
SOMMAIRE

**Ont participé
à cet ouvrage**

Philippe Clarenc
Jean-François Côte
Alain David
Jacques Frigitt
Philippe Gallot
Stéphane Gregoir
Anne Laferrère
Adélia Nobre
Catherine Rougerie
Nelly Tauzin

**Directeur
de la publication**
Jean-Luc Tavernier

© Insee 2014
ISSN : 1142-3080
ISBN : 978-2-11-138698-3

Résumé	3
1 - Introduction	5
2 - Théorie des indices hédoniques	9
2.1. Le pourquoi des approches hédoniques	9
2.1.1. Modèle hédonique avec stratification a priori	9
2.1.2. Des démarches alternatives problématiques	10
2.2. Utilisation du modèle hédonique	10
2.2.1. Estimation des coefficients correctifs	11
2.2.2. Estimation de la valeur de référence de la date t	11
2.2.3. Construction d'un système de valorisation	12
2.2.4. Construction d'une batterie d'indices	12
2.3. Rendre plus robuste l'approche hédonique	12
2.3.1. Recherche des scores sous-jacents	13
2.3.2. Regroupement de strates	13
2.4. Surveiller la spécification	13
2.5. Extension du modèle de base	14
3 - Calculs mis en œuvre	17
3.1. Parcs de référence et d'estimation	17
3.2. Le modèle servant à l'estimation	18
3.3. Prix courant du logement de référence	19
3.4. Valeur courante du parc de référence	21
3.5. La méthode de calcul	21
3.5.1. Renouvellement des parcs de référence et d'estimation tous les deux ans	22
3.5.2. Calcul des indices élémentaires en période courante	23
3.5.3. Calcul des indices agrégés en période courante	25
3.6. Un exemple numérique pour novembre 2012	26
4 - Les bases de donnée notariales	29
4.1. Description des bases	29
4.2. Délai d'intégration des transactions	29
4.3. Couverture	31
4.4. Champ des indices qualité des données	33
4.4.1. Délimitation de champ	33
4.4.2. Transactions retenues pour le calcul	33
4.4.3. Les variables disponibles	34
4.4.4. Traitement des informations manquantes	36
4.4.5. Structure des parcs d'estimation	37
5 - Estimation de l'équation hédonique, mise en œuvre pratique	43
5.1. Estimation de l'équation hédonique en deux étapes	43
5.2. Exemples numériques	44
5.2.1. Exemples de régressions (parc 2009-2010)	44
5.2.2. Qualité des régressions	55
6 - Désaisonnalisation des indices, suivi et publication	61
6.1. Désaisonnalisation des indices	61
6.1.1. Méthode de calcul des corrections pour variations saisonnières	61
6.1.2. Prise en compte des corrections pour jours ouvrables	61

6.1.3. Nature des données désaisonnalisées	61
6.1.4. Fréquence de calcul des coefficients CVS	61
6.1.5. Méthode directe ou indirecte	62
6.1.6. Calcul	62
6.1.7. Résultats	65
6.2. Tableau de bord	66
6.2.1. Objectifs et composition du tableau de bord	66
6.2.2. Exemples de tableaux	66
6.3. Publication	69

7 - Comparaisons de la V2 avec la V3 73

7.1. Les différentes sources d'écart	73
7.2. Écart global entre les versions 2 et 3	73
7.3 Décomposition des écarts entre les versions 2 et 3	78
7.3.1. Écarts dus à l'utilisation de données super-définitives	78
7.3.2. Impact de l'actualisation des zonages	79
7.3.3. Impact de la nouvelle méthode de calcul	80
7.3.4. Impact de la correction d'une erreur	82
7.3.5. Impact de la nouvelle méthode de correction des variations saisonnières	83

Annexe 1 : La transmission des actes de ventes aux bases notariales et l'estimation des taux de couverture	85
Annexe 2 : Le découpage géographique	89
Annexe 3 : Mise à jour de la spécification des modèles	107
Annexe 4 : Tests de stabilité et durée de la période de référence	115
Annexe 5 : Les méthodes de calcul alternatives	121
Annexe 6 : Les indices de prix avancés	127
Annexe 7 : Convention avec PNS	129
Annexe 8 : Convention avec le Conseil supérieur du notariat et Min.not	135
Annexe 9 : Lexique	141
Annexe 10 : Bibliographie	145
Annexe 11 : Actualité des indices	147

Résumé

Connaître l'évolution des prix des logements est un préalable au bon fonctionnement et à la transparence de leurs marchés. Une méthode a été mise au point par l'Insee en liaison avec le notariat pour produire des indices de prix réguliers et fiables des logements anciens, les indices Notaires-Insee. Elle a été exposée dans sa première version dans le numéro 98 d'Insee Méthodes, (David et al., 2002), puis dans une version révisée, la « version 2 », dans le numéro 111 d'Insee Méthodes (Beauvois et al., 2005). Le présent numéro d'Insee-Méthodes présente une nouvelle version, la « version 3 », des indices Notaires-Insee.

La difficulté à estimer des indices de prix pour les logements tient au fait que le prix d'un logement donné est rarement observé et que la qualité des logements évolue au cours du temps. Pour traiter cette difficulté, la méthode retenue repose sur des modèles économétriques (modèles hédoniques), fondés sur l'estimation des prix relatifs des caractéristiques des logements. Dans chaque zone élémentaire, un modèle est estimé à partir d'un parc de logements, dit parc d'estimation ; il est supposé s'appliquer pour une durée de quelques années. Il permet, à chaque période courante, d'estimer la valeur d'un parc fixe de logements (« parc de référence »), à partir de la totalité des transactions observées à la période courante, en calculant la moyenne des prix reconstitués d'un bien de référence qui leur correspond. Le rapport de la valeur estimée du parc de référence à la période courante sur sa valeur observée à la période de base constitue l'indice.

Les données sont fournies par les notaires, qui rédigent les contrats pour la quasi-totalité des ventes de logements anciens en France. Ces données couvrent à l'heure actuelle environ les deux tiers de l'ensemble du champ. Les caractéristiques des logements comportent leur superficie, leur nombre de pièces, de salles de bains, l'étage, la présence d'un ascenseur, d'un garage, etc.

Comme prévu initialement, les modèles initiaux ont été révisés après quatre ans. Ceci a donné lieu à une « version 2 » des indices, en vigueur de juillet 2004 à octobre 2011. Toujours conformément aux engagements initiaux, une nouvelle révision a été effectuée (version 3), en vigueur depuis novembre 2011. La mise en place des nouveaux indices s'accompagne de la publication plus systématique d'indices corrigés des variations saisonnières, pour un meilleur suivi des tendances du marché.

Cette version comprend 293 modèles hédoniques de base, estimés dans des zones géographiques appelées strates, homogènes en termes de prix, séparément pour les maisons et les appartements : 15 modèles pour les appartements et 7 pour les maisons en Île-de-France, 97 pour les appartements et 174 pour les maisons en province.

Les indices couvrent le territoire métropolitain et sont validés sous l'appellation « indices Notaires-Insee » ; leur conception est surveillée par un Conseil scientifique. Ils sont publiés chaque trimestre dans une première version provisoire deux mois après la fin du trimestre, puis dans une version définitive six mois après la fin du trimestre (cinq mois pour les indices d'Île-de-France). Les indices trimestriels des prix des logements anciens en Île-de-France ont été labellisés par l'Autorité de la statistique publique (ASP).

Le présent document décrit les enrichissements apportés à la méthode de calcul et inclut des annexes détaillées sur les caractéristiques techniques des nouveaux modèles.

Chapitre 1 : Introduction

L'information et la transparence sont des conditions nécessaires au bon fonctionnement d'un marché concurrentiel. Il est donc crucial que les acteurs économiques puissent disposer d'informations fiables sur les prix des logements et leur évolution, d'autant que le logement occupe une place essentielle dans l'économie, en termes de production, de patrimoine, de budget et d'endettement des ménages.

Un logement vendu peut être neuf (jamais occupé auparavant) ou ancien (« d'occasion »). Depuis janvier 2013, un indice du prix des logements neufs est diffusé sur le site www.insee.fr ainsi qu'un indice de prix sur l'ensemble des logements (neufs et anciens). *Le présent document ne porte que sur les indices des logements anciens*¹.

Le prix d'un logement est rarement observé

La fabrication d'un indice de prix des logements pose les mêmes problèmes que celle de tout indice de prix : comment séparer des évolutions « pures » de prix et les évolutions de la qualité des logements ? Mais elle pose aussi des problèmes spécifiques.

D'une part, deux logements ne sont jamais tout à fait identiques. Un logement est une combinaison de caractéristiques (on emploiera aussi le terme de qualités) dont seul l'ensemble constitue le « bien logement ». Il est en effet à la fois un emplacement géographique (et donc l'accès à un voisinage caractérisé par des biens publics locaux), une construction (caractérisée par son type, sa taille et un niveau de confort multidimensionnel) et souvent un bien affectif, lieu de souvenirs familiaux.

D'autre part, un logement ne change de main que rarement. On peut estimer qu'en moyenne il est vendu tous les 35 ans² et subit une mutation à titre gratuit ou onéreux tous les 25 ans. L'observation du prix d'un logement donné est donc rare. Ce grand décalage temporel entre achat et revente d'un même logement complique la fabrication d'un indice de prix. Certains indices se basent sur le prix affiché ou le prix demandé, d'autres sur des estimations, mais il est préférable de disposer de prix de marché, c'est-à-dire ceux des transactions effectivement réalisées.

Comment mesurer l'évolution du prix d'un logement donné ou, plus généralement, d'un ensemble de logements, alors qu'on n'observe que quelques transactions à chaque période ?

Quels sont les prix observés ?

Si l'on veut construire un indice d'évolution du prix du stock de tous les logements, on ne peut se contenter de faire la moyenne des prix des transactions courantes pour la comparer à celle de la période précédente. Une telle comparaison mélangerait des effets de prix à ceux de l'évolution de la qualité et de la non-représentativité des transactions. Le problème a en effet deux composantes : d'une part les transactions observées à chaque période ne sont pas tirées au hasard dans le stock de logements ; d'autre part le stock évolue et sa qualité ne reste pas constante d'une période à l'autre.

Faisons abstraction du second problème et supposons que le stock de logements n'évolue pas, ou peu, c'est-à-dire qu'il n'y a pas de construction neuve et que l'entretien compense exactement l'usure des immeubles. Si les transactions d'une période donnée étaient suffisamment nombreuses et tirées au hasard dans ce stock de logements, on pourrait se contenter de faire la moyenne des prix des transactions courantes pour la comparer à celle de la période précédente et obtenir ainsi un indice des prix du stock de logements. Mais si, ce qui est fréquent en pratique, les transactions d'un trimestre ne sont pas un échantillon représentatif du stock de tous les

¹ Dans les années qui ont suivi la deuxième guerre mondiale, le neuf a prédominé dans les achats de résidences principales (il était encore majoritaire au début des années 1980 : 58 % des transactions réalisées entre 1980 et 1984, selon l'enquête Logement de 1984. Il ne représentait plus, en 2006, qu'un petit tiers des achats effectués au cours des quatre années précédentes (source : enquête Logement de 2006).

² Cette estimation fait abstraction du parc détenu par des personnes morales (organismes de logements sociaux principalement, mais aussi investisseurs institutionnels et sociétés foncières), qui est le plus souvent vendu en bloc. Ces durées moyennes entre deux mutations sont calculées sur la base de 800 000 mutations à titre onéreux et 330 000 mutations à titre gratuit (pour lesquelles la personne de référence du ménage propriétaire change) par an, rapportées à un parc de 28 millions de logements détenus autrement qu'en société (source : CGEDD).

logements, une telle comparaison mélange des évolutions de qualité de l'échantillon aux évolutions des prix³. Pour le comprendre, supposons le prix de chaque logement absolument constant dans le temps ; un trimestre donné on observe surtout des ventes dans les beaux quartiers puis, le trimestre suivant, surtout des transactions de bas de gamme. Le prix moyen va baisser, sans rapport avec l'évolution réelle, nulle, qui est pourtant celle que veut saisir un indice de prix pur.

La qualité des logements évolue

Passons maintenant au second problème. Si nous supposons à l'inverse disposer à chaque date d'un échantillon de transactions représentatif du stock de logements, mais que ce stock n'est pas fixe, nous ne pouvons pas non plus construire d'indice de prix, puisque la comparaison mélangera un effet de prix et l'évolution de la qualité du stock au fur et à mesure que les logements les plus anciens s'useront, ou seront détruits et remplacés par des neufs⁴.

Le problème posé par l'évolution de la qualité est familier des fabricants d'indices de prix. Si le prix d'une ampoule électrique passe de 1 à 2 euros, alors que sa durée de vie augmente, son prix a apparemment doublé. Mais si la satisfaction du consommateur d'ampoule est la même pour une ampoule de durée de vie longue que pour deux ampoules de durée de vie courte, on dira que la qualité a aussi été multipliée par deux, et donc que le prix « pur » est resté constant. Pour les biens de consommation, ces changements de qualité sont fréquents ; parfois la comparaison même des biens est délicate quand on est face à un nouveau produit.

En matière de logement, la situation est à la fois plus simple et plus compliquée. Plus compliquée parce que, comme on l'a dit, la qualité ne peut être mesurée selon une seule dimension, comme par exemple la durée de vie pour l'ampoule électrique. Le nombre de caractéristiques à prendre en compte est grand⁵. Ceci implique le recours à des techniques économétriques dites « hédoniques ». Mais la situation du logement est aussi plus simple pour le statisticien parce que le nombre de révolutions techniques est limité ; sur une période donnée de quelques années, les changements de qualité (au sens de l'apparition d'une nouvelle caractéristique d'un logement) sont lents. Il s'agira par exemple d'un changement dans le nombre moyen de salles de bain par logement ou des évolutions technologiques liées au respect de normes environnementales nouvelles.

On fait donc une première hypothèse : les caractéristiques des logements susceptibles d'influencer leur prix, telles que la qualité du quartier, le nombre de pièces, la présence d'un ascenseur ou l'étage, sont en nombre fini, et ce nombre est constant pendant la durée du calcul de l'indice. Si une nouvelle caractéristique de qualité apparaît, qui fait varier le prix des logements, disons l'existence de climatiseurs, ou l'obligation d'un contrôle de l'amiante, dans un premier temps elle ne sera pas prise en compte dans l'estimation des indices comme une qualité nouvelle, mais comme une variation de prix. En d'autres termes, on admet que la satisfaction des consommateurs n'est pas encore vraiment influencée par la climatisation ou le contrôle d'amiante, et qu'ils les considèrent comme une hausse de prix. Ces changements de qualité seront intégrés par les statisticiens grâce à l'adaptation à périodicité régulière de la spécification des modèles.

La méthode hédonique

Les méthodes économétriques sont appliquées dès qu'un bien est composite, que ce soit en matière d'indice de prix à la consommation (par exemple pour les ordinateurs, dont la valeur des caractéristiques évolue, sans que de nouvelles caractéristiques apparaissent aussi fréquemment⁶), de coût d'un service (comme la banque⁷, ou l'hôpital), ou dans les études du marché du travail où ce sont les caractéristiques des travailleurs qui sont hétérogènes et multidimensionnelles⁸.

Les techniques hédoniques sont nées aux Etats-Unis entre les deux guerres mondiales, en pleine crise économique. Entre 1925 et 1935, selon les indices de prix officiels, estimés à partir de prix moyens, les prix de l'automobile avaient augmenté de 45% et la pression était forte pour que General Motors baisse ses prix pour

³ Ce problème de l'évolution de la qualité des transactions observées d'une période à l'autre est parfois qualifié de problème de pondération. Il est bien posé par Triplett (1983) qui distingue indices de prix d'input et d'output.

⁴ Ce second problème est curieusement peu évoqué dans la littérature sur les prix des logements.

⁵ En réalité de nombreux biens de consommation ou d'équipement présentent aussi ce caractère composite et multidimensionnel. Qu'on songe à un ordinateur, un voyage en avion ou même un simple paquet de sucre dont l'emballage peut varier.

⁶ Sur l'utilisation des méthodes hédoniques appliquées aux ordinateurs, voir Moreau (1996), Triplett (2004).

⁷ Dans ce cas, c'est plutôt le nombre des caractéristiques qui évolue.

⁸ Zvi Griliches a beaucoup travaillé sur cette technique ; voir par exemple, Griliches (1971).

maintenir la consommation et l'emploi. C'est alors qu'en 1939 Andrew Court, qui travaillait pour l'Automobile Manufacturers Association et savait que la qualité des automobiles avait aussi changé (verre de sécurité, changement de vitesse, moteurs plus puissants...), mit au point une méthode pour tenir compte de l'évolution des caractéristiques des automobiles dans les calculs d'indices des prix⁹. Court définit une automobile type, de caractéristiques de vitesse, **sécurité**, vitres, largeur de sièges, etc., données, qui servirait d'étalon du confort ou du *plaisir*¹⁰ de l'utilisateur. Ensuite il divisa le prix moyen des véhicules par cet étalon, son « contenu hédonique ». Il trouva que le prix de l'automobile type avait baissé de 55% sur la période¹¹.

En matière de prix des logements, outre la première hypothèse (on peut définir chaque logement par la combinaison d'un nombre limité de caractéristiques), on suppose aussi, seconde hypothèse, que la relation entre le prix au mètre carré et les caractéristiques d'un logement est, pour un type de logement et une zone élémentaire donnés, fixe pour la durée du calcul de l'indice. Dans une zone élémentaire, l'indice de prix des logements est défini comme le rapport entre la valeur d'un parc de référence fixe pendant la période courante et sa valeur pendant la période de base de l'indice. La valeur de chaque logement de ce parc est estimée chaque trimestre à l'aide des prix des transactions observées effectivement et de relations économétriques estimées sur une période de temps déterminée.

Les données disponibles et la mise en œuvre de la méthode

En France, les mutations immobilières s'effectuent devant un notaire, officier public nommé par l'État, qui assure la publication des actes au fichier immobilier et collecte certaines taxes pour le compte de l'État. Si chaque acte doit comprendre l'identification de l'assiette foncière sur la matrice cadastrale, cette obligation ne concerne pas le descriptif du logement vendu. Depuis plusieurs années, les notaires, conscients de l'importance de ces informations, font cependant l'effort de les collecter.

A Paris, un indice du prix des appartements anciens avait été calculé régulièrement depuis 1984 selon une méthode de stratification mise au point avec l'Insee. En 1997, le Conseil supérieur du notariat a voulu créer un indice de prix pour les logements situés en province. Ce fut l'occasion pour l'Insee de mettre au point une méthode rigoureuse de calcul d'un indice de prix de l'ensemble des logements anciens, en suivant la route ouverte par Court et Griliches¹². Cette méthode a été présentée pour la première fois en 2002 dans « Les indices de prix des logements anciens », document publié dans Insee Méthodes, n°98¹³. Une seconde version de ce document a été réalisée en 2005¹⁴. La présente publication (*version 3*) constitue donc une deuxième mise à jour.

Depuis la première version des modèles, la collecte des données sur les transactions immobilières a progressé : l'augmentation du nombre de transactions enregistrées a permis de passer à des indices trimestriels « purs », contrairement aux premiers indices, qui étaient calculés sur une base « glissante », annuelle ou semestrielle. En outre, les indices publiés sont aujourd'hui plus nombreux et plus rapidement disponibles : des indices provisoires sont publiés environ huit semaines après la fin du trimestre, soit des délais similaires à nos homologues des Etats-Unis et de la plupart des pays européens. Sont publiés au même moment les indices définitifs du trimestre précédent¹⁵. Aujourd'hui, la chambre interdépartementale des notaires de Paris (CINP), au travers de son association Paris notaires service (PNS), calcule chaque trimestre les indices franciliens en version brute et corrigée des variations saisonnières (CVS). La société Min.not, filiale du groupe ADSN (Association pour le développement du service notarial) fait de même pour la province et calcule les indices France entière. L'Insee calcule tous les ans les coefficients CVS et valide les indices avant chaque publication trimestrielle.

L'Assemblée nationale a adopté le 28 mars 2011 la loi n° 2011-331 de modernisation des professions judiciaires ou juridiques et certaines professions réglementées. Les articles 15 et 16 de cette loi instaurent, à la charge de la profession notariale, une mission de service public consistant à collecter, transmettre, centraliser et diffuser l'information sur les mutations d'immeubles à titre onéreux (ce qui inclut non seulement les ventes de logements anciens, mais aussi les ventes en l'état futur d'achèvement de logements neufs, les ventes de terrains, ou encore

⁹ A la même époque le Bureau of Labor Statistics essayait de comparer les qualités des tracteurs et des camions, pour avoir une idée de ce qui était changement de qualité et changement de prix.

¹⁰ D'où le mot hédonique pour qualifier la méthode, qui est en réalité une méthode de régression économétrique tout à fait banale.

¹¹ L'histoire est rapportée par Warsh (1999).

¹² Ont travaillé à la mise au point de la méthode utilisée : Alain David, François Dubujet, Christian Gouriéroux, Anne Laferrère et Claude Taffin.

¹³ David et al. (2002).

¹⁴ Beauvois et al. (2005).

¹⁵ Cependant, la Corse et les DOM restent exclus du champ des indices nationaux.

celles de locaux à usage non résidentiel). Le décret d'application a été publié le 3 septembre 2013¹⁶. L'amélioration du taux de couverture des bases qui devrait en découler ouvre la voie à un enrichissement de l'information régionale, départementale et locale, et notamment à la mise en place d'indices infranationaux supplémentaires, sur le modèle de ceux qui sont déjà publiés pour les régions Rhône-Alpes, Provence-Alpes-Côte d'Azur et, plus récemment, Nord-Pas de Calais.

Les indices en nouvelle base (version 3) décrits dans le présent ouvrage ont été mis en service fin 2011. Leur mise en place s'accompagne de la publication plus systématique d'indices corrigés des variations saisonnières, pour un meilleur suivi des tendances du marché. De plus, les perspectives ouvertes par le développement de la télétransmission des actes par les études notariales devraient permettre de raccourcir la durée de leur traitement. Les indices trimestriels des prix des logements anciens en Île-de-France ont été labellisés par l'Autorité de la statistique publique (ASP) en juin 2011¹⁷. Cette dernière veille au respect du principe d'indépendance professionnelle dans la conception, la production et la diffusion de statistiques publiques et accorde la labellisation à des statistiques produites à des fins d'intérêt général, par des administrations, organismes publics et organismes privés chargés d'une mission de service public. Les indices portant sur la province pourront être examinés par l'ASP dans un second temps.

Des prolongements variés

La méthode de calcul de l'indice des prix des logements peut avoir des prolongements variés.

Une application peut être de créer un système expert d'évaluation des prix des logements. L'objectif est de voir si un bien proposé à la vente est sous ou surévalué par rapport à un autre bien qui sert de référence. En outre, l'outil peut utiliser les coefficients obtenus lors de l'estimation de l'équation hédonique pour transformer le prix de la transaction en prix équivalent bien de référence.

Une autre application serait de calculer des indices de prix selon les modalités d'une variable d'intérêt. Des premières investigations sont en cours et la possibilité de décliner les indices selon le nombre de pièces du logement est à l'étude.

Enfin, pour avoir une connaissance du marché plus réactive, il a été décidé construire une nouvelle génération d'indices, des indices dits « avancés » des prix de logements anciens. Il s'agit d'utiliser non pas les mutations réalisées au cours d'un trimestre donné mais les avant-contrats de ventes. En effet, la plupart des avant-contrats se concrétisent en transactions définitives et pourraient constituer de bons indicateurs des tendances du marché.

Contenu de cet ouvrage

A la suite de ce chapitre introductif, un second chapitre développe la théorie des indices hédoniques. Le troisième détaille les calculs mis en œuvre et les nouveautés apportées lors de cette deuxième mise à jour. Le chapitre 4 décrit les bases de données utilisées. Au chapitre 5, la méthode d'estimation de l'équation hédonique ainsi que quelques exemples sont présentés. La désaisonnalisation des indices, les outils de leur suivi trimestriel, les publications sont détaillés au chapitre 6. Enfin, au chapitre 7, nous comparons les séries d'indices obtenues en version 2 et celles obtenues en version 3.

¹⁶ JORF n°0206 du 5 septembre 2013 page 14976, texte n° 2, Décret n° 2013-803 du 3 septembre 2013 relatif aux bases de données notariales portant sur les mutations d'immeubles à titre onéreux.

¹⁷ L'ASP a été créée par la loi n°2008-776 du 4 août 2008. Elle veille au respect du principe d'indépendance professionnelle dans la conception, la production et la diffusion de statistiques publiques ainsi que des principes d'objectivité, d'impartialité, de pertinence et de qualité des données produites (art. 144). Elle a notifié la labellisation de la statistique trimestrielle du prix du logement en Île-de-France, produite conjointement par la chambre des notaires de Paris et l'Insee, pour une durée de cinq ans, dans son avis n° 2011-01 du 21 juin 2011.

Chapitre 2 : Théorie des indices hédoniques

2.1 Le pourquoi des approches hédoniques

Les classes d'une nomenclature de biens et services, même restreintes, comportent des éléments de caractéristiques très disparates. Les logements (les biens considérés ici) se distinguent ainsi par leur localisation, leur superficie, leur nombre de pièces, le fait d'être occupés ou pas, leur ancienneté... tandis que les locations de logements (les services) se différencient en plus par les durées de bail, les possibilités de renouvellement, les clauses de sorties anticipées.

Cette hétérogénéité se retrouve sur les marchés correspondants, entraînant des taux de rotation ou d'occupation des logements variés, et des prix très différents. Elle entraîne diverses difficultés pour l'analyse des prix.

La principale vient du fait que le prix d'un logement ne peut être observé qu'au moment des transactions, assez rares (observabilité partielle) ; de façon analogue un loyer n'est observable que s'il y a utilisation du service. En dehors de ces situations, ces valeurs, prix et loyers, n'ont d'ailleurs pas d'existence au sens économique.

Une façon classique de contourner cette difficulté est de supposer des *valeurs implicites*, appelées également dans notre contexte « estimation du prix » du logement et « valeur locative ». Ces valeurs implicites ne sont connues que lorsqu'elles coïncident avec des prix de transaction ou des loyers et ne pourront donc être reconstituées qu'à partir de modèles décrivant la composition des prix ou des loyers, et leur évolution.

Les approches hédoniques reposent sur de tels modèles et expliquent comment les utiliser pour construire les valeurs non observées et définir des ensembles cohérents d'indices de prix.

2.1.1 Modèle hédonique avec stratification a priori

Nous allons présenter l'approche hédonique à partir d'une formulation simplifiée. Nous supposons que les divers biens peuvent être regroupés en strates, définies a priori et telles que les évolutions de prix soient approximativement parallèles à l'intérieur d'une même strate. Elles peuvent en revanche différer sensiblement d'une strate à l'autre.

Dans notre application, les strates seront des zones géographiques élémentaires où se trouvent les logements. Nous désignons ces strates par $s, s=1, \dots, S$. Étant donné un logement i de la strate s , de caractéristiques $z_{i,s,t}$ (superficie, nombre de pièces, etc.), nous estimons, grâce à un modèle de régression, sa valeur implicite $p^*_{i,s,t}$ à la date t .

A l'intérieur de la strate s , les valeurs implicites sont supposées telles que :

$$p^*_{i,s,t} \approx c(s, z_{i,s,t}) p^*_{s,t} \quad (2.1)$$

- où $p^*_{s,t}$ est une valeur implicite de référence pour la strate s à la date t ,
- $c(s, z_{i,s,t})$ un coefficient correctif tenant compte des caractéristiques du bien et dont la valeur peut dépendre de la strate,
- \approx signifie « peu différent de ».

Le *correctif* et la *valeur de référence* sont définis à un scalaire multiplicatif près (ce qui est un problème d'identification). Il est alors possible de fixer la valeur de référence comme correspondant à un bien de qualité pré-spécifiée (logement de référence), z_0 :

$$p^*_{0,s,t} \approx p^*_{s,t} \Leftrightarrow c(s, z_0) = 1$$

disons le deux-pièces, ancien, se trouvant au rez-de chaussée, etc.

$\frac{p^*_{i,s,t}}{c(s, z_{i,s,t})}$ sera nommé « prix équivalent bien de référence » du logement (i,s) à la date t .

L'approximation (2.1) peut-être rendue plus proche d'un modèle économétrique en introduisant de façon appropriée des termes d'erreur et en spécifiant une forme paramétrée pour le correctif. Une telle spécification est par exemple :

$$\log p_{i,s,t}^* = \sum_{k=1}^K \beta_{k,s} X_k(z_{i,s,t}) + \log p_{s,t}^* + \varepsilon_{i,s,t}^* \quad (2.2)$$

où les termes d'erreur $\varepsilon_{i,s,t}^*$ sont supposés indépendants, de moyenne nulle et de variance $\eta_{s,t}^2$ dépendant éventuellement de la strate et de la date. Les $X_k(z_{i,s,t})$ sont des variables explicatives en nombre K , fonctions des caractéristiques du logement ou de croisements de telles ou telles caractéristiques.

La formulation (2.2) présente l'avantage de recourir à un modèle linéaire dans les paramètres $\hat{c}_0(s, z) = \exp\left[\sum_{k=1}^K \hat{\beta}_{k,s} X_k(z)\right]$, qui vont permettre de définir le correctif, et $\log p_{s,t}^*$, qui va fournir la valeur de référence de la strate.

On notera que le modèle comporte des effets croisés de la strate et des autres caractéristiques du logement, les coefficients $\beta_{k,s}$ étant spécifiques de la strate.

2.1.2 Des démarches alternatives problématiques

Peut-on reconstituer des évolutions de prix sans utiliser un modèle hédonique ? Deux démarches alternatives sont possibles : l'une, la méthode des ventes répétées, est difficile à mettre en œuvre rigoureusement, et l'autre, basée sur l'observation des prix moyens, conduit à des résultats biaisés.

Même si un bien donné i est rarement échangé à deux dates successives données $t, t+1$ (on parle alors de données répétées), on peut cependant espérer mesurer des évolutions de prix en comparant deux biens analogues. Si les variables explicatives $X_k(z)$ sont par exemple qualitatives, il peut exister plusieurs biens de la strate ayant approximativement les mêmes niveaux de prix et pas seulement des évolutions parallèles. On utilisera ces prix pour calculer une évolution. Cependant, dès que les effets de qualité sont suffisamment importants et nombreux, le marché peut ne pas être assez liquide pour que ceci soit réalisable. Par ailleurs on néglige une grande partie de l'information contenue dans les données de transaction en ne retenant que les prix correspondant aux ventes répétées.

Une seconde démarche souvent proposée consiste à comparer le prix moyen des biens de la strate s échangés en $t+1$, au prix moyen des biens de cette même strate échangés en t , avec l'espoir d'approcher ainsi l'évolution de la valeur de référence. Cette approche est biaisée, car la structure par qualité des biens échangés n'est pas stable dans le temps. Pour illustrer cette difficulté, considérons un cas, où un seul bien est échangé à chaque date dans la strate, le bien étant de qualité z_t à la date t . Le rapport des prix observés serait approximativement :

$$\frac{c(s, z_{t+1}) p_{s,t+1}^*}{c(s, z_t) p_{s,t}^*}$$

et diffère de l'évolution de la valeur de référence à cause de la modification du terme correctif.

2.2 Utilisation du modèle hédonique

L'utilisation de modèles permet de pallier l'inobservabilité partielle et de corriger les effets de qualité. La démarche comprend plusieurs étapes selon le schéma ci-dessous.

- Étape 1 : estimation des coefficients correctifs grâce à un ensemble prédéfini de données de transactions, appelé dans la suite *parc d'estimation*. Choix d'un ensemble de logements, appelé *parc de référence*, dont on suivra les prix.
- Étape 2 : à chaque date t , utilisation des données de transactions effectives et des correctifs estimés à l'étape 1 pour reconstituer les valeurs des logements du parc de référence.

- Étape 3 : utilisation des valeurs de référence et des correctifs pour construire une batterie d'indices de prix et un système expert de valorisation.

2.2.1 Estimation des coefficients correctifs

On considère un parc d'estimation constitué d'un ensemble de transactions durant une période prédéfinie $t=1, \dots, T_0$ (dite *période d'estimation*). Les transactions fournissent des couples (prix $p_{j,s,t}$, qualité $z_{j,s,t}$), $j=1, \dots, J_{s,t}$, $s=1, \dots, S$, $t=1, \dots, T_0$, où $J_{s,t}$ désigne le nombre de transactions dans la strate s à la période t .

On estime les paramètres $\beta_{k,s}$, $k=1, \dots, K$ dans chaque strate, par moindres carrés ordinaires, grâce à l'équation (2.2). On en déduit une estimation du terme correctif de la strate :

$$\hat{c}_0(s, z) = \exp \left[\sum_{k=1}^K \hat{\beta}_{k,s} X_k(z) \right]$$

ainsi qu'une fourchette, éventuellement, pour chacun de ces termes, en considérant la précision des coefficients estimés :

$$[\hat{c}_1(s, z), \hat{c}_2(s, z)]$$

Ces correctifs seront conservés inchangés pendant une période future, dont la longueur doit être précisée.

2.2.2 Estimation de la valeur de référence de la date t

Considérons maintenant une date t , postérieure à la période d'estimation et des données de transaction :

$p_{j,s,t}$, $z_{j,s,t}$, pour chacun des biens j de la strate s à la date t . Nous avons :

$$\begin{aligned} \log p_{j,s,t} &= \log c(s, z_{j,s,t}) + \log p_{s,t}^* + \varepsilon_{j,s,t} \\ &\cong \log \hat{c}_0(s, z_{j,s,t}) + \log p_{s,t}^* + \varepsilon_{j,s,t} \end{aligned}$$

après remplacement des correctifs par leurs approximations obtenues à la première étape.

On obtient $J_{s,t}$ estimations du prix du bien de référence de la strate s à la date t , $p_{s,t}^*$, et on en déduit

l'approximation par moindres carrés ordinaires de $p_{s,t}^*$:

$$\begin{aligned} \log \hat{p}_{s,t}^* &= \frac{1}{J_{s,t}} \sum_{j=1}^{J_{s,t}} [\log p_{j,s,t} - \log \hat{c}_0(s, z_{j,s,t})] \\ \Leftrightarrow \hat{p}_{s,t}^* &= \prod_{j=1}^{J_{s,t}} \left[\frac{p_{j,s,t}}{\hat{c}_0(s, z_{j,s,t})} \right]^{\frac{1}{J_{s,t}}} \end{aligned}$$

$p_{s,t}^*$ est donc estimé par la moyenne géométrique des prix équivalents bien de référence. La prise en compte des écarts-types permet également de proposer une fourchette pour la valeur de référence. Plus précisément, notons :

$\hat{\eta}_{s,t}^2$ la variance empirique des valeurs $\log(p_{j,s,t} / \hat{c}_0(s, z_{j,s,t}))$, $j=1, \dots, J_{s,t}$.

Cette variance empirique approche celle $\eta_{s,t}^2$ du terme d'erreur. La valeur de référence admet un logarithme compris entre $\log \hat{p}_{s,t}^* - 2\hat{\eta}_{s,t}$ et $\log \hat{p}_{s,t}^* + 2\hat{\eta}_{s,t}$, et la fourchette est :

$$(\hat{p}_{1,s,t}^* = \exp(-2\hat{\eta}_{s,t}) \hat{p}_{s,t}^*, \hat{p}_{2,s,t}^* = \exp(2\hat{\eta}_{s,t}) \hat{p}_{s,t}^*)$$

2.2.3 Construction d'un système de valorisation

On peut alors estimer les valeurs implicites de tout bien de la date t , de qualité z , non nécessairement échangé. L'estimation est :

$$\hat{p}_{s,t}^* \hat{c}_0(s, z)$$

la fourchette peut être prise égale à :

$$(\hat{p}_{1,s,t}^* \hat{c}_1(s, z), \hat{p}_{2,s,t}^* \hat{c}_2(s, z))$$

2.2.4 Construction d'une batterie d'indices

La méthode hédonique a permis de définir des indices élémentaires par strate :

$$I_{s,t} = \hat{p}_{s,t}^*, s = 1, \dots, S, t = 1, \dots, T$$

qui peuvent servir de base à la construction d'indices synthétiques. Cette dernière est menée de façon classique, en suivant par exemple une approche de type indice de Laspeyres. Pour cela, on définit dès la période initiale un panier de biens (un parc de logements), appelé parc de référence, en précisant les qualités z et strates s concernées. Z désigne les qualités différentes introduites dans le panier (parc de référence) et $N_{z,s}$ le nombre de biens (logements) de caractéristiques z dans la strate s .

L'indice synthétique est défini en suivant la valeur de ce panier :

$$I_t = \sum_{s=1}^S \sum_{z \in Z} N_{z,s} I_{s,t}$$

Ce panier peut aussi servir à construire des indices synthétiques désagrégés de façon cohérente. On peut par exemple introduire un indice « deux pièces », en considérant le sous-parc restreint aux seuls deux pièces :

$$I_t(\text{deux pièces}) = \sum_{s=1}^S \sum_{z \in Z} N_{z,s} I_{s,t} \quad (\text{deux pièces})$$

ou un indice pour une région donnée, calculé sur un sous-ensemble de strates :

$$I_t(\text{région}) = \sum_{s \in \text{région}} \sum_{z \in Z} N_{z,s} I_{s,t}$$

De façon standard, on peut prendre par convention une année de base, $t = 0$. La normalisation doit alors être effectuée indice par indice, conduisant à des indices base 100 à $t = 0$ donnés par :

$$I_{t/0} = 100 I_t / I_0$$

$$I_{t/0}(\text{deux pièces}) = 100 I_t(\text{deux pièces}) / I_0(\text{deux pièces})$$

2.3 Rendre plus robuste l'approche hédonique

L'approche, reposant sur des estimations, peut être sensible à la précision de celles-ci ou à l'instabilité des paramètres. Il est utile de faire éventuellement quelques regroupements de variables de qualité ou de strates permettant d'accroître la significativité des résultats.

2.3.1 Recherche des scores sous-jacents

Considérons les termes correctifs. Pour chaque strate, nous avons estimé un ensemble de paramètres

$$\hat{\beta}_{1,s}, \dots, \hat{\beta}_{K,s} \text{ définissant le score de la strate } s : \sum_{k=1}^K \hat{\beta}_{k,s} X_k.$$

On peut examiner, si ces S scores ne dépendent pas d'un nombre plus faible de scores sous-jacents. Pour cela, une démarche est la suivante :

1. On définit la matrice \hat{B} , de taille (K, S) , dont les colonnes sont les vecteurs $(\hat{\beta}_{1,s}, \dots, \hat{\beta}_{K,s})$;
2. On détermine les valeurs propres $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_{S_0}$ et les vecteurs propres associés $\hat{\alpha}_1, \dots, \hat{\alpha}_{S_0}$ de la matrice $\hat{B}'\hat{B}$, où \hat{B}' désigne la transposée de \hat{B} .
3. Le nombre de valeurs propres S_0 significativement différentes de zéro donne le nombre de scores indépendants sous-jacents et ceux-ci sont donnés par :

$$Z_l = \hat{\gamma}_l' X = \hat{\alpha}_l' \hat{B}' X, l = 1, \dots, S_0$$

4. La robustification consiste à contraindre le terme correctif à être de la forme :

$$c(s, z) = \exp \left[\sum_{l=1}^{S_0} \lambda_{l,s} Z_l \right]$$

Comme S_0 est inférieur à la fois à K et S , souvent assez petit, il y a beaucoup moins de paramètres à estimer dans cette forme contrainte de terme correctif.

Quitte à désagréger en partie les scores Z_l déduits de l'approche précédente, il est usuel de retenir des sous-scores ne comportant que des variables avec des interprétations de même type. Par exemple un sous-score effectuera la correction pour les caractéristiques physiques du logement, un second pour son aménagement, un troisième pour la qualité de l'environnement, un quatrième pour sa localisation... On obtient ainsi une structure hiérarchique des effets des variables, qui facilite la mise en place des systèmes experts de valorisation, leur compréhension et leur mise à jour.

2.3.2 Regroupement de strates

De façon analogue, on peut examiner les strates par l'intermédiaire de l'évolution des indices $I_{s,t}$ correspondants. Des analyses des corrélations empiriques entre ces séries temporelles peuvent permettre de repérer des strates dont les valeurs de référence évoluent de façon parallèle. Elles peuvent alors, si ceci est interprétable, être regroupées.

2.4 Surveiller la spécification

Les valeurs des paramètres peuvent se modifier dans le temps et les correctifs, calculés à la période d'estimation, se détériorer. Il est important de mettre en place des instruments de suivi de la qualité du modèle afin de repérer le moment où il devra être réestimé et avoir des idées sur les modifications à apporter. On pourra procéder à un examen approprié des résidus d'estimation.

A chaque date t de tels résidus sont :

$$\hat{\varepsilon}_{j,s,t} = \log p_{j,s,t} - \log \hat{c}_0(s, z_{j,s,t}) - \log p_{s,t}^*, j = 1, \dots, J_{s,t}$$

Ils doivent être agrégés pour à la fois éliminer des effets qualités et orienter la surveillance vers les paramètres $\beta_{k,s}$ susceptibles d'être affectés. Pour cela, on peut considérer diverses caractéristiques marginales, par

exemple : deux-pièces (d.p.), logement ancien (a.), ..., et pour chacune d'entre elles moyenner les résidus correspondants¹⁸ :

$\hat{\varepsilon}_{s,t}(d.p.) =$ moyenne des $\hat{\varepsilon}_{j,s,t}$ sur les deux pièces de la strate s et de la date t ,

$\hat{\varepsilon}_{s,t}(a.) =$ moyenne analogue sur les logements anciens...

Si le modèle est bien spécifié, de telles moyennes devraient varier autour de zéro. On va donc rechercher des écarts systématiques. Par exemple constater des valeurs $\hat{\varepsilon}_{s,t}(a.)$ trop souvent positives pour la strate s_0 , à partir d'une certaine date t_0 , peut nécessiter la réévaluation d'un paramètre relatif à une variable explicative X_{k_0} en rapport avec l'ancienneté. La mise en évidence d'une tendance dans la suite $t \rightarrow \hat{\varepsilon}_{s_0,t}(a.)$ peut signifier que l'hypothèse de proportionnalité d'évolution des prix à l'intérieur de la strate s n'est plus satisfaite et que cette strate doit être décomposée.

2.5 Extension du modèle de base

Le modèle hédonique utilisé jusqu'à maintenant à titre d'illustration présente l'inconvénient de définir a priori des strates homogènes pour les évolutions des prix. Ce modèle peut être généralisé de la façon suivante. Commençons par écrire la formulation (2.2) sous une forme résumée :

$$\log p_{i,s,t}^* = \sum_{s_0=1}^S \sum_{k=1}^K \xi_{s_0}(i,s,t) \beta_{k,s_0} X_k(z_{i,s,t}) + \sum_{s_0=1}^S \log p_{s_0,t}^* \xi_{s_0}(i,s,t) + \varepsilon_{i,s,t}^*$$

où ξ_{s_0} désigne la variable indicatrice de la strate s_0 , c'est-à-dire celle qui vaut 1, si l'observation se trouve dans la strate s_0 , 0 sinon. Sous cette forme, le modèle s'applique à l'ensemble des données toutes strates confondues, et reste linéaire dans les divers paramètres :

$$\beta_{k,s_0}, k=1, \dots, K, s_0=1, \dots, S, \log p_{s_0,t}^*, s_0=1, \dots, S, t=1, \dots, T$$

Ses limites apparaissent clairement dans les expressions des variables explicatives : la partie « terme constant » inclut des effets croisés strate \times qualité très spécifiques, alors que les parties donnant la dynamique des valeurs de référence : $p_{s_0,t}^*$ ne font pas intervenir les effets de qualité autres que l'indicatrice de la strate.

Un modèle élargi est alors :

$$\log p_{i,t}^* = c_0(z_{i,t}; \theta_0) + \sum_{l=1}^L c_l(z_{i,t}; \theta_l) f_{l,t} + \varepsilon_{i,t}^* \quad (2.3)$$

où $f_{1,t}, \dots, f_{L,t}$ sont des facteurs dynamiques à déterminer, $c_0(z_{i,t}; \theta_0), \dots, c_L(z_{i,t}; \theta_L)$ des termes correctifs donnant par exemple les sensibilités aux facteurs, avec des paramètres $\theta_0, \dots, \theta_L$ à estimer. Il n'y a plus lieu ici de distinguer la strate s qui est réintégrée parmi les autres caractéristiques du logement.

La démarche hédonique peut être appliquée à partir d'un tel modèle. Par exemple, en situation courante de la date t , les valeurs des facteurs seront estimées par moindres carrés ordinaires sur le modèle approché :

$$\log p_{j,t} \cong \hat{c}_0(z_{j,t}) + \sum_{l=1}^L \hat{c}_l(z_{j,t}) f_{l,t}$$

où les $\hat{c}_l(z_{j,t})$ ont été déterminés à la période d'estimation. Nous ne discuterons pas de façon plus approfondie les procédures d'estimation correspondantes. Signalons seulement que le modèle (2.3) permet de déterminer les meilleures prévisions des facteurs en supposant tous les prix implicites observables. Ces prévisions apparaissent

¹⁸ Si le nombre de transactions se révèle insuffisant pour calculer de telles moyennes, on peut augmenter la période sur laquelle la moyenne est calculée en considérant deux ou trois dates consécutives.

comme des combinaisons linéaires des logarithmes de prix, avec des coefficients dont la somme peut être normalisée à 1 :

$$f_{l,t}^* \cong \sum_{i=1}^N \pi_l(z_{it}) \log p_{i,t}^*$$

$$\Leftrightarrow \exp f_{l,t}^* = \prod_{i=1}^N [p_{i,t}^*]^{\pi_l(z_{i,t})}$$

À chaque facteur correspond ainsi implicitement un panier (parc) de composition $(\pi_l(z_{i,t}), i \text{ variant})$, variant dans le temps, tel que l'évolution de f_l soit proche de celle de la valeur du panier (du parc). On parle de *panier* (parc) *mimétique* du facteur (Huberman, Kandel, Stambaugh, 1987). Ainsi il existe des choix adaptés des indices désagrégés dont les évolutions vont reproduire celles des facteurs sous-jacents.

Chapitre 3 : Calculs mis en œuvre

L'indice élémentaire est défini comme le rapport de la valeur courante d'un parc de logements de référence à sa valeur de la période de base. La méthode décrite en détail dans ce chapitre est la même pour tous les indices, sous deux réserves :

- les variables représentatives de la qualité des biens (caractéristiques) ne sont pas exactement identiques pour les appartements et pour les maisons, et selon que l'on considère l'Île-de-France ou la province ;
- la manière de découper le champ géographique en strates varie selon que l'on considère l'Île-de-France ou la province.

A l'occasion du passage de la version 2 à la version 3, la méthode de calcul des indices a été modifiée. Afin de mieux prendre en compte les évolutions du marché, les parcs d'estimation et de référence (cf. définitions ci-dessous) seront actualisés plus fréquemment (tous les deux ans). Conséquence de ce changement, les indices sont désormais chaînés. Trois formules de chaînage sont mises en œuvre : la première concerne les indices élémentaires et les deux autres les indices agrégés.

3.1 Parcs de référence et d'estimation

Définition 1 : le parc et la période de référence

Le parc de référence constitue le « panier de biens » ou « portefeuille » dont on va mesurer la variation de prix. On s'assure ainsi que l'indice retrace bien l'évolution du prix des mêmes logements et qu'il n'est pas sensible aux variations de la structure du marché.

Il s'agit donc de valoriser les logements du parc de référence à la date courante, alors qu'ils n'ont pas été vendus, c'est-à-dire que l'on n'a pas pu observer leur prix. Pour cela, on utilise un modèle économétrique qui relie le prix d'un logement¹⁹ à ses caractéristiques physiques et à sa localisation.

Le *parc de référence* est constitué de toutes les transactions d'une période - dite *période de référence* -, entrant dans le champ de l'indice, sauf celles situées aux extrêmes de la distribution des valeurs²⁰.

Dans la version 3 des indices, la période de référence est constituée des années $n-3$ et $n-2$ pour les indices des années n et $n+1$ (avec n pair), puisque le parc est mis à jour tous les deux ans à partir de 2008 (voir plus loin 3.5.1).

Définition 2 : le parc et la période d'estimation

Le parc d'estimation regroupe les transactions qui serviront à l'estimation des modèles. Il est constitué de toutes les transactions de la période d'estimation entrant dans le champ de l'indice, sauf celles pour lesquelles la valeur estimée par le modèle s'écarte de la valeur réelle de plus de deux écarts-types. Les parcs de référence et d'estimation sont découpés en zones élémentaires de calcul ou strates, au nombre de 22 en Île-de-France et 271 en province (tableau 3.2)²¹.

En province, certaines strates sont composées d'une seule ville, lorsque celle-ci possède un marché immobilier de taille suffisante (en pratique, plus de 110 transactions par trimestre). Dans la plupart des strates, les modèles peuvent faire apparaître, via des indicatrices, des zones plus détaillées de prix homogènes. Les autres strates de province sont définies par l'ensemble des villes appartenant au croisement des critères : région, département, unité urbaine ou d'autres distinctions géographiques (espace rural, communes balnéaires ou massifs alpins), ou encore selon le revenu moyen par habitant (année 2006) pour d'autres communes. Les strates d'Île-de-France sont définies sur des critères géographiques et de prix²².

¹⁹ Plus exactement le logarithme du prix.

²⁰ Il s'agit du prix au mètre carré dans le cas des appartements et du prix total en ce qui concerne les maisons. On rabote 1/20^{ème} à chaque extrémité de la distribution.

²¹ Pour la constitution des strates, voir annexe 2.

²² Pour le détail des zones de calcul, voir l'annexe 2, page 47.

3.2 Le modèle servant à l'estimation

Un modèle économétrique est estimé pour chaque zone (strate), donc sur un marché supposé homogène, à partir des transactions du parc d'estimation.

Le modèle est de la forme²³ :

$$\log p_i = \log p_0 + \sum_{a=1}^2 \mu_a Y_{a,i} + \sum_{m=1}^{12} \theta_m M_{m,i} + \sum_{k=1}^K \beta_k X_{k,i} + \varepsilon_i \quad (3.1)$$

avec les notations suivantes :

p_i : prix au m^2 (pour les appartements) ou prix total (pour les maisons) du bien i

p_0 : prix du logement de référence du parc d'estimation

$Y_{a,i}$: indicatrice de l'année de mutation du bien i

$M_{m,i}$: indicatrice du mois de mutation du bien i

$X_{k,i}$: caractéristiques physiques du bien i , indexées par $k=1$ à K

Les $X_{k,i}$ sont des variables calculées à partir des modalités des caractéristiques initiales disponibles, Z_i , soit l'époque de construction, le nombre de pièces, le nombre de salles de bains, le nombre de garages ou parkings, l'état du bien ; et de plus, pour les appartements seulement, l'étage, la présence d'un balcon, d'une cave, d'un ascenseur, la surface moyenne par pièce ; et pour les maisons seulement, la présence d'un sous-sol, le nombre de niveaux, la surface du terrain et la surface habitable...²⁴

Le cas échéant, la localisation du logement à l'intérieur de la strate (quartier), est considérée comme une de ses caractéristiques, au même titre que le nombre de pièces ou sa surface.

Les $X_{k,i}$ peuvent inclure des effets croisés. Elles peuvent être des variables indicatrices, ou des variables continues dans le cas des surfaces. Chacune des caractéristiques a un prix, évalué à la période de base, par rapport à celles d'un bien de référence.

Définition 3 : le logement de référence du parc d'estimation

Le *logement de référence* est celui qui possède les caractéristiques prises comme références dans les régressions (tableau 3.1).

Les caractéristiques non observables ou inobservées des logements sont supposées non corrélées avec ces caractéristiques de qualité observées. Le marché des maisons et celui des appartements sont supposés distincts : par exemple, on suppose qu'une pièce supplémentaire peut ne pas valoir le même prix en appartement et en maison. Mais elle vaut le même prix dans un appartement, quel que soit l'étage par exemple. Dans les strates, les prix relatifs des caractéristiques sont fixes. En d'autres termes, si un appartement au troisième étage vaut x% de plus qu'au rez-de-chaussée, ce rapport est le même pour tous les appartements de la strate de calcul. En revanche, il est susceptible d'être différent d'une strate à l'autre.

²³ Pour simplifier les notations, on omet l'indice s de la strate.

²⁴ Toutes ces variables ne sont pas présentes dans chacune des deux bases ni toujours exploitées dans les modèles. Pour le détail des modèles mis en œuvre, voir les chapitres 3 et 4.

Tableau 3.1 - Caractéristiques du logement de référence (parc d'estimation)²⁵

Indice	Logement de référence en Île-de-France	Logement de référence en province
Appartements	<ul style="list-style-type: none"> - 3 pièces, - rez-de-chaussée, - surface moyenne par pièce pour les studios comprise entre 20 et 30 m², pour les deux pièces entre 17 et 24 m², pour les trois pièces entre 18 et 22 m² et pour les quatre pièces et plus entre 17 et 21 m², - sans garage, - sans cave, - 1 salle de bains, - construit entre 1948 et 1969, - vendu le 12^e mois de la deuxième année de la période de référence, 	<ul style="list-style-type: none"> - sans terrasse ou balcon**, - en bon état**.
Maisons	<ul style="list-style-type: none"> - 4 pièces, - 2 niveaux, - époque de construction inconnue, - 1 garage, - 1 salle de bains, - 100 m² de surface habitable et 610 m² de surface de terrain, - vendue le 12^e mois de la deuxième année de la période de référence, - 1 bâtiment*, - pas de cave*. 	<ul style="list-style-type: none"> - pas de sous-sol**, - état du bien inconnu**.

* Variable non présente dans la base en province

** Variable non présente dans la base en Île-de-France

Le plus souvent, le modèle fait intervenir une variable indicatrice de quartier à l'intérieur de la strate. Dans ce cas, le logement de référence appartient par construction au quartier omis dans la régression. Le quartier est introduit lorsque la strate concerne une agglomération ou une ville, et que les volumes de transactions sont suffisants (au moins 110 transactions par an). Les quartiers ne correspondent pas nécessairement à des découpages administratifs.

Le tableau 3.2 présente le nombre de strates de la version 3 : 22 en Île-de-France contre 62 en version 2, 271 en province contre 234 en version 2. Le nombre de quartiers pris en compte a été réduit en Île-de-France (de 230 à 102) et augmenté en province (de 1125 à 1509).

Tableau 3.2 - Nombre de strates élémentaires de calcul et nombre de quartiers

Indice	Nombre de strates	Nombre de quartiers
<i>Île-de-France</i>	22	102
Appartements	15	66
Maisons	7	36
<i>Province</i>	271	1 509
Appartements	97	516
Maisons	174	993
Total	293	1 611

3.3 Prix courant du logement de référence

Le même type de modèle peut être utilisé en période courante. Le bien de référence possède les caractéristiques décrites au paragraphe précédent, mais correspond à la période de mutation courante t . Son prix est noté $p_{0,t}$.

L'évolution du prix du bien de référence est le fondement de la construction de l'indice. Pour cette raison, elle doit incorporer les effets saisonniers, les tendances et les cycles. Ceci explique pourquoi les coefficients saisonniers et annuels ne figurent plus dans le modèle courant.

²⁵ Les variables sont décrites au chapitre 4, paragraphe 4.4.2.

Le prix au m² d'un bien j vendu au cours d'une période t peut s'exprimer à l'aide du modèle :

$$\log(p_{j,t}) = \log(p_{0,t}) + \sum_{k=1}^K \beta_{k,t} X_{k,j,t} + \varepsilon_{j,t}$$

On remarquera que le modèle (3.1) utilisé dans la phase d'estimation est compatible avec ce dernier modèle, utilisé pour les valeurs courantes. L'introduction des indicatrices d'années et de mois vient du fait que le parc d'estimation couvre une période de plusieurs mois. De façon équivalente, le prix de référence pour le parc d'estimation et le mois (a,m) serait :

$$\log(p_{0,a,m}) = \log p_0 + \mu_a + \theta_m$$

Il faut maintenant expliquer comment reconstituer le prix du bien de référence à partir des données de transactions courantes.

Plaçons-nous d'abord dans l'hypothèse où les coefficients $\beta_{k,t}$ sont connus et introduisons $\tilde{p}_{j,t}$ le « prix équivalent bien de référence » du bien j, t :

$$\log(\tilde{p}_{j,t}) = \log(p_{j,t}) - \sum_{k=1}^K \beta_{k,t} X_{k,j,t}$$

Le modèle peut être réécrit :

$$\log(\tilde{p}_{j,t}) = \log(p_{0,t}) + \varepsilon_{j,t}$$

Donc, si on connaît les coefficients $\beta_{k,t}$ le logarithme du prix du bien de référence $\log(p_{0,t})$ sera estimé par les moindres carrés ordinaires, c'est-à-dire par la moyenne :

$$\log(\hat{p}_{0,t}) = \frac{1}{J_t} \sum_{j=1}^{J_t} \log(\tilde{p}_{j,t})$$

où J_t est le nombre de transactions de la période t .

Si on suppose que le modèle estimé sur la période de référence est stable dans le temps (pour la période du calcul d'un indice, qui est maintenant de deux ans), on peut remplacer dans le calcul des prix équivalents bien de référence, les coefficients $\beta_{k,t}$ par les valeurs $\hat{\beta}_k$ estimées sur la période de référence :

$$\log(\tilde{p}_{j,t}) \cong \log(p_{j,t}) - \sum_{k=1}^K \hat{\beta}_k X_{k,j,t} = \log \left[\frac{p_{j,t}}{\exp \left(\sum_{k=1}^K \hat{\beta}_k X_{k,j,t} \right)} \right]$$

On estime alors le prix au m² du bien de référence à la période t , $p_{0,t}$, par une moyenne géométrique des prix équivalents bien de référence des J_t biens vendus au cours de la période t :

$$\log \hat{p}_{0,t} = \frac{1}{J_t} \sum_{j=1}^{J_t} \log \tilde{p}_{j,t} = \frac{1}{J_t} \log \left(\prod_{j=1}^{J_t} \tilde{p}_{j,t} \right)$$

soit :

$$\hat{p}_{0,t} = \left(\prod_{j=1}^{J_t} \tilde{p}_{j,t} \right)^{\frac{1}{J_t}}$$

Dans la suite nous notons $\hat{\alpha}_{0,t} = \log \hat{p}_{0,t}$.

3.4 Valeur courante du parc de référence

Connaissant une valeur approchée du bien de référence (donnée par l'estimation du prix équivalent du logement de référence à la période t), on peut alors reconstituer les valeurs approchées des biens figurant dans le parc de référence et, par agrégation, la valeur du parc lui-même. Les calculs sont menés strate par strate. Pour cette raison, nous réintroduisons l'indice s de la strate.

On estime la valeur d'un bien i du parc de référence de la strate s à la période courante t à partir de ses caractéristiques $X_{k,i,s}$, qui, rappelons-le, sont indépendantes de la date t , par définition même du parc de référence dont la composition est stable dans le temps²⁶.

Dans le cas des appartements, par exemple, la valeur approchée est :

$$\hat{p}_{i,s,t} = \exp \left(\hat{\alpha}_{0,s,t} + \sum_{k=1}^K \hat{\beta}_{k,s} X_{k,i,s} \right) A_{i,s}$$

où $A_{i,s}$ désigne la surface du bien i,s .

En sommant la valeur courante estimée des N_s biens du parc de référence de la strate s , on obtient la valeur du parc de référence à la date t :

$$\hat{W}_{s,t} = \sum_{i=1}^{N_s} \hat{p}_{i,s,t} = \sum_{i=1}^{N_s} \exp \left(\hat{\alpha}_{0,s,t} + \sum_{k=1}^K \hat{\beta}_{k,s} X_{k,i,s} \right) A_{i,s}$$

On estime de la même façon, la valeur du parc de référence dans la strate s , à la période 0 , dite *période de base* de l'indice :

$$\hat{W}_{s,0} = \sum_{i=1}^{N_s} \exp \left(\hat{\alpha}_{0,s,0} + \sum_{k=1}^K \hat{\beta}_{k,s} X_{k,i,s} \right) A_{i,s}$$

3.5 La méthode de calcul

Trois changements majeurs sont introduits par rapport à la version précédente des indices.

- Les parcs de référence et d'estimation sont désormais actualisés tous les deux ans ; les vecteurs de prix relatifs des caractéristiques, et donc la valeur des parcs de référence à la période de base, sont réestimés tous les deux ans également. Le Conseil scientifique des indices Notaires-Insee²⁷ en a décidé ainsi après avoir constaté que, sur la période initialement envisagée comme nouvelle référence (2002-2006), les modèles hédoniques n'étaient pas stables²⁸. Du fait de l'actualisation régulière des coefficients et des pondérations, une méthode de chaînage est introduite. La spécification des modèles et la définition des strates devraient continuer à être revues tous les six ans environ.

- L'introduction d'indicateurs mensuelles dans l'équation (3.1) permet désormais de calculer des indices mensuels. PNS et Min.not calculent ces indices mais seul PNS les publie.

- En version 2, on utilisait des moyennes arithmétiques pour agréger les indices élémentaires. En version 3, on utilise des moyennes géométriques pour les niveaux infra-départementaux et départementaux. Ce mode de calcul est utilisé dans la construction des indices statistiques lorsque l'on pense qu'il peut exister un arbitrage entre différents biens de nature voisine. Dans notre cas, il peut y

²⁶ La qualité des logements ne varie donc pas.

²⁷ La composition du Conseil scientifique des indices Notaires-Insee, chargé de la surveillance des indices, est décrite dans les conventions passées avec le notariat. Cf. annexes 8 et 9.

²⁸ Voir l'annexe 4, Tests de stabilité et durée de la période de référence.

avoir un choix d'acquisition entre des biens appartenant à des strates géographiques voisines. On utilise des moyennes arithmétiques pour les niveaux plus agrégés que les niveaux départementaux car l'acheteur a moins de chance d'arbitrer entre des strates géographiques plus éloignées pour s'installer (région, province, France entière).

3.5.1 Renouvellement des parcs de référence et d'estimation tous les deux ans

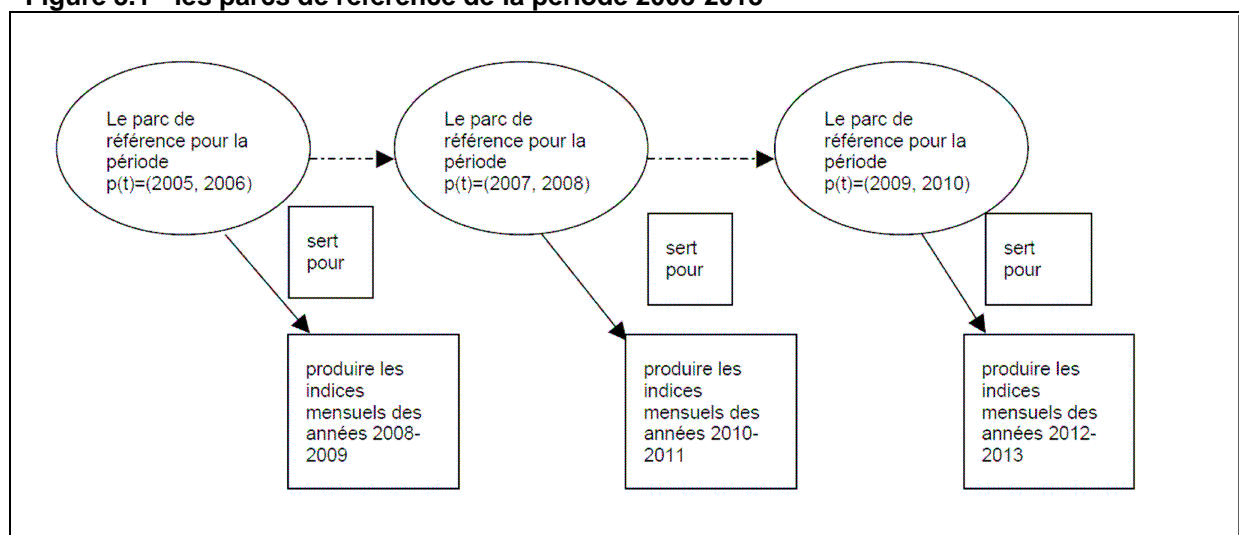
Les parcs de référence

Dans la version 1 des indices, le parc de référence incluait les transactions de trois à cinq années selon les indices (1994-96 pour les appartements de province, 1994-97 pour les maisons, et 1992-96 pour les appartements de Paris et Petite Couronne). Dans la version 2, le parc contenait les transactions des quatre années 1998-2001.

Dans la version 3, les parcs de référence sont remis à jour tous les deux ans. La période de référence $p(t)$ est également fixée à deux années de transactions, donc plus courte qu'auparavant. Elle est définie par rapport au début de la période courante de production des indices de la façon suivante : pour la production courante des indices sur le couple d'année $(n, n+1)$ avec n pair, $p(t)$ débute deux ans avant, soit $p(t) = (n-3, n-2)^{29}$.

La figure 3.1 permet de visualiser la succession des différents parcs de référence utilisés pour le calcul des indices mensuels publiés sur les années 2008-2013.

Figure 3.1 - les parcs de référence de la période 2008-2013



Période de base et base de publication

Les parcs de référence sont valorisés au dernier trimestre $q(t)$ de la période de référence, donc au quatrième trimestre de l'année paire $n-2$ pour les indices diffusés sur la période $(n, n+1)^{30}$. Les poids servant à l'agrégation des indices élémentaires sont calculés à partir de cette valorisation du parc.

Le trimestre de valorisation du parc de référence est appelé *période de base* des indices sur la période $(n, n+1)$. Il ne doit pas être confondu avec la *base 100* de la série publiée, qui a été fixée au premier trimestre 2010.

Notations

La formule de calcul de la valeur du parc de référence doit être généralisée pour tenir compte du renouvellement du parc de référence tous les deux ans par l'introduction d'un indicage supplémentaire, - noté $p(t)$, le parc de référence utilisé - : la valeur estimée du parc devient $\widehat{W}_{s,t}^{p(t)}$. De même, on introduit la valeur du parc de

²⁹ Par exemple, la première période courante de production des indices est le couple d'années 2008 et 2009 et $p(t) = (2005, 2006)$. Au moment de l'estimation, $n-1$ n'est pas disponible.

³⁰ Le quatrième trimestre 2006 sera donc par exemple le trimestre de valorisation du parc de référence 2005-2006.

référence à la période de base $q(t)$, servant au chaînage des indices sur la période de référence $p(t)$. Selon les étapes, elle sera notée $\hat{W}_{s,q(t),t}^{p(t)}$ ou plus simplement, si le contexte le permet : $\hat{W}_{s,t}$ ou $\hat{W}_{s,q(t),t}$ selon les cas.

En période courante, la valeur du parc de référence devient :

$$\hat{W}_{s,t} = \sum_{i=1}^{N_s} \exp\left(\hat{\alpha}_{0,s,t} + \sum_{k=1}^K \hat{\beta}_{k,s} X_{k,i,s}\right) A_{i,s}$$

avec les mêmes notations qu'au paragraphe 3.4 :

$\hat{W}_{s,t}$: la valeur courante du parc de référence dans la strate s ,

$\hat{\alpha}_{0,s,t}$: l'estimation du logarithme du prix au m² du bien de référence de la strate s à la date t ; les coefficients

sont estimés sur la période $p(t)$,

N_s : le nombre de transactions de la strate s au cours de la période de référence,

$\sum_{k=1}^K \hat{\beta}_{k,s} X_{k,i,s}$: la correction à appliquer à l'estimation du logarithme du prix du bien de référence de la strate s à

la date t , pour estimer à cette même date le logarithme du prix du bien i de la strate s échangé au cours de la période de référence,

$A_{i,s}$: la surface du bien i de la strate s échangé au cours de la période de référence.

3.5.2 Calcul des indices élémentaires en période courante

Calcul des évolutions de prix

Les prix « équivalents bien de référence » sont calculés à l'aide du vecteur de prix des caractéristiques, comme indiqué au paragraphe 3.3. A l'aide de ces prix estimés, on valorise le parc de référence au trimestre t .

L'introduction d'indicatrices mensuelles dans les modèles (3.1) à la place des indicatrices trimestrielles autorise désormais la publication d'indices mensuels. Cependant on ne dispose pas de suffisamment d'observations pour calculer des indices mensuels « purs ». Les transactions mises à contribution dans ce calcul restent relatives à un trimestre. L'indice mensuel est calculé avec les données du dernier trimestre disponible incluant ce mois : il s'agit d'un indice mensuel en trimestre « glissant ».

Chaînage

L'évolution des prix entre les trimestres $t-1$ et t est obtenue en rapportant la valeur du parc de référence du trimestre t à la valeur de parc de référence du trimestre $t-1$ observé trois mois plus tôt ; t inclut les mois $m-2$, $m-1$ et m et $t-1$ inclut les mois $m-5$, $m-4$ et $m-3$. La formule de calcul de l'évolution des prix entre les trimestres $t-1$ et t est donc la suivante :

$$I_{t/0}(s) = \frac{\hat{W}_{s,t}^{p(t)}}{\hat{W}_{s,t-1}^{p(t)}} \times I_{t-1/0}(s)$$

avec t , le trimestre glissant et 0 le trimestre de référence.

L'évolution des prix entre les trimestres 0 et t est obtenue en chaînant les évolutions des valeurs du parc de référence entre 0 et t , selon la formule :

$$I_{t/0}(s) = \left[\prod_{u=1}^t \frac{\hat{W}_{s,u}^{p(u)}}{\hat{W}_{s,u-1}^{p(u)}} \right]$$

L'introduction de ce chaînage est rendue nécessaire par l'actualisation du parc de référence tous les deux ans, qui constitue une nouveauté par rapport à la précédente version des indices. En version 3, les indices antérieurs à 2008 ont été calculés en appliquant les évolutions des indices de la version 2.

Mise en œuvre du chaînage

Les indices mensuels étant calculés en trimestre glissant, il faut une base de début de chaînage pour les trois séries d'indices élémentaires correspondant aux premier, deuxième et troisième mois des trimestres civils. La série d'indices commence en janvier 2008. La base de début de chaînage est fixée à octobre, novembre et décembre 2007.

On calcule donc trois valeurs de départ pour permettre le calcul des trois premiers indices (ceux de janvier, février et mars 2008). On estime la valeur du parc de référence $p(t)=(2007,2008)$ au mois d'octobre 2007 (trimestre glissant composé des ventes d'août, septembre et octobre 2007), au mois de novembre et au mois de décembre 2007. L'indice d'octobre 2007 est ensuite fixé à 100. Celui de novembre est obtenu en rapportant la valeur en novembre du parc de référence $p(t)=(2007,2008)$ à sa valeur en octobre. Celui de décembre est obtenu en rapportant la valeur en décembre du parc de référence $p(t)=(2007,2008)$ à sa valeur en octobre (tableau 3.3).

A titre d'exemple, le calcul des indices de janvier 2008, décembre 2009 et janvier 2010 d'une strate s sera effectué de la façon suivante :

$$I_{janvier2008/0}(s) = \frac{\hat{W}_{s,janvier2008}^{parc2005-2006}}{\hat{W}_{s,octobre2007}^{parc2005-2006}} * I_{octobre2007/0}(s)$$

$$I_{décembre2009/0}(s) = \frac{\hat{W}_{s,décembre2009}^{parc2005-2006}}{\hat{W}_{s,septembre2009}^{parc2005-2006}} * I_{septembre2009/0}(s)$$

$$I_{janvier2010/0}(s) = \frac{\hat{W}_{s,janvier2010}^{parc2007-2008}}{\hat{W}_{s,octobre2009}^{parc2007-2008}} * I_{octobre2009/0}(s)$$

Tableau 3.3 - Calcul des indices de prix pour les trois derniers mois de l'année 2007

Mois du 4 ^{ème} trimestre 2007	Valeur estimée du parc de référence 2007 - 2008 * au mois de publication	Formule de calcul	Valeur de l'indice
1 ^{er} mois : octobre	500	100,0	100,0
2 ^{ème} mois : novembre	510	510/500 X 100	102,0
3 ^{ème} mois : décembre	513	513/500 X 100	102,6

* en millions d'euros

Ensuite, on peut commencer le chaînage des indices élémentaires pour les mois de publication du premier trimestre 2008 en utilisant la formule suivante :

$$I_{t/0}(s) = \frac{\hat{W}_{s,t}}{\hat{W}_{s,t-1}} \times I_{t-1/0}(s)$$

avec t le trimestre glissant et $p(t)=(2005,2006)$ comme défini en 3.5.1.

Les indices élémentaires ainsi obtenus peuvent être maintenant agrégés pour obtenir les indices de niveaux supérieurs, tels que département, villes, régions... Pour la publication, la base 100 des indices a été fixée au premier trimestre 2010.

3.5.3 Calcul des indices agrégés en période courante

Pour corriger la non-exhaustivité de la base des notaires, on introduit un coefficient de redressement δ . Ce coefficient est estimé pour chaque département et chaque année de mutation, à partir de données fiscales (montants des droits de mutation établis par la Direction générale des Finances publiques, DGFIP). Il est obtenu en divisant le montant des transactions dans le département d pour l'année a , estimé d'après les données fiscales, par ce même montant enregistré dans la base notariale.

δ ne distingue pas les appartements et les maisons. Il est appliqué à tous les biens du croisement (département x année de mutation)³¹. Il modifie donc le terme qui pondère les indices élémentaires des strates, c'est-à-dire leur poids en valeur dans le parc de référence.

Premier niveau d'agrégation : calcul des indices infra-départementaux et départementaux

Pour les indices infra-départementaux et départementaux, la formule utilisée est une moyenne géométrique :

$$\frac{I_{t/0}}{I_{t-1/0}} = \prod_{s=1}^n \left(\frac{I_{t/0}(s)}{I_{t-1/0}(s)} \right)^{\left(\frac{\delta_s \hat{W}_{s,q(t)}}{\sum_{u=1}^n \delta_u \hat{W}_{u,q(t)}} \right)}$$

avec :

t , le trimestre

s, u , la strate

$I_{t/0}$, l'indice de l'agrégat en t par rapport au trimestre de référence 0,

$I_{t/0}(s)$, l'indice de la strate s en t par rapport au trimestre de référence 0,

$q(t)$, le dernier trimestre de l'année paire $n-2$ (période de base),

$\hat{W}_{s,q(t)}$, la valeur des biens de la strate s échangés au cours de la période $p(t)$, estimée au trimestre $q(t)$,

δ_s , le coefficient de correction pour non exhaustivité de la base des notaires.

Deuxième niveau d'agrégation : calcul des indices supra-départementaux

Pour les indices d'un agrégat A supra-départemental, on utilise la moyenne arithmétique :

$$\frac{I_{t/0}(A)}{I_{t_0/0}(A)} = \sum_{d \in A} \left(\frac{\delta_d \hat{W}_{d,q(t)}}{\sum_{e \in A} \delta_e \hat{W}_{e,q(t)}} \right) \left(\frac{I_{t/0}(d)}{I_{t_0/0}(d)} \right)$$

avec, outre les notations précédentes :

t_0 , le dernier trimestre de l'année $n-1$,

d, e , département ou partie de département

$I_{t/0}(A)$, l'indice de l'agrégat A au trimestre t par rapport au trimestre de référence 0,

³¹ Ce coefficient est l'inverse du taux de couverture des bases notariales (présenté au chapitre 4, paragraphe 4.3). Toutefois, lorsque le taux de couverture est inférieur à 20%, le coefficient utilisé est plafonné et n'est donc jamais supérieur à 5. De même que dans la version 2 des indices, on utilise un rapport de montants et non un rapport de volumes (nombre de transactions), comme c'était le cas dans la version 1 des indices.

$I_{t/0}(d)$, l'indice du département d au trimestre t par rapport au trimestre de référence 0,

$\hat{W}_{d,q(t)}$, la valeur des biens du département échangés au cours de la période $p(t)$ estimée au trimestre $q(t)$.

3.6 Un exemple numérique pour novembre 2012

Voici un exemple de calcul d'un indice élémentaire de prix pour novembre 2012. Cet exemple concerne les appartements, dans la strate 2 d'Île-de-France. Les coefficients utilisés sur la période 2012-2013 sont estimés sur les transactions de 2009-2010.

Étape 1 : extraction de la base

L'indice de novembre 2012 est calculé à partir de l'ensemble des mutations de septembre, octobre et novembre 2012. 222 mutations, correspondant au champ de l'indice et comportant les informations nécessaires au calcul des indices, ont été enregistrées dans la base pour la strate et le trimestre considérés. Les prix au m² s'échelonnent de 1 731 €/m² à 5 225 €/m².

Étape 2 : calcul du prix équivalent bien de référence

La première des 222 mutations porte sur un trois-pièces de 64 m², avec une salle de bain, deux garages et une cave, situé au rez-de-chaussée d'un immeuble d'époque de construction inconnu ; le logement est situé dans le quartier n°2. Le prix de la vente a été de 190 000 €, soit 2 969 €/m².

Il faut enlever les coefficients des caractéristiques spécifiques au bien pour se ramener à son prix « équivalent bien de référence »³². On a donc :

$$\begin{aligned}\text{Log}(\tilde{p}_{j,\tau}) &= \text{Log}(2\,969), \text{ soit } 7,9959 \\ &\quad +0,0598 \text{ (quartier 2} \rightarrow \text{quartier de référence)} \\ &\quad -0,0227 \text{ (1 cave et plus} \rightarrow \text{sans cave)} \\ &\quad -0,0670 \text{ (2 garages et plus} \rightarrow \text{sans garage)} \\ &\quad -0,1573 \text{ (époque de construction inconnue} \rightarrow \text{1948-1969)} \\ &= 7,8088\end{aligned}$$

Le prix « équivalent bien de référence » $\tilde{p}_{j,\tau}$ de cette première mutation s'obtient comme $e^{7,8088}$, soit 2 462 €/m². On réitère l'opération pour les 221 autres mutations.

Dans chaque quartier, les mutations dont le prix équivalent bien de référence est inférieur au 2^e centile ou supérieur au 98^e centile de la distribution sont retirées. Dans notre exemple, le quartier ne compte que 36 transactions pour le trimestre. Dans ce cas, on retire le bien le plus cher et le bien le moins cher. Sur l'ensemble des cinq quartiers de la strate 2, on supprime ainsi 14 références dont les prix équivalents au bien de référence sont extrêmes (6 dans le quartier 1 et 2 dans chacun des 4 autres).

Le prix du bien de référence, égal à l'exponentielle de la moyenne arithmétique des logarithmes des 208 prix équivalents bien de référence des mutations conservées, est $\text{Log}(7,8314)$, soit 2 518 €/m².

Étape 3 : calcul de l'indice mensuel d'une strate en novembre 2012

Pour déterminer l'indice mensuel provisoire³³ d'une strate en novembre 2012, on calcule le prix au m² définitif de la strate en août 2012 (composé des transactions de juin, juillet et août 2012), comme moyenne géométrique des prix au m² obtenus à l'étape 2 pour chaque bien de la strate, soit 2 498 €/m². Ce prix est comparé au prix provisoire trois mois plus tard (2 518 €/m²) pour déterminer l'évolution trimestrielle de la strate. C'est l'évolution de ce prix au m² par rapport au prix au m² définitif estimé dans la strate en novembre 2011 (période

³² Cf. tableau 3.1 en page 19.

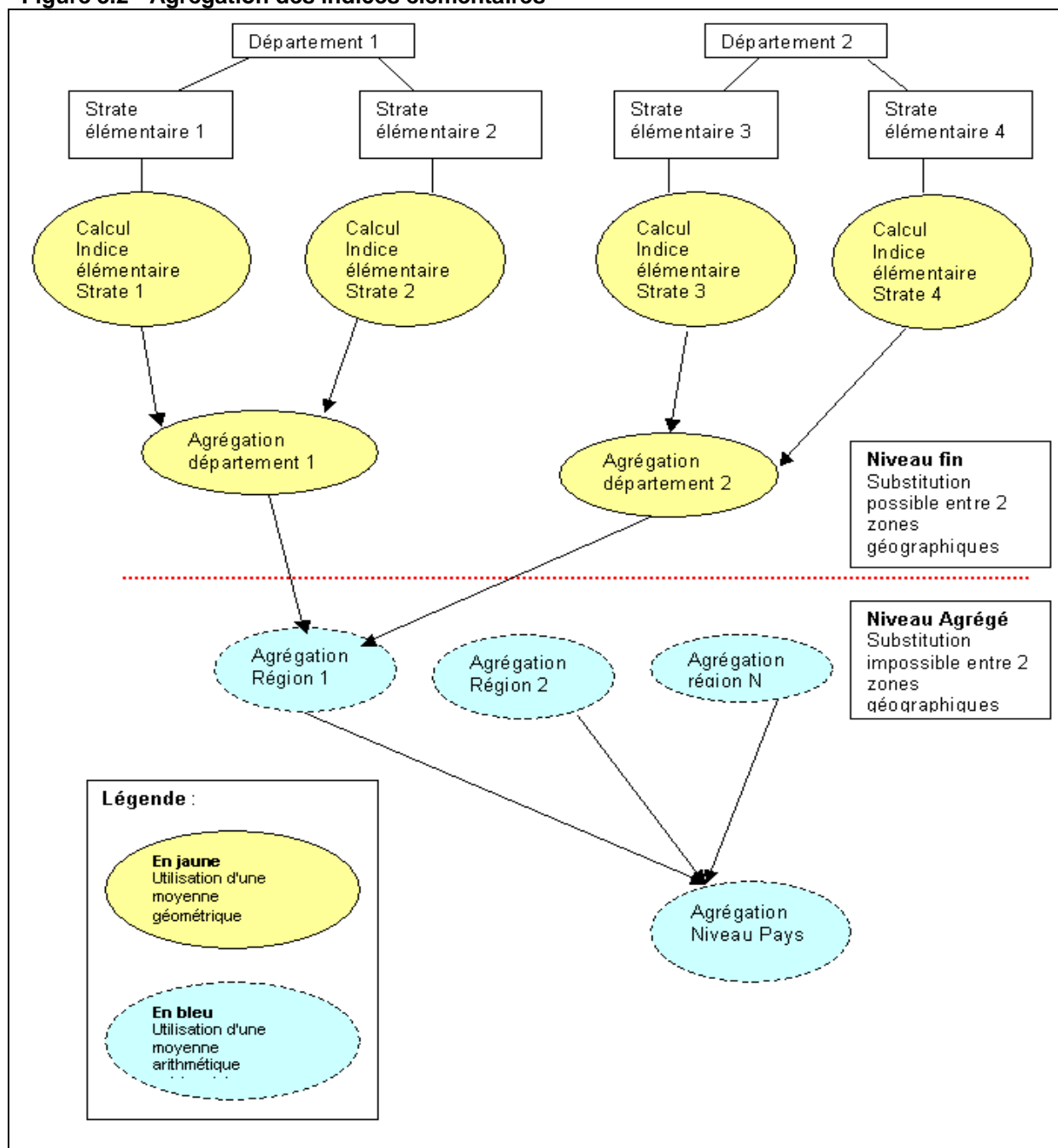
³³ L'indice définitif est publié trois mois plus tard.

de référence des évolutions en 2012 et 2013) qui sera agrégée au niveau départemental et régional pour déterminer les évolutions des prix par département et pour la région.

Des indices élémentaires à l'indice national : agrégation des indices élémentaires

La figure 3.2 ci-dessous montre l'enchaînement des étapes pour passer de l'indice élémentaire à l'indice national. Les départements 1 et 2 appartiennent à la région 1. La moyenne géométrique est appliquée pour les niveaux infra-départementaux et départementaux. On utilise la moyenne arithmétique pour les niveaux plus agrégés³⁴.

Figure 3.2 - Agrégation des indices élémentaires



³⁴ Cf. en page 21.

Chapitre 4 : Les bases de données notariales

Les indices Notaires-Insee sont calculés à partir des prix des transactions enregistrées dans les bases de données notariales³⁵.

4.1 Description des bases

Ces bases sont au nombre de deux :

- La base BIEN, gérée par PNS (Paris Notaires Services), couvre l'Île-de-France. Elle a été mise en place à partir de 1989 pour Paris, 1991 pour la petite couronne et 1996 pour la grande couronne ; elle est exploitable pour les indices depuis 1991.
- La base Perval, gérée par l'ADSN (Association pour le développement du service notarial, dépendant du notariat), couvre la province et les DOM. Elle a été mise en place à partir de 1994.

Ces bases enregistrent les transactions de tous les types de biens : appartements et maisons, mais aussi immeubles, locaux d'activités, terrains, garages, vignobles et autres biens agricoles. En août 2013, elles contenaient 14,2 millions d'enregistrements correspondant aux mutations des années 1990 à 2012 (*tableau 4.1*). Chaque année, elles s'enrichissent d'environ 500 000 nouvelles mutations d'appartements et de maisons anciens (25% en Île-de-France et 75% en province) susceptibles d'être utilisées pour le calcul des indices Notaires-Insee. L'utilisation des bases notariales et le calcul des indices sont régis par des conventions entre la Chambre interdépartementale des notaires de Paris ou le Conseil supérieur du notariat et l'Insee³⁶.

Les articles 15 et 16 de la loi n°2011-331 du 28 mars 2011 confient une nouvelle mission de service public au notariat. Cette loi modifie ainsi le dispositif juridique des deux lois qui encadrent l'organisation et le statut du notariat³⁷. Jusqu'en 2011, la mission de service public du notariat était définie par l'article 1 de l'ordonnance n° 45-2590 du 2 novembre 1945 relative au statut du notariat : « Les notaires sont les officiers publics, établis pour recevoir tous les actes et contrats auxquels les parties doivent ou veulent faire donner le caractère d'authenticité attaché aux actes de l'autorité publique, et pour en assurer la date, en conserver le dépôt, en délivrer des grosses et expéditions ». Dorénavant, la transmission des informations relatives aux mutations d'immeubles à titre onéreux au Conseil supérieur du notariat devient pour les notaires une mission de service public. En outre, cette même loi, reconnaît aussi comme mission de service public, l'activité de centralisation et diffusion des données collectées par le Conseil supérieur du notariat.

La transmission des données en provenance des études notariales s'effectuait jusqu'en 2009 sur support papier. En 2010 s'est engagé le basculement vers la voie électronique, qui, à terme, accroîtra la flexibilité du dispositif. En 2013, environ la moitié des enregistrements reçus par les bases le sont par voie électronique. La mise en œuvre de la transmission électronique a permis de gagner sur les délais d'envoi des actes. Elle présente en revanche un inconvénient majeur pour le moment : la codification, réalisée par les études et non plus par les administrateurs des bases, nécessite un volume de corrections plus important que les actes papiers.

4.2 Délai d'intégration des transactions

Le délai entre la signature d'un acte et l'intégration de la transaction dans les bases notariales conditionne évidemment la réactivité des indices. Le délai visé en 2013 en Île-de-France est de 30 jours pour la saisie des données et leur entrée dans la base, hors réception des actes. En 2012, le délai total moyen était de 82 jours. En province, en 2010, le délai de traitement a augmenté par rapport à 2009 (127 contre 112), la forte remontée du nombre de transactions ayant été plus rapide que l'augmentation des moyens de traitement. Mais le retard pris dans le traitement des actes a été résorbé fin 2011 (89 jours au 2^{ème} semestre 2011). En 2012, en moyenne, il fallait 76 jours (54 pour la réception des actes, 22 pour la saisie des données et leur intégration en base).

Les notaires se sont fixé par ailleurs des seuils minimum de stocks d'actes non codifiés, pour assurer aux codificateurs une quantité de travail dans la continuité. Le stock minimum est de l'ordre de 30 000 actes en province et devrait descendre à 20 000 d'ici la fin de l'année 2013 ; en Île-de-France, le volume des stocks non codifiés doit se situer en régime de croisière entre 3 000 et 5 000 actes papiers et entre 2 000 et 5 000 actes électroniques.

³⁵ Pour le détail de la transmission des données issues des actes notariés, cf. annexe 1.

³⁶ Les conventions sont reproduites en annexes 8 et 9.

³⁷ Cf. annexe 7.

Tableau 4.1 - Enregistrements présents dans les bases de données notariales, par année de mutation

France métropolitaine					
Année de la mutation	Nombre d'enregistrements	Dont appartements et maisons	Dont appartements et maisons anciens	Appartements anciens	Maisons anciennes
1990	105 186	82 475	54 227	41 027	13 200
1991	154 485	114 236	90 180	52 737	37 443
1992	279 572	195 304	165 453	85 085	80 368
1993	327 631	227 823	195 929	95 225	100 704
1994	394 505	281 141	236 020	114 531	121 489
1995	402 567	280 963	242 414	109 178	133 236
1996	557 948	412 273	365 524	159 694	205 830
1997	555 070	397 629	345 942	150 894	195 048
1998	645 241	467 347	409 326	183 282	226 044
1999	787 764	571 815	493 771	227 086	266 685
2000	778 810	560 500	510 719	236 243	274 476
2001	780 481	571 173	515 391	242 854	272 537
2002	754 277	556 753	496 082	238 902	257 180
2003	802 454	588 543	515 064	250 533	264 531
2004	823 556	598 580	517 971	254 352	263 619
2005	851 706	623 377	533 046	264 876	268 170
2006	853 088	621 774	528 903	265 762	263 141
2007	855 971	621 867	528 480	264 472	264 008
2008	740 125	521 952	441 499	218 401	223 098
2009	627 281	454 096	370 111	180 816	189 295
2010	778 846	591 965	489 837	237 719	252 118
2011	719 914	534 713	463 821	224 390	239 431
2012	626 746	458 612	406 714	190 773	215 941
Total	14 203 224	10 334 911	8 916 424	4 288 832	4 627 592

Source : Bases BIEN et Perval - août 2013

Tableau 4.2 - Enregistrements présents dans la base Perval, par année de mutation

Province					
Année de mutation	Nombre d'enregistrements	Dont appartements et maisons	Dont appart. et maisons anciens	Appartements anciens	Maisons anciennes
1990	41 476	27 500	23 619	11 341	12 278
1991	97 448	67 126	56 323	24 275	32 048
1992	220 889	147 032	123 075	51 224	71 851
1993	265 316	176 123	150 620	58 891	91 729
1994	314 822	214 935	179 791	69 771	110 020
1995	330 619	222 636	192 180	70 194	121 986
1996	436 646	310 901	276 769	97 984	178 785
1997	433 432	296 838	259 422	91 109	168 313
1998	503 638	349 206	307 319	112 883	194 436
1999	606 950	421 348	364 425	138 231	226 194
2000	605 940	416 751	381 083	146 593	234 490
2001	607 127	426 318	385 757	151 799	233 958
2002	576 602	406 180	361 322	143 985	217 337
2003	615 992	429 736	375 242	151 370	223 872
2004	622 696	428 530	368 987	147 973	221 014
2005	643 823	447 509	380 739	155 639	225 100
2006	653 110	453 308	381 721	160 336	221 385
2007	661 782	458 936	385 306	161 772	223 534
2008	581 206	389 982	326 399	135 229	191 170
2009	491 910	340 026	273 998	112 013	161 985
2010	596 982	434 569	359 958	146 287	213 671
2011	564 361	402 584	351 476	145 198	206 278
2012	493 752	346 968	310 787	122 682	188 105
Total	10 966 519	7 615 042	6 576 318	2 606 779	3 969 539

Source : Base Perval - août 2013

Tableau 4.3 - Enregistrements présents dans BIEN, par année de mutation

Île-de-France					
Année de la mutation	Nombre d'enregistrements	Dont appartements et maisons	Dont appart. et maisons anciens	Appartements anciens	Maisons anciennes
1990	63 710	54 975	30 608	29 686	922
1991	57 037	47 110	33 857	28 462	5 395
1992	58 683	48 272	42 378	33 861	8 517
1993	62 315	51 700	45 309	36 334	8 975
1994	79 683	66 206	56 229	44 760	11 469
1995	71 948	58 327	50 234	38 984	11 250
1996	121 302	101 372	88 755	61 710	27 045
1997	121 638	100 791	86 520	59 785	26 735
1998	141 603	118 141	102 007	70 399	31 608
1999	180 814	150 467	129 346	88 855	40 491
2000	172 870	143 749	129 636	89 650	39 986
2001	173 354	144 855	129 634	91 055	38 579
2002	177 675	150 573	134 760	94 917	39 843
2003	186 462	158 807	139 822	99 163	40 659
2004	200 860	170 050	148 984	106 379	42 605
2005	207 883	175 868	152 307	109 237	43 070
2006	199 978	168 466	147 182	105 426	41 756
2007	194 189	162 931	143 174	102 700	40 474
2008	158 919	131 970	115 100	83 172	31 928
2009	135 371	114 070	96 113	68 803	27 310
2010	181 864	157 396	129 879	91 432	38 447
2011	155 553	132 129	112 345	79 192	33 153
2012	132 994	111 644	95 927	68 091	27 836
Total	3 236 705	2 719 869	2 340 106	1 682 053	658 053

Source : Base BIEN - août 2013

4.3 Couverture

Les bases notariales, alimentées jusqu'à présent sur la base du volontariat, ne sont pas exhaustives ; le taux de couverture, défini ici comme le rapport entre les montants connus dans les bases et les montants réellement échangés, n'est donc pas de 100%.

Connaître le taux de couverture est important, pour deux raisons principales. D'une part, si un certain type de transactions était sous-enregistré et que l'évolution de leur prix présentait un caractère particulier, l'indice serait biaisé³⁸. Cela se produirait si les notaires variaient dans leur comportement d'envoi des informations dans les bases, en fonction des caractéristiques d'évolution des prix des biens. Il importe donc de connaître le taux de couverture par type de bien et zone géographique. D'autre part il est intéressant en soi de connaître le nombre des transactions immobilières pour en suivre la conjoncture. On ne peut le faire avec les transactions enregistrées dans les bases notariales qu'à condition que leur taux de couverture soit constant. Or il fluctue et est appelé à augmenter, en raison notamment de l'inscription dans la loi de l'obligation d'alimenter les bases.

En l'absence de données nationales exhaustives sur le nombre de ventes de logements anciens par type de bien, on est obligé d'utiliser des approches indirectes. On peut penser à trois approches : s'appuyer sur une enquête spécifique, utiliser un décompte d'actes notariaux ou utiliser les données fiscales.

- Enquête spécifique

L'enquête Existan, menée annuellement par le ministère de l'Équipement auprès des services fiscaux au cours des années 90, a constitué un outil précieux. Elle se fondait sur un échantillonnage de documents sur support papier détenus par les services fiscaux. Jusqu'en 2000, elle couvrait l'ensemble du territoire. Elle a été abandonnée lors de l'informatisation des services fiscaux (qui a entraîné la disparition des supports papier), pour

³⁸ Le choix d'un parc de référence dont la structure est proche du stock tel qu'on peut le connaître aux recensements, minimise ce biais potentiel. Deux questions s'entremêlent : le parc de référence doit refléter (sans biais) l'ensemble des transactions ; c'est le problème de la représentativité des bases notariales, dont le taux de couverture est un élément ; il doit aussi être une approximation acceptable du stock de logements, puisqu'on cherche à mesurer l'évolution de la valeur (théorique) du stock de logements (voir chapitre 1) et pas seulement la valeur du flux des biens échangés. Ce deuxième aspect justifie la comparaison (en structure) du parc de référence avec le recensement. Voir les tableaux du paragraphe 4.4.3.

des raisons de coût et dans l'attente de l'exploitation de la base patrimoniale des services fiscaux. Cette source distinguait les différents types de biens (appartements, maisons, immeubles entiers) et les caractéristiques des acheteurs et des vendeurs. Elle permettait de comparer, par département, le nombre de transactions figurant dans les bases notariales avec ceux de l'enquête, donc d'estimer des taux de couverture.

- Décompte des actes notariaux

Pour l'Île-de-France, la base BIEN est couplée avec une enquête sur l'activité des études notariales qui décompte, mensuellement et par département, le nombre d'actes notariaux, donc de transactions. On estime le taux de couverture de la base comme le rapport entre le nombre d'actes enregistrés dans la base et le nombre total d'actes de vente signés. Ainsi, en 2010, le taux de couverture global était de 82% en Île-de-France (86% pour Paris, 82% pour la petite couronne et 79% en grande couronne). Cependant le décompte des actes ne distingue pas les transactions dans le neuf et dans l'ancien, ni les logements des autres types de biens ; de plus la localisation prise en compte est celle de l'étude, non celle du bien. Pour la province, on ne dispose pas de décompte du nombre d'actes ; il faut donc estimer différemment le taux de couverture.

- Données fiscales

Depuis l'abandon d'Existan, c'est la source fiscale qui sert pour estimer les taux de couverture³⁹. L'information de base est le montant des droits de mutation perçus, par département et par taux des droits départementaux (actuellement 3,80% ou 0,60%). L'assiette des droits à 0,60% inclut des mutations à titre gratuit (donations) et n'est donc pas utilisable. En revanche, l'assiette des droits à 3,80% (division du produit des droits par 3,80%) peut être comparée à l'assiette des mêmes droits figurant dans les bases notariales. Leur quotient⁴⁰ est égal au taux de couverture (en montant) des bases notariales pour l'ensemble des biens taxés à 3,80% : logements anciens, immobilier d'entreprise ancien et terrains non soumis à TVA, sous réserve de quelques exceptions. Sous certaines approximations, on en déduit un taux de couverture pour les seuls logements anciens⁴¹.

Ainsi estimé, le taux de couverture des bases notariales pour les logements anciens en 2010 était de 63% pour l'ensemble de la France (Corse et départements d'outre-mer inclus), 80% pour l'Île-de-France et 56% pour la province. Il est très variable d'un département à l'autre : en 2010, hors Corse et départements et territoires d'outre-mer, il était inférieur à 30% dans 7 départements, compris entre 30 et 50% dans 34 départements, compris entre 50% et 70% dans 40 départements et supérieur à 70% dans 13 départements (*figure 4.1*).

Que le taux de couverture soit faible ne présente pas en soi de problème statistique pourvu que le nombre d'enregistrements soit supérieur à un certain seuil ; plus ennuyeux est le fait qu'on ne sache pas si les actes enregistrés dans la base sont représentatifs de l'ensemble des transactions ou s'ils présentent des biais, ce qui se produirait par exemple si certaines études notariales n'envoyaient jamais d'actes sur des marchés particuliers ou si certains notaires n'envoyaient pas les actes d'un certain type de clients⁴². Les comparaisons partielles effectuées jusqu'à présent ne donnent aucune raison de penser qu'il puisse y avoir biais significatif, sauf dans les quelques départements où le taux de couverture est le plus faible.

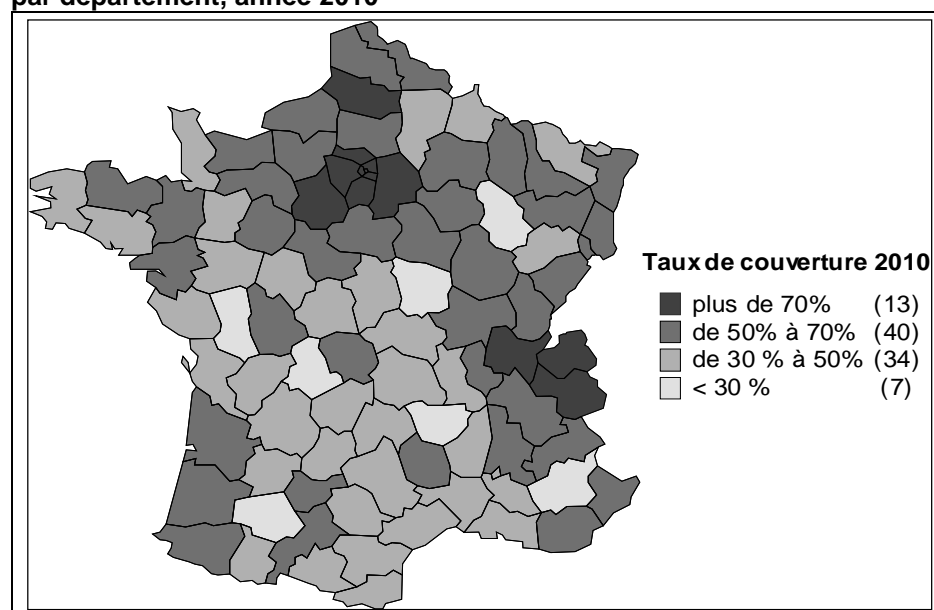
³⁹ Les résultats obtenus avec les deux sources ont été comparés sur l'année 1997. Le nombre et les montants de transactions sont identiques à 1% près sur l'ensemble de la France ; des écarts plus importants sont constatés sur certains départements.

⁴⁰ La signature de l'acte est antérieure de 1,7 mois en moyenne à l'enregistrement des droits par les conservations des hypothèques. On compare le montant des actes présents dans les bases notariales pour le mois n au montant des droits enregistrés par la direction générale des finances publiques (DGFIP) pour le mois $n+2$.

⁴¹ Les calculs et les résultats sont détaillés en annexe 1.

⁴² En province, selon les administrateurs des bases, les remontées seraient plutôt de bonne qualité pour les maisons et les appartements. Il n'en est pas toujours de même pour les autres types de bien et certains offices n'alimentent pas ou peu les actes concernant les biens ruraux (terrains agricoles et viticoles), mais ces biens n'entrent pas dans la constitution des indices. En Île-de-France, il se pourrait, toujours selon les administrateurs des bases, que les biens exceptionnels (très chers par exemple) soient moins transmis que les biens plus communs (du fait de l'acheteur ou même du notaire).

Figure 4.1 - Taux de couverture estimé des bases de données notariales, logements anciens, par département, année 2010



Lecture : Les taux de couverture sont calculés en montants de transactions.

Source : Bases Perval et BIEN

4.4 Champ des indices, qualité des données

4.4.1 Délimitation de champ

Le champ des indices Notaires-Insee est limité aux logements anciens au sens fiscal (appartements ou maisons soumis aux droits d'enregistrements et non à la TVA). Les logements neufs sont donc exclus. On exclut aussi les logements jugés atypiques tels que chambres, greniers, lofts, ateliers, loges de gardien, châteaux, grandes propriétés, hôtels particuliers.

Le logement doit être libre d'occupation au moment de la vente, destiné à un usage strict d'habitation et acquis en pleine propriété par un particulier ou une société civile immobilière⁴³. On enlève donc les logements lorsque la période d'occupation par un tiers ou par le vendeur excède six mois, en considérant que, compte tenu de la réglementation en matière de baux locatifs, ces logements subissent en général une décote. Enfin, on ne retient que les ventes réalisées de gré à gré, donc hors adjudications.

4.4.2 Transactions retenues pour le calcul

De plus, les transactions retenues doivent respecter les caractéristiques suivantes :

- pour les appartements :
 - le nombre de pièces est inférieur à 9,
 - la surface habitable est comprise entre 10 m² et 200 m²,
 - le prix de vente est compris entre 1 500 et 5 000 000 €,
 - le prix au m² est inférieur à 25 000 €.
- pour les maisons :
 - le nombre de pièces est inférieur à 13,
 - la surface habitable est comprise entre 20 m² et 300 m²,

⁴³ Le nombre de logements vendus selon la nature de l'acquéreur est détaillé en annexe 10, tableaux A10.1 et A10.2.

- la surface du terrain (y compris surface au sol de la maison) est supérieure à 9 m²,
- le prix de vente est supérieur à 1 500 €,
- (en province seulement) le prix de vente est inférieur à 15 000 000 €⁴⁴.

Tableau 4.4 - Nombre d'observations exclues à chaque étape (parc de référence 2009-2010)

	Île-de-France		Province	
	Appartements	Maisons	Appartements	Maisons
Nombre de transactions en entrée	116 544	40 857	205 791	224 696
Bien occupé (partiellement ou totalement)	10 454 (6,6%)		18 827 (4,4%)	
Usage autre que d'habitation	2 413 (1,5%)		10 419 (2,4%)	
Propriété autre que pleine	1 321 (0,8%)		232 (0,1%)	
Type de mutation autre que de gré à gré	1 081 (0,7%)		174 (0,0%)	
Acquéreur hors champ (=autres que particuliers, SCI et inconnus)	2 733 (1,7%)		7 530 (1,7%)	
Bien non entièrement ancien	27 590 (17,5%)		72 494 (16,8%)	
Sous-total	80 672 (69,2%)	35 665 (87,3%)	130 761 (63,5%)	202 419 (90,1%)
Appartement hors champ (=autre que studio, duplex, triplex, appartement standard, type non renseigné)	1 007 (1,2%)		913 (0,7%)	
Appartement sans nombre de pièces ni surface	aucun ⁴⁵		4 921 (3,8%)	
Appartement de 9 pièces et plus	30 (0,0%)		73 (0,1%)	
Appartement de 1 à 9 m ²	23 (0,0%)		145 (0,1%)	
Appartement de 201 m ² ou plus	308 (0,4%)		290 (0,2%)	
Appartement de prix supérieur ou égal à 25 000 €/m ²	9 (0,0%)		32 (0%)	
Appartement de prix inférieur ou égal à 1 500 €	aucun		299 (0,2%)	
Appartement de prix supérieur ou égal à 5 000 000 €	14 (0,0%)		3 (0,0%)	
Maison hors champ (=autre que maison de ville, pavillon, villa, ferme, inconnu)		412 (0,5%)		8182 (4%)
Maison sans nombre de pièces ni surface		aucun		19 749 (9,8%)
Maison de 13 pièces et plus		13 (0,0%)		346 (0,2%)
Maison de 1 à 19 m ²		7 (0,0%)		95 (0%)
Maison de 301 m ² ou plus		72 (0,2%)		665 (0,3%)
Maison avec un terrain de 0 à 9 m ²		8 (0,0%)		1 885 (0,9%)
Maison de prix inférieur ou égal à 1 500 €		aucun		516 (0,3%)
Maison de prix supérieur ou égal à 15 000 000 €				2 (0%)
Nombre de transactions valides pour le calcul des indices	79 264 (68%)	32 057 (78,5%)	178 657 (86,8%)	180 568 (80,3%)

4.4.3 Les variables disponibles

De plus en plus d'actes sont codifiés dans les études et transmis sous forme électronique. Les autres sont encore envoyés sous format papier : photocopies des extraits d'actes ou des DHN (Documents Hypothécaires Normalisés), ou des actes complets, complétés de certaines informations non présentes dans l'élément envoyé.

⁴⁴ En Île-de-France, la sélection est faite en amont, lors de la codification. Tous les actes de vente de maisons dont le prix est supérieur à 3,5 fois le prix médian courant (obtenu en multipliant le prix médian au T4 2000, période de base de la V2 des indices Notaires – Insee, par l'évolution de l'indice) sont vérifiés par un codificateur ; de plus toutes les maisons dont le prix est supérieur à 7 millions d'euros sont vérifiées par un codificateur.

⁴⁵ En Île-de-France, le nombre de pièces est estimé lorsqu'il n'est pas renseigné (cf. tableau 4.8)

Ces actes sont alors codifiés par les administrateurs des bases. Certaines variables sont mal renseignées et ne sont pas utilisées dans les modèles hédoniques. Les variables explicatives utilisées dans les modèles sont décrites dans le tableau 4.5.

Tableau 4.5 - Les variables utilisées dans les modèles pour définir la qualité des logements

Variable	Nombre de modalités	Modalités
- le nombre de pièces (appartements)	5	« 1 pièce » à « 5 pièces et plus »
- le nombre de pièces (maisons)	5	« moins de 3 pièces » à « 7 pièces et plus »
- l'étage et la présence d'un ascenseur (appartements uniquement)	6	rez-de-chaussée, 1 ^{er} étage, 2 ^{ème} étage, 3 ^{ème} étage, 4 ^{ème} étage ou plus sans ascenseur, 4 ^{ème} étage ou plus avec ascenseur
- le nombre de niveaux (maisons uniquement)	3	1, 2, 3 ou plus
- la surface habitable moyenne par pièce par taille des biens (appartements uniquement)		pour les studios : de 20 à 30 m ² , pour les deux pièces : de 17 à 24 m ² , pour les trois pièces : de 18 à 22 m ² , pour les quatre pièces et plus : de 17 à 21 m ²
- la surface habitable (maisons uniquement),		en m ²
- la surface du terrain (maisons uniquement)		en m ²
- le nombre de garages ou parkings	3	0, 1, 2 et plus
- le nombre de salles de bains	4	0, 1, 2 et plus, indéterminé
- l'époque de construction	10	avant 1850, 1850-1913, 1914-1947, 1948-1969, 1970-1980, 1981-1991, 1992-2000, 2001-2010, 2011-2020 ⁴⁶ , non déterminée
- la date de la mutation		
- la présence d'un balcon ou d'une terrasse (appartements de province uniquement)	2	oui, non
- le nombre de bâtiments (maisons d'Île-de-France uniquement)	2	oui, non
- l'existence d'une cave (appartements),	3	oui, non, indéterminé
- l'existence d'une cave (maisons en Île-de-France),	3	0, 1, 2 et plus
- l'existence d'un sous-sol (maisons en province),	2	oui, non
- l'état du bien (province uniquement)	3	état bon, travaux à prévoir, logement à rénover

Enfin, certaines variables, récentes ou mal renseignées, ne sont pas utilisées pour le moment mais pourraient être intéressantes à exploiter :

- le mode de chauffage (présence ou non d'un chauffage central),
- l'étiquette énergie (modalités A à G selon l'échelle de consommation d'énergie primaire),
- l'étiquette climat (modalités A à G selon l'échelle d'émission de quantité de gaz à effet de serre émis),
- le statut de l'acquéreur (primo-acquéreur ou non),
- le type de négociation (avec un notaire, avec une agence, de particulier à particulier),
- le type d'énergie renouvelable,
- les matériaux utilisés pour la construction (béton, pierre de taille, matériaux métalliques, pierre, brique, bois, terre, autres),
- le fait que le bâtiment bénéficie d'un label de construction (BBC, HQE...),
- Première vente du bien depuis son achèvement (oui, non),
- Bien occupé à titre de résidence principale (oui, non), etc.

⁴⁶ Cette modalité a été créée au début de la décennie 2010. Elle sera intégrée dans les modèles hédoniques lorsqu'il y aura assez d'observations.

4.4.4 Traitement des informations manquantes

Pour pouvoir être mobilisée dans les calculs, une mutation doit être accompagnée au minimum de son prix, du lieu de la transaction et de sa date, ainsi que de la nature du bien.

L'une ou moins des variables de surface ou nombre de pièces doit être renseignée.

Dans les autres cas, les variables peuvent être imputées (par estimation ou recodage, selon les règles précisées dans le tableau 4.8), ou bien la valeur manquante est traitée en tant que telle (tableaux 4.7 et 4.8).

Tableau 4.7 - Part des observations faisant l'objet d'une imputation

	Surface habitable	Nombre de pièces	Époque de construction	Nombre de garages ou parkings	Nombre de salles de bains	Étage ou nombre de niveaux	Ascenseur
Collectif							
Paris	9,2	2,5	11,4	0,6	11,6	2,1	44,7
Petite couronne	9,1	1,3	17,6	0,3	7,8	2,9	46,8
Grande couronne	9,4	1,1	18,4	0,2	4,8	4,0	47,3
Province	6,9	*	44,6	26,8	4,9	5,5	62,4
Individuel							
Paris, petite couronne	59,4	2,0	69,8	5,4	6,6	0,7	-
Grande couronne	51,7	1,1	52,3	4,0	3,7	0,6	-
Province	41,9	*	36,9	22,2	8,8	17	-

Champ : ensemble des transactions valides pour le calcul de l'indice au sens du tableau 4.4, y compris les observations non retenues dans les parcs d'estimation et de référence ; années 2007-2008

"*" : pas d'imputation (rejet)

"-" : sans objet

Tableau 4.8 - Méthode de traitement de la non-réponse selon les variables

Type de non réponse	Champ géographique	Modalités attendues	Action	Valeur si recodage/commentaire
Prix	Île-de-France, Province	Valeur numérique non nulle	Rejet	
Nature et usage du bien	Île-de-France, Province	Appartement ou maison	Rejet	
Lieu de la transaction	Île-de-France, Province		Rejet	
Surface terrain (maisons)	Île-de-France, Province	Valeur numérique non nulle	Rejet	
Surface habitable et nombre de pièces	Île-de-France, Province		Rejet	Rejet si les deux variables sont manquantes simultanément
Surface habitable (appartements)	Île-de-France	Valeur numérique non nulle	Imputée	Estimation en fonction du nombre de salles de bains, du nombre de pièces, de l'époque de construction, du nombre de niveaux (ou d'étages) et du type de bien
Surface habitable (appartements)	Province	Valeur numérique non nulle	Imputée	Rejet pour la construction des parcs d'estimation ou de référence ; estimation en campagne courante en fonction du nombre de salles de bains, du nombre de pièces, de l'époque de construction
Surface habitable (maisons)	Île-de-France	Valeur numérique non nulle	Imputée	Estimation en fonction du nombre de salles de bains, du nombre de pièces, de l'époque de construction, du nombre de niveaux (ou d'étage), du type de biens et de la surface du terrain
Surface habitable (maisons)	Province	Valeur numérique non nulle	Imputée	Estimation en fonction du nombre de salles de bains, du nombre de pièces, de l'époque de construction et du nombre de niveaux
Nombre de bâtiments (maisons)	Île-de-France	Valeur numérique non nulle		1 bâtiment

Type de non réponse	Champ géographique	Modalités attendues	Action	Valeur si recodage/commentaire
Nombre de salles de bain	Île-de-France, Province	Valeur numérique non nulle	Recodée	Nombre de salles de bain inconnu
Nombre de pièces	Île-de-France	Valeur numérique non nulle	Imputée	Estimation en fonction de la surface et de la strate de calcul et du type de bien
Nombre de pièces	Province	Valeur numérique non nulle	Rejet	
Époque	Île-de-France, Province	10 tranches	Recodée	1948-1969 pour les appartements ; époque inconnue pour les maisons
Présence d'un ascenseur (appartements)	Île-de-France, Province	Oui/non	Recodée	Oui
Étage (appartements)	Île-de-France, Province	Valeur numérique non nulle	Recodée	Rez-de-chaussée
Nombre de niveaux (maisons)	Île-de-France, Province	Valeur numérique non nulle	Recodée	2 niveaux
Nombre de caves (appartements)	Île-de-France	Valeur numérique	Recodée	0 (pas de cave)
Nombre de caves (appartements)	Province	Valeur numérique	Recodée	Nombre de caves inconnues
Nombre de caves (maisons)	Île-de-France	Valeur numérique	Recodée	0 (pas de cave)
Présence d'un sous-sol (maisons)	Province	Oui/non	Recodée	Non
Nombre de garages (appartements)	Île-de-France, Province	Valeur numérique	Recodée	0 (Pas de garage)
Nombre de garages (maisons)	Île-de-France, Province	Valeur numérique	Recodée	1 garage pour Île-de-France 0 (Pas de garage) pour Province
Présence d'un balcon (appartements)	Province	Oui/non	Recodée	Non
Présence d'une terrasse (appartements)	Province	Oui/non	Recodée	Non
État du bien	Province	3 modalités caractère	Recodée	État du bien inconnu

4.4.5 Structure des parcs d'estimation

On trouvera dans les tableaux qui suivent la structure des parcs d'estimation, selon les principales variables utilisées dans les régressions. A titre de comparaison, les pourcentages issus du recensement de la population de 2008 sont indiqués lorsque c'est possible.

Tableau 4.9 - Structure du parc d'estimation 2007-2008 et comparaison avec le recensement de la population ; Île-de-France (appartements)

	Parc d'estimation 2007-2008		Parc au recensement de 2008 (%)	
	Effectif	%		
Total	146 089			
Nombre de pièces				
1	22 329	15%	17%	
2	43 849	30%	29%	
3	42 627	29%	28%	
4	26 601	18%	17%	
5 ou plus	10 683	7%	9%	
Surface du logement				
moins de 40 m2	42 421	29%	31%	
de 40 à moins de 70 m2	64 673	44%	40%	
de 70 à moins de 100 m2	30 553	21%	22%	
de 100 à moins de 150 m2	7 329	5%	6%	
150 m2 ou plus	1 113	1%	1%	
Époque de construction				tranches au recensement
Non renseigné	22 039	15%		
Avant 1914	23 091	16%		
1914-1947	19 341	13%	38%	avant à 1949
1947-1969	33 055	23%	32%	de 1949 à 1974
1970-1980	25 830	18%	11%	de 1975 à 1981
1981-1991	10 123	7%	5%	de 1982 à 1989
Après 1991	12 610	9%	14%	en 1990 et plus
Salles de bains				
Non renseigné	11 005	8%		
0	2 432	2%		
1	123 703	85%		
2 ou plus	8 949	6%		
Garage, parking				
Non renseigné	463	0%		
0	83 600	57%	52%	non
1	54 726	37%	48%	oui
2 ou plus	7 300	5%		

Tableau 4.10 - Structure du parc d'estimation et comparaison avec le recensement de la population ; Île-de-France (maisons)

	Parc d'estimation 2007-2008		Parc au recensement de 2008 (%)	
	Effectif	%		
Total	55 792			
Nombre de pièces				
1 à 3	9 299	17%	17%	
4	13 732	25%	26%	
5	15 691	28%	28%	
6 ou plus	17 070	31%	29%	
Total				
Surface du logement				
moins de 40 m2	215	0%	2%	
de 40 à moins de 70 m2	7 109	13%	13%	
de 70 à moins de 100 m2	20 842	37%	37%	
de 100 à moins de 150 m2	22 470	40%	35%	
150 m2 ou plus	5 156	9%	13%	
Époque de construction				tranches au recensement
Non renseigné	31 411	56%		
Avant 1914	2 524	5%		
1914-1947	4 818	9%	30%	avant à 1949
1947-1969	4 114	7%	27%	de 1949 à 1974
1970-1980	5 423	10%	13%	de 1975 à 1981
1981-1991	4 364	8%	13%	de 1982 à 1989
Après 1991	3 138	6%	17%	en 1990 et plus
Salles de bains				
Non renseigné	1 756	3%		
0	464	1%		
1	35 468	64%		
2 ou plus	18 104	32%		
Garage, parking				
Non renseigné	2 213	4%		
0	15 425	28%	19%	non
1	33 545	60%	81%	oui
2 ou plus	4 609	8%		

Tableau 4.11 - Structure du parc d'estimation 2007-2008 et comparaison avec le recensement de la population ; province (appartements)

	Parc d'estimation 2007-2008		Parc au recensement de 2008 (%)
	Effectif	%	
Total	222 026		
Nombre de pièces			
1	37 103	17%	13%
2	59 507	27%	26%
3	63 289	29%	32%
4	44 909	20%	20%
5 ou plus	17 218	8%	9%
Surface du logement			
moins de 40 m ²	57 044	26%	23%
de 40 à moins de 70 m ²	92 236	42%	42%
de 70 à moins de 100 m ²	57 146	26%	28%
de 100 à moins de 150 m ²	13 926	6%	6%
150 m ² ou plus	1 674	1%	1%
Époque de construction			
Non renseignée	84 066	38%	
Avant 1914	10 008	5%	
1914-1947	13 600	6%	30%
1948-1969	38 941	18%	32%
1970-1980	34 231	15%	11%
1981-1991	22 384	10%	7%
après 1991	18 796	8%	20%
Garage, parking			
0 ou non renseigné	118 872	54%	51%
1	91 965	41%	49%
2 ou plus	11 189	5%	
Salles de bains			
0 ou non renseigné	9 860	4 %	
1	202 873	91 %	
2 ou plus	9 293	4 %	
Étage			
0 ou non renseigné	50 308	23%	
1	52 787	24%	
2	45 397	20%	
3	32 600	15%	
4 ou plus	40 934	18%	
Ascenseur (étage > 3)			
Non	3 043	7%	
Oui ou non renseigné	37 891	93%	
Cave			
0 ou non renseigné	99 225	45%	
1 ou plus	122 801	55%	
Terrasse ou balcon			
Non ou non renseigné	147 416	66%	
Oui	74 610	34%	
État du bien			
Non renseigné	113 718	51%	
Bon	90 128	41%	
Travaux à prévoir	14 973	7%	
A rénover	3 207	1%	

tranches au recensement
avant à 1949
de 1949 à 1974
de 1975 à 1981
de 1982 à 1989
en 1990 et plus
non
oui

Tableau 4.12 - Structure du parc d'estimation 2007-2008 et comparaison avec le recensement de la population ; province (maisons)

	Parc d'estimation 2007-2008		Parc au recensement de 2008 (%)	
	Effectif	%		
Total	328 847			
Nombre de pièces				
1 à 3	65 115	20%	15%	
4	83 402	25%	28%	
5	87 586	27%	30%	
6 ou plus	92 744	28%	27%	
Total				
Surface du logement				
Moins de 40 m ²	2 703	1%	2%	
De 40 à moins de 70 m ²	34 764	11%	11%	
De 70 à moins de 100 m ²	115 433	35%	40%	
De 100 à moins de 150 m ²	136 532	42%	35%	
150 m ² ou plus	39 415	12%	12%	
Époque de construction				tranches au recensement
Non renseigné	105 280	32%		
Avant 1914	41 975	13%		
1914-1947	45 173	14%	35%	avant à 1949
1947-1969	38 267	12%	21%	de 1949 à 1974
1970-1980	42 099	13%	13%	de 1975 à 1981
1981-1991	28 212	9%	11%	de 1982 à 1989
Après 1991	27 841	8%	20%	en 1990 et plus
Garage, parking				
0 ou non renseigné	124 530	38%	21%	non
1	170 351	52%	79%	oui
2 ou plus	33 966	10%		
Salle de bains				
0 ou non renseigné	21 389	7 %		
1	235 118	71 %		
2 ou plus	72 340	22 %		
Niveaux				
0 ou non renseigné	29 536	9%		
1	79 051	24%		
2	183 452	56%		
3 ou plus	36 808	11%		
Surface de terrain				
Moins de 500 m ²	135 690	41%		
De 500 à 1 000 m ²	89 983	27%		
De 1 000 à 1 500 m ²	39 356	12%		
De 1 500 à 2 500 m ²	30 269	9%		
Plus de 2 500 m ²	33 549	10%		
État du bien				
Non renseigné	144 590	44%		
Bon	122 674	37%		
Travaux à prévoir	42 631	13%		
A rénover	18 952	6%		
Sous-sol				
Non ou non renseigné	251 234	76%		
Oui	77 613	24%		

Chapitre 5 : Estimation de l'équation hédonique, mise en œuvre pratique

5.1 Estimation de l'équation hédonique en deux étapes

Le vecteur des prix relatifs des caractéristiques s'obtient après deux étapes d'estimation, la seconde réduisant l'influence des valeurs extrêmes. L'objectif est de parvenir à une estimation plus robuste des coefficients des prix relatifs.

On trouvera en annexe 3 les nouveautés de la version 3 pour la spécification des modèles.

Première étape : détermination du parc d'estimation

Une estimation du modèle de base sur le parc de référence est réalisée dans chaque strate, comme mentionné au chapitre 3.

A l'issue de cette première estimation, on élimine les observations extrêmes, associées à un grand résidu en valeur absolue et peu compatibles avec le modèle postulé. L'objectif poursuivi est de produire une estimation robuste des coefficients du modèle, moins sensible aux caractéristiques particulières de l'échantillon de données exploité. Ceci permet aussi de rapprocher la distribution empirique des résidus d'une loi normale $(0, \sigma^2)$, pour laquelle la fréquence des observations très grandes en proportion de l'écart-type (plus de 3 écarts-type) et en valeur absolue est négligeable.

Il existe plusieurs méthodes de détection de tels points. On utilise ici la méthode des résidus standardisés (standardized residuals en anglais). On standardise le résidu estimé $\hat{\epsilon}_i$ à l'aide de l'écart-type de la variance estimée de $\hat{\epsilon}_i$. Cette méthode rend possible la détection des observations atypiques dans le cas où le résidu estimé \hat{r}_i de l'observation i serait trop grand.

La formule de calcul est donc pour l'observation i :

$$\hat{r}_i = \frac{\hat{\epsilon}_i}{\hat{s} \sqrt{1 - h_{ii}}} \quad (5.1)$$

avec \hat{s} : la racine carrée de la variance estimée de $\hat{\epsilon}_i$ égale à $\hat{s}^2 = \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{n - (k + 1)}$

n : le nombre d'observations

$k+1$: le nombre de paramètres de l'équation économétrique

$\hat{\epsilon}_i$, : les résidus estimés de la régression

$h_{ii} = x_i'(X'X)^{-1}x_i$: l'effet levier de l'observation i

La quantité (5.1) suit une loi Bêta si le numérateur et le dénominateur sont indépendants. Cette quantité illustre le caractère atypique ou non du résidu standardisé \hat{r}_i et sert de statistique de test pour les hypothèses :

H_0 : l'observation i n'est pas atypique ;

H_1 : l'observation i est atypique.

Après avoir réalisé la régression, on enlève du parc d'estimation toutes les observations ayant un résidu standardisé \hat{r}_i non compris dans l'intervalle $-2 < \hat{r}_i < 2$, pour un niveau de risque $\alpha=0,05$. A la fin de cette étape, on obtient un nouveau parc d'estimation permettant une estimation plus robuste du modèle hédonique.

Deuxième étape : affinement de la spécification afin de la rendre parcimonieuse

Une nouvelle estimation de l'équation économétrique est réalisée sur le parc d'estimation, issu de l'étape 1. L'objectif de l'étape 2 est de réduire le nombre de variables explicatives et d'accroître ainsi la validité et la robustesse du modèle. On réduit aussi par la même occasion le risque de colinéarité entre les variables, risque qui s'accroît avec un nombre de variables plus grand et qui augmente la dépendance des valeurs estimées à l'échantillon disponible. La logique de la sélection des variables est de ne faire entrer dans l'équation que les variables les plus corrélées avec la variable expliquée. On recherche donc une combinaison de variables qui, associées avec la variable expliquée, donnent un R^2 élevé lors de l'estimation du modèle. Il existe diverses méthodes de sélection des variables ; on a utilisé la méthode dite ascendante (*méthode Forward*).

La méthode Forward est une méthode de pas à pas, où l'on injecte une variable après l'autre. On utilise, à chaque pas, un test d'entrée de la variable dans l'équation (test de Fisher). On initialise le processus avec un modèle sans variable au pas 0. Au pas 1, on choisit la variable qui a la plus forte participation à l'augmentation de la variance expliquée. La procédure s'arrête lorsqu'aucune des variables restantes ne satisfait les conditions d'entrée (seuil d'entrée fixé à 0,50) ou, autrement dit, dès qu'aucune variable n'a une valeur de test au-dessus de ce seuil. Le seuil d'entrée de la $j^{\text{ème}}$ variable correspond à un test de Fisher : on compare la somme des carrés (SSE_{M1}) du modèle complet ($M1$) comprenant les j premières variables, avec la somme des carrés (SSE_{M2}) du modèle restreint ($M2$) aux $j-1$ premières variables, soit :

$$\text{Seuil}_{\text{Variable}_j} = \frac{(SSE_{M1} - SSE_{M2})}{SSE_{M2} / (n - k - 1)}$$

où n est le nombre d'observations et k , le nombre de variables du modèle.

5.2 Exemples numériques

On présente ci-après six exemples de modèles, successivement pour les appartements et les maisons, dans trois strates différentes à chaque fois.

Dans chaque modèle et pour chaque variable descriptive, un bien de référence a été choisi, qui définit la modalité de référence. Dans la seconde étape « robuste », la procédure d'estimation comme indiqué dans le paragraphe précédent repose sur une approche Forward. Les variables sont introduites au fur et à mesure dans le modèle en fonction de leur pouvoir explicatif (c'est-à-dire de leur contribution à la réduction de la variance empirique des termes d'erreur de la régression).

A l'issue de cette étape, un certain nombre de variables n'ont pas été introduites dans l'équation du fait de leur influence non significativement différente de la modalité de référence qui leur est associée. Les lignes marquées n.s. dans les tableaux ci-dessous correspondent à ces modalités de variables non sélectionnées. Ainsi par exemple, dans le tableau 5.1, le modèle final est obtenu après l'introduction de 49 variables en plus de la constante en 48 étapes exploratoires. Cette modélisation est satisfaisante dans la mesure où 73,1% de la variance des prix au m^2 (en logarithme) est associée à ces 49 variables descriptives des caractéristiques des biens et donc prise en compte par le modèle.

5.2.1 Exemples de régressions (parc 2009-2010)

Appartements

Tableau 5.1 - Appartements de la strate 3 en Île-de-France (hors Paris)

($R^2 = 0,731$; $s = 0,190^{47}$; nombre d'observations : 8059 – 48 étapes FORWARD)

Variables	Coefficient	Écart-type	P-valeur
(Constante)	7,844	0,009	0,000
Année 2009	-0,045	0,004	0,000
Année 2010		Référence	

⁴⁷ s est l'erreur moyenne quadratique (ou racine carrée de l'écart quadratique moyen).

Variables	Coefficient	Écart-type	P-valeur
janvier	-0,025	0,008	0,002
février	-0,033	0,007	0,000
mars	-0,052	0,007	0,000
avril	-0,042	0,007	0,000
mai	-0,038	0,008	0,000
juin	-0,018	0,007	0,006
juillet	-0,023	0,006	0,000
août	ns	-	-
septembre	ns	-	-
octobre	ns	-	-
novembre	ns	-	-
décembre		Référence	
Avant 1850	ns	-	-
1850 – 1913	0,108	0,034	0,002
1914 – 1947	0,078	0,022	0,000
1948 – 1969		Référence	
1970 – 1980	-0,032	0,005	0,000
1981 – 1991	0,022	0,005	0,000
1992 – 2010	0,115	0,006	0,000
Époque inconnue	ns	-	-
0 salle de bains	ns	-	-
1 salle de bains		Référence	
2 salles de bains et plus	0,018	0,008	0,032
0 cave		Référence	
1 cave et plus	0,041	0,004	0,000
0 garage		Référence	
1 garage	0,071	0,006	0,000
2 garages et plus	0,113	0,008	0,000
Rez-de-chaussée		Référence	
1 ^{er} étage	ns	-	-
2 ^{ème} étage	ns	-	-
3 ^{ème} étage	ns	-	-
4 ^{ème} étage avec ascenseur	-0,034	0,005	0,000
4 ^{ème} étage sans ascenseur	ns	-	-
Surface par pièce des studios <20 m²	-0,050	0,017	0,003
Surface par pièce des studios entre 20 et 30 m²		Référence	
Surface par pièce des studios >30 m²	-0,103	0,013	0,000
Surface par pièce des 2 pièces <17 m²	0,157	0,018	0,000
Surface par pièce des 2 pièces entre 17 et 24 m²		Référence	
Surface par pièce des 2 pièces >24 m²	-0,124	0,008	0,000
Surface par pièce des 3 pièces <18 m²	0,061	0,017	0,000
Surface par pièce des 3 pièces entre 18 et 22 m²		Référence	
Surface par pièce des 3 pièces >22 m²	-0,081	0,007	0,000
Surface par pièce des 4 pièces et plus <17 m²	ns	-	-
Surface par pièce des 4 pièces entre 17 et 21 m²		Référence	
Surface par pièce des 4 pièces et plus >21 m²	-0,059	0,007	0,000
Studio dans le quartier 1	0,293	0,014	0,000
2 pièces dans le quartier 1	0,145	0,010	0,000
3 pièces dans le quartier 1		Référence	
4 pièces dans le quartier 1	-0,102	0,010	0,000
5 pièces et plus dans le quartier 1	-0,253	0,016	0,000
Studio dans le quartier 2	0,141	0,018	0,000
2 pièces dans le quartier 2	-0,043	0,012	0,000
3 pièces dans le quartier 2	-0,209	0,009	0,000
4 pièces dans le quartier 2	-0,308	0,010	0,000
5 pièces et plus dans le quartier 2	-0,419	0,015	0,000
Studio dans le quartier 3	0,338	0,015	0,000
2 pièces dans le quartier 3	0,181	0,011	0,000
3 pièces dans le quartier 3	0,033	0,010	0,001
4 pièces dans le quartier 3	-0,091	0,011	0,000
5 pièces et plus dans le quartier 3	-0,193	0,022	0,000
Studio dans le quartier 4	0,346	0,017	0,000
2 pièces dans le quartier 4	0,174	0,012	0,000
3 pièces dans le quartier 4	0,019	0,009	0,039
4 pièces dans le quartier 4	-0,132	0,010	0,000

Variables	Coefficient	Écart-type	P-valeur
5 pièces et plus dans le quartier 4	-0,232	0,015	0,000
Studio dans le quartier 5	0,410	0,030	0,000
2 pièces dans le quartier 5	0,302	0,016	0,000
3 pièces dans le quartier 5	0,214	0,011	0,000
4 pièces dans le quartier 5	0,079	0,014	0,000
5 pièces et plus dans le quartier 5	ns	-	-

Tableau 5.2 - Appartements de la strate 7502 à Paris

(R² = 0,280 ; s=0,190 ; nombre d'observations : 13702 - 58 étapes FORWARD)

Variables	Coefficient	Écart-type	P-valeur
(Constante)	8,717	0,008	0,000
Année 2009	-0,118	0,003	0,000
Année 2010		Référence	
janvier	-0,101	0,007	0,000
février	-0,101	0,007	0,000
mars	-0,115	0,007	0,000
avril	-0,097	0,007	0,000
mai	-0,098	0,007	0,000
juin	-0,079	0,006	0,000
juillet	-0,056	0,006	0,000
août	-0,048	0,007	0,000
septembre	-0,036	0,006	0,000
octobre	-0,032	0,006	0,000
novembre	-0,019	0,007	0,006
décembre		Référence	
Avant 1850	0,072	0,011	0,000
1850 – 1913	0,035	0,004	0,000
1914 – 1947	0,034	0,005	0,000
1948 – 1969		Référence	
1970 – 1980	0,020	0,006	0,001
1981 – 1991	0,054	0,010	0,000
1992 – 2010	0,070	0,010	0,000
Époque inconnue	0,037	0,005	0,000
0 salle de bains	-0,056	0,007	0,000
1 salle de bains		Référence	
2 salles de bains et plus	0,045	0,009	0,000
nb de salles de bain inconnu	-0,027	0,005	0,000
0 cave		Référence	
1 cave et plus	ns	-	-
0 garage		Référence	
1 garage	ns	-	-
2 garages et plus	0,062	0,019	0,001
Rez-de-chaussée		Référence	
1 ^{er} étage	0,013	0,005	0,017
2 ^{ème} étage	0,037	0,005	0,000
3 ^{ème} étage	0,041	0,005	0,000
4 ^{ème} étage avec ascenseur	0,056	0,005	0,000
4 ^{ème} étage sans ascenseur	0,045	0,009	0,000
Surface par pièce des studios <20 m ²	0,036	0,007	0,000
Surface par pièce des studios entre 20 et 30 m ²		Référence	
Surface par pièce des studios >30 m ²	-0,025	0,009	0,007
Surface par pièce des 2 pièces <17 m ²	ns	-	-
Surface par pièce des 2 pièces entre 17 et 24 m ²		Référence	
Surface par pièce des 2 pièces >24 m ²	ns	-	-
Surface par pièce des 3 pièces <18 m ²	-0,024	0,006	0,000
Surface par pièce des 3 pièces entre 18 et 22 m ²		Référence	
Surface par pièce des 3 pièces >22 m ²	ns	-	-
Surface par pièce des 4 pièces et plus <17 m ²	ns	-	-
Surface par pièce des 4 pièces entre 17 et 21 m ²		Référence	
Surface par pièce des 4 pièces et plus >21 m ²	0,023	0,008	0,003
Studio dans le quartier 1	-0,010	0,007	0,167
2 pièces dans le quartier 1	-0,025	0,006	0,000
3 pièces dans le quartier 1		Référence	
4 pièces dans le quartier 1	ns	-	-

Variables	Coefficient	Écart-type	P-valeur
5 pièces et plus dans le quartier 1	ns	-	-
Studio dans le quartier 2	0,237	0,014	0,000
2 pièces dans le quartier 2	0,219	0,011	0,000
3 pièces dans le quartier 2	0,223	0,014	0,000
4 pièces dans le quartier 2	0,193	0,021	0,000
5 pièces et plus dans le quartier 2	0,285	0,032	0,000
Studio dans le quartier 3	0,073	0,008	0,000
2 pièces dans le quartier 3	0,074	0,006	0,000
3 pièces dans le quartier 3	0,092	0,007	0,000
4 pièces dans le quartier 3	0,081	0,010	0,000
5 pièces et plus dans le quartier 3	0,089	0,014	0,000
Studio dans le quartier 4	0,173	0,014	0,000
2 pièces dans le quartier 4	0,131	0,010	0,000
3 pièces dans le quartier 4	0,129	0,013	0,000
4 pièces dans le quartier 4	0,101	0,019	0,000
5 pièces et plus dans le quartier 4	ns	-	-
Studio dans le quartier 5	0,051	0,010	0,000
2 pièces dans le quartier 5	0,036	0,007	0,000
3 pièces dans le quartier 5	0,092	0,008	0,000
4 pièces dans le quartier 5	0,098	0,011	0,000
5 pièces et plus dans le quartier 5	0,110	0,017	0,000
Studio dans le quartier 6	-0,022	0,011	0,035
2 pièces dans le quartier 6	-0,037	0,008	0,000
3 pièces dans le quartier 6	-0,035	0,010	0,001
4 pièces dans le quartier 6	-0,066	0,015	0,000
5 pièces et plus dans le quartier 6	-0,071	0,025	0,004

Tableau 5.3 - Appartements à Toulouse

(R² = 0,57 ; s=0,17 ; nombre d'observations : 5 538 - 70 étapes FORWARD)

Variables	Coefficient	Écart-type	P-valeur
(Constante)	7,505	0,013	0,000
Année 2009	-0,053	0,005	0,000
Année 2010		Référence	
janvier	-0,023	0,009	0,011
février	-0,051	0,010	0,000
mars	-0,038	0,009	0,000
avril	-0,048	0,010	0,000
mai	-0,026	0,009	0,007
juin	ns	-	-
juillet	ns	-	-
août	ns	-	-
septembre	ns	-	-
octobre	ns	-	-
novembre	ns	-	-
décembre		Référence	
Avant 1850	0,103	0,024	0,000
1850 - 1913	0,144	0,018	0,000
1914 - 1947	0,130	0,017	0,000
1948 - 1969		Référence	
1970 - 1980	0,001	0,010	0,883
1981 - 1991	0,084	0,010	0,000
1992 - 2010	0,080	0,009	0,000
Époque inconnue	0,076	0,009	0,000
0 salle de bains	0,026	0,013	0,044
1 salle de bains		Référence	
2 salles de bains et plus	0,084	0,015	0,000
Nombre de salle de bain inconnu	ns	-	-
0 cave		Référence	
1 cave et plus	-0,064	0,006	0,000
Nombre de cave inconnu	ns	-	-
0 garage		Référence	
1 garage	0,077	0,006	0,000
2 garages et plus	0,155	0,011	0,000

Variables	Coefficient	Écart-type	P-valeur
Rez-de-chaussée		Référence	
1er étage	ns	-	-
2ème étage	ns	-	-
3ème étage	ns	-	-
4ème étage avec ascenseur	-0,019	0,006	0,003
4ème étage sans ascenseur	ns	-	-
Surface par pièce des studios <20 m²	0,144	0,014	0,000
Surface par pièce des studios entre 20 et 30 m²		Référence	
Surface par pièce des studios >30 m²	-0,103	0,013	0,000
Surface par pièce des 2 pièces <17 m²	0,082	0,012	0,000
Surface par pièce des 2 pièces entre 17 et 24 m²		Référence	
Surface par pièce des 2 pièces >24 m²	-0,047	0,010	0,000
Surface par pièce des 3 pièces <18 m²	ns	-	-
Surface par pièce des 3 pièces entre 18 et 22 m²		Référence	
Surface par pièce des 3 pièces >22 m²	0,023	0,009	0,016
Surface par pièce des 4 pièces et plus <17 m²	ns	-	-
Surface par pièce des 4 pièces entre 17 et 21 m²		Référence	
Surface par pièce des 4 pièces et plus >21 m²	0,058	0,011	0,000
Bon état		Référence	
Travaux à prévoir	-0,062	0,010	0,000
A rénover	-0,187	0,019	0,000
État inconnu	-0,037	0,006	0,000
Sans terrasse ni balcon		Référence	
Avec terrasse (s) ou balcon (s)	0,049	0,005	0,000
Studio dans le quartier 1	0,176	0,021	0,000
2 pièces dans le quartier 1	0,010	0,019	0,597
3 pièces dans le quartier 1		Référence	
4 pièces dans le quartier 1	-0,249	0,019	0,000
5 pièces et plus dans le quartier 1	-0,248	0,030	0,000
Studio dans le quartier 2	0,303	0,018	0,000
2 pièces dans le quartier 2	0,194	0,018	0,000
3 pièces dans le quartier 2	0,096	0,021	0,000
4 pièces dans le quartier 2	ns	-	-
5 pièces et plus dans le quartier 2	ns	-	-
Studio dans le quartier 3	0,534	0,021	0,000
2 pièces dans le quartier 3	0,522	0,022	0,000
3 pièces dans le quartier 3	0,520	0,022	0,000
4 pièces dans le quartier 3	0,482	0,032	0,000
5 pièces et plus dans le quartier 3	0,468	0,029	0,000
Studio dans le quartier 4	0,277	0,027	0,000
2 pièces dans le quartier 4	0,219	0,020	0,000
3 pièces dans le quartier 4	0,114	0,020	0,000
4 pièces dans le quartier 4	0,148	0,022	0,000
5 pièces et plus dans le quartier 4	0,096	0,044	0,031
Studio dans le quartier 5	0,233	0,029	0,000
2 pièces dans le quartier 5	0,145	0,020	0,000
3 pièces dans le quartier 5	0,056	0,017	0,001
4 pièces dans le quartier 5	0,002	0,020	0,916
5 pièces et plus dans le quartier 5	ns	-	-
Studio dans le quartier 6	0,185	0,027	0,000
2 pièces dans le quartier 6	0,063	0,015	0,000
3 pièces dans le quartier 6	-0,017	0,015	0,254
4 pièces dans le quartier 6	-0,097	0,019	0,000
5 pièces et plus dans le quartier 6	-0,149	0,044	0,001
Studio dans le quartier 7	0,274	0,028	0,000
2 pièces dans le quartier 7	0,188	0,018	0,000
3 pièces dans le quartier 7	0,084	0,017	0,000
4 pièces dans le quartier 7	-0,043	0,019	0,020
5 pièces et plus dans le quartier 7	-0,126	0,032	0,000
Studio dans le quartier 8	0,202	0,026	0,000
2 pièces dans le quartier 8	0,104	0,020	0,000
3 pièces dans le quartier 8	0,047	0,020	0,021
4 pièces dans le quartier 8	-0,066	0,024	0,006
5 pièces et plus dans le quartier 8	-0,131	0,040	0,001

Variables	Coefficient	Écart-type	P-valeur
Studio dans le quartier 9	0,384	0,015	0,000
2 pièces dans le quartier 9	0,313	0,014	0,000
3 pièces dans le quartier 9	0,277	0,014	0,000
4 pièces dans le quartier 9	0,252	0,016	0,000
5 pièces et plus dans le quartier 9	0,322	0,021	0,000

Tableau 5.4 - Appartements à Lyon

(R² = 0,44 ; s=0,17 ; nombre d'observations : 7 168 – 69 étapes FORWARD)

Variables	Coefficient	Écart-type	P-valeur
(Constante)	7,947	0,010	0,000
Année 2009	-0,086	0,004	0,000
Année 2010	Référence		
janvier	-0,050	0,009	0,000
février	-0,054	0,009	0,000
mars	-0,062	0,008	0,000
avril	-0,047	0,008	0,000
mai	-0,044	0,008	0,000
juin	-0,025	0,007	0,000
juillet	ns	-	-
août	ns	-	-
septembre	ns	-	-
octobre	ns	-	-
novembre	ns	-	-
décembre	Référence		
Avant 1850	0,102	0,014	0,000
1850 - 1913	0,079	0,009	0,000
1914 - 1947	0,041	0,009	0,000
1948 - 1969	Référence		
1970 - 1980	-0,031	0,008	0,000
1981 - 1991	0,061	0,008	0,000
1992 - 2010	0,132	0,008	0,000
Époque inconnue	0,050	0,006	0,000
0 salle de bains	ns	-	-
1 salle de bains	Référence		
2 salles de bains et plus	0,040	0,009	0,000
Nombre de salle de bain inconnu	-0,041	0,009	0,000
0 cave	Référence		
1 cave et plus	ns	-	-
Nombre de cave inconnu	ns	-	-
0 garage	Référence		
1 garage	0,044	0,005	0,000
2 garages et plus	0,092	0,010	0,000
Rez-de-chaussée	Référence		
1er étage	0,035	0,007	0,000
2ème étage	0,053	0,007	0,000
3ème étage	0,045	0,007	0,000
4ème étage avec ascenseur	0,056	0,006	0,000
4ème étage sans ascenseur	ns	-	-
Surface par pièce des studios <20 m ²	0,097	0,018	0,000
Surface par pièce des studios entre 20 et 30 m ²	Référence		
Surface par pièce des studios >30 m ²	-0,127	0,011	0,000
Surface par pièce des 2 pièces <17 m ²	0,038	0,015	0,011
Surface par pièce des 2 pièces entre 17 et 24 m ²	Référence		
Surface par pièce des 2 pièces >24 m ²	-0,057	0,008	0,000
Surface par pièce des 3 pièces <18 m ²	0,060	0,012	0,000
Surface par pièce des 3 pièces entre 18 et 22 m ²	Référence		
Surface par pièce des 3 pièces >22 m ²	ns	-	-
Surface par pièce des 4 pièces et plus <17 m ²	ns	-	-
Surface par pièce des 4 pièces entre 17 et 21 m ²	Référence		
Surface par pièce des 4 pièces et plus >21 m ²	0,032	0,007	0,000
Bon état	Référence		
Travaux à prévoir	-0,104	0,007	0,000
A rénover	-0,199	0,014	0,000
État inconnu	-0,020	0,005	0,000

Variables	Coefficient	Écart-type	P-valeur
Sans terrasse ni balcon		Référence	
Avec terrasse (s) ou balcon (s)	0,023	0,005	0,000
Studio dans le quartier 1	0,134	0,021	0,000
2 pièces dans le quartier 1	0,036	0,014	0,014
3 pièces dans le quartier 1		Référence	
4 pièces dans le quartier 1	ns	-	-
5 pièces et plus dans le quartier 1	ns	-	-
Studio dans le quartier 2	0,258	0,025	0,000
2 pièces dans le quartier 2	0,145	0,020	0,000
3 pièces dans le quartier 2	0,048	0,018	0,007
4 pièces dans le quartier 2	0,064	0,021	0,002
5 pièces et plus dans le quartier 2	0,073	0,029	0,010
Studio dans le quartier 3	0,069	0,015	0,000
2 pièces dans le quartier 3	-0,049	0,011	0,000
3 pièces dans le quartier 3	-0,096	0,010	0,000
4 pièces dans le quartier 3	-0,121	0,011	0,000
5 pièces et plus dans le quartier 3	-0,134	0,015	0,000
Studio dans le quartier 4	0,155	0,022	0,000
2 pièces dans le quartier 4	0,059	0,015	0,000
3 pièces dans le quartier 4	ns	-	-
4 pièces dans le quartier 4	ns	-	-
5 pièces et plus dans le quartier 4	ns	-	-
Studio dans le quartier 5	-0,081	0,022	0,000
2 pièces dans le quartier 5	-0,111	0,017	0,000
3 pièces dans le quartier 5	-0,199	0,014	0,000
4 pièces dans le quartier 5	-0,260	0,016	0,000
5 pièces et plus dans le quartier 5	-0,286	0,022	0,000
Studio dans le quartier 6	0,123	0,026	0,000
2 pièces dans le quartier 6	ns	-	-
3 pièces dans le quartier 6	-0,155	0,013	0,000
4 pièces dans le quartier 6	-0,227	0,012	0,000
5 pièces et plus dans le quartier 6	-0,199	0,016	0,000
Studio dans le quartier 7	0,232	0,019	0,000
2 pièces dans le quartier 7	0,139	0,014	0,000
3 pièces dans le quartier 7	0,114	0,012	0,000
4 pièces dans le quartier 7	0,085	0,014	0,000
5 pièces et plus dans le quartier 7	0,100	0,015	0,000
Studio dans le quartier 8	ns	-	-
2 pièces dans le quartier 8	-0,072	0,012	0,000
3 pièces dans le quartier 8	-0,134	0,011	0,000
4 pièces dans le quartier 8	-0,199	0,014	0,000
5 pièces et plus dans le quartier 8	-0,161	0,019	0,000
Studio dans le quartier 9	0,057	0,020	0,004
2 pièces dans le quartier 9	-0,096	0,013	0,000
3 pièces dans le quartier 9	-0,201	0,012	0,000
4 pièces dans le quartier 9	-0,273	0,013	0,000
5 pièces et plus dans le quartier 9	-0,332	0,019	0,000

Maisons

Tableau 5.5 - Maisons dans la strate 5 en Île-de-France ⁴⁸

(R² = 0,678 ; s=0,190 ; nombre d'observations : 7120 - 34 étapes FORWARD)

Variables	Coefficient	Écart-type	P-valeur
(constante)	9,995	0,052	0,000
Surface habitable (m²)	0,353	0,012	0,000
Surface de terrain (m²)	0,148	0,003	0,000
Année 2009	-0,049	0,003	0,000
Année 2010		Référence	

⁴⁸ Dans cet exemple, la variable « nombre de bâtiments » ne ressort pas. Cette variable n'est présente que pour une strate seulement.

Variables	Coefficient	Écart-type	P-valeur
janvier	ns	-	-
février	-0,019	0,007	0,005
mars	-0,016	0,007	0,024
avril	-0,019	0,007	0,005
mai	-0,016	0,007	0,014
juin	ns	-	-
juillet	ns	-	-
août	ns	-	-
septembre	ns	-	-
octobre	ns	-	-
novembre	ns	-	-
décembre		Référence	
Avant 1913	ns	-	-
1914 – 1947	ns	-	-
1948 – 1969	-0,019	0,008	0,000
1970 – 1980	-0,018	0,005	0,000
Après 1980	0,037	0,004	0,000
Époque de construction inconnue		Référence	
0 salle de bains	-0,163	0,032	0,000
1 salle de bains		Référence	
2 salles de bains	0,050	0,004	0,000
3 salles de bains et plus	-0,078	0,013	0,000
0 cave		Référence	
1 cave	0,026	0,005	0,000
2 caves et plus	ns	-	-
0 garage	-0,036	0,005	0,000
1 garage		Référence	
2 garages et plus	0,039	0,006	0,000
1 niveau	-0,041	0,005	0,000
2 niveaux		Référence	
3 niveaux et plus	ns	-	-
3 pièces et moins dans le quartier 1	-0,059	0,009	0,000
4 pièces dans le quartier 1		Référence	
5 pièces dans le quartier 1	ns	-	-
6 pièces dans le quartier 1	0,024	0,007	0,000
7 pièces et plus dans le quartier 1	0,042	0,009	0,000
3 pièces et moins dans le quartier 2	ns	-	-
4 pièces dans le quartier 2	-0,033	0,007	0,000
5 pièces dans le quartier 2	-0,037	0,006	0,000
6 pièces dans le quartier 2	-0,028	0,008	0,000
7 pièces et plus dans le quartier 2	-0,040	0,011	0,000
3 pièces et moins dans le quartier 3	0,120	0,021	0,000
4 pièces dans le quartier 3	0,169	0,010	0,000
5 pièces dans le quartier 3	0,158	0,008	0,000
6 pièces dans le quartier 3	0,188	0,012	0,000
7 pièces et plus dans le quartier 3	0,182	0,016	0,000
3 pièces et moins dans le quartier 4	0,139	0,017	0,000
4 pièces dans le quartier 4	0,112	0,013	0,000
5 pièces dans le quartier 4	0,129	0,012	0,000
6 pièces dans le quartier 4	0,153	0,014	0,000
7 pièces et plus dans le quartier 4	0,188	0,022	0,000

Tableau 5.6 - Maisons de la banlieue de Lille

(R² = 0,85 ; s=0,17; nombre d'observations : 6 702 - 75 étapes FORWARD)

Variables	Coefficient	Écart-type	P-valeur
(Constante)	8,689	0,047	0,000
Surface habitable (m²)	0,531	0,011	0,000
Surface de terrain (m²)	0,180	0,004	0,000
Année 2009	-0,053	0,004	0,000
Année 2010		Référence	

Variables	Coefficient	Écart-type	P-valeur
janvier	-0,040	0,009	0,000
février	ns	-	-
mars	-0,021	0,009	0,018
avril	-0,033	0,009	0,000
mai	-0,026	0,009	0,004
juin	-0,016	0,007	0,026
juillet	ns	-	-
août	ns	-	-
septembre	ns	-	-
octobre	ns	-	-
novembre	ns	-	-
décembre		Référence	
Avant 1913	-0,064	0,011	0,000
1914 – 1947	-0,039	0,009	0,000
1948 – 1969	-0,008	0,009	0,421
1970 – 1980	0,041	0,010	0,000
Après 1980	0,137	0,010	0,000
Époque de construction inconnue		Référence	
0 salle de bains	ns	-	-
1 salle de bains		Référence	
2 salles de bains et plus	0,123	0,007	0,000
Présence d'un sous-sol	0,019	0,005	0,000
Pas de sous-sol		Référence	
0 garage	-0,055	0,005	0,000
1 garage		Référence	
2 garages et plus	0,026	0,008	0,002
1 niveau	ns	-	-
2 niveaux		Référence	
3 niveaux et plus	-0,016	0,005	0,002
Bon état	0,029	0,006	0,000
Travaux à prévoir	-0,138	0,007	0,000
A rénover	-0,357	0,012	0,000
État inconnu		Référence	
3 pièces et moins dans le quartier 1	ns	-	-
4 pièces dans le quartier 1		Référence	
5 pièces dans le quartier 1	0,116	0,017	0,000
6 pièces dans le quartier 1	0,170	0,023	0,000
7 pièces et plus dans le quartier 1	0,252	0,034	0,000
3 pièces et moins dans le quartier 2	ns	-	-
4 pièces dans le quartier 2	ns	-	-
5 pièces dans le quartier 2	0,076	0,019	0,000
6 pièces dans le quartier 2	ns	-	-
7 pièces et plus dans le quartier 2	ns	-	-
3 pièces et moins dans le quartier 3	ns	-	-
4 pièces dans le quartier 3	-0,039	0,017	0,021
5 pièces dans le quartier 3	ns	-	-
6 pièces dans le quartier 3	ns	-	-
7 pièces et plus dans le quartier 3	-0,142	0,072	0,047
3 pièces et moins dans le quartier 4	-0,088	0,029	0,003
4 pièces dans le quartier 4	ns	-	-
5 pièces dans le quartier 4	-0,092	0,022	0,000
6 pièces dans le quartier 4	-0,201	0,040	0,000
7 pièces et plus dans le quartier 4	ns	-	-
3 pièces et moins dans le quartier 5	0,096	0,038	0,013
4 pièces dans le quartier 5	0,197	0,021	0,000
5 pièces dans le quartier 5	0,190	0,018	0,000
6 pièces dans le quartier 5	0,308	0,028	0,000
7 pièces et plus dans le quartier 5	0,344	0,032	0,000
3 pièces et moins dans le quartier 6	ns	-	-
4 pièces dans le quartier 6	ns	-	-
5 pièces dans le quartier 6	-0,064	0,023	0,006
6 pièces dans le quartier 6	ns	-	-
7 pièces et plus dans le quartier 6	ns	-	-
3 pièces et moins dans le quartier 7	0,129	0,047	0,006
4 pièces dans le quartier 7	0,170	0,027	0,000

Variables	Coefficient	Écart-type	P-valeur
5 pièces dans le quartier 7	0,236	0,024	0,000
6 pièces dans le quartier 7	0,335	0,032	0,000
7 pièces et plus dans le quartier 7	0,472	0,044	0,000
3 pièces et moins dans le quartier 8	0,164	0,032	0,000
4 pièces dans le quartier 8	0,212	0,016	0,000
5 pièces dans le quartier 8	0,242	0,016	0,000
6 pièces dans le quartier 8	0,234	0,020	0,000
7 pièces et plus dans le quartier 8	0,421	0,025	0,000
3 pièces et moins dans le quartier 9	0,150	0,035	0,000
4 pièces dans le quartier 9	0,154	0,022	0,000
5 pièces dans le quartier 9	0,219	0,024	0,000
6 pièces dans le quartier 9	0,191	0,030	0,000
7 pièces et plus dans le quartier 9	0,295	0,036	0,000
3 pièces et moins dans le quartier 10	ns	-	-
4 pièces dans le quartier 10	0,045	0,021	0,032
5 pièces dans le quartier 10	ns	-	-
6 pièces dans le quartier 10	ns	-	-
7 pièces et plus dans le quartier 10	0,230	0,078	0,003
3 pièces et moins dans le quartier 11	-0,295	0,023	0,000
4 pièces dans le quartier 11	-0,200	0,013	0,000
5 pièces dans le quartier 11	-0,187	0,013	0,000
6 pièces dans le quartier 11	-0,158	0,020	0,000
7 pièces et plus dans le quartier 11	-0,195	0,024	0,000
3 pièces et moins dans le quartier 12	0,074	0,032	0,020
4 pièces dans le quartier 12	0,144	0,019	0,000
5 pièces dans le quartier 12	0,084	0,020	0,000
6 pièces dans le quartier 12	0,175	0,029	0,000
7 pièces et plus dans le quartier 12	0,135	0,045	0,003
3 pièces et moins dans le quartier 13	-0,114	0,025	0,000
4 pièces dans le quartier 13	-0,086	0,014	0,000
5 pièces dans le quartier 13	-0,107	0,017	0,000
6 pièces dans le quartier 13	-0,128	0,030	0,000
7 pièces et plus dans le quartier 13	-0,144	0,058	0,014
3 pièces et moins dans le quartier 14	-0,063	0,023	0,006
4 pièces dans le quartier 14	ns	-	-
5 pièces dans le quartier 14	-0,030	0,012	0,015
6 pièces dans le quartier 14	ns	-	-
7 pièces et plus dans le quartier 14	ns	-	-
3 pièces et moins dans le quartier 15	ns	-	-
4 pièces dans le quartier 15	ns	-	-
5 pièces dans le quartier 15	ns	-	-
6 pièces dans le quartier 15	ns	-	-
7 pièces et plus dans le quartier 15	ns	-	-
3 pièces et moins dans le quartier 16	-0,062	0,019	0,001
4 pièces dans le quartier 16	ns	-	-
5 pièces dans le quartier 16	ns	-	-
6 pièces dans le quartier 16	ns	-	-
7 pièces et plus dans le quartier 16	ns	-	-
3 pièces et moins dans le quartier 17	ns	-	-
4 pièces dans le quartier 17	0,059	0,013	0,000
5 pièces dans le quartier 17	0,118	0,011	0,000
6 pièces dans le quartier 17	0,132	0,015	0,000
7 pièces et plus dans le quartier 17	0,153	0,021	0,000

Tableau 5.7 - Maisons de la banlieue de Bordeaux

(R² = 0,67 ; s=0,19; nombre d'observations : 3 818 - 53 étapes FORWARD)

Variables	Coefficient	Écart-type	P-valeur
(Constante)	9,414	0,074	0,000
Surface habitable (m ²)	0,534	0,017	0,000
Surface de terrain (m ²)	0,074	0,005	0,000
Année 2009	-0,063	0,006	0,000
Année 2010	Référence		

Variables	Coefficient	Écart-type	P-valeur
janvier	-0,030	0,014	0,034
février	-0,054	0,014	0,000
mars	-0,078	0,013	0,000
avril	-0,044	0,013	0,001
mai	-0,058	0,013	0,000
juin	-0,026	0,011	0,015
juillet	ns	-	-
août	0,022	0,011	0,040
septembre	ns	-	-
octobre	ns	-	-
novembre	ns	-	-
décembre		Référence	
Avant 1913	0,073	0,021	0,001
1914 – 1947	0,067	0,012	0,000
1948 – 1969	ns	-	-
1970 – 1980	ns	-	-
Après 1980	0,066	0,008	0,000
Époque de construction inconnue		Référence	
0 salle de bains	ns	-	-
1 salle de bains		Référence	
2 salles de bains et plus	0,082	0,009	0,000
Présence d'un sous-sol	0,036	0,015	0,013
Pas de sous-sol		Référence	
0 garage	-0,030	0,007	0,000
1 garage		Référence	
2 garages et plus	0,038	0,013	0,004
1 niveau	0,017	0,007	0,013
2 niveaux		Référence	
3 niveaux et plus	ns	-	-
Bon état	0,014	0,007	0,041
Travaux à prévoir	-0,102	0,013	0,000
A rénover	-0,324	0,020	0,000
État inconnu		Référence	
3 pièces et moins dans le quartier 1	ns	-	-
4 pièces dans le quartier 1		Référence	
5 pièces dans le quartier 1	0,091	0,021	0,000
6 pièces dans le quartier 1	0,117	0,027	0,000
7 pièces et plus dans le quartier 1	0,151	0,034	0,000
3 pièces et moins dans le quartier 2	ns	-	-
4 pièces dans le quartier 2	0,029	0,019	0,128
5 pièces dans le quartier 2	ns	-	-
6 pièces dans le quartier 2	ns	-	-
7 pièces et plus dans le quartier 2	ns	-	-
3 pièces et moins dans le quartier 3	ns	-	-
4 pièces dans le quartier 3	ns	-	-
5 pièces dans le quartier 3	ns	-	-
6 pièces dans le quartier 3	ns	-	-
7 pièces et plus dans le quartier 3	-0,090	0,042	0,031
3 pièces et moins dans le quartier 4	0,068	0,026	0,010
4 pièces dans le quartier 4	0,106	0,026	0,000
5 pièces dans le quartier 4	0,086	0,030	0,004
6 pièces dans le quartier 4	0,156	0,043	0,000
7 pièces et plus dans le quartier 4	ns	-	-
3 pièces et moins dans le quartier 5	-0,117	0,031	0,000
4 pièces dans le quartier 5	-0,065	0,022	0,003
5 pièces dans le quartier 5	-0,069	0,022	0,002
6 pièces dans le quartier 5	-0,108	0,042	0,011
7 pièces et plus dans le quartier 5	ns	-	-
3 pièces et moins dans le quartier 6	-0,103	0,019	0,000
4 pièces dans le quartier 6	-0,161	0,015	0,000
5 pièces dans le quartier 6	-0,145	0,018	0,000
6 pièces dans le quartier 6	-0,126	0,028	0,000
7 pièces et plus dans le quartier 6	-0,239	0,045	0,000
3 pièces et moins dans le quartier 7	-0,282	0,023	0,000
4 pièces dans le quartier 7	-0,181	0,015	0,000

Variables	Coefficient	Écart-type	P-valeur
5 pièces dans le quartier 7	-0,166	0,015	0,000
6 pièces dans le quartier 7	-0,112	0,022	0,000
7 pièces et plus dans le quartier 7	-0,180	0,035	0,000
3 pièces et moins dans le quartier 8	-0,062	0,022	0,004
4 pièces dans le quartier 8	-0,064	0,015	0,000
5 pièces dans le quartier 8	0,011	0,016	0,480
6 pièces dans le quartier 8	ns	-	-
7 pièces et plus dans le quartier 8	0,074	0,028	0,008
3 pièces et moins dans le quartier 9	-0,136	0,028	0,000
4 pièces dans le quartier 9	-0,076	0,017	0,000
5 pièces dans le quartier 9	-0,044	0,015	0,004
6 pièces dans le quartier 9	-0,053	0,019	0,005
7 pièces et plus dans le quartier 9	0,033	0,025	0,184

5.2.2 Qualité des régressions

La qualité des régressions hédoniques est habituellement mesurée par le coefficient de détermination R^2 , compris entre 0 et 1. Plus ce coefficient est élevé, plus la régression a un fort pouvoir prédictif. Pour des données individuelles en coupe, de bonnes valeurs de R^2 sont de l'ordre de 0,25-0,40 pour 1 000 à 3 000 observations et une vingtaine de variables explicatives. C'est cette qualité qui est observée ici et valable pour les diverses strates. Rappelons que la variable dépendante est le logarithme (népérien) du prix au m² pour les appartements ou du prix pour les maisons.

