

LA CONSTRUCTION DE L'INDICE DES PROPRIETAIRES- OCCUPANTS (OOH) ET DE L'INDICE DES PRIX DES LOGEMENTS (HPI) : LA NECESSITE DE L'ELABORATION D'UN INDICE DES PRIX DES LOGEMENTS NEUFS

Thomas BALCONE ()*

() Insee, Division des Prix à la Consommation*

Introduction

Dans le cadre du projet OOH (Owner Occupied Housing) conduit par Eurostat, des travaux ont été menés afin de construire un indice permettant de suivre les dépenses des ménages propriétaires-occupants liées à l'acquisition et à l'utilisation de leur logement (indice OOH), et un indice des prix des logements (HPI). L'indice des propriétaires-occupants (indice OOH) est calculé comme la moyenne pondérée des quatre indices suivants :

- l'Indice du Coût de la Construction (ICC) pour l'individuel pur
- l'Indice de Prix des travaux d'Entretien-Amélioration de logements (IPEA)
- l'Indice des Prix à la Consommation (IPC) de l'assurance-habitation
- un indice des prix des logements neufs

L'indice des prix des logements (HPI) est, quant à lui, calculé comme la moyenne pondérée de l'indice Notaires-Insee des logements anciens et d'un indice des prix des logements neufs.

Ainsi, dans le cadre du projet OOH, nous avons dû créer un indice afin de suivre l'évolution des prix des logements neufs. Pour établir cet indice, nous avons utilisé l'Enquête sur la Commercialisation des Logements Neufs (ECLN).

Le marché du logement neuf étant en constante évolution (un logement neuf vendu disparaissant du marché) et différent d'une zone géographique à une autre, calculer l'indice des prix des logements neufs comme le rapport des prix moyens des logements neufs vendus entre deux trimestres consécutifs n'est pas judicieux. En effet, un tel indice ne prendrait pas en compte l'effet qualité lié à l'évolution des caractéristiques des logements neufs. Pour tenir compte de cet effet qualité, une modélisation des prix des logements neufs relevés dans l'enquête ECLN est nécessaire.

Ainsi, nous avons essayé d'établir un modèle hédonique expliquant le prix d'un logement neuf en fonction de ses caractéristiques physiques (surface habitable du logement, nombre de pièces, présence d'un balcon, ...) et des caractéristiques géographiques de la commune de résidence (taille de l'unité urbaine, zone d'études et d'aménagement du territoire (ZEAT), ...). Trois modèles hédoniques ont été testés afin de calculer un indice des prix des logements neufs:

- le modèle à indicatrices temporelles
- le modèle à « période de référence » (hedonic re-pricing method)
- le modèle sur périodes adjacentes

Dans le cadre de cette présentation, nous commencerons par présenter brièvement le projet OOH. Puis, nous nous intéresserons à la source de données utilisée pour calculer un indice des prix des logements neufs et au champ de cet indice. Ensuite, après avoir montré les limites d'une approche naïve, on présentera la démarche suivie pour construire un indice des prix des logements neufs à qualité constante. Enfin, nous nous focaliserons sur l'indice des prix des logements français.

Sommaire

1. PRÉSENTATION DU PROJET OOH (OWNER OCCUPIED HOUSING).....	3
2. LA SOURCE UTILISÉE ET LE CHAMP DE L'INDICE DES PRIX DES LOGