communes de banlieue des unités urbaines de 10 000 habitants ou plus : tout ce qui n'est pas ville-centre est banlieue.

Bien de référence

Bien dont on suit le prix pour calculer les indices. C'est le bien dont les caractéristiques sont les modalités de référence des variables explicatives du modèle de prix des transactions (par exemple, maisons de 4 pièces, à deux niveaux, avec un garage et une salle de bains). Le bien de référence est présenté au chapitre 3 (*tableau 3.1*).

Collectif

Les logements collectifs sont assimilés à des appartements (studio, appartement, duplex, triplex). Sont exclus les chambres, greniers, lofts, ateliers, logements de gardiens.

État d'occupation et destination

Les logements retenus dans les indices sont libres d'occupation au moment de la vente, destinés à un usage strict d'habitation et acquis en pleine propriété. On ne retire pas les appartements loués par l'acquéreur avant la vente. On retire les appartements occupés par un tiers ou par le vendeur lorsque la privation de jouissance excède six mois et lorsqu'il y a un droit d'usage et d'habitation ou une réserve d'usufruit.

Individuel

Les logements sont appelés logements individuels si ce sont des maisons, qu'elles soient isolées ou groupées. Elles disposent d'une entrée indépendante privative directe sur l'extérieur. On rejette pour des raisons d'homogénéité les types de biens suivants : grandes propriétés, châteaux, hôtels particuliers, tours, moulins... On retient donc les fermes, maisons de ville ou de village, pavillons et villas. On retient aussi les cas où la nature de la maison n'est pas précisée.

Nature de la mutation

Les seules mutations prises en compte sont les ventes de gré à gré, réalisées directement entre vendeur et acquéreur ou par l'intermédiaire d'un professionnel de l'immobilier. Les ventes sur adjudication volontaire du Marché Immobilier des Notaires sont donc exclues.

Nature du vendeur et de l'acquéreur

Le vendeur peut être un particulier, un professionnel, ou une société. Seuls les logements acquis par un particulier ou par une société civile immobilière (SCI) entrent dans le calcul des indices. Les logements acquis par des professionnels sont hors champ. Les non-réponses, beaucoup plus nombreuses, sont prises dans le champ de l'indice.

Tableau A9.1 - Nombre de logements vendus selon la nature de l'acquéreur en province

Nature de l'acquéreur	Appartements						Maisons	
	Villes-centres des UU >10 000		Banlieues des UU >10 000		Rural et UU <10 000		Nombre	%
	Nombre	%	Nombre	%	Nombre	%		
Particuliers	78 763	89%	34 883	90%	14 789	88%	193 190	91%
S.C.I.	4 602	5%	1 714	4%	808	5%	8 377	4%
Marchands de biens	546	1%	210	1%	53	0%	1 288	1%
Autres (non professionnels)*	949	1%	480	1%	210	1%	2 912	1%
Non-réponse	3 830	4%	1 632	4%	959	6%	7 310	3%
Total	88 690	100%	38 919	100%	16 819	100%	213 077	100%

* Administrations, entreprises, etc.

Champ : mutations de logements anciens des bases de données notariales, 2010.

Tableau A9.2 - Nombre de logements vendus selon la nature de l'acquéreur en Île-de-France (appartements)

Nature de l'acquéreur	Appartements							
	Paris		Petite couronne		Grande couronne		Total	
	Nombre	%	Nombre	%	Nombre	%	Nombre	%
Particuliers	25 376	85%	33 125	92%	24 092	95%	82 593	90%
SCI	2 140	7%	1 471	4%	661	3%	4 272	5%
Marchands de biens	231	1%	92	0%	28	0%	351	0%
Autres	691	2%	599	2%	187	1%	1 477	2%
Non-réponse	1 435	5%	859	2%	449	2%	2 743	3%
Total	29 873	100%	36 146	100%	25 417	100%	91 436	100%

Tableau A9.3 - Nombre de logements vendus selon la nature de l'acquéreur en Île-de-France (maisons)

Nature de l'acquéreur	Maisons					
	Paris et petite couronne		Grande couronne		Total	
	Nombre	%	Nombre	%	Nombre	%
Particuliers	9 303	90%	26 476	94%	35 779	93%
SCI	468	5%	779	3%	1 247	3%
Marchands de biens	82	1%	147	1%	229	1%
Autres	229	2%	268	1%	497	1%
Non-réponse	257	2%	438	2%	695	2%
Total	10 339	100%	28 108	100%	38 447	100%

Champ : mutations de logements anciens des bases de données notariales, 2010.

Parc de référence

Parc dont l'évolution de la valeur constitue l'indice (cf. chapitre 3). Pour un quartier ou une ville, il se compose de l'ensemble des transactions de la période de référence entrant dans le champ de l'indice, sauf celles dont le prix est jugé extrême (1/20^{ème} à chaque extrémité de la distribution)⁷⁸. C'est le panier ou portefeuille de biens de l'indice.

Parc d'estimation

Parc de logements dont les valeurs servent à estimer les prix relatifs des caractéristiques des biens. Il est constitué de tous les logements vendus au cours de la période de référence, entrant dans le champ de l'indice. On exclut les transactions jugées aberrantes, c'est-à-dire celles dont les résidus sont supérieurs à deux écart-types, c'est-à-dire situés en dehors de l'intervalle $[\bar{x} - 2\sigma; \bar{x} + 2\sigma]$. Le parc d'estimation et le parc de référence sont donc des sous-ensembles de l'ensemble des transactions de la période d'estimation entrant dans le champ de

⁷⁸ Dans la version 1, par précaution et dans l'ignorance de la qualité des données, on avait éliminé 1/6^{ème} des transactions extrêmes (David et al., 2002). Des tests ultérieurs ont montré que les indices étaient robustes à une élimination plus parcimonieuse des extrêmes.

l'indice (tableau A9.4). La non-réponse et le traitement des observations manquantes sont détaillés au chapitre 4 (tableau 4.8).

Par exemple : pour la strate « banlieue de Lille en maisons anciennes », on a 10 337 transactions en 2007 et 2008. Parmi celles-ci, 2 831 comportent des variables manquantes ou mal renseignées, et 340 ont un prix estimé supérieur ou inférieur au prix observé de plus de deux écart-types. Finalement, il reste 7 166 transactions dans le parc d'estimation. Le parc de référence est quant à lui constitué de 6 772 transactions : les 7 506 transactions d'origine du parc d'estimation (7 166 + 340), diminuées des références extrêmes.

Tableau A9.4 - La taille du parc de référence et du parc d'estimation des indices Notaires-Insee

	Nombre de transactions du parc de référence*	Nombre de transactions du parc d'estimation*
France Métropolitaine	723 691	764 084
Île-de-France		
Appartements	146 089	154 318
Paris	45 811	48 269
Petite Couronne	58 673	62 126
Grande Couronne	41 605	43 923
Maisons (total)	55 792	58 893
Petite Couronne (+Paris)	14 923	15 733
Grande Couronne	40 869	43 160
Province		
Appartements	209 824	222 026
UU > 10 000 centre	128 992	137 162
UU > 10 000 banlieue	56 502	59 827
Rural et UU < 10 000	24 330	25 037
Maisons	311 986	328 847

* Parc 2007-2008

Tableau A9.5 - Paris : comparaison des structures du parc total, du parc de référence et des transactions annuelles, en %

	Nombre de logements*	Nombre de transactions du parc de référence**	Nombre de transactions en 2010	Nombre de transactions en 2011
Strate				
1	20,0	20,4	19,7	20,2
2	30,0	33,0	32,4	31,8
3	26,0	23,3	23,6	24,1
4	10,8	10,1	10,5	10,2
5	13,2	13,1	13,8	13,6
Total	100,0	100,0	100,0	100,0
Époque de construction				
Avant 1992	95,0	86,8	79,8	75,6
Depuis 1992	5,0	2,2	2,6	2,5
Non renseigné	0,0	11,0	17,6	21,7
Total	100,0	100	100,0	100,0
Taille				
1 ou 2 pièces	55,8	57,0	55,7	57,9
3 pièces et plus	44,2	41,2	42,5	40,6
Non renseigné	0,0	1,8	1,8	1,5
Total	100,0	100,0	100,0	100,0

* Source : recensement de la population de 2008

** Parc 2007-2008

Le tableau A9.5 compare, pour Paris, la structure du parc de référence (2007 - 2008) à celles du parc des logements au recensement de la population de 2008, et des transactions des années 2010 et 2011. Une proportion non négligeable des variables ne sont pas renseignées dans les bases (modalité « non renseigné ») et par ailleurs les transactions excluent le parc de logements neufs. Les comparaisons sont donc approximatives. Globalement cependant, ces structures témoignent d'une bonne représentation du parc total par les transactions.

La structure détaillée des parcs d'estimation pour les appartements et les maisons d'Île-de-France et de province est donnée en fin d'annexe.

Période de base, trimestre de base de l'indice

Le 1^{er} trimestre 2010.