Appartements

Tableau 5.8 - Qualité des régressions et nombre d'observations du parc d'estimation selon la strate pour les appartements en Île-de-France

Strate	Étapes	R ²	Parc d'Estimation
1	45	0,478	5 348
2	48	0,635	2 860
3	48	0,731	4 615
4	52	0,601	9 287
5	52	0,508	11 833
6	58	0,519	17 154
7	28	0,681	3 225
8	51	0,640	8 667
9	39	0,292	11 139
10	50	0,415	16 746
7501	47	0,501	8 463
7502	58	0,280	13 702
7503	56	0,397	9 895
7504	32	0,393	4 251
7505	38	0,338	5 626
Total			132 811

Tableau 5.9 - Qualité des régressions et nombre d'observations du parc d'estimation selon la strate pour les appartements en province

Strate	Étapes	R ²	Parc d'estimation
Agglomération de Genève - Annemasse (partie française)	42	0,47	1 876
Agglomération de Bayonne	53	0,54	2 253
Banlieue de Toulon	57	0,60	3 895
Banlieue de Lyon	64	0,54	4 993
Banlieue de Marseille-Aix-en-Provence	49	0,60	1 939
Antibes	32	0,36	1 480
Cannes	49	0,49	2 470
Cannet	23	0,39	1 121
Menton	16	0,27	761
Nice	69	0,43	8 251
Banlieue de Nice	71	0,46	4 671
Marseille	86	0,59	6 289

Strate	Étapes	R ²	Parc d'estimation
Caen	23	0,48	993
Dijon	58	0,54	2 434
Besançon	32	0,46	1 806
Brest	45	0,43	1 750
Toulouse	70	0,57	5 538
Bordeaux	55	0,37	2 537
Banlieue de Bordeaux	46	0,60	2 058
Agde	23	0,41	1 412
Montpellier	48	0,45	1 947
Rennes	47	0,59	2 507
Tours	30	0,54	1 181
Grenoble	60	0,56	2 351
Banlieue de Grenoble	42	0,53	2 163
Saint-Etienne	17	0,32	1 490
Nantes	56	0,45	3 472
Agglomération de Saint-Nazaire	26	0,74	865
Banlieue de Nantes	32	0,65	1 081
Orléans	39	0,52	1 259
Reims	41	0,57	1 558
Nancy	25	0,25	1 484
Banlieue de Nancy	33	0,43	1 181
Metz	31	0,40	1 308
Lille	47	0,59	2 366
Banlieue de Lille (partie française)	40	0,59	2 032
Clermont-Ferrand	26	0,42	1 284
Pau	19	0,31	1 375
Strasbourg	39	0,50	2 422
Banlieue de Strasbourg (partie française)	30	0,43	1 302
Agglomération de Mulhouse	18	0,42	944
Lyon	69	0,44	7 168
Villeurbanne	30	0,46	1 854
Agglomération de Chambéry	25	0,47	1 244
Annecy	30	0,52	1 186
Banlieue de Annecy	35	0,63	1 269
Le Havre	34	0,50	1 269
Rouen	40	0,44	1 356
Banlieue de Rouen	25	0,56	795
Toulon	41	0,43	1 960
Littoral Grand Nord (Régions 31,22)	45	0,74	2 042
Littoral de Normandie (Région 25, 23)	49	0,65	2 249
Littoral Bretagne Est (dép. 35,56)	50	0,66	1 974
Littoral Bretagne Ouest (dép 22,29)	41	0,65	1 735
Littoral Pays de la Loire (Région 52)	28	0,38	1 069
Littoral Sud-Ouest (Régions 54,72)	49	0,46	2 695
Littoral Languedoc-Roussillon Sud (dép 11,66 et 34)	38	0,44	2 252
Littoral Languedoc-Roussillon Sud (dép. 34, 30)	40	0,58	1 612
Littoral Provence-Alpes-Côte d'Azur Ouest (dép. 13, 83)	28	0,25	1 374
Littoral Provence-Alpes-Côte d'Azur Est (dép. 83,06)	43	0,62	2 324
Stations de ski de Tarentaise	34	0,66	1 887
Stations de ski de Maurienne, Val d'Arly, Beaufortin, les Bauges, Chablais, Giffre, Les Bornes et Chartreuse.	37	0,45	1 809
Stations de ski de Mont-Blanc et Aravis.	40	0,65	1 417
Stations de ski de l'Isère	29	0,57	1 149
Autres communes du Nord sauf Picardie, à revenu faible	24	0,28	1 758
Autres communes du Nord sauf Picardie, à revenu moyen	49	0,47	1 880
Autres communes du Nord sauf Picardie, à revenu élevé	34	0,37	1 514
Autres communes de l'Ouest sauf Pays de la Loire, à revenu faible	44	0,51	2 563
Autres communes de l'Ouest sauf Pays de la Loire, à revenu élevé	23	0,38	1 747
Autres communes du Limousin et d'Auvergne, à revenu faible	33	0,52	1 605
Autres communes du Limousin et d'Auvergne, à revenu élevé	35	0,44	1 620
Autres communes de Picardie, à revenu faible	44	0,62	2 029
Autres communes de Picardie, à revenu élevé	24	0,38	1 222
Autres communes du Centre, à revenu faible	40	0,59	1 999

Strate	Étapes	R ²	Parc d'estimation
Autres communes du Centre, à revenu élevé	34	0,49	1 546
Autres communes de Bourgogne, à revenu faible	36	0,50	1 407
Autres communes de Bourgogne, à revenu élevé	27	0,42	1 336
Autres communes de Lorraine, à revenu faible	38	0,44	1 768
Autres communes de Lorraine, à revenu élevé	27	0,28	1 713
Autres communes d'Alsace, à revenu faible	32	0,48	1 861
Autres communes d'Alsace, à revenu élevé	23	0,36	1 316
Autres communes de Franche-Comté, à revenu faible	51	0,49	1 578
Autres communes de Franche-Comté, à revenu élevé	28	0,33	1 341
Autres communes du Pays de la Loire, à revenu faible	36	0,40	1 696
Autres communes du Pays de la Loire, à revenu élevé	35	0,62	1 518
Autres communes du Midi-Pyrénées, à revenu faible	51	0,70	2 414
Autres communes du Midi-Pyrénées, à revenu élevé	26	0,43	1 655
Autres communes Du Rhône-Alpes-Est, à revenu faible	27	0,37	2 068
Autres communes Du Rhône-Alpes-Est, à revenu moyen	32	0,52	1 852
Autres communes Du Rhône-Alpes-Est, à revenu élevé	36	0,38	2 456
Autres communes Du Rhône-Alpes-Ouest, à revenu faible	41	0,57	2 210
Autres communes Du Rhône-Alpes-Ouest, à revenu élevé	45	0,53	2 368
Autres communes du Sud-Est (Languedoc-Roussillon et PACA), à revenu faible	34	0,38	2 133
Autres communes du Sud-Est (Languedoc-Roussillon et PACA), à revenu élevé	39	0,38	3 250
Autres communes de Provence-Alpes-Côte d'Azur Nord, à revenu faible	44	0,51	1 738
Autres communes de Provence-Alpes-Côte d'Azur Nord, à revenu moyen	34	0,60	1 598
Autres communes de Provence-Alpes-Côte d'Azur Nord, à revenu élevé	35	0,50	1 448
TOTAL			197 995

Maisons

Tableau 5.10 - Qualité des régressions et nombre d'observations du parc d'estimation selon la strate pour les maisons en Île-de-France

Strate	Étapes	R ²	Parc d'Estimation
1	53	0,795	5 977
2	24	0,693	1 652
3	41	0,654	11 314
4	37	0,681	5 423
5	34	0,678	7 120
6	37	0,644	9 974
7	39	0,622	11 115
Total			52 575

Tableau 5.11 - Qualité des régressions et nombre d'observations du parc d'estimation selon la strate pour les maisons en province

Strate	Étapes	R ²	Parc d'estimation
Unité Urbaine d'Avignon	19	0,73	953
Unité Urbaine de Béthune	29	0,72	2 005
Unité Urbaine de Metz	23	0,66	1 254
Unité Urbaine de Douai-Lens	32	0,72	3 436
Unité Urbaine de Toulon	33	0,69	2 136
Unité Urbaine de Lyon	26	0,68	4 003
Banlieue de Marseille-Aix-en-Provence	35	0,72	2 067
Unité Urbaine de Nice	36	0,67	2 895
Marseille	24	0,78	910
Unité Urbaine de Dijon	16	0,74	892
Unité Urbaine de Brest	17	0,64	1 001
Toulouse	14	0,67	860
Banlieue de Toulouse	22	0,73	2 221
Bordeaux	25	0,78	1 087
Banlieue de Bordeaux	53	0,67	3 818
Unité Urbaine de Montpellier	13	0,71	701
Unité Urbaine de Rennes	19	0,70	1 059
Unité Urbaine de Tours	24	0,75	1 424
Unité Urbaine de Grenoble	19	0,72	933

Strate	Étapes	R ²	Parc d'estimation
Nantes	17	0,74	1 107
Unité Urbaine de Saint-Nazaire	18	0,67	835
Banlieue de Nantes	35	0,78	2 514
Unité Urbaine de Orléans	26	0,76	1 685
Unité Urbaine de Reims	13	0,76	715
Unité Urbaine de Nancy	26	0,67	1 954
Lille	18	0,77	825
Unité Urbaine de Maubeuge (partie française)	18	0,72	1 049
Tourcoing	12	0,73	1 011
Unité Urbaine de Dunkerque	21	0,69	760
Unité Urbaine de Valenciennes (partie française)	31	0,70	2 357
Banlieue de Lille (partie française)	75	0,85	6 702
Unité Urbaine de Calais	14	0,73	1 040
Unité Urbaine de Clermont-Ferrand	17	0,75	963
Le Mans	24	0,77	1 258
Le Havre	24	0,71	864
Unité Urbaine de Rouen	36	0,81	2 332
Amiens	20	0,80	1 196
Littoral du Finistère et des Côtes d'Armor, à revenu très faible	15	0,60	1 266
Littoral du Finistère et des Côtes d'Armor, à revenu faible	17	0,59	1 062
Littoral du Finistère et des Côtes d'Armor, à revenu moyen	13	0,59	981
Littoral du Finistère et des Côtes d'Armor, à revenu élevé	15	0,62	1 087
Littoral d'Ille-et-Vilaine, et du Morbihan, à revenu très faible	24	0,61	1 084
Littoral d'Ille-et-Vilaine, et du Morbihan, à revenu faible	17	0,63	1 063
Littoral d'Ille-et-Vilaine, et du Morbihan, à revenu moyen	18	0,59	919
Littoral d'Ille-et-Vilaine, et du Morbihan, à revenu élevé	14	0,57	1 073
Littoral du Nord-Pas-de-Calais et des deux Normandie, à revenu très faible	26	0,68	1 297
Littoral du Nord-Pas-de-Calais et des deux Normandie, à revenu faible	16	0,58	1 315
Littoral du Nord-Pas-de-Calais et des deux Normandie, à revenu moyen	18	0,58	1 366
Littoral du Nord-Pas-de-Calais et des deux Normandie, à revenu élevé	19	0,66	1 215
Littoral Méditerranéen, à revenu faible	24	0,72	1 457
Littoral Méditerranéen, à revenu moyen	18	0,75	1 054
Littoral Méditerranéen, à revenu élevé	30	0,70	1 106
Littoral de Loire-Atlantique, Vendée et Charente-Maritime, à revenu faible	35	0,65	2 211
Littoral de Loire-Atlantique, Vendée et Charente-Maritime, à revenu élevé	23	0,67	1 742
Littoral de Gironde, des Landes et des Pyrénées-Atlantiques, à revenu faible	28	0,63	1 558
Littoral de Gironde, des Landes et des Pyrénées-Atlantiques, à revenu élevé	24	0,66	1 111
Stations de Ski d'Isère, Savoie et Haute-Savoie	31	0,68	1 664
Autres communes d'Indre et d'Indre-et-Loire, à revenu faible	18	0,70	859
Autres communes d'Indre et d'Indre-et-Loire, à revenu moyen	24	0,72	1 165
Autres communes d'Indre et d'Indre-et-Loire, à revenu élevé	22	0,70	1 247
Autres communes d'Eure-et-Loir, à revenu faible	35	0,66	1 544
Autres communes d'Eure-et-Loir, à revenu moyen	25	0,67	1 557
Autres communes d'Eure-et-Loir, à revenu élevé	30	0,66	1 639
Autres communes du Loiret, à revenu faible	23	0,68	1 369
Autres communes du Loiret, à revenu moyen	19	0,68	1 278
Autres communes du Loiret, à revenu élevé	22	0,68	1 410
Autres communes du Loir et Cher, à revenu faible	26	0,73	1 532
Autres communes du Loir et Cher, à revenu élevé	26	0,70	1 518
Autres communes du Cher, à revenu faible	20	0,74	995
Autres communes du Cher, à revenu élevé	34	0,78	1 042
Autres communes de Savoie et de Haute-Savoie, à revenu faible	30	0,71	1 636
Autres communes de Savoie et de Haute-Savoie, à revenu moyen	29	0,64	1 594
Autres communes de Savoie et de Haute-Savoie, à revenu élevé	29	0,61	1 529
Autres communes de Rhône-Alpes Ouest, à revenu très faible	26	0,69	1 905
Autres communes de Rhône-Alpes Ouest, à revenu faible	35	0,69	2 079
Autres communes de Rhône-Alpes Ouest, à revenu moyen	25	0,67	1 936
Autres communes de Rhône-Alpes Ouest, à revenu élevé	24	0,61	1 998
Autres communes de l'Ain, à revenu faible	25	0,65	2 028
Autres communes de l'Ain, à revenu élevé	29	0,70	2 315
Autres communes de l'Isère, à revenu faible	22	0,61	1 910
Autres communes de l'Isère, à revenu élevé	35	0,64	2 006
Autres communes de Champagne-Ardenne Nord, à revenu faible	27	0,69	1 377
Autres communes de Champagne-Ardenne Nord, à revenu moyen	40	0,64	1 558
Autres communes de Champagne-Ardenne Nord, à revenu élevé	28	0,68	1 583
Autres communes de Champagne-Ardenne Sud, à revenu faible	31	0,68	1 505

Strate	Étapes	R ²	Parc d'estimation
Autres communes de Champagne-Ardenne Sud, à revenu élevé	28	0,68	1 533
Autres communes de la Somme, à revenu faible	25	0,67	1 225
Autres communes de la Somme, à revenu moyen	20	0,70	1 459
Autres communes de la Somme, à revenu élevé	28	0,74	1 547
Autres communes de l'Aisne, à revenu très faible	28	0,74	1 564
Autres communes de l'Aisne, à revenu faible	33	0,73	1 716
Autres communes de l'Aisne, à revenu moyen	26	0,68	1 678
Autres communes de l'Aisne, à revenu élevé	33	0,74	1 794
Autres communes de l'Oise, à revenu faible	26	0,71	1 270
Autres communes de l'Oise, à revenu moyen	21	0,64	1 357
Autres communes de l'Oise, à revenu élevé	16	0,68	1 376
Autres communes de Seine-Maritime, à revenu faible	25	0,73	1 544
Autres communes de Seine-Maritime, à revenu moyen	23	0,71	1 811
Autres communes de Seine-Maritime, à revenu élevé	29	0,70	1 600
Autres communes de l'Eure, à revenu faible	35	0,65	1 856
Autres communes de l'Eure, à revenu moyen	26	0,64	1 698
Autres communes de l'Eure, à revenu élevé	17	0,66	1 779
Autres communes de Basse-Normandie, à revenu très faible	31	0,68	1 862
Autres communes de Basse-Normandie, à revenu faible	31	0,67	2 082
Autres communes de Basse-Normandie, à revenu moyen	46	0,73	2 166
Autres communes de Basse-Normandie, à revenu élevé	30	0,68	2 251
Autres communes de Bourgogne Est (21,71), à revenu faible	25	0,67	2 043
Autres communes de Bourgogne Est (21,71), à revenu moyen	24	0,61	2 116
Autres communes de Bourgogne Est (21,71), à revenu élevé	30	0,71	2 237
Autres communes de Bourgogne Ouest (89,58), à revenu faible	21	0,69	1 058
Autres communes de Bourgogne Ouest (89,58), à revenu moyen	26	0,62	1 257
Autres communes de Bourgogne Ouest (89,58), à revenu élevé	21	0,62	1 564
Autres communes du Pas-de-Calais, à revenu faible	28	0,67	1 776
Autres communes du Pas-de-Calais, à revenu moyen	28	0,70	1 617
Autres communes du Pas-de-Calais, à revenu élevé	28	0,72	1 932
Autres communes du Nord, à revenu très faible	25	0,73	1 595
Autres communes du Nord, à revenu faible	27	0,72	1 624
Autres communes du Nord, à revenu moyen	33	0,71	1 705
Autres communes du Nord, à revenu élevé	28	0,78	1 706
Autres communes de Lorraine Ouest (55,88), à revenu faible	28	0,68	1 807
Autres communes de Lorraine Ouest (55,88), à revenu élevé	30	0,64	1 836
Autres communes de Lorraine Est (54,57), à revenu faible	31	0,56	1 813
Autres communes de Lorraine Est (54,57), à revenu moyen	37	0,56	1 913
Autres communes de Lorraine Est (54,57), à revenu élevé	38	0,62	2 098
Autres communes d'Alsace, à revenu faible	24	0,61	1 174
Autres communes d'Alsace, à revenu moyen	32	0,62	1 316
Autres communes d'Alsace, à revenu élevé	25	0,69	1 436
Autres communes de Franche-Comté, à revenu faible	25	0,64	2 097
Autres communes de Franche-Comté, à revenu moyen	27	0,64	2 181
Autres communes de Franche-Comté, à revenu élevé	29	0,64	2 386
Autres communes du Pays de la Loire Nord, à revenu très faible	35	0,72	2 241
Autres communes du Pays de la Loire Nord, à revenu faible	28	0,71	2 275
Autres communes du Pays de la Loire Nord, à revenu moyen	47	0,76	2 502
Autres communes du Pays de la Loire Nord, à revenu élevé	41	0,76	2 427
Autres communes de Loire-Atlantique, à revenu faible	27	0,66	2 202
Autres communes de Loire-Atlantique, à revenu élevé	33	0,75	2 163
Autres communes de Vendée, à revenu faible	24	0,70	1 413
Autres communes de Vendée, à revenu élevé	28	0,71	1 525
Autres communes de Bretagne du Nord, à revenu faible	31	0,68	2 195
Autres communes de Bretagne du Nord, à revenu moyen	43	0,70	2 342
Autres communes de Bretagne du Nord, à revenu élevé	32	0,73	2 243
Autres communes de Bretagne du Sud, à revenu faible	32	0,71	2 364
Autres communes de Bretagne du Sud, à revenu élevé	30	0,69	2 739
Autres communes de Poitou-Charentes Sud, à revenu faible	33	0,67	2 389
Autres communes de Poitou-Charentes Sud, à revenu élevé	41	0,70	2 921
Autres communes de Poitou-Charentes Nord, à revenu faible	20	0,69	1 505
Autres communes de Poitou-Charentes Nord, à revenu moyen	30	0,70	1 654
Autres communes de Poitou-Charentes Nord, à revenu élevé	32	0,71	1 682
Autres communes d'Aquitaine du Nord-est, à revenu faible	20	0,66	2 016
Autres communes d'Aquitaine du Nord-est, à revenu élevé	28	0,67	1 923
Autres communes d'Aquitaine du Sud-ouest, à revenu faible	44	0,60	2 276

Strate	Étapes	R ²	Parc d'estimation
Autres communes d'Aquitaine du Sud-ouest, à revenu élevé	31	0,64	2 448
Autres communes de Gironde, à revenu faible	33	0,69	1 671
Autres communes de Gironde, à revenu élevé	29	0,65	1 971
Autres communes de Midi-Pyrénées du Nord, à revenu faible	14	0,69	1 113
Autres communes de Midi-Pyrénées du Nord, à revenu moyen	16	0,71	1 462
Autres communes de Midi-Pyrénées du Nord, à revenu élevé	23	0,67	1 731
Autres communes de Midi-Pyrénées du Sud, à revenu faible	23	0,71	1 470
Autres communes de Midi-Pyrénées du Sud, à revenu moyen	24	0,66	1 553
Autres communes de Midi-Pyrénées du Sud, à revenu élevé	32	0,71	1 805
Autres communes du Limousin, à revenu faible	22	0,73	1 894
Autres communes du Limousin, à revenu élevé	34	0,71	1 998
Autres communes d'Auvergne, à revenu faible	18	0,65	1 408
Autres communes d'Auvergne, à revenu moyen	25	0,64	1 738
Autres communes d'Auvergne, à revenu élevé	28	0,71	2 067
Autres communes du Languedoc-Roussillon, à revenu très faible	22	0,72	1 724
Autres communes du Languedoc-Roussillon, à revenu faible	16	0,74	1 530
Autres communes du Languedoc-Roussillon, à revenu moyen	23	0,72	1 909
Autres communes du Languedoc-Roussillon, à revenu élevé	29	0,70	1 717
Autres communes de Provence-Alpes-Côte d'Azur du Nord, à revenu faible	19	0,68	1 517
Autres communes de Provence-Alpes-Côte d'Azur du Nord, à revenu élevé	21	0,72	1 621
Autres communes de Provence-Alpes-Côte d'Azur du Sud, à revenu faible	20	0,74	1 354
Autres communes de Provence-Alpes-Côte d'Azur du Sud, à revenu moyen	14	0,76	1 368
Autres communes de Provence-Alpes-Côte d'Azur du Sud, à revenu élevé	19	0,71	1 322
TOTAL			293 410

Chapitre 6 : Désaisonnalisation des indices, suivi et publication

Les principaux indices de prix des logements anciens, analysés sur une période longue et reconnus de bonne qualité, ont reçu l'appellation d'indices « Notaires-Insee ». Ils ont, au préalable, été examinés par un Conseil scientifique, créé dans le cadre des conventions passées entre le notariat et l'Insee⁴⁹. En outre, l'Autorité de la statistique publique (ASP) a labellisé les indices de l'Île-de-France produits par Paris Notaires Services (avis n° 2011-01 de l'Autorité de la statistique publique du 21 juin 2011 sur la labellisation de la statistique trimestrielle du prix du logement en Île-de-France). Chacun de ces indices se décline en un indice brut et un indice corrigé des variations saisonnières.

Des indices supplémentaires couvrant des régions, des départements, des grandes villes ou des unités urbaines pourront recevoir l'appellation indices « Notaires-Insee » dans la mesure où ils seront jugés stables et s'appuieront sur des données suffisamment nombreuses, selon des critères établis par le Conseil scientifique des indices Notaires-Insee.

Depuis novembre 2011, de nouvelles séries (basées sur une actualisation régulière des modèles hédoniques, en base 100 au 1er trimestre 2010) ont remplacé les anciennes.

Les indices font l'objet d'un suivi régulier par le Conseil scientifique des indices Notaires-Insee, via des tableaux de bord.

6.1 Désaisonnalisation des indices

6.1.1 Méthode de calcul des corrections pour variations saisonnières

En version 3, la méthode de désaisonnalisation retenue est la méthode CENSUS X12. La méthode utilisée en version 2 était la méthode CENSUS X11. X12 permet d'améliorer l'estimation des coefficients CVS. En particulier, le calcul de moyennes mobiles a pour conséquence de raccourcir la longueur des séries. Avec X12, il est possible de prolonger les séries à l'aide de modèles Arima pour estimer les trimestres ou mois manquants. La méthode X12 est disponible dans les logiciels SAS et Démétra. C'est le logiciel Démétra, qui a été retenu pour la désaisonnalisation des indices Notaires-Insee.

6.1.2 Prise en compte des corrections pour jours ouvrables

La composition en jours ouvrables des trimestres n'a pas d'impact notable sur la formation des prix. Les indices de prix ne font pas l'objet d'un traitement visant à corriger l'effet des jours ouvrables.

6.1.3 Nature des données désaisonnalisées

Les indices sont à base trimestrielle et publiés chaque trimestre, couvrant un trimestre calendaire, par exemple, le premier trimestre 2013 (janvier 2013, février 2013 et mars 2013). Il est possible également de calculer des indices « trimestriels glissants », à base trimestrielle mais publiés chaque mois (couvrant par exemple la période de trois mois : février 2013, mars 2013 et avril 2013). Ils correspondent pour un mois donné à la moyenne des données du mois de référence et des deux mois précédents⁵⁰. Ces calculs sont effectués par les notaires. La désaisonnalisation a été conçue en tenant compte de cette particularité : elle s'applique aux indices trimestriels glissants.

6.1.4 Fréquence de calcul des coefficients CVS

Les coefficients saisonniers sont révisés une fois par an. Leur mise à jour a lieu en août, une fois disponibles les données définitives du 4^{ème} trimestre de la dernière année écoulée. Une mise à jour annuelle des CVS apparaît préférable à une mise à jour trimestrielle ou mensuelle. Elle permet de limiter les révisions de coefficients, d'autant plus que ces derniers peuvent se révéler fragiles s'ils sont estimés à partir de données non

⁴⁹ Voir la composition du Conseil scientifique des indices Notaires-Insee dans les conventions relatives au partenariat entre l'Insee et le Notariat (annexes 7 et 8).

⁵⁰ Cf. chapitre 3, paragraphe 3.5.2.

définitives. A l'occasion de la révision annuelle des coefficients CVS, les modèles Arima utilisés pour la prolongation des séries brutes sont également actualisés.

Les coefficients CVS utilisés pour le calcul d'indices postérieurs à la période de désaisonnalisation résultent de projections. Ces projections sont faites sur deux ans.

6.1.5 Méthode directe ou indirecte

Dans une première méthode, dite directe, on désaisonnalise chaque série, quel que soit son niveau d'agrégation. La méthode indirecte, au contraire, consiste à désaisonnaliser d'abord les séries élémentaires puis à agréger les séries désaisonnalisées.

D'un point de vue théorique, aucune méthode n'est meilleure que l'autre. Cependant, la méthode indirecte présente l'avantage, adapté au cas des indices Notaires-Insee, d'assurer la cohérence entre les évolutions de niveaux d'agrégation différents. La méthode directe peut, quant à elle, conduire à des résultats contradictoires : il est possible que toutes les composantes CVS d'un agrégat (par exemple les départements) évoluent dans un sens alors que la CVS de l'agrégat (par exemple la région) évolue dans l'autre. La communication serait dans ce cas délicate.

Les désaisonnalisations sont donc ici réalisées de manière indirecte. Pour l'agrégation, les indices CVS élémentaires sont pondérés par les valeurs de parc de référence à la période de base $q(t)$, servant au chaînage des indices sur la période de référence $p(t)$ ⁵¹. Par conséquent, les pondérations retenues sont les mêmes que celles des indices bruts.

6.1.6 Calcul

Pour le calcul des coefficients de désaisonnalisation, on peut choisir un modèle additif ou multiplicatif. Le modèle additif suppose que les composantes de la série sont indépendantes les unes des autres. Ainsi, le niveau des variations saisonnières est indépendant du niveau de la série. Le modèle multiplicatif suppose, quant à lui, que les composantes de la série sont dépendantes les unes des autres. Pour les prix des logements, c'est *a priori* le schéma multiplicatif qui s'impose dans tous les cas parce que le niveau des variations saisonnières croît et décroît avec le niveau de la série.

Dans la procédure X12, on peut aussi définir la longueur de la moyenne mobile. En particulier, si la saisonnalité évolue rapidement, il peut être préférable de réduire la longueur des moyennes mobiles.

Concrètement, deux découpages territoriaux sont utilisés, le premier pour les appartements et le second pour les maisons. Pour chaque type de logement, les désaisonnalisations sont dans un premier temps réalisées au niveau élémentaire. Au-dessus de ce niveau élémentaire, les indices CVS sont obtenus par agrégation.

Découpage pour les appartements :

- départements d'Île-de-France (8 séries) ;
- ensemble des villes-centre de province de plus de 10 000 habitants ;
- ensemble des villes de banlieue de province de plus de 10 000 habitants ;
- ensemble des villes de province de moins de 10 000 habitants et des communes rurales de province ;
- régions Rhône-Alpes, Provence-Alpes-Côte d'Azur et Nord-Pas-de-Calais⁵².

Découpage pour les maisons :

- départements d'Île-de-France à l'exception de Paris (7 séries) ;
- ensemble de la province ;
- régions Rhône-Alpes, Provence-Alpes-Côte d'Azur et Nord-Pas-de-Calais⁵³.

Le nombre de séries à désaisonnaliser s'élève à 28 : 15 pour l'Île-de-France et 13 pour la province.

⁵¹ Cf. chapitre 3, paragraphe 3.5.1.

⁵² Les indices CVS régionaux ne sont pas utilisés pour le calcul des indices CVS relatifs à la France métropolitaine. Par contre, ils le sont pour le calcul des indices CVS des prix régionaux de l'ensemble des logements.

⁵³ Cf. note 52.

Tableau 6.1 - Liste des séries « Notaires-Insee » désaisonnalisées

Appartements de Paris
Appartements de Seine-et-Marne
Maisons de Seine-et-Marne
Appartements des Yvelines
Maisons des Yvelines
Appartements de l'Essonne
Maisons de l'Essonne
Appartements des Hauts-de-Seine
Maisons des Hauts-de-Seine
Appartements de Seine-Saint-Denis
Maisons de Seine-Saint-Denis
Appartements du Val-de-Marne
Maisons du Val-de-Marne
Appartements du Val-d'Oise
Maisons du Val-d'Oise
Appartements des villes-centres des agglomérations de plus de 10 000 habitants
Appartements des banlieues des agglomérations de plus de 10 000 habitants
Appartements des agglomérations de moins de 10 000 et rural
Maisons de province
Appartements de Provence-Alpes-Côte d'Azur
Maisons de Provence-Alpes-Côte d'Azur
Appartements de Marseille
Appartements de Rhône-Alpes
Maisons de Rhône-Alpes
Appartements de Lyon
Appartements du Nord-Pas-de-Calais
Maisons du Nord-Pas-de-Calais
Maisons de Lille

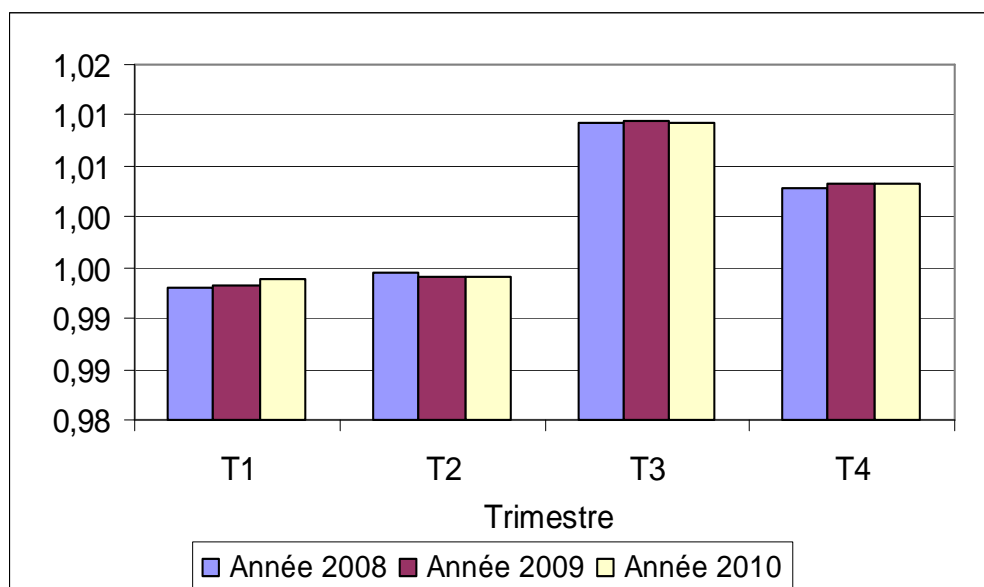
Tableau 6.2 - Correspondance entre les séries élémentaires et les agrégats

Séries agrégées	Séries élémentaires correspondantes
Province	
Appartements des agglomérations de plus de 10 000 habitants	- Appartements villes-centre - Appartements banlieue
Appartements	- Appartements villes-centre - Appartements banlieue - Appartements rural
Appartements et maisons	- Appartements villes-centre - Appartements banlieue - Appartements rural - Maisons
Appartements et maisons : Rhône-Alpes	- Appartements Rhône-Alpes - Maisons Rhône-Alpes
Appartements et maisons : Provence-Alpes-Côte d'Azur	- Appartements Provence-Alpes-Côte d'Azur - Maisons Provence-Alpes-Côte d'Azur
Appartements et maisons : Nord-Pas-de-Calais	- Appartements Nord-Pas-de-Calais - Maisons Nord-Pas-de-Calais
Île-de-France	
Appartements	- Appartements par département
Appartements Île-de-France hors Paris	- Appartements par département
Appartements grande couronne	- Appartements par département (dép. 77, 78, 91 et 95)
Appartements petite couronne	- Appartements par département (dép. 92, 93 et 94)
Maisons	- Maisons par département
Maisons grande couronne	- Maisons par département (dép. 77, 78, 91 et 95)
Maisons petite couronne	- Maisons par département (dép. 92, 93 et 94)
Appartements et maisons	- Appartements par département - Maisons par département
France métropolitaine	
Appartements	- Appartements villes-centre - Appartements banlieue - Appartements rural - Appartements Île-de-France par départements
Maisons	- Maisons Province - Maisons Île-de-France par départements
Appartements et maisons	- Appartements villes-centre - Appartements banlieue - Appartements rural - Maisons Province - Appartements Île-de-France par départements - Maisons Île-de-France par départements

6.1.7 Résultats

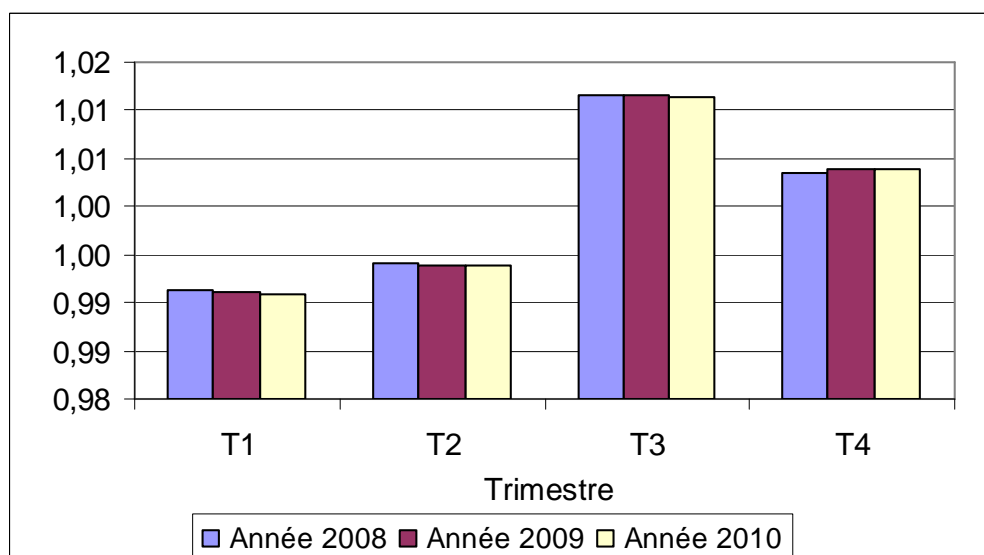
Les prix des appartements comme ceux des maisons connaissent des fluctuations saisonnières. Toutes choses égales par ailleurs, ils sont légèrement plus élevés au troisième trimestre. Cet effet saisonnier est plus marqué pour les maisons que pour les appartements (*figures 6.1 et 6.2*).

Figure 6.1 - Coefficients CVS relatifs aux indices des appartements pour l'ensemble de la France métropolitaine



Lecture : c'est au troisième trimestre que les prix des appartements sont les plus élevés. Par ailleurs, la saisonnalité évolue peu au cours du temps.

Figure 6.2 - Coefficients CVS relatifs aux indices des maisons pour l'ensemble de la France métropolitaine



Lecture : comme pour les appartements, c'est au troisième trimestre que les prix des maisons sont les plus élevés. Les coefficients saisonniers relatifs aux maisons sont également relativement stables au cours du temps.

6.2 Tableau de bord

Chacun des indices trimestriels est disponible en données brutes et en données CVS, en données provisoires avancées, provisoires, semi-définitives et définitives⁵⁴.

6.2.1 Objectifs et composition du tableau de bord

Le tableau de bord sert à la fois à commenter les indices, à déceler d'éventuelles anomalies et finalement à valider la publication. Par souci de rapidité, l'Insee ne valide que les indices ayant l'appellation «Notaires-Insee».

Le principe de la validation repose sur le suivi des bases de données (volumes, taux de couverture), l'analyse des révisions, des évolutions et la comparaison des évolutions de prix moyens et des indices, pour déceler des effets de structure et en trouver les causes.

On analyse les indices provisoires avancés et les indices provisoires du trimestre courant, les indices semi-définitifs et les indices définitifs du trimestre précédent, ainsi que les volumes de transactions correspondants. Les principaux chiffres à vérifier sont ceux qui sont publiés par l'Insee dans la collection « Informations rapides » et diffusés dans la banque de données macro-économiques (BDM).

6.2.2 Exemples de tableaux

Afin de faciliter l'identification de données suspectes, le tableau de bord réunit des informations sur les niveaux d'indices, les évolutions, les révisions, les contributions aux évolutions d'ensemble, les contributions aux révisions d'ensemble, les volumes de transactions et les taux de couverture, sous forme de tableaux de synthèse.

La contribution d'une zone donnée d aux évolutions trimestrielles des agrégats supra-départementaux est fournie par la formule suivante :

$$C_{d,t} = \frac{I_{A,t/0}}{I_{A,t-1/0}} * \left(\frac{\delta_d \hat{W}_{d,p(t),q(t)}}{\sum_{e \in A} \delta_e \hat{W}_{e,p(t),q(t)}} \right) * \frac{I_{t/0}(d) - I_{t-1/0}(d)}{I_{t_0/0}(d)}$$

où :

t représente une date mensuelle postérieure à décembre 2007 correspondant à une année a appartenant à l'ensemble $\{N, N+1\}$ avec N pair,

t_0 la date correspondant au même mois (1er, 2ème ou 3ème mois du trimestre) du dernier trimestre de l'année $N-1$ avec N pair,

$C_{d,t}$: la contribution de la zone d à l'évolution de l'agrégat A entre $t-1$ et t ,

$I_{A,t/0}$: l'indice de l'agrégat A à la date t par rapport au trimestre de référence 0 ,

$I_{A,t_0/0}$: l'indice de l'agrégat A à la date t_0 par rapport au trimestre de référence 0 ,

$I_{t/0}(d)$: l'indice de la zone d (ou d'une partie de la zone d si l'agrégat ne comprend pas toute la zone) à la date t par rapport au trimestre de référence 0 ,

$I_{t_0/0}(d)$: l'indice de la zone d (ou d'une partie de la zone d si l'agrégat ne comprend pas toute la zone) à la date t_0 par rapport au trimestre de référence 0 ,

$\hat{W}_{d,p(t),q(t)}$: la valeur des biens relatifs à la zone d (ou à une partie de la zone d si l'agrégat ne comprend pas toute la zone) échangés au cours de la période $p(t)$ estimée à la date $q(t)$, $p(t)$ correspondant aux années $N-3$ et $N-2$ avec N pair et $q(t)$ étant le dernier trimestre de l'année $N-2$ avec N pair,

⁵⁴ Cf. annexe 11.

δ_d , δ_e les coefficients appliqués pour tenir compte de la non-exhaustivité de la base des notaires au cours de l'année $N-2$ avec N pair.

$\hat{W}_{e,p(t),q(t)}$: la valeur des biens relatifs à la zone e (ou à une partie de la zone e si l'agrégat ne comprend pas toute la zone) échangés au cours de la période $p(t)$ estimée à la date $q(t)$, $p(t)$ correspondant aux années $N-3$ et $N-2$ avec N pair et $q(t)$ étant le dernier trimestre de l'année $N-2$ avec N pair.

Des jeux de tableaux sont dédiés au contrôle des évolutions et au contrôle des révisions, au niveau agrégé et détaillé. Les contributions aux évolutions et aux révisions permettent d'identifier rapidement les séries fines à l'origine d'évolutions ou de révisions marquées de certains agrégats (tableau 6.3).

Les tableaux de synthèse comprennent également les informations suivantes :

- les évolutions trimestrielles ;
- les contributions aux évolutions trimestrielles d'ensemble ;
- les révisions ;
- les contributions aux révisions d'ensemble.

Tableau 6.3 - Synthèse des évolutions relatives aux appartements de province (exemple au 3^{ème} trimestre 2012)

Libellé	Pond. (en %)	Indices CVS		Variations trimestrielles CVS		Glissements annuels CVS	Volumes		Évolutions des volumes		Taux de couverture	
		T-1	T	T-1/T-2	T/T-1		T-1	T	T/T-1	T/T-4	T-1	T
France		114,0	113,8	-0,4%	-0,2%	-0,7%					50,3%	42,0%
Province	100,0%	107,2	106,6	-0,5%	-0,6%	-1,3%	22 936	15 480	-32,5%	-50,2%	44,5%	34,8%
Rural	9,7%	104,1	104,3	-1,6%	0,2%	-2,3%	2 810	1 871	-33,4%	-46,1%		
Centre	59,7%	107,7	107,2	-0,5%	-0,4%	-1,2%	13 547	9 233	-31,8%	-51,1%		
Banlieue	30,6%	107,6	106,3	-0,2%	-1,3%	-1,3%	6 579	4 376	-33,5%	-50,1%		
31 Nord-Pas de Calais	3,3%	111,4	112,0	1,2%	0,5%	2,5%	734	534	-27,2%	-50,6%	45,3%	41,1%
82 Rhône-Alpes	22,4%	110,5	110,0	0,3%	-0,5%	-0,2%	5 253	3 850	-26,7%	-45,3%	51,3%	40,8%
69123 Lyon	5,1%	118,9	118,1	0,6%	-0,6%	1,0%	769	715	-7,0%	-44,2%		
93 PACA	27,4%	105,6	105,1	-1,8%	-0,5%	-2,5%	5 247	3 020	-42,4%	-55,8%	51,2%	34,0%
13055 Marseille	4,3%	105,5	105,4	-1,2%	-0,1%	-3,7%	798	577	-27,7%	-48,7%		

Lecture : L'indice provisoire des prix des appartements relatif à l'ensemble de la France métropolitaine est de 113,8 au 3^{ème} trimestre 2012 (donnée CVS). Il s'établit à un niveau plus faible en province (106,6 points). Entre le 2^{ème} et le 3^{ème} trimestre 2012, les prix sont orientés à la baisse. Le repli est un peu plus marqué en province qu'au niveau national. Sur un an, les prix sont également en baisse (-0,7 % pour l'ensemble de la France métropolitaine).

PACA= Provence-Alpes-Côte d'Azur

Les taux de couverture sont estimés pour la France et les départements (cf. chapitre 4).

Tableau 6.4 - Plus fortes évolutions provisoires en valeur absolue (au niveau de la strate)

strate	Libellé de la strate	Pondé- ration (en %)	Indice T-1	Indice T	variation trimestrielle à la date T *	Volumes T	Contribution à la variation trimestrielle de province à la date T
54701	Banlieue de Nancy	0,4%	101,5	112,7	11,1%	76	0,000404015
97009	Littoral PACA Ouest (dép. 13, 83)	1,0%	102,3	113,2	10,7%	53	0,001073172
97004	Littoral Bretagne Ouest (dép. 22,29)	0,7%	102,8	113,2	10,2%	131	0,000672242
99521	Autres communes des Pays de la Loire, à revenu faible	0,7%	97,4	105,3	8,2%	110	0,00056283
99222	Autres communes de Picardie, à revenu élevé	0,4%	96,8	104,2	7,6%	74	0,000326289
99082	Autres communes du Limousin et d'Auvergne, à revenu élevé	0,6%	101,1	108,2	7,1%	78	0,000395908
99431	Autres communes de Franche-Comté, à revenu faible	0,5%	102,5	109,4	6,7%	91	0,000342599
74601	Banlieue d'Annecy	0,7%	113,6	120,4	6,0%	94	0,000423566
97010	Littoral PACA Est (dép 83,06)	1,8%	102,5	108,7	6,0%	123	0,001052136
651	Agglomération de Genève-Annemasse (partie française)	1,0%	120,2	126,6	5,3%	141	0,000530451
67701	Banlieue de Strasbourg (partie française)	0,7%	108,1	113,2	4,7%	69	0,000305825
33063	BORDEAUX	1,4%	118,0	123,2	4,4%	237	0,000589698
59350	LILLE	1,2%	116,5	121,5	4,3%	154	0,000508159
99911	Autres communes du Sud-Est (Languedoc-Roussillon et PACA), à revenu faible	1,3%	105,3	109,8	4,3%	230	0,000531111
29019	BREST	0,7%	104,4	108,5	3,9%	138	0,000262074
97008	Littoral Languedoc-Roussillon Sud (dép. 34, 30)	0,9%	108,8	113,0	3,9%	189	0,000349547
99933	Autres communes De PACA Nord, à revenu élevé	0,7%	104,4	108,5	3,9%	54	0,000251746
99412	Autres communes de Lorraine, à revenu élevé	0,5%	99,5	103,4	3,9%	125	0,000195157
64445	PAU	0,5%	95,0	98,7	3,8%	67	0,00017471
35238	RENNES	1,2%	112,4	116,7	3,8%	224	0,000461684
...

Lecture : La plus forte variation entre le trimestre T-1 et le trimestre T concerne la banlieue de Nancy (+11,1 %). L'évolution de cette série ne contribue que très faiblement à l'évolution de l'indice d'ensemble (0,04 point de pourcentage).

PACA= Provence-Alpes-Côte d'Azur

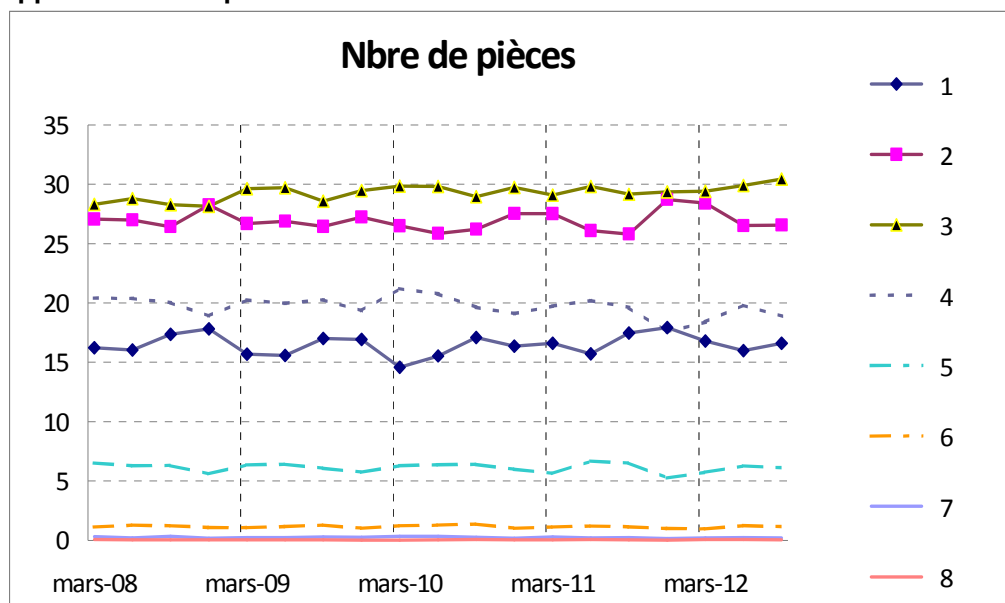
Ces données sont classées par ordre décroissant en valeur absolue, pour faciliter la détection d'éventuelles erreurs. Par exemple, dans la banlieue de Nancy, les prix augmentent de 11,1 % par rapport au trimestre précédent. Compte tenu de l'ampleur de la hausse, une vérification a été faite basée sur l'examen des indices provisoires avancés et des volumes correspondants.

Outre les éléments de synthèse, le tableau de bord contient des informations thématiques sur les indices, leurs évolutions, leurs révisions, les prix moyens, les volumes, les taux de couverture et les pondérations.

Le tableau de bord comprend de plus une page de graphiques permettant de visualiser les indices Notaires-Insee, les indices de prix moyens, les volumes de transactions, les taux de couverture, etc.

Par ailleurs, le tableau de bord permet de suivre l'évolution de la répartition des modalités des variables utilisées pour l'estimation des prix. Il s'agit de repérer le cas échéant des changements ou des erreurs systématiques de codification des modalités (voir par exemple figure 6.3, le cas de la variable « nombre de pièces »).

Figure 6.3 - Évolution de la structure des transactions en nombre de pièces pour les appartements de province



Lecture : la structure des transactions en nombre de pièces apparaît relativement stable dans le temps.

6.3 Publication

Depuis juillet 2004, les indices Notaires-Insee sont publiés quatre fois par an par l'Insee sur le support « Informations Rapides ». Ils sont également disponibles dans la Banque de données macro-économiques (BDM) de l'Insee. Les séries figurent sur le site de l'Insee, <http://www.insee.fr>.

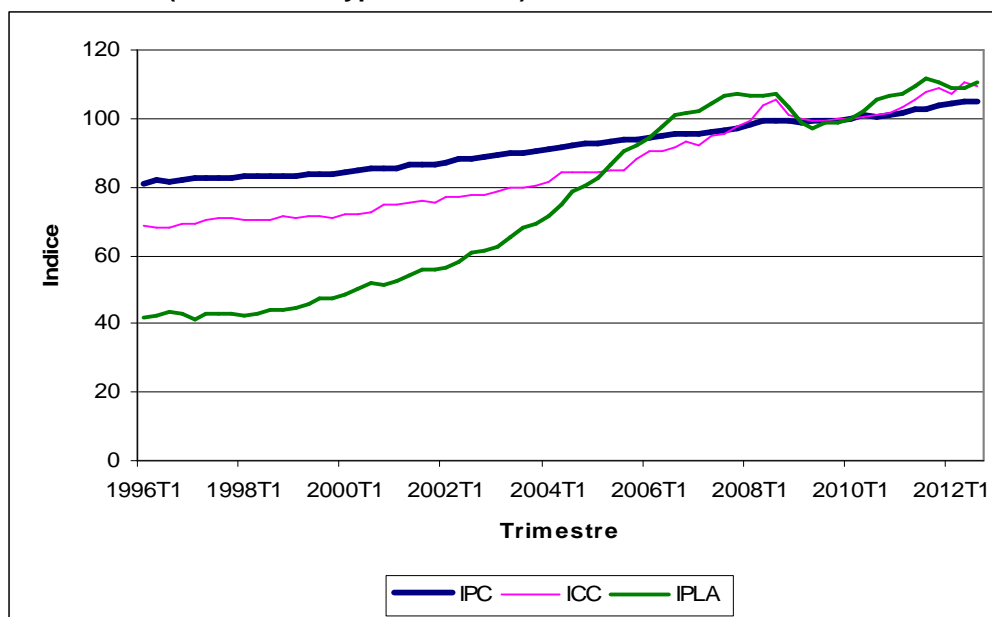
La liste des indices publiés s'est enrichie en 2011 avec la publication des indices du Nord-Pas de Calais, compte tenu d'une bonne couverture régionale des bases notariales constatée sur les années précédentes, dans le cadre d'une procédure probatoire.

Les indices sont également diffusés par les notaires, sur leurs propres sites, <http://www.paris.notaires.fr> pour l'Île-de-France et <http://www.immoprix.com> pour la province.

En règle générale, la première publication des indices d'un trimestre donné intervient environ deux mois après la fin du trimestre. Les critères de diffusion et le type d'indices publiés sont détaillés en annexe 12.

Les graphiques ci-dessous présentent l'évolution des principaux indices Notaires-Insee depuis le 1^{er} trimestre 1996 (date de début des séries relatives à la France métropolitaine) jusqu'au 1^{er} trimestre 2012.

Figure 6.4 - Indices de prix des logements anciens, des prix à la consommation et du coût de la construction (France tous types de biens)



Lecture : Entre 1998 et 2011, les prix des logements anciens ont progressé beaucoup plus vite que le coût de la construction, lequel a lui-même augmenté de façon beaucoup plus marquée que les prix à la consommation.

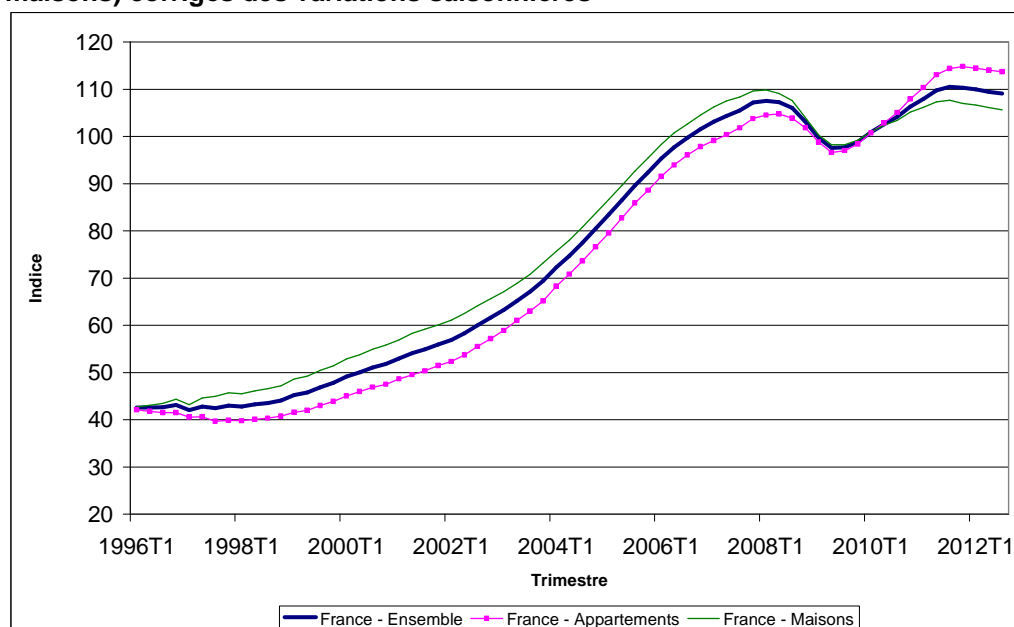
IPLA : indice de prix des logements anciens

IPC : indice des prix à la consommation

ICC : indice du coût de la construction

Base 100=2010T1

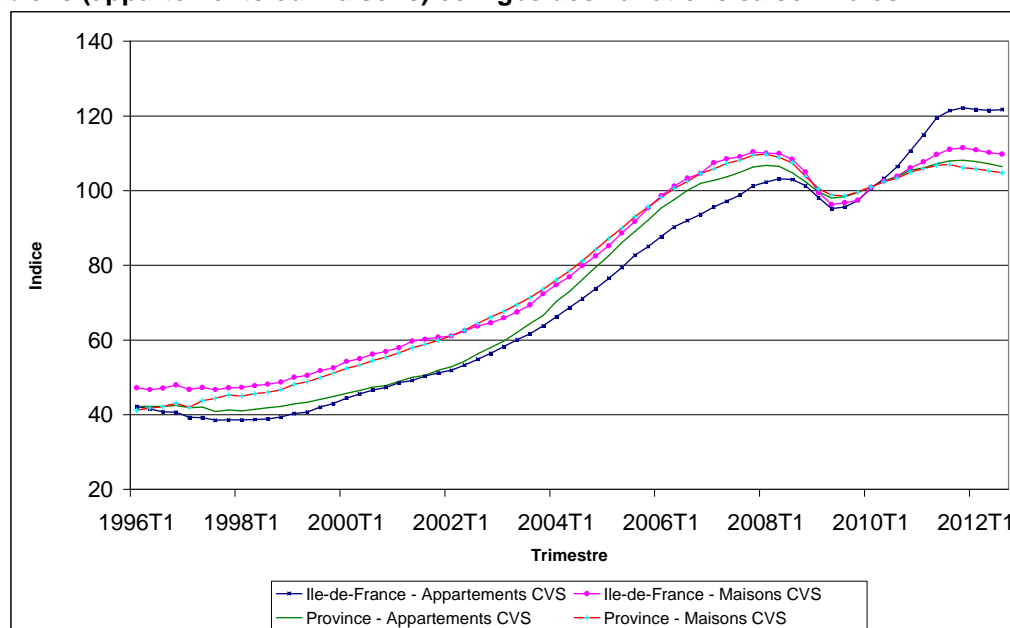
Figure 6.5 - Indices de prix des logements anciens par type de biens (appartements ou maisons) corrigés des variations saisonnières



Lecture : entre 1998 et 2011, les prix des appartements ont progressé plus rapidement que ceux des maisons.

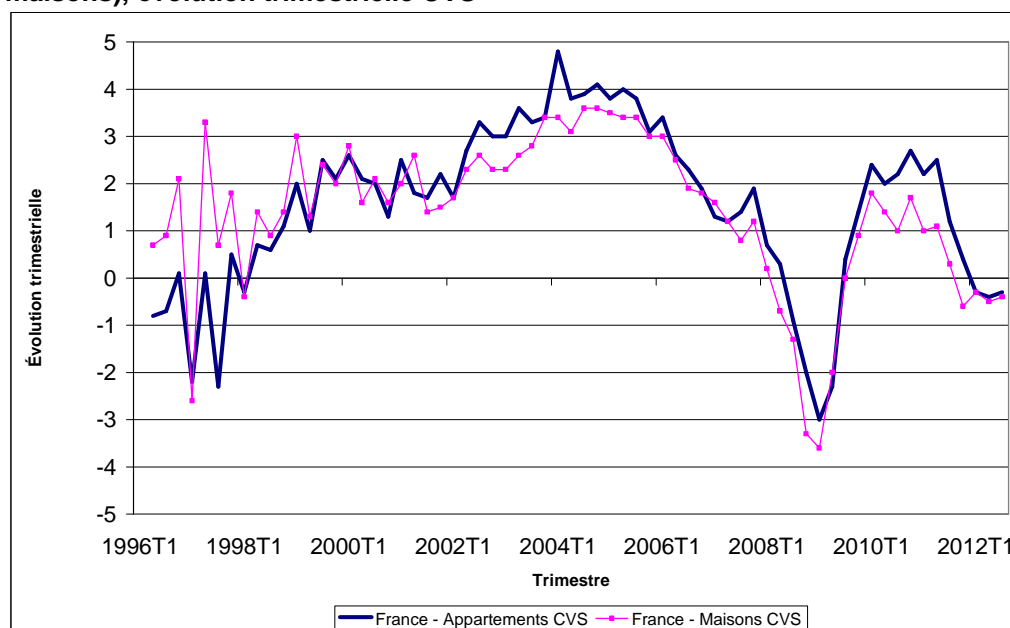
Base 100=2010T1

Figure 6.6 - Indices de prix des logements anciens en Île-de-France et en province par type de biens (appartements ou maisons) corrigés des variations saisonnières



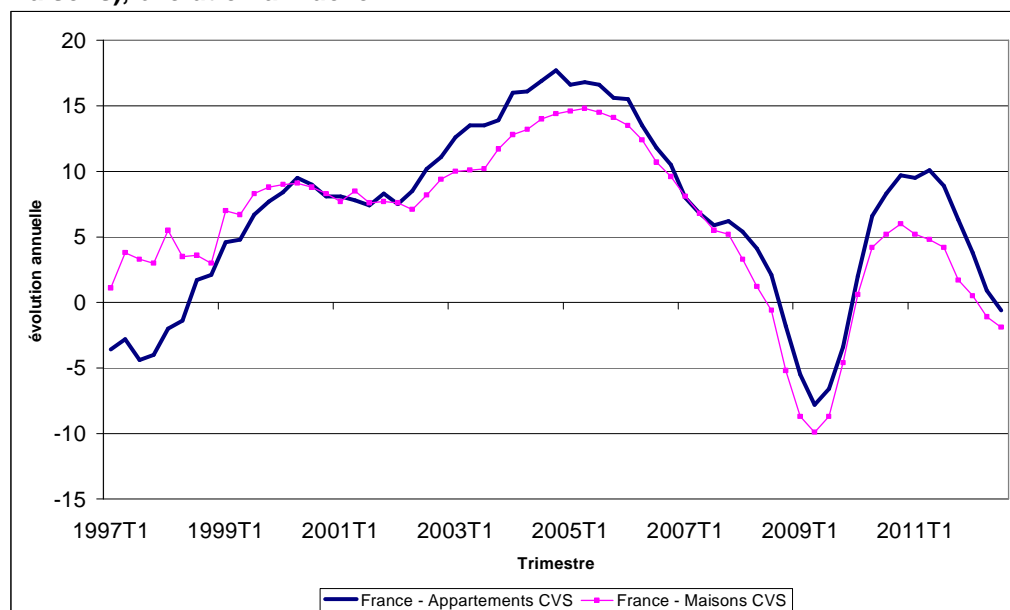
Lecture : depuis 1998, ce sont les prix des appartements franciliens qui ont le plus augmenté.
Base 100=2010T1

Figure 6.7 - Indices de prix des logements anciens par type de biens (appartements ou maisons), évolution trimestrielle CVS



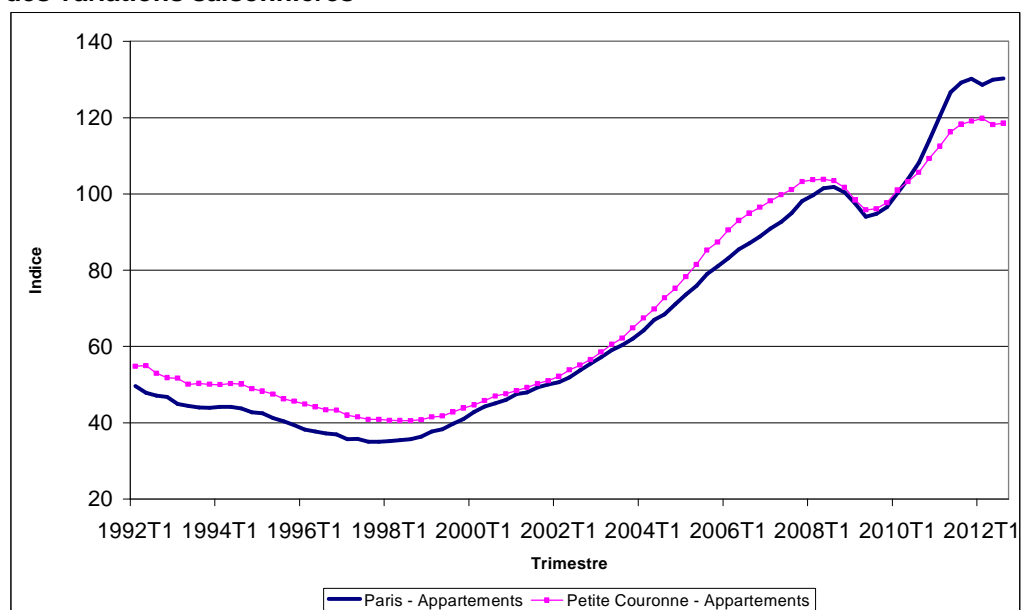
Lecture : depuis 2001, dans la plupart des cas, les évolutions trimestrielles de prix des appartements ont dépassé celles des maisons.
Base 100=2010T1

Figure 6.8 - Indices de prix des logements anciens par type de biens (appartements ou maisons), évolution annuelle



Lecture : depuis 2001, les évolutions annuelles de prix des appartements ont été supérieures à celles des maisons.
Base 100=2010T1

Figure 6.9 - Indices de prix des appartements anciens pour Paris et la Petite couronne, corrigés des variations saisonnières



Lecture : entre 1998 et 2012, les prix des appartements parisiens ont augmenté plus rapidement que ceux des appartements de la Petite couronne. En particulier, entre 2009 et 2011, le rebond a été beaucoup plus marqué à Paris qu'en Petite couronne.
Base 100=2010T1

Chapitre 7 : Comparaison de la V2 avec la V3

Les méthodes de calcul des indices en version 2 sont présentées dans Beauvois et al. (2005). On présente dans ce chapitre une comparaison entre les indices calculés selon la version 2 et ceux de la version 3. Pour cela, on décompose les différents changements intervenus pour en évaluer les impacts respectifs. Ces changements résident d'une part dans la constitution de nouvelles strates (zonage), et d'autre part dans un certain nombre de modifications que l'on regroupera sous le terme de nouvelle méthode de calcul, à savoir : les nouvelles spécifications des modèles, l'actualisation plus fréquente des coefficients correcteurs, le chaînage des indices ainsi que le passage à l'agrégation par moyenne géométrique.

Les indices des appartements sont relativement peu impactés par le nouveau zonage et le changement de méthode. A l'inverse, l'introduction de nouvelles strates en province et d'un chaînage ont un impact significatif sur les indices de prix des maisons.

Pour le recalcul des indices postérieurs à 2007 et pour la révision des indices des maisons antérieurs à 2008, on a utilisé des données plus complètes que lors de la publication initiale. Ceci a pour effet d'accentuer l'évolution de l'indice en période haussière.

La correction d'une erreur dans le codage du nombre de garages des maisons de province a un effet plus important mais heureusement concentré sur 2010 et 2011.

7.1 Les différentes sources d'écart

Pour les appartements, la nouvelle méthode a été appliquée à partir de 2008 et les indices antérieurs à 2008 ont été raccordés avec les nouveaux indices. Concernant les maisons, les indices ont été intégralement recalculés, du fait d'une nouvelle spécification des surfaces habitables et des surfaces des terrains (cf. annexe 3). Les indices CVS ont quant à eux été recalculés, la méthode X11 ayant été remplacée par la méthode X12-Arima.

Il existe quatre sources d'écarts avec les indices de la version 2 :

- la prise en compte de données « super-définitives » lorsque les indices ont été recalculés : des actes continuent à être enregistrés après la publication des indices ; on appelle « super-définitives » les données disponibles aujourd'hui et incluant par conséquent des actes qui n'ont pas participé au calcul de l'indice au moment de sa publication ;
- l'utilisation d'un nouveau zonage : afin de tenir compte de la déformation au cours du temps des prix relatifs des caractéristiques des logements, les contours géographiques des strates ont été entièrement revus. L'utilisation de méthodes hédoniques repose en effet sur l'hypothèse que les évolutions de prix sont homogènes au sein des strates ;
- l'application d'une nouvelle méthode de calcul ; la spécification des modèles économétriques a été affinée. De plus, afin de mieux prendre en compte les évolutions du marché, le choix a été fait d'actualiser plus fréquemment les coefficients servant au calcul des indices. Ce changement a impliqué la mise en place d'un chaînage des indices. De surcroît, la méthode d'agrégation des indices élémentaires a été revue. C'est en effet une méthode géométrique qui a été retenue au niveau départemental. Une nouvelle méthode de désaisonnalisation a également été mise en œuvre.
- la correction d'une erreur : le passage à la version 3 des indices Notaires-Insee coïncide avec la correction d'une erreur concernant les indices des années 2009 à 2011. Cette erreur était due à une rupture dans la codification des données de la province relatives à une variable. Elle est survenue à la suite du déploiement d'un nouveau logiciel de codification.

Après avoir présenté l'écart global entre les versions 2 et 3 (7.2), ce chapitre se concentre sur l'effet de chacune des sources d'écart (7.3).

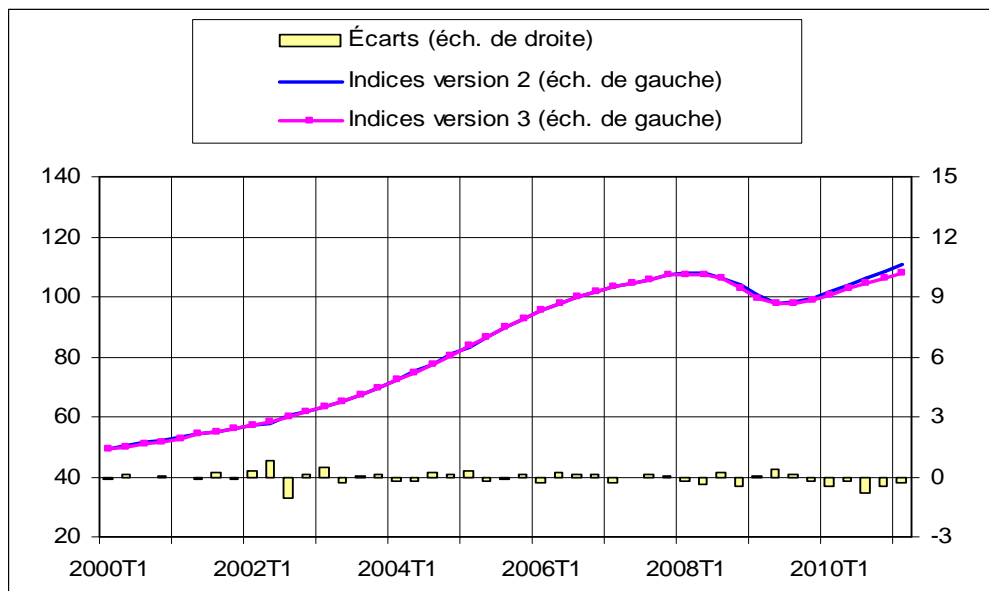
7.2 Écart global entre les versions 2 et 3

Le passage de la version 2 à la version 3 a un impact significatif sur les évolutions trimestrielles de prix relatives aux années 2010 et 2011. L'évolution des prix entre le quatrième trimestre 2009 et le quatrième trimestre 2010

est ainsi révisée à la baisse de 2,0 points. A l'inverse, les évolutions de prix antérieures à 2008 sont faiblement révisées (*figure 7.1*)⁵⁵, si l'on excepte le deuxième et le troisième trimestre 2002.

Avertissement : jusqu'à la fin de ce chapitre, les écarts sont calculés comme la différence (en points de %) entre les évolutions par rapport au trimestre précédent, obtenues avec chacune des versions.

Figure 7.1 - Indices CVS en versions 2 et 3, France métropolitaine, base 100=2010T1



S'agissant des appartements, les révisions sont très faibles sur l'ensemble de la période (*figure 7.2*). Les révisions de prix antérieures à 2008 sont uniquement liées à l'application d'une nouvelle méthode de désaisonnalisation (utilisation de X12-Arima au lieu de X11). A partir de 2008, outre la nouvelle méthode de correction des variations saisonnières, les révisions sont liées à l'utilisation de données super-définitives, à la mise en œuvre du nouveau zonage et à l'application de la nouvelle méthode de calcul (nouvelle modélisation, actualisation régulière des coefficients, moyenne géométrique et chaînage).

Les révisions relatives aux maisons sont faibles pour les années antérieures à 2010 si l'on excepte le deuxième et le troisième trimestre 2002. A l'inverse, elles sont marquées pour les années 2010 et 2011 (*figure 7.3*). Pour ces deux années, outre l'application d'une nouvelle méthode de calcul des indices, d'une nouvelle méthode de correction des variations saisonnières et l'utilisation de données super-définitives, les révisions sont liées à la correction de l'erreur mentionnée plus haut. L'évolution du prix des maisons entre fin 2009 et fin 2010 est ainsi révisée à la baisse de 3,7 points.

⁵⁵ Pour faciliter les comparaisons entre les indices de la version 2 et ceux de la version 3, ce sont des indices version 2 « recalculés » qui sont présentés dans les figures 7.1 à 7.9. Un coefficient multiplicateur a été appliqué à chaque série d'indices de la V2 afin que les indices du 4^{ème} trimestre 2007 soient égaux dans les deux versions. Le choix du 4^{ème} trimestre 2007 permet de mieux appréhender les écarts dus au recalcul des indices en V2 et ceux dus au passage à la V3.

Figure 7.2 - Indices CVS en versions 2 et 3, France métropolitaine, appartements, base 100=2010T1

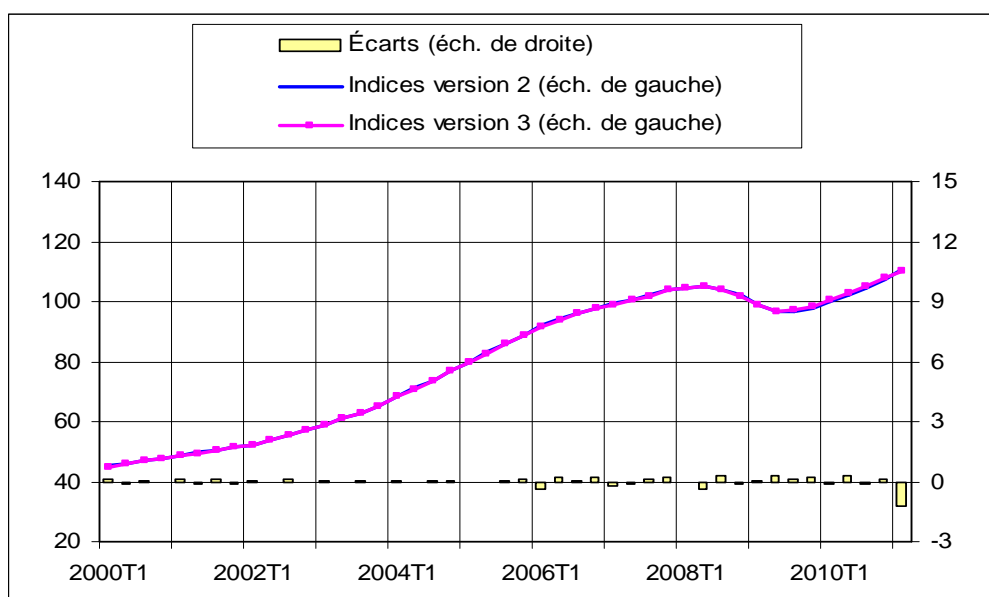
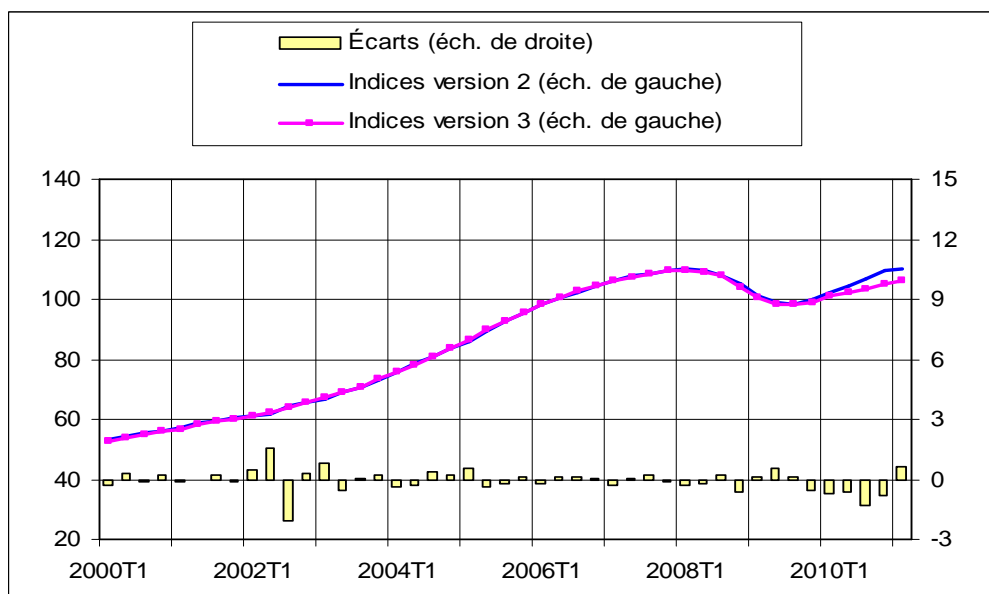


Figure 7.3 - Indices CVS en versions 2 et 3, France métropolitaine, maisons, base 100=2010T1



Les évolutions relatives à l'Île-de-France sont faiblement révisées (*figure 7.4*). Les révisions concernant les appartements sont également faibles (*figure 7.5*). Celles concernant les maisons sont un peu plus marquées au début des années 2000, du fait notamment d'un changement de spécification des modèles. Les surfaces habitables des maisons et les surfaces des terrains sont en effet mieux prises en compte (*figure 7.6*).

Figure 7.4 - Indices CVS en versions 2 et 3, Île-de-France, base 100=2010T1

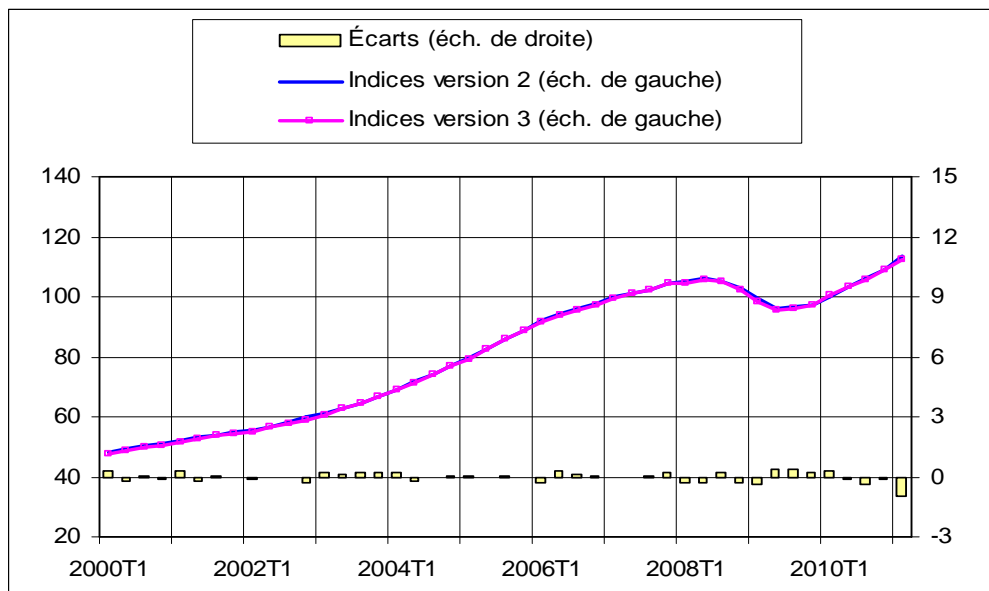


Figure 7.5 - Indices CVS en versions 2 et 3, Île-de-France, appartements, base 100=2010T1

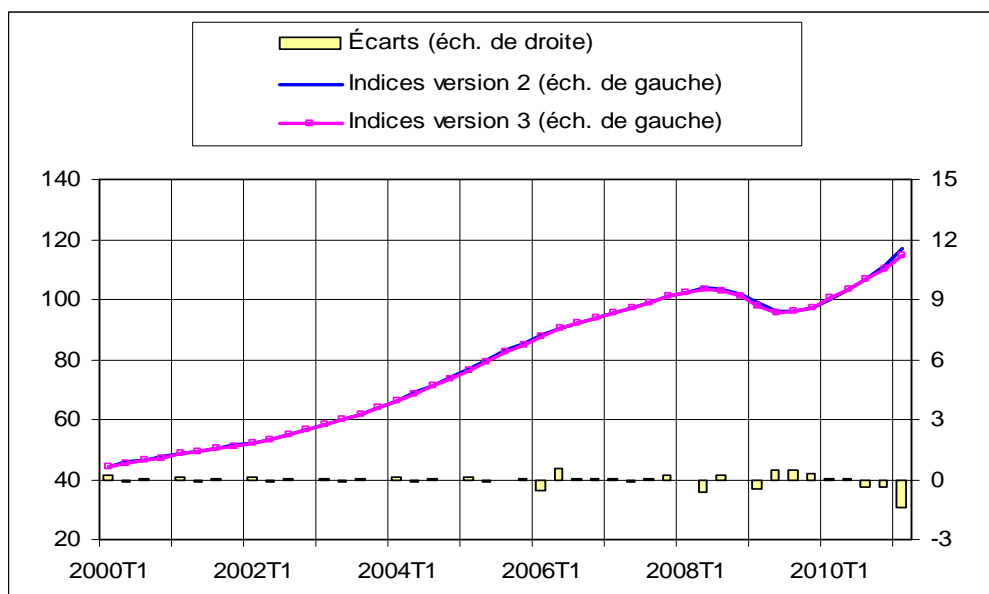
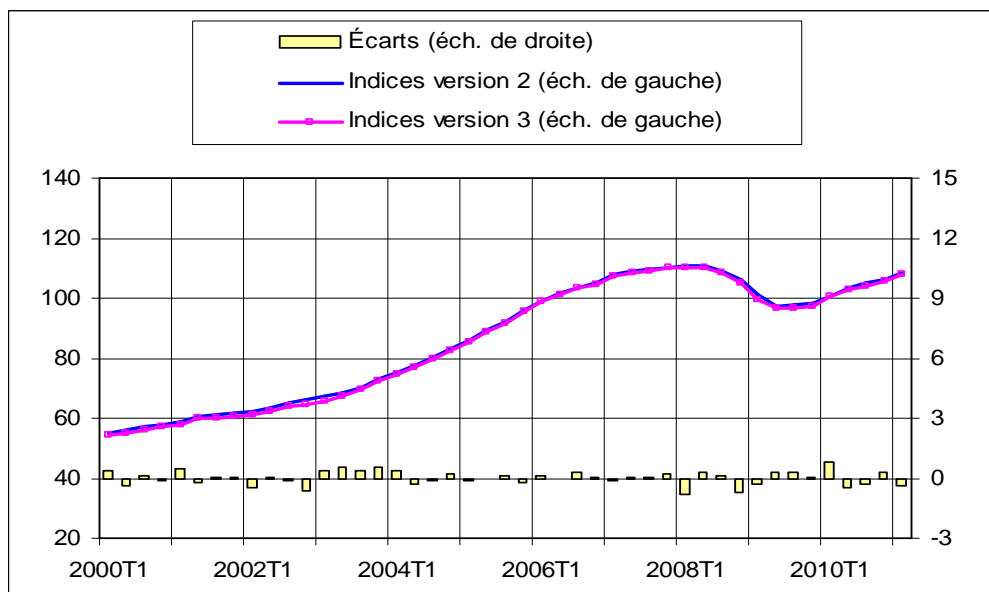


Figure 7.6 - Indices CVS en versions 2 et 3, Île-de-France, maisons, base 100=2010T1



Les évolutions de prix en province sont révisées de façon notable en 2010 et 2011 (*figure 7.7*), du fait de la correction de l'erreur concernant l'indice des maisons (*figure 7.9*). Les révisions concernant les appartements de province sont quant à elles modérées (*figure 7.8*).

Figure 7.7 - Indices CVS en versions 2 et 3, province, base 100=2010T1

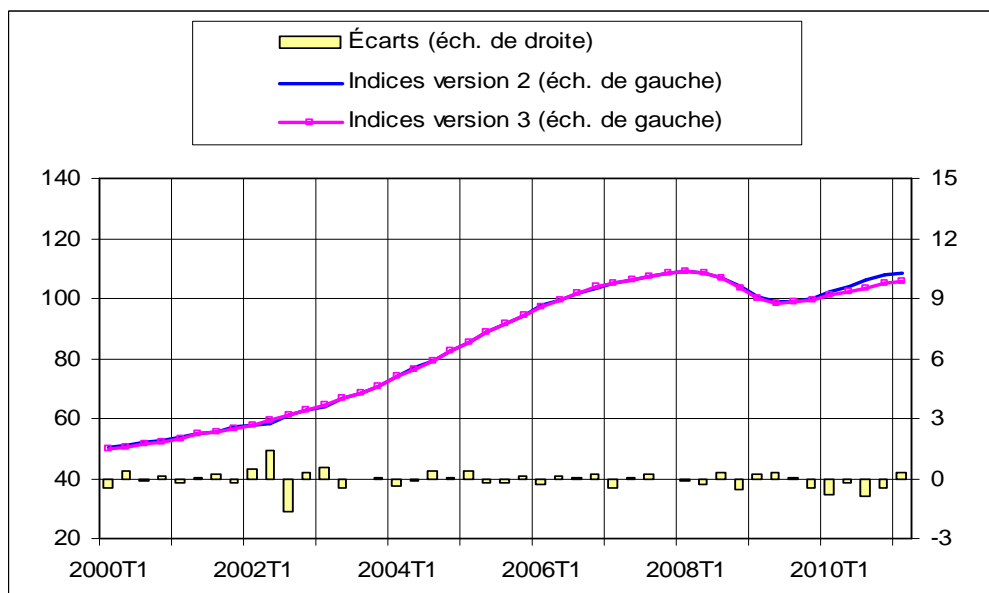


Figure 7.8 - Indices CVS en versions 2 et 3, province, appartements, base 100=2010T1

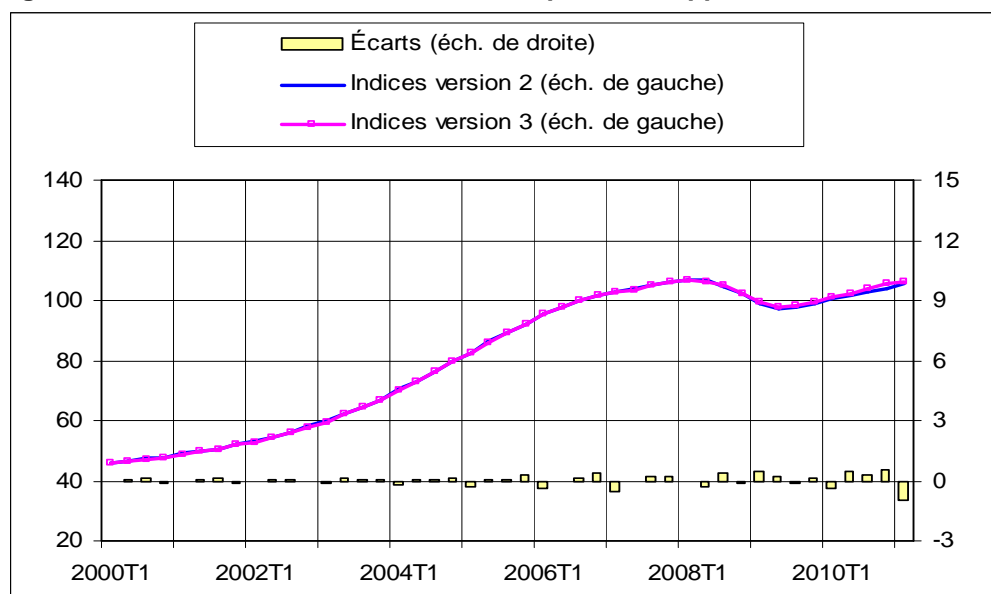
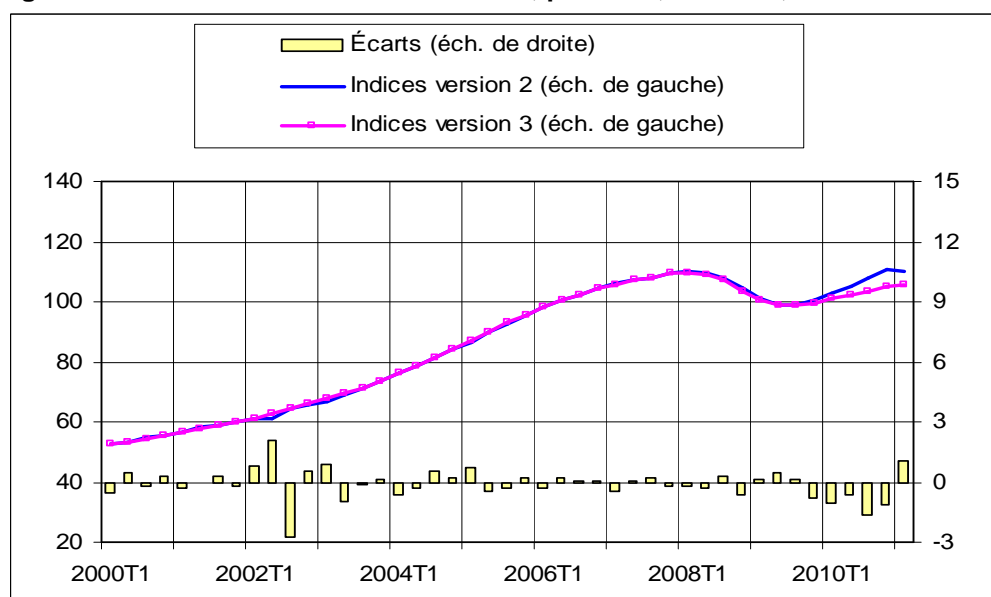


Figure 7.9 - Indices CVS en versions 2 et 3, province, maisons, base 100=2010T1



7.3 Décomposition des écarts entre les versions 2 et 3

7.3.1 Écarts dûs à l'utilisation de données super-définitives

Pour les appartements, les écarts entre indices définitifs et indices super-définitifs ont un impact sur la V3 seulement à partir de 2008. Les indices V3 antérieurs à 2008 reposent en effet sur des données V2 définitives (seules disponibles à l'époque du calcul des indices). À l'inverse, dans le cas des maisons, comme les indices V3 postérieurs à 2007, les indices V3 antérieurs à 2008 ont été recalculés à partir de données super-définitives.