Période d'estimation ou période de référence

Période sur laquelle on estime les modèles de base, c'est-à-dire les prix relatifs des caractéristiques des biens. Cette période est de deux ans, dans la version 3 des indices.

Prix du bien

Prix net vendeur (commission d'agence déduite si elle figure dans l'acte), donc hors taxes et frais de notaire.

Strate, quartier

Les strates sont des zones géographiques à l'intérieur desquelles les prix sont homogènes (cf. chapitre 3). Pour la méthode de détermination des strates, cf. annexe 2. Pour le découpage communal des strates, cf. classeur Excel, obtenu sur demande auprès des notaires.

Une strate correspond donc au champ d'application d'un modèle hédonique distinct. Un quartier est une zone plus fine dont l'influence se traduit par des indicatrices appropriées à l'intérieur d'un modèle hédonique. Les quartiers ne correspondent pas nécessairement à des découpages administratifs.

Unité urbaine

On définit une unité urbaine comme un ensemble d'habitations telles qu'aucune ne soit séparée de la plus proche de plus de 200 mètres, abritant au moins 2 000 habitants. Les communes répondant à ces critères forment des unités urbaines (UU), les autres sont dites rurales. La notion d'unité urbaine, fondée sur la continuité du bâti et sur le seuil de 2 000 habitants dits agglomérés, est donc une notion plutôt visuelle, basée sur la démographie et l'habitat.

Ville-centre

Une commune est à elle seule ville-centre si elle représente plus de la moitié de la population de l'agglomération. Sinon, toutes les communes qui ont une population supérieure à la moitié de la population de la commune la plus importante sont villes-centre, avec cette dernière.

Si le nombre annuel de mutations est inférieur à 110, la ville-centre n'a pas d'indice propre.

Annexe 10 : Bibliographie

Pour une bibliographie plus complète, voir Beauvois et al. (2005), <http://www.insee.fr/fr/publications-et-services/sommaire.asp?id=91&nivgeo=0> disponible en ligne sur ce site.

Bailey M., Muth R., Nourse H. : "A Regression Method for Real Estate Price Index Construction", *Journal of the American Statistical Association*, 58, 933-942, 1963.

Beauvois M., David A., Dubujet F., Friggit, J., Gouriéroux C., Laferrère A., Massonnet S., Vrancken E. : "Les indices Notaires-Insee de prix des logements anciens, Version 2 des modèles hédoniques", *Insee Méthode*, n° 111, 2005

Case B., Pollakowski H., Wachter S. : "Frequency of Transaction and House Price Modelling", *JREFE*, 14, 173-188, 1997.

Clapp J., Giacotto C. : "House Price Indices Based on All Transactions Compared to Repeat Subsamples", *AREUEA*, 19, 270-285, 1991.

Court A. : "Hedonic Price Indexes with Automotive Examples", p. 99-117, *The Dynamics of Automobile Demand, New-York, General Motors Company*, 1939.

David A., Dubujet F., Gouriéroux C. et Laferrère A. : "Les indices de prix des logements anciens", *Insee méthodes*, n° 98, 2002.

Duon, G., "Documents sur le problème du logement", Ministère de l'Économie, Service National de la Statistique, *Études Économiques*, n°1, 1946.

Eichholtz Piet M.A., "A Long Run House Price Index : the Herengracht Index, 1629-1973", *REE*, 25 (2), 175-192, Amsterdam, August 1996.

Grebler, Blank et Winnick, Annexe C de "Capital Formation in Residential Real Estate: Trends and Prospects", *National Bureau of Economic Research*, Princeton University Press, Princeton, 1956.

Griliches Z. : "Price Indices and Quality Change", Cambridge, Harvard University Press, 1971.

Mark J., Goldberg M. : "Alternative Housing Price Indices : An Evaluation", *AREUEA*, 12, 30-49, 1984.

Moreau A. : "Methodology of the Price Index for Microcomputers and Printers in France", dans *Industry Productivity : International Comparison and Measurement Issues*, OECD Proceedings, 99-118, 1996.

Triplett Jack E. : "Concept of Quality in Input and Output Price Measures: A Resolution of the User-Value Resource Cost Debate", dans *The US Income and Product Accounts: Selected Topics*, éd M.F. Foss, NBER Studies in Income and Wealth, University of Chicago Press, Chicago, 1983.

Triplett Jack E. : "Handbook on Hedonic Indexes and Quality Adjustments in Prices Indexes: Special Application to Information Technology Products", *STI Working papers 2004/9*, OECD, 2004.

Warsh D. : "Economics principals, Roots and Wings", *Globe Columnist*, 14 mars 1999.