Pour la province, les écarts entre indices définitifs et super-définitifs apparaissent faibles (*figures 7.10 et 7.11*). En niveau, de 2002 à 2009, l'écart moyen est de 0,1 point pour les appartements et de 0,6 point pour les maisons. L'impact haussier du passage à des données super-définitives s'explique par le mouvement de hausse des prix. Les nouveaux actes pris en compte étant principalement des actes signés en fin de trimestre, leurs prix sont plus élevés que les prix moyens trimestriels obtenus à partir de données définitives. Par ailleurs, les écarts de

variations notables au deuxième et au troisième trimestre 2002 concernant les prix des maisons de province sont dus pour une large part à la prise en compte de nouvelles données au deuxième trimestre 2002 (conséquence des aléas de collecte à l'époque : sous-estimation sur un trimestre, rattrapage au trimestre suivant).

Figure 7.10 - Indices bruts définitifs et super-définitifs en version 2, province, appartements, base 100=2000T4

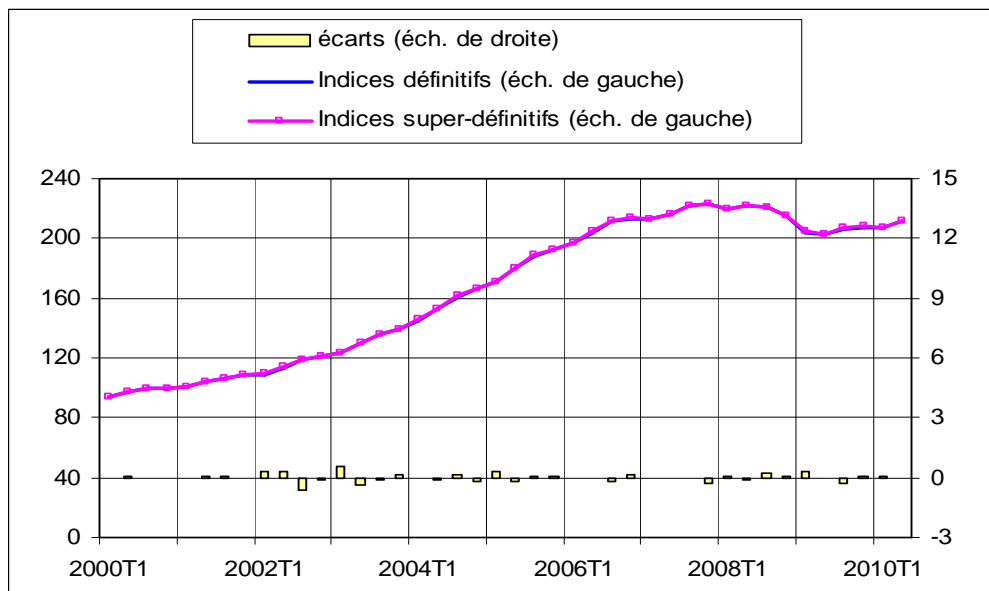
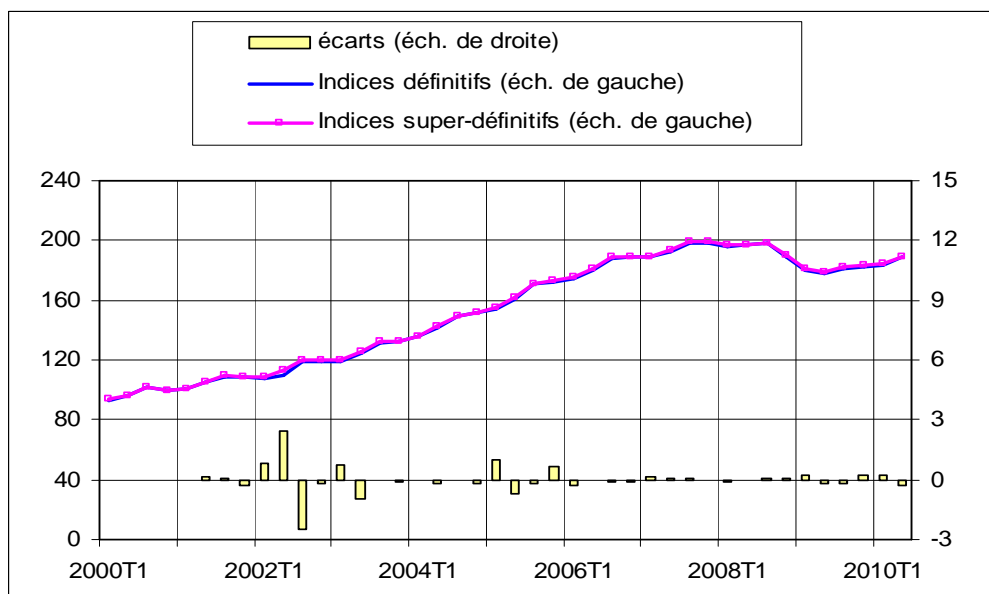


Figure 7.11 - Indices bruts définitifs et super-définitifs en version 2, province, maisons, base 100=2000T4



7.3.2 Impact de l'actualisation des zonages

Pour les appartements comme pour les maisons, l'actualisation des zonages a un impact sur les indices V3 à partir de 2008. Les indices V3 antérieurs à 2008 sont en effet établis à partir de zonages V2.

Pour la province, l'impact de l'actualisation du zonage est un peu plus marqué que l'effet du passage de données définitives à des données super-définitives. En niveau, de 2002 à 2009, l'écart moyen est de 0,3 point pour les

appartements et de 1,4 point pour les maisons (*figures 7.12 et 7.13*). En particulier, dans le cas des maisons de province, le nombre de strates est en augmentation en version 3.

Figure 7.12 - Indices bruts super-définitifs en version 2 avec zonage version 2 et zonage version 3, province, appartements, base 100=2000T4

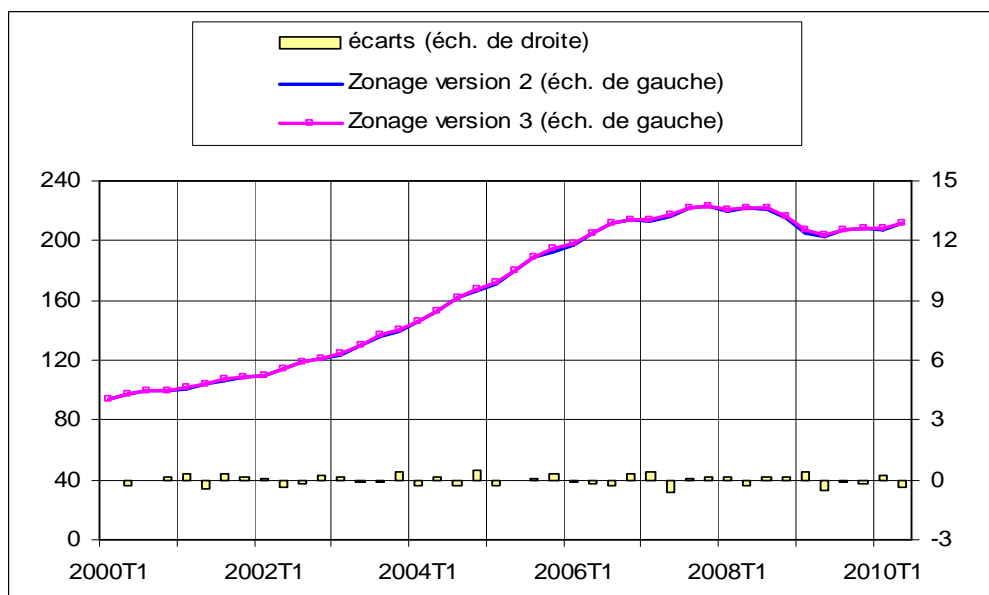
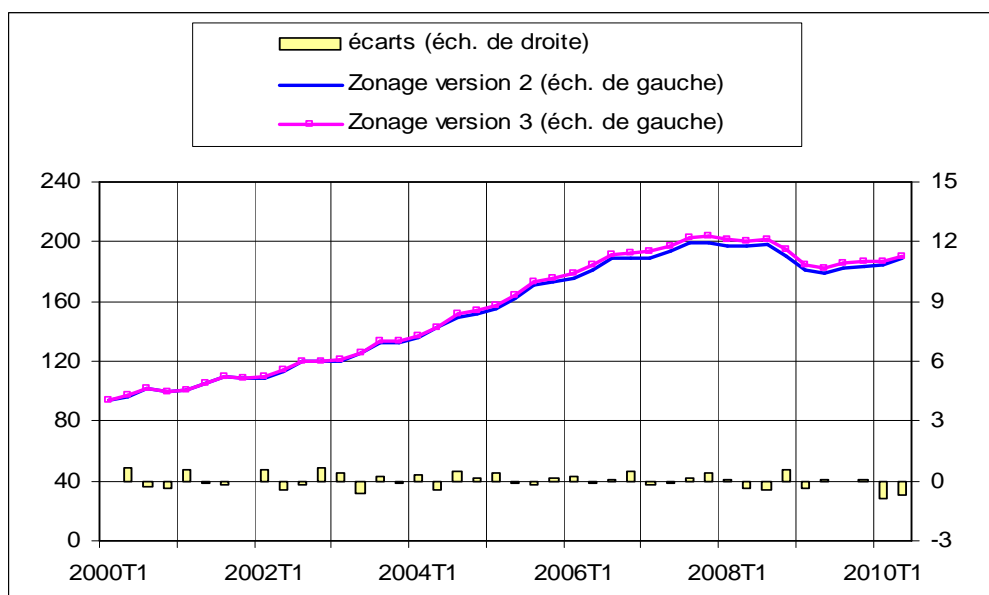


Figure 7.13 - Indices bruts super-définitifs en version 2 avec zonage version 2, province, maisons, base 100=2000T4



7.3.3 Impact de la nouvelle méthode de calcul

La nouvelle méthode de calcul induit des révisions sur les indices de la V3 essentiellement à partir de 2008. Pour les indices antérieurs à 2008, seules les spécifications de la surface habitable des maisons et de la surface des terrains sont modifiées.

Le passage de la moyenne arithmétique à la moyenne géométrique a peu d'impact sur les résultats que ce soit pour les appartements ou les maisons de province (*figures 7.14 et 7.15*). Dans le premier cas, toutes les

agrégations sont faites arithmétiquement. Dans le second cas, les données départementales sont agrégées de façon géométrique et les données supra-départementales de façon arithmétique.

Figure 7.14 - Indices bruts super-définitifs en version 2 avec agrégation arithmétique et agrégation géométrique, province, appartements, base 100=2000T4

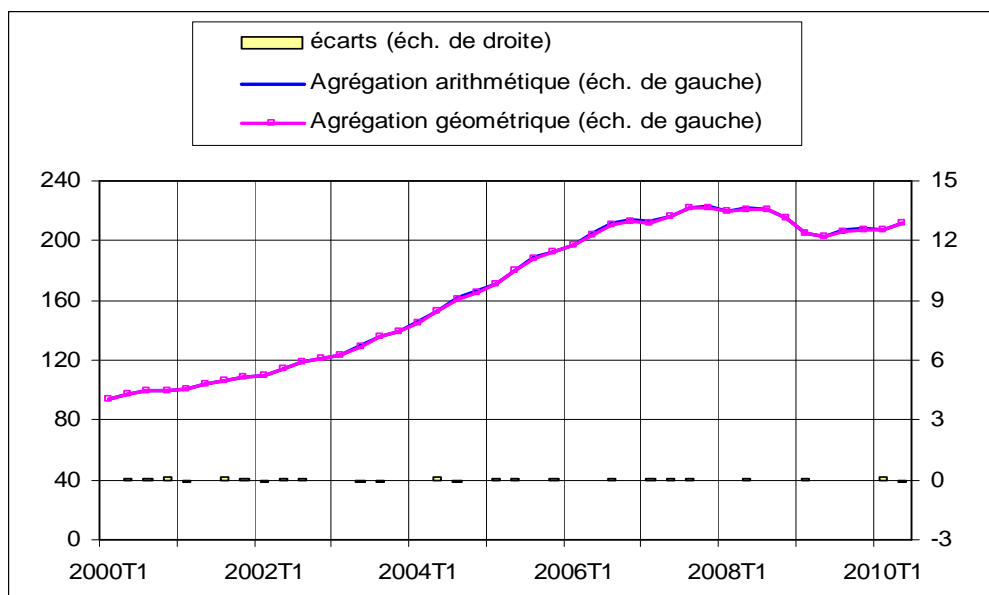
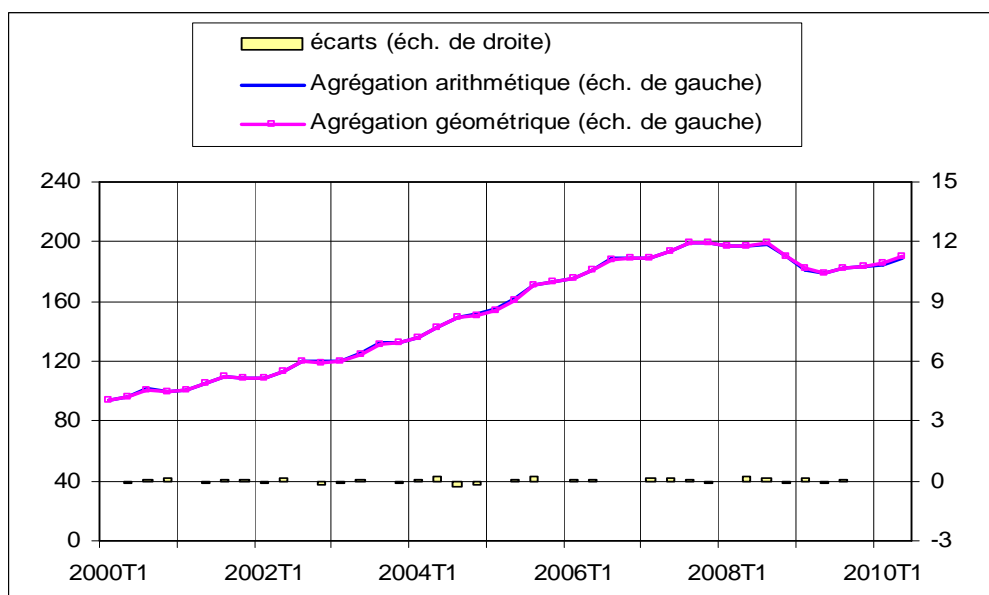


Figure 7.15 - Indices bruts super-définitifs en version 2 avec agrégation arithmétique et agrégation géométrique, province, maisons, base 100=2000T4



L'introduction d'un chaînage, conséquence de l'actualisation plus fréquente des coefficients correcteurs, a quant à elle un effet significatif sur les indices des maisons (*figures 7.16 et 7.17*). En niveau, de 2002 à 2009, l'écart moyen entre les indices chaînés et non chaînés est de -0,2 point pour les appartements et de 1,5 point pour les maisons. Pour le calcul des indices chaînés, les parcs de référence et d'estimation sont actualisés tous les deux ans. Les agrégations d'indices sont quant à elles faites arithmétiquement. Les zonages retenus sont les zonages V2.

Figure 7.16 - Indices bruts super-définitifs en version 2 avec et sans chaînage, province, appartements, base 100=2000T4

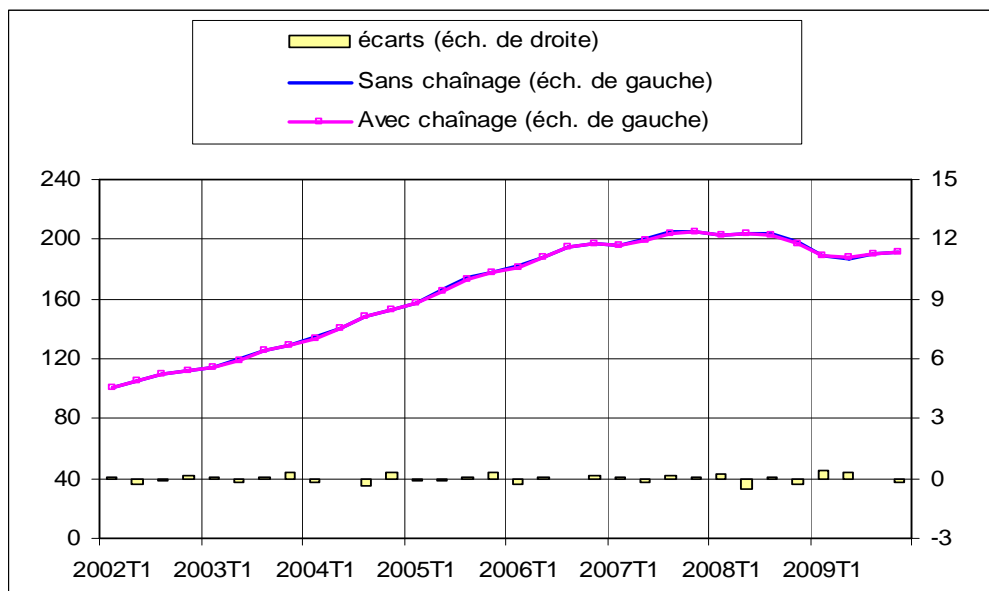
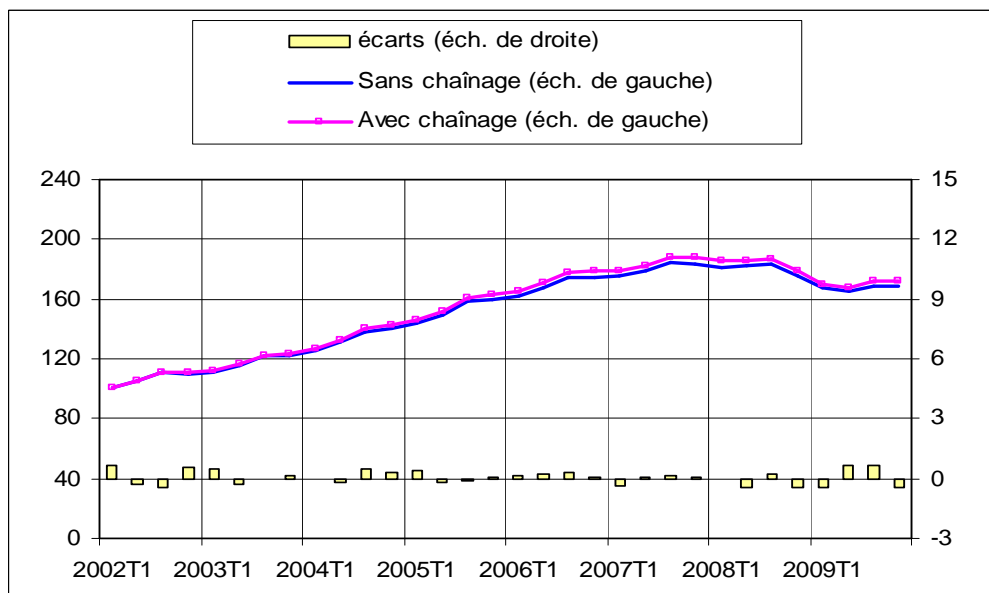


Figure 7.17 - Indices bruts super-définitifs en version 2 avec et sans chaînage, province, maisons, base 100=2000T4

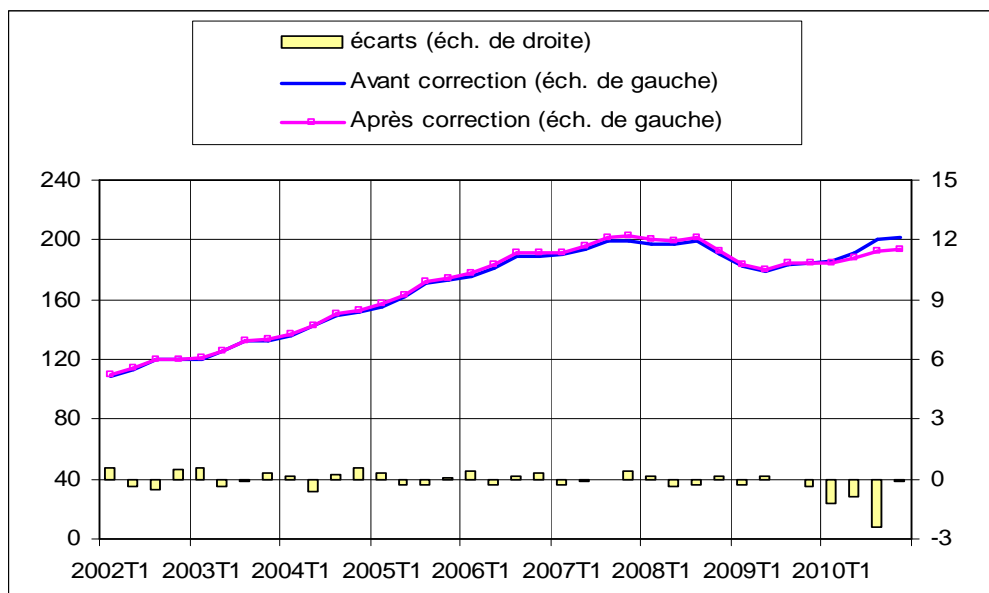


7.3.4 Impact de la correction d'une erreur

Les indices de prix des maisons de province publiés en version 2 étaient entachés d'une erreur à partir de la mi-2009, du fait d'une rupture dans la codification du nombre de garages. Jusqu'à la mi-2009, les données manquantes relatives à cette variable n'étaient pas recodées, ce qui revenait implicitement à les assimiler à la modalité de référence, qui est la modalité « 1 garage ». A la suite du déploiement d'un nouveau logiciel, les

données manquantes relatives à cette variable ont été remplacées par la modalité « pas de garage ». La part de cette dernière a ainsi augmenté progressivement, passant de moins de 5% à la mi-2009 à plus de 35% fin 2010. Après correction de l'erreur, l'évolution des prix entre le quatrième trimestre 2009 et le quatrième trimestre 2010 est ainsi révisée à la baisse de 4 % (*figure 7.18*). En effet, l'erreur d'imputation conduisait à sous-estimer l'évolution de la qualité des maisons.

Figure 7.18 - Indices super-définitifs des prix des maisons anciennes relatifs à la province en version 2 avant et après correction de l'erreur, base 100=2000T4

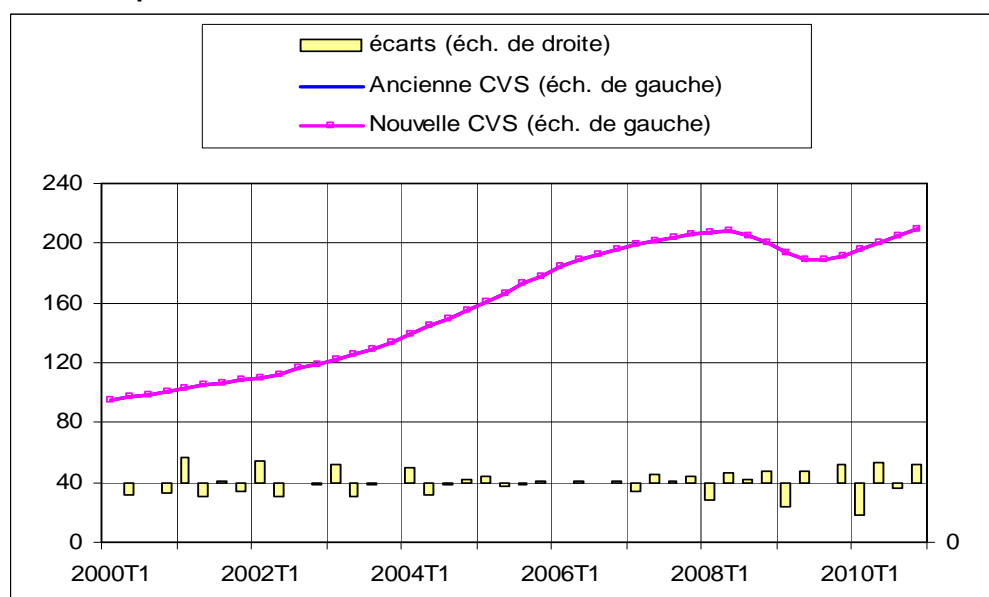


7.3.5 Impact de la nouvelle méthode de correction des variations saisonnières

A l'occasion du passage à la version 3 des indices Notaires-Insee, la méthode de désaisonnalisation des indices a été améliorée. On est passé de la méthode X11 à la méthode X12-Arima. En particulier, désormais, avant désaisonnalisation, les séries brutes sont prolongées par des modèles Arima, ce qui permet d'améliorer l'estimation des coefficients saisonniers.

Le passage à la nouvelle méthode de désaisonnalisation a cependant très peu d'effet sur les indices, les écarts n'excédant pas 0,1 point de pourcentage pour l'ensemble de la France métropolitaine (*figure 7.19*). Les écarts sont plus forts en début et en fin de période, en ligne avec les effets attendus du prolongement des séries par la méthode X12-Arima.

Figure 7.19 - Indices des prix des logements anciens relatifs en France métropolitaine en version 2 après ancienne et nouvelle correction des variations saisonnières⁵⁶



⁵⁶ N. B. : l'échelle de droite du graphique 7.19 n'est pas la même que celle des graphiques précédents.

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 souhaiteront.

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*).

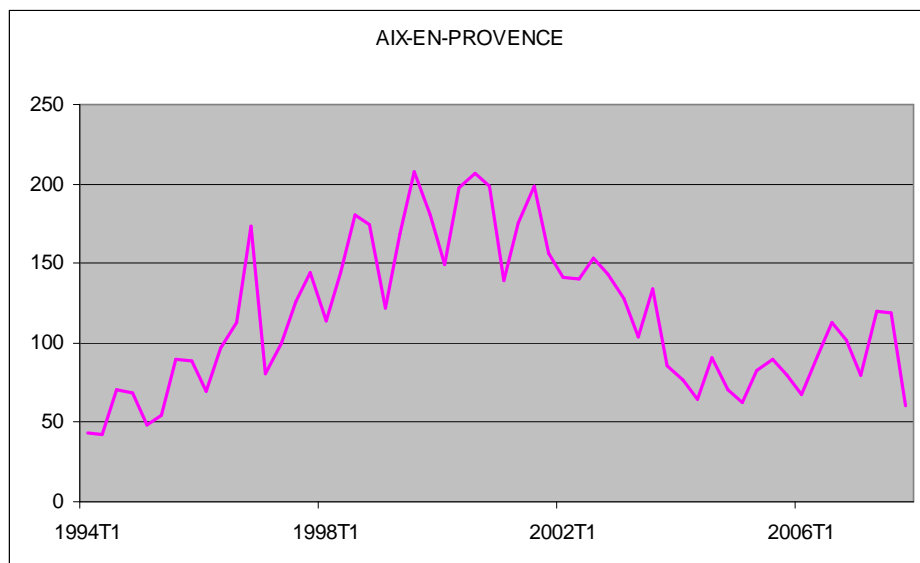
On traite successivement les unités urbaines, les communes littorales et les communes appartenant aux massifs montagneux les plus touristiques, et les communes rurales. Ces dernières n'ont pas un nombre de ventes suffisant par trimestre pour être sélectionnées en tant que telles et n'ont pas de caractéristique 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).

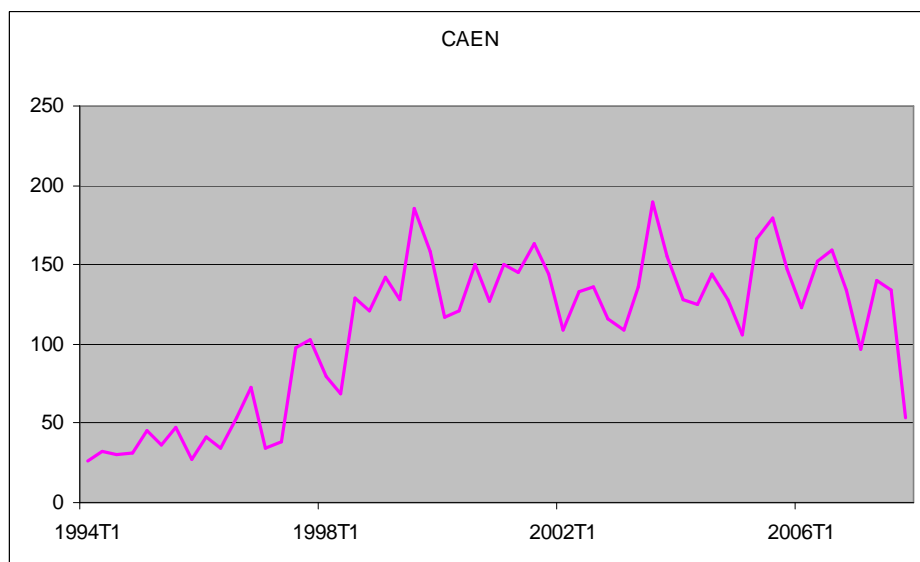
⁶³ Il n'y a pas de strate pour les arrondissements. Les arrondissements de Lyon et Marseille sont traités via des indicatrices dans les régressions.

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)

(maisons)

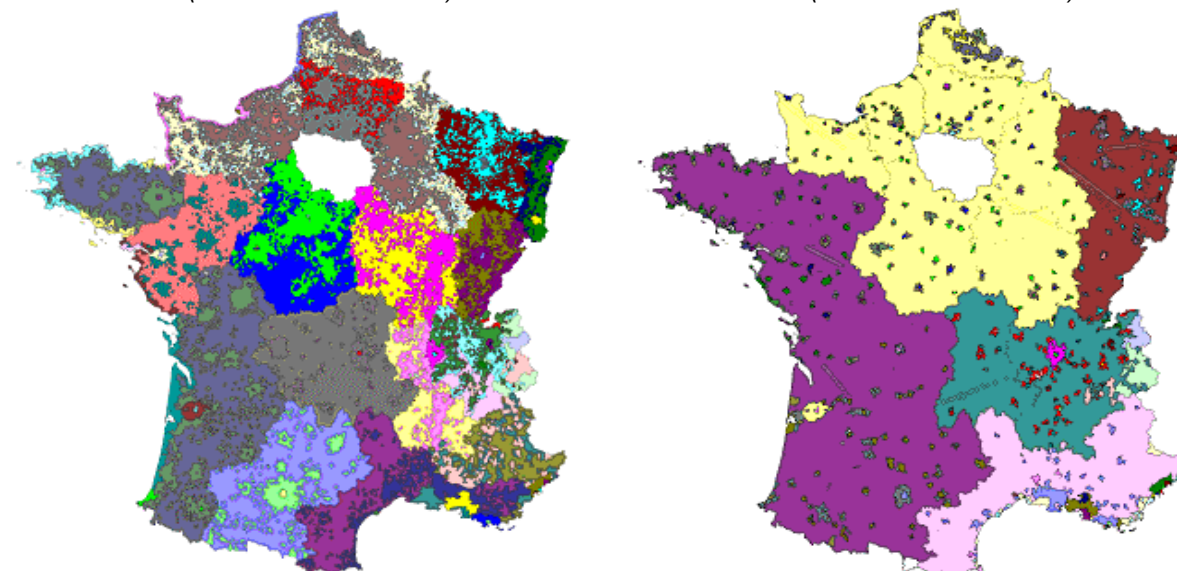
Code de la strate	Département	Région	Borne inférieure de revenu pour la strate (en €)	Borne supérieure de revenu pour la strate (en €)	Nombre total de trimestres où le seuil de 110 logements est dépassé	Nombre total de ventes de 1998 à 2007	
99210	Regroupement des départements Ardennes et Marne	Champagne 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	8 848	
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)



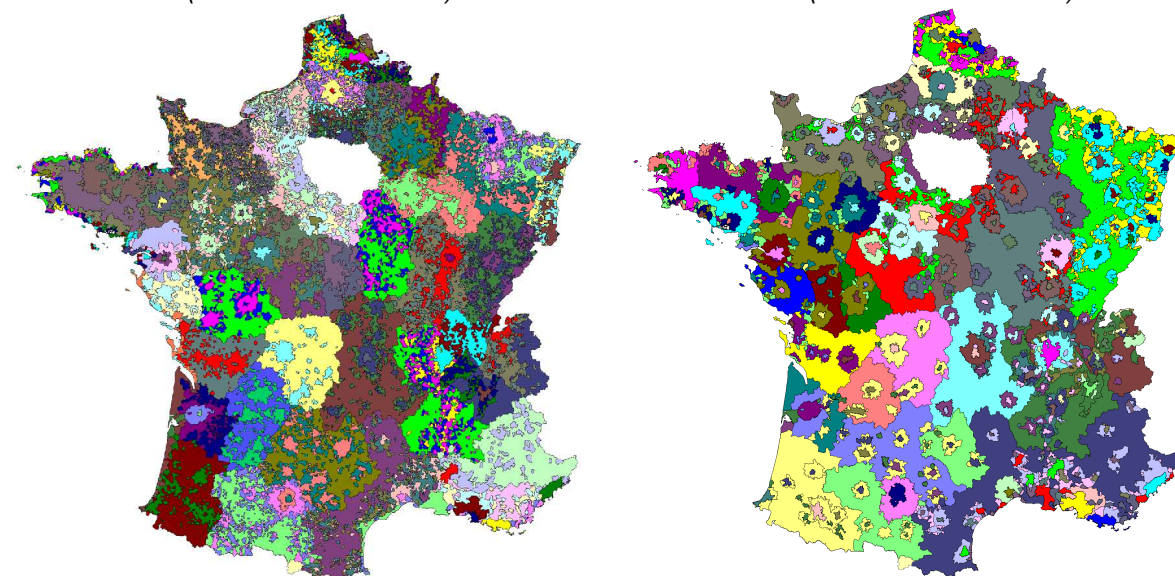
Source : base Perval

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 :

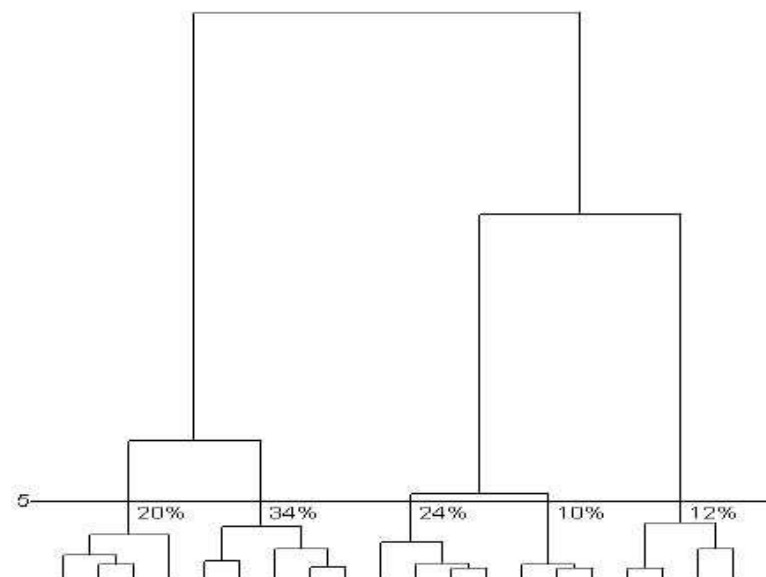
- Prix par m² élevé,
- époque de construction des années 1950 à 1970,
- appartement de grande surface,
- appartement avec 2 salles de bain ou plus,
- appartement avec ascenseur.

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.

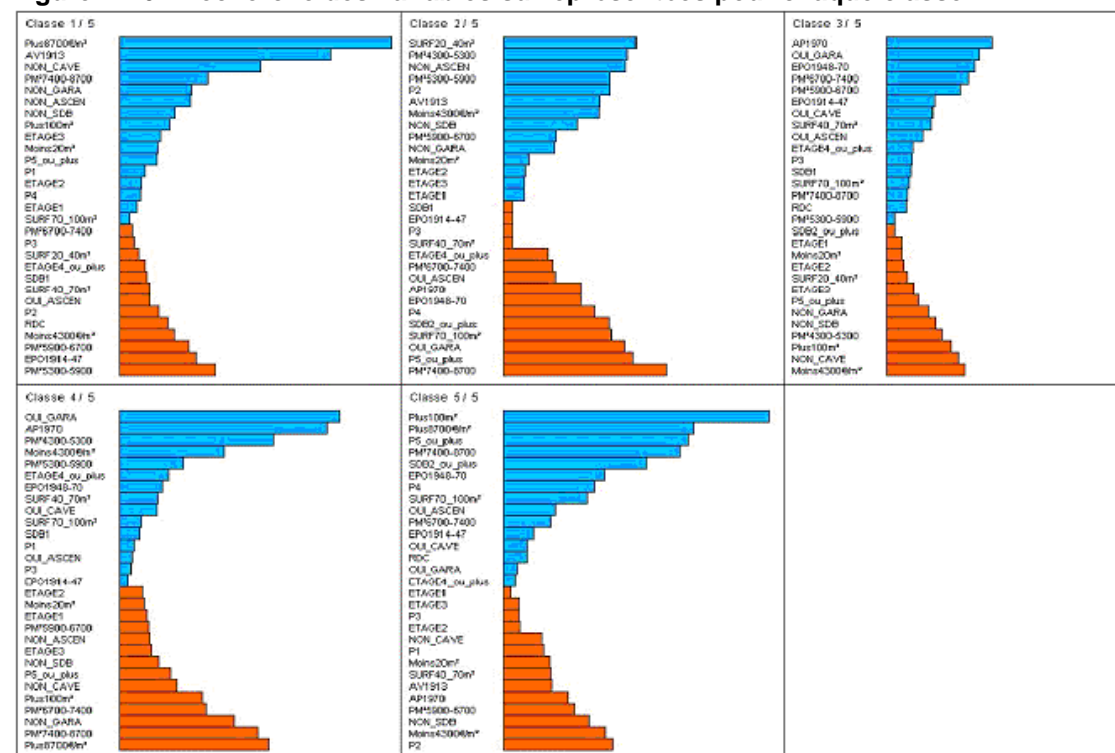
Figure A2.5 - Arbre hiérarchique pour les appartements de Paris



Source : Base BIEN

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

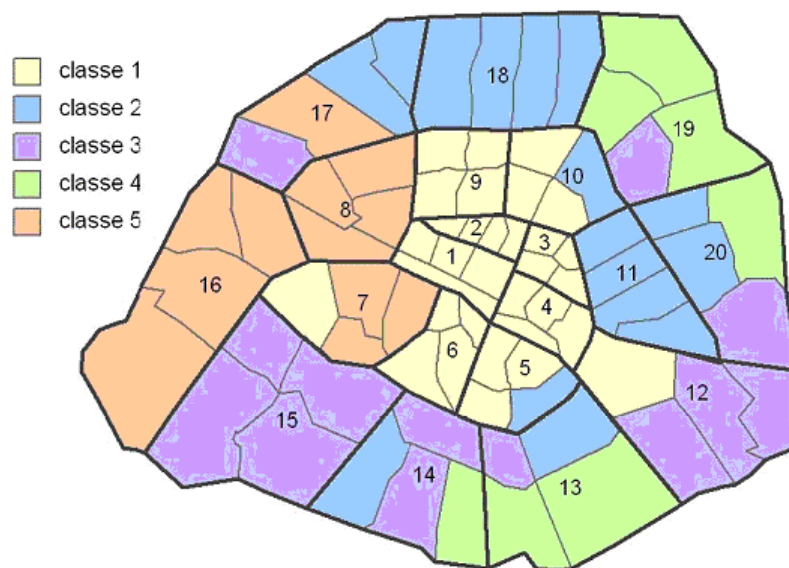
Figure A2.6 - Recherche des variables surreprésentées pour chaque classe



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 1	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	113 257	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

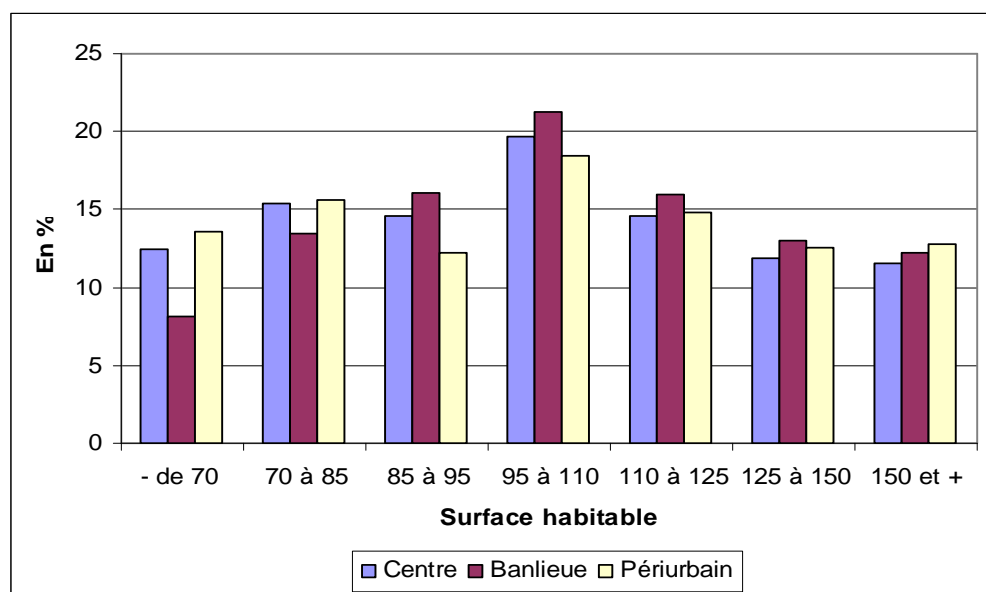
Analyse des variables de surface

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



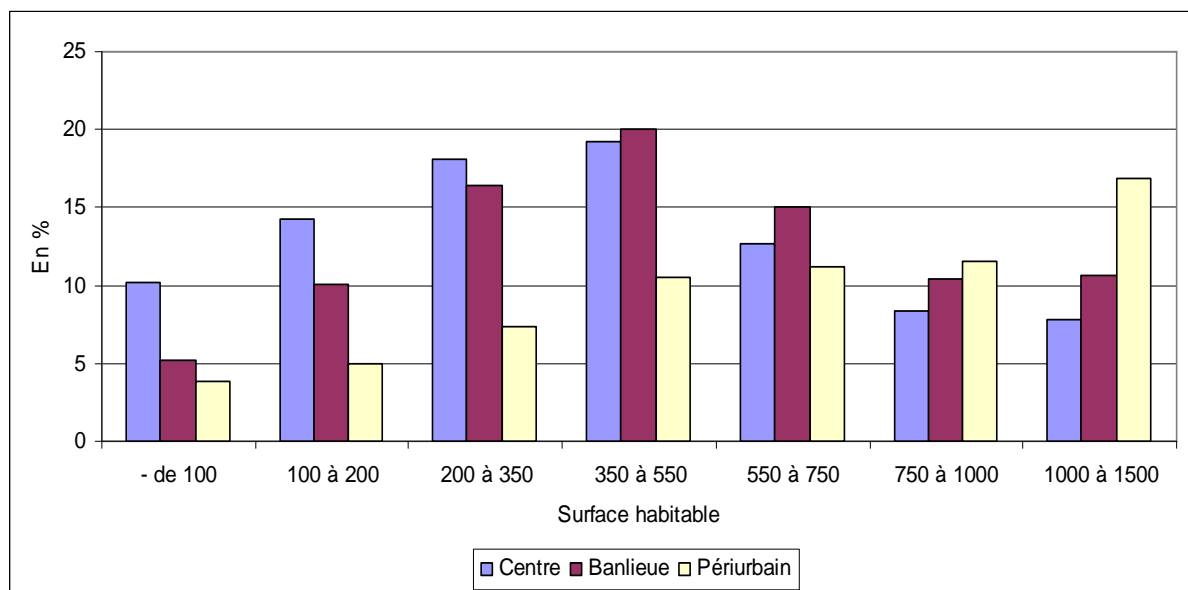
Source : Base Perval, transactions observées entre 2004 et 2008

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



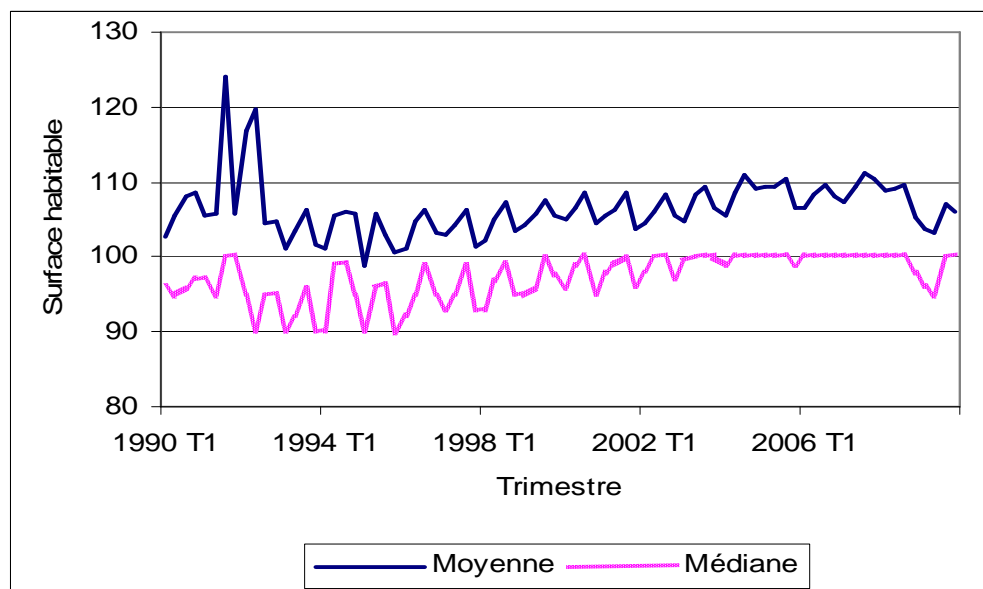
Source : Base Perval, transactions observées entre 2004 et 2008

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).

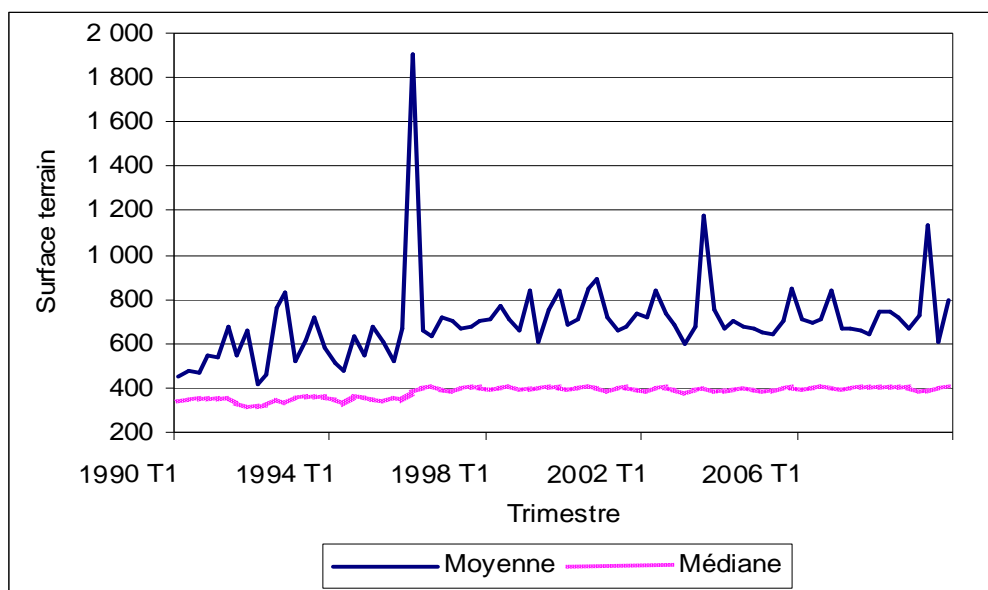
Concernant la surface du terrain, la médiane fluctue peu alors que la moyenne est sujette à une plus grande volatilité (figure A3.5).

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



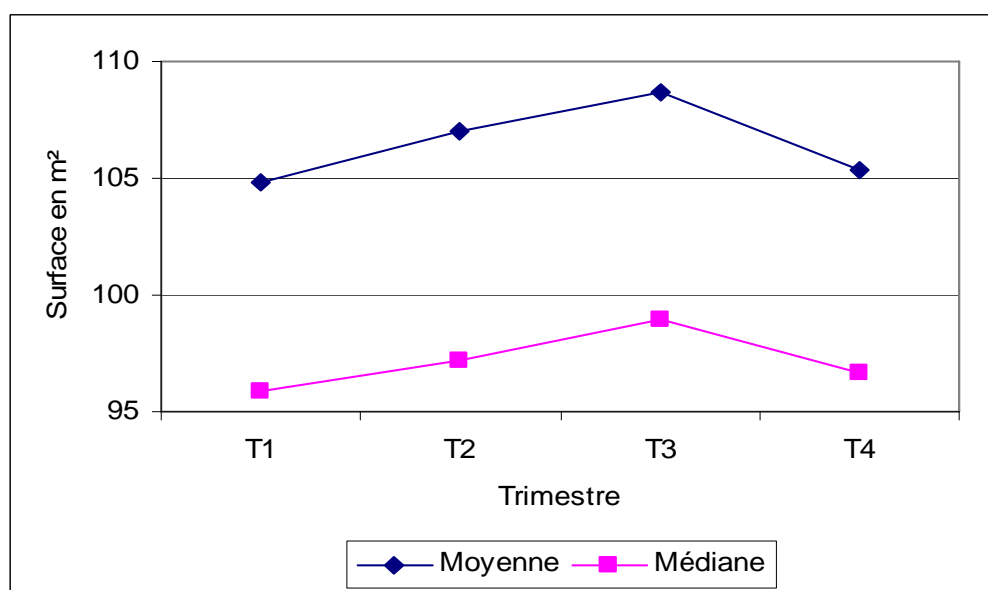
Source : Base BIEN, transactions observées entre le 1^{er} trimestre 1990 et le 4^{ème} trimestre 2009

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



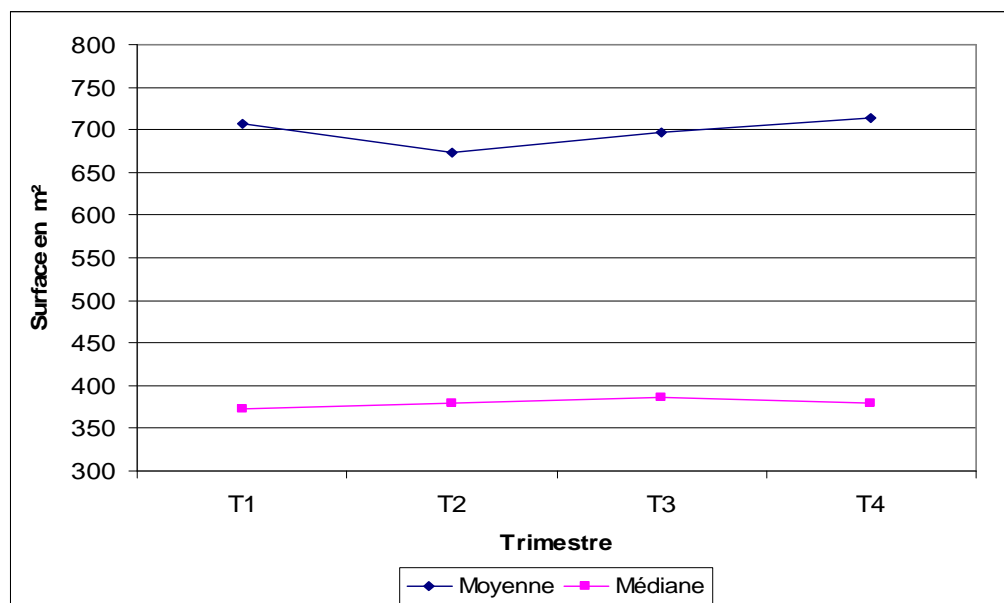
Source : Base BIEN, transactions observées entre le 1^{er} trimestre 1990 et le 4^{ème} trimestre 2009
 Remarque : l'évolution des séries est affectée par l'ajout de la grande couronne à partir de 1996.

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



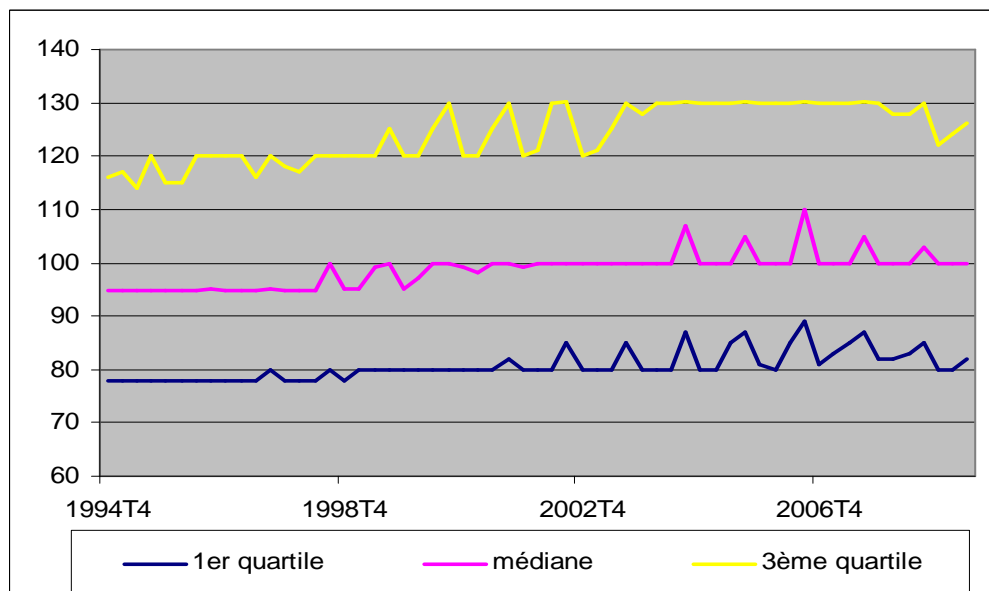
Source : Base BIEN, transactions observées par trimestre (entre le 1^{er} trimestre 1990 et le 4^{ème} trimestre 2009)

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.

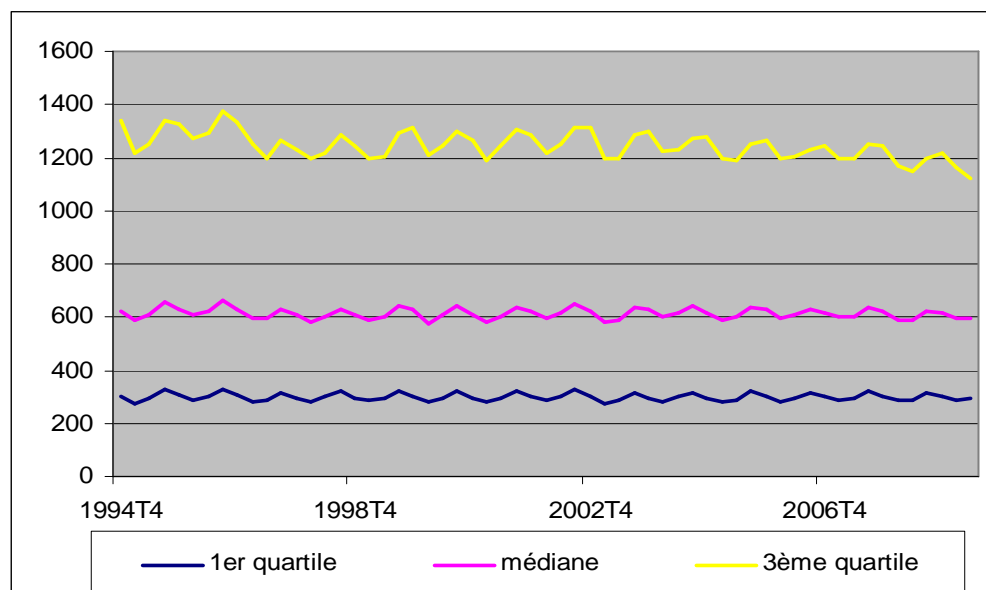
Figure A3.7 - Surface habitable des maisons de province selon trois quantiles



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

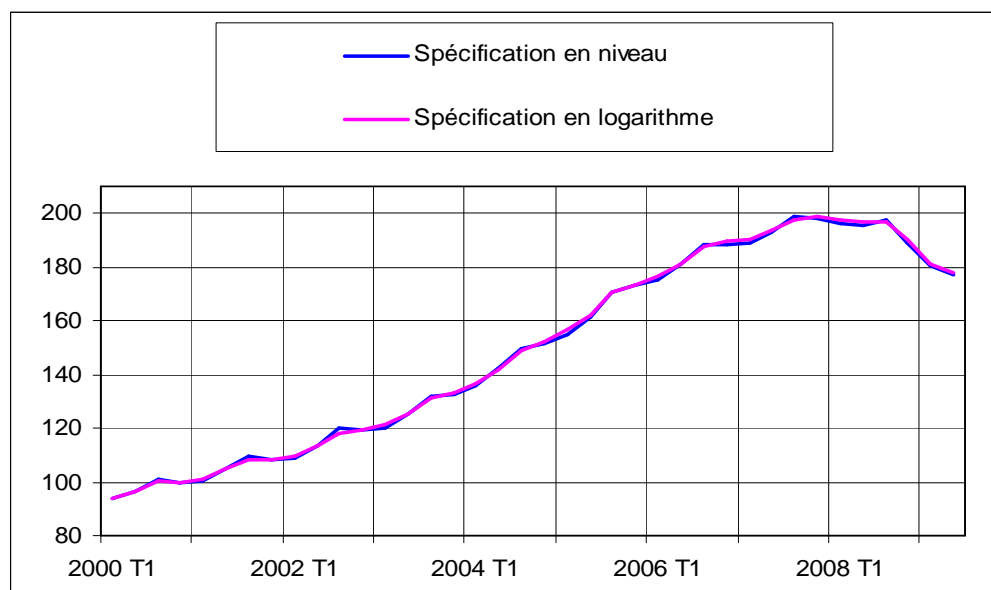
Nous avons également étudié l'impact de la nouvelle méthode sur la détermination des coefficients du vecteur des caractéristiques et sur les séries d'indices de prix. Nous présentons les résultats pour la province.

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.

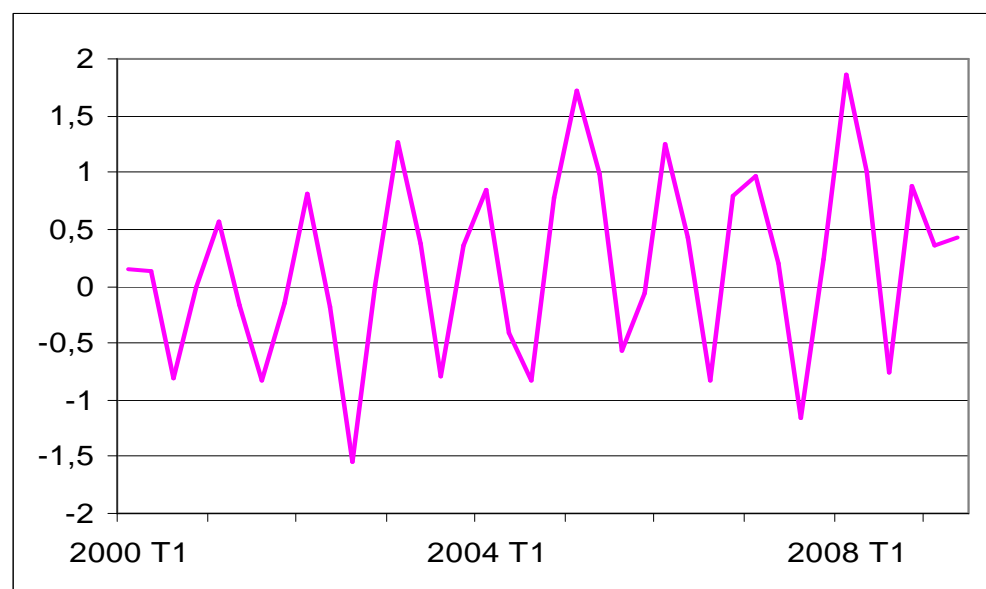
Figure A3.9 - Indice de prix des maisons en province



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 $\bar{\mathcal{E}}_{t,1}$ 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 $\bar{\mathcal{E}}_{1,1}$, $\bar{\mathcal{E}}_{2,1}$, ..., $\bar{\mathcal{E}}_{16,1}$ 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.

⁶⁷ Voir définitions du chapitre 3.

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.

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 détecter 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.

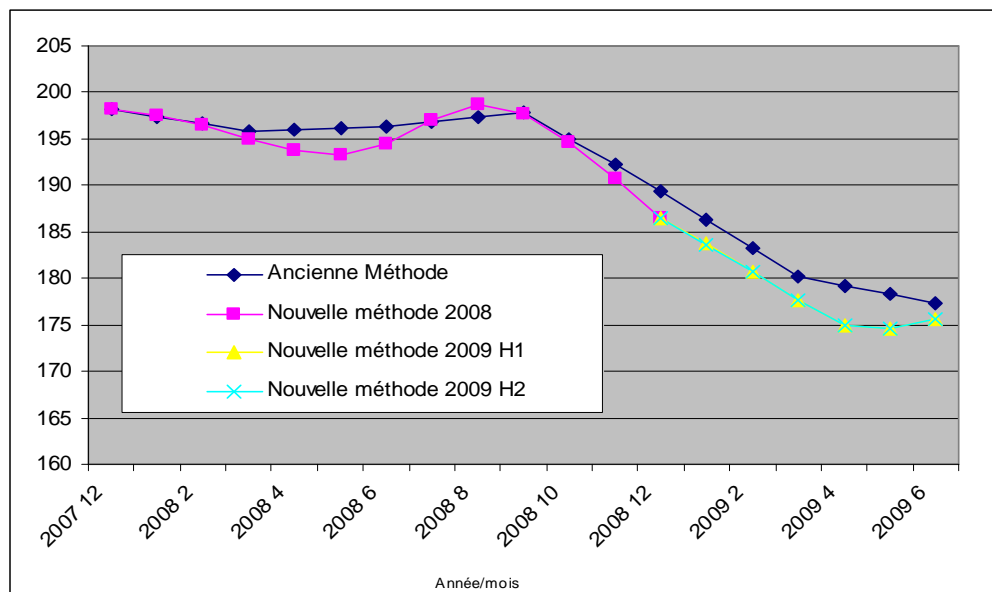
Choix des périodes de référence et de mise à jour du vecteur des prix relatifs des caractéristiques

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

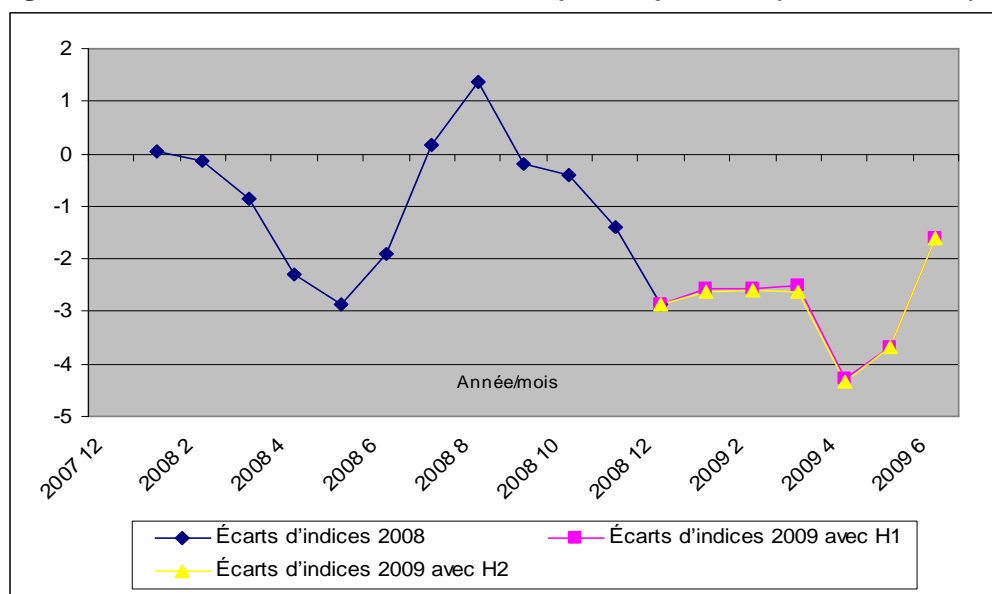


Source : Base Perval

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 $\hat{e}_{i,t}$. 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 :

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

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 recoupent, 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 HAUTBAS, 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'adjoint 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	Pour l'INSEE, le Directeur Général
<i>Signé : Maître Bénasse</i>	<i>Signé : Monsieur Jean-Luc Tavernier</i>
<i>Signé : Maître Hautebas</i>	
<i>Signé : Maître Savoure</i>	
<i>Signé : Maître Lemoine</i>	
<i>Signé : Maître Herrnberger</i>	
<i>Signé : Maître Cauro</i>	

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 » 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 » 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 ».

En outre, les indices de province feront l'objet d'une procédure de labellisation par l'Autorité de la statistique publique.

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'adjoint 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.

Annexe 11 : Actualité des indices

Date de mise à jour : mars 2014

Liste des indices labellisés par l'Autorité de la statistique publique

L'autorité de la statistique publique (ASP) a labellisé les indices de l'Île-de-France produits par Paris Notaires Services (Avis n° 2011-01 de l'Autorité de la statistique publique du 21 juin 2011 sur la labellisation de la statistique trimestrielle du prix du logement en Île-de-France).

Chacun des indices trimestriels est disponible en données brutes et en données CVS et se décline en :

- un indice provisoire (*calculé environ un mois et demi après la fin du trimestre auquel il se rapporte*),
- un indice définitif (*calculé environ quatre mois et demi après la fin du trimestre auquel il se rapporte*).

Liste des indices labellisés :

- Appartements et maisons :

- ensemble de l'Île-de-France,
- ensemble de l'Île-de-France hors Paris,
- ensemble des trois départements de la petite couronne,
- ensemble des quatre départements de la grande couronne.
- chacun des sept départements d'Île-de-France en dehors de Paris,

- Appartements :

- ensemble de l'Île-de-France,
- ensemble de l'Île-de-France hors Paris,
- ensemble des trois départements de la petite couronne,
- ensemble des quatre départements de la grande couronne.
- chacun des huit départements d'Île-de-France,

- Maisons :

- ensemble de l'Île-de-France,
- ensemble de l'Île-de-France hors Paris,
- ensemble des trois départements de la petite couronne,
- ensemble des quatre départements de la grande couronne.
- chacun des sept départements d'Île-de-France en dehors de Paris,

Liste des indices Notaires-Insee

Deux entités différentes diffusent des indices sous l'appellation « indice Notaires - Insee » :

- pour l'Île-de-France : Paris Notaires Services (PNS),
- pour la province et la France métropolitaine : Min.not.

Île-de-France

Les indices diffusés par PNS sont couverts par une convention entre PNS et l'Insee⁷⁹. Ils sont disponibles en données brutes et en données CVS et se déclinent en :

- un indice provisoire (*calculé environ un mois et demi après la fin du trimestre auquel il se rapporte*),
- un indice définitif (*calculé environ quatre mois et demi après la fin du trimestre auquel il se rapporte*).

Liste des indices Notaires-Insee pour l'Île-de-France :

- Appartements et maisons :
 - ensemble de l'Île-de-France,
 - ensemble de l'Île-de-France hors Paris,
 - ensemble des trois départements de la petite couronne,
 - ensemble des quatre départements de la grande couronne.
 - chacun des sept départements d'Île-de-France en dehors de Paris,
- Appartements :
 - ensemble de l'Île-de-France,
 - ensemble de l'Île-de-France hors Paris,
 - ensemble des trois départements de la petite couronne,
 - ensemble des quatre départements de la grande couronne.
 - chacun des huit départements d'Île-de-France,
- Maisons :
 - ensemble de l'Île-de-France,
 - ensemble de l'Île-de-France hors Paris,
 - ensemble des trois départements de la petite couronne,
 - ensemble des quatre départements de la grande couronne.
 - chacun des sept départements d'Île-de-France en dehors de Paris,

Province et France métropolitaine

Min.not diffuse les indices de province et les agrégations nationales. Ces indices sont couverts par une convention entre le Conseil supérieur du notariat, Min.not et l'Insee⁸⁰. Ils sont disponibles en données brutes et en données CVS et se déclinent en :

- un indice provisoire avancé (*calculé environ un mois et demi après la fin du trimestre auquel il se rapporte*⁸¹),
- 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*),
- un indice définitif (*calculé environ six mois après la fin du trimestre auquel il se rapporte*).

⁷⁹ Cf. annexe 8.

⁸⁰ Cf. annexe 9.

⁸¹ Cet indice est diffusé pour l'ensemble de la province et l'ensemble de la France métropolitaine uniquement.

Liste des indices Notaires-Insee pour la province et la France métropolitaine :

- Appartements et maisons :

- métropole,
- province,
- région Nord-Pas-de-Calais,
- région Provence-Alpes-Côte-D'Azur,
- région Rhône-Alpes.

- Appartements :

- métropole,
- province,
- agglomérations de plus de 10 000 habitants en province,
- villes-centres des agglomérations de plus de 10 000 habitants en province,
- agglomérations de plus de 10.000 habitants en province,
- communes rurales et les communes des agglomérations de moins de 10 000 habitants en province,
- région Nord-Pas-de-Calais⁸²,
- région Provence-Alpes-Côte-D'Azur⁸³,
- région Rhône-Alpes⁸⁴,
- commune de Lyon⁸⁵,
- commune de Marseille⁸⁶.

- Maisons :

- métropole,
- province,
- région Nord-Pas-de-Calais,
- région Provence-Alpes-Côte-D'Azur,
- région Rhône-Alpes.
- agglomération de Lille (indices semi-définitif et définitif seulement).

⁸² Indices provisoire, semi-définitif et définitif seulement.

⁸³ Idem.

⁸⁴ Idem.

⁸⁵ Idem.

⁸⁶ Idem.

Tableau A11.1 - Les identifiants des séries dans la Base de données macro-économiques (BDM) sur le site insee.fr

	Identifiant CVS	Identifiant Brut
Commune de Marseille - Appartements	1587626	1587625
Agglomération de Lille - Maisons	1587618	1587617
Commune de Lyon - Appartements	1587634	1587633
Paris - Appartements	1587636	1587635
Seine-et-Marne - Appartements	1587638	1587637
Seine-et-Marne - Maisons	1587640	1587639
Yvelines - Appartements	1587642	1587641
Yvelines - Maisons	1587644	1587643
Essonne - Appartements	1587646	1587645
Essonne - Maisons	1587648	1587647
Hauts-de-Seine - Appartements	1587650	1587649
Hauts-de-Seine - Maisons	1587652	1587651
Seine Saint Denis - Appartements	1587654	1587653
Seine Saint Denis - Maisons	1587656	1587655
Val-de-Marne - Appartements	1587658	1587657
Val-de-Marne - Maisons	1587660	1587659
Val-d'Oise - Appartements	1587662	1587661
Val-d'Oise - Maisons	1587664	1587663
France métropolitaine - Appartements	1587576	1587575
France métropolitaine - Ensemble	1587580	1587579
France métropolitaine - Maisons	1587578	1587577
Province - Aggl. de plus de 10 000 habitants - Appartements	1587588	1587587
Province - Aggl. de plus de 10 000 hab. - Banlieues	1587592	1587591
Province - Aggl. de plus de 10 000 hab. - Villes centres - Appartements	1587590	1587589
Province - Aggl. de moins de 10 000 hab. et rural - Appartements	1587594	1587593
Province - Appartements	1587582	1587581
Province - Ensemble	1587586	1587585
Province - Maisons	1587584	1587583
Île-de-France - Grande Couronne - Appartements	1587608	1587607
Île-de-France - Grande Couronne - Maisons	1587610	1587609
Île-de-France hors Paris - Appartements	1587596	1587602
Île-de-France - Petite Couronne - Appartements	1587604	1587603
Île-de-France - Petite Couronne - Maisons	1587606	1587605
Île-de-France - Appartements	1587597	1587595
Île-de-France - Ensemble	1587601	1587600
Île-de-France - Maisons	1587599	1587598
Nord-Pas-de-Calais - Appartements	1587612	1587611
Nord-Pas-de-Calais - Ensemble	1587616	1587615
Nord-Pas-de-Calais - Maisons	1587614	1587613
Rhône-Alpes - Appartements	1587628	1587627
Rhône-Alpes - Ensemble	1587632	1587631
Rhône-Alpes - Maisons	1587630	1587629
Provence-Alpes-Côte d'Azur - Appartements	1587620	1587619
Provence-Alpes-Côte d'Azur - Ensemble	1587624	1587623
Provence-Alpes-Côte d'Azur - Maisons	1587622	1587621
Île-de-France hors Paris - Ensemble	1634730	1634729
Île-de-France - Petite Couronne - Ensemble	1634732	1634731
Île-de-France - Grande Couronne - Ensemble	1634734	1634733
Île-de-France - Seine et Marne - Ensemble	1634736	1634735
Île-de-France - Yvelines - Ensemble	1634738	1634737
Île-de-France - Essonne - Ensemble	1634740	1634739
Île-de-France - Hauts de Seine - Ensemble	1634742	1634741
Île-de-France - Seine Saint Denis - Ensemble	1634744	1634743
Île-de-France - Val de Marne - Ensemble	1634746	1634745
Île-de-France - Val d'Oise - Ensemble	1634748	1634747

Aggl. = agglomérations ; hab.=habitants

Critères de diffusion des indices trimestriels Notaires-Insee

Pour bénéficier de l'appellation « indice Notaires-Insee », un indice doit respecter un certain nombre de critères surveillés par le Conseil scientifique des indices Notaires-Insee : les taux de couverture, l'ampleur des révisions, les volumes d'actes pris en compte, les délais d'intégration des actes dans les bases.

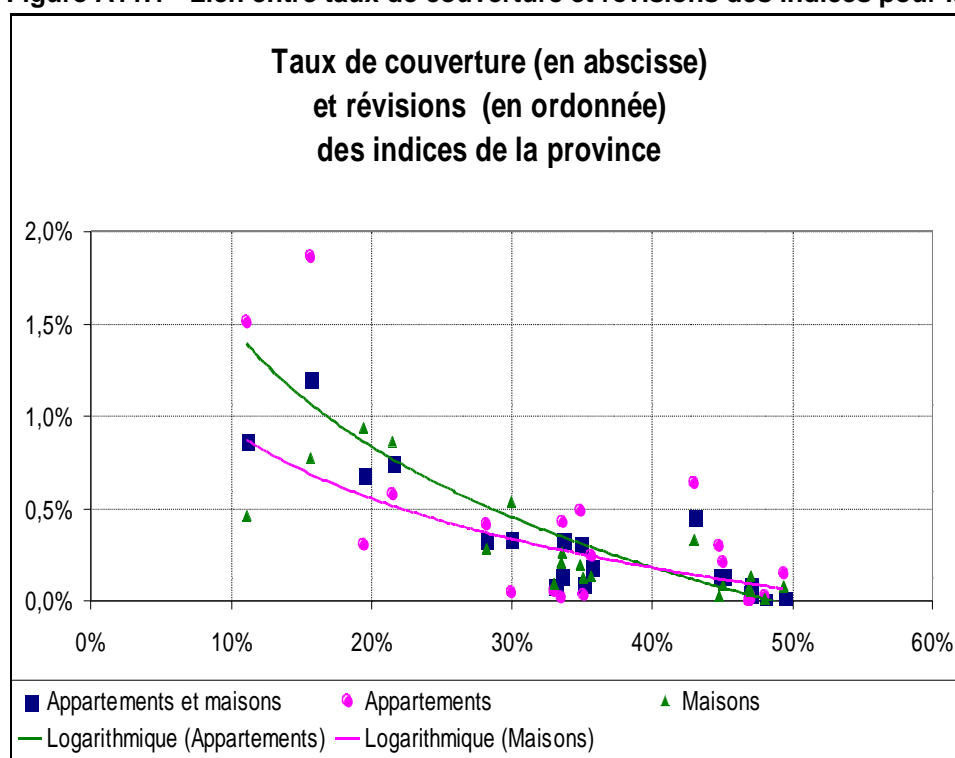
Des taux de couverture minimum ont été déterminés, par type d'indice (*tableau A11.2*). Ils ont été fixés de façon à garantir que les révisions de l'indice soient faibles entre sa première diffusion (indice « provisoire avancé ») et la dernière (indice « définitif »). Le choix des seuils s'est fait sur la base d'une analyse des liens entre les taux de couverture et l'ampleur des révisions d'indices pour les indices de province (*figure A11.1*). Si ces seuils ne sont pas respectés, la diffusion de l'indice est interrompue.

Tableau A11.2 - Taux de couverture minimum pour la région et la métropole

Selon le niveau géographique	Provisoire Avancé	Provisoire	Semi-définitif	Définitif
Métropole	20%	20%	20%	20%
Province	30%	30%	30%	30%
Autres indices supra régionaux	Non diffusé	50%	50%	50%
Région	Non diffusé	50%	50%	50%

N. B. : les volumes de transactions et les taux de couverture en l'Île-de-France dépassent toujours largement les seuils minimum fixés pour la province.

Figure A11.1 - Lien entre taux de couverture et révisions des indices pour la province



Période de référence pour les parcs de référence et d'estimation

Les parcs de référence et d'estimation ont la même période de référence dans toutes les versions pour un indice élémentaire.

Période de référence avant la version V3

La période de référence des parcs de référence et d'estimation est la même pour la durée de vie de la base.

Tableau A11.3 - Période de référence des parcs de référence et d'estimation en V1 et V2

Version		Type de logement	Période de référence
Version 1	Province	Appartement	1994 - 1996
	Province	Maison	1994 - 1997
	Paris	Appartement	1992 - 1996
	Petite couronne (Île-de-France)	Appartement	1992 - 1996
Version 2	Province	Appartement	1998 - 2001
	Province	Maison	1998 - 2001
	Île-de-France	Appartement	1998 - 2001
	Île-de-France	Maison	1998 - 2001

Période de référence en V3

En version 3, la période de référence n'est pas la même que la durée de vie de la base. Les parcs de référence et d'estimation sont renouvelés tous les deux ans. On prend le couple d'année N-3 et N-2 pour produire les indices de prix des années N et N+1. Les périodes de références sont les mêmes pour la province et l'Île-de-France et elles ne dépendent pas du type de logement.

Tableau A11.4 - Période de référence et période de production en V3

Période de référence	Période de production des indices
2005 - 2006	2008 - 2009
2007 - 2008	2010 - 2011
2009 - 2010	2012 - 2013
2011 - 2012	2014 - 2015
2013 - 2014	2016 - 2017
2015 - 2016	2018 - 2019

Liens utiles

Insee

- Bulletin statistique : <http://www.insee.fr/fr/bases-de-donnees/bsweb/>

- Banque de données macro-économique (BDM) : <http://www.bdm.insee.fr/bdm2/index.action>

- Information Rapide trimestriel sur les prix des logements anciens :

<http://www.insee.fr/fr/themes/indicateur.asp?id=96>

- Les indices de prix des logements anciens, version 2 des modèles hédoniques, Insee Méthodes n°111, décembre 2005.

<http://www.insee.fr/fr/publications-et-services/sommaire.asp?codesage=IMET111&nivgeo=0>

<http://www.insee.fr/en/publications-et-services/sommaire.asp?codesage=IMET111&nivgeo=0>

- Les indices de prix des logements anciens, Insee Méthodes n° 98, septembre 2002

<http://www.insee.fr/fr/publications-et-services/sommaire.asp?codesage=IMET098&nivgeo=0>

Chambre interdépartementale des notaires de Paris (CINP) : <http://www.paris.notaires.fr/>

- Historique des conférences de presse des Notaires d'Île-de-France :

<http://www.paris.notaires.fr/presse/conferences-presse-immobilieres>

- Dernières tendances du marché présentées par les notaires :

<http://www.paris.notaires.fr/outil/immobilier/prix-et-nombre-de-ventes-paris-idf>

- Note de conjoncture immobilière des notaires de France :

<http://www.notaires.fr/notaires/communiqués-de-presse>

Conseil général de l'Environnement et du Développement durable

- Statistiques historiques et analyses sur l'évolution du marché immobilier résidentiel sur le long terme :

http://www.cgedd.developpement-durable.gouv.fr/rubrique.php?id_rubrique=138

- Nombre et montant des ventes immobilières, estimation :

<http://www.cgedd.developpement-durable.gouv.fr/nombre-et-montant-des-ventes-a1>

- Indices séculaires du prix des logements :

http://www.cgedd.developpement-durable.gouv.fr/IMG/doc/house-price-index-Paris-and-others-secular_cle7fed11.doc

Eurostat

Handbook on Residential Property Prices Indices (RPPIs):

http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/methodology/hps/rppi_handbook

OCDE

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

<http://www.oecd.org/dataoecd/37/31/33789552.pdf>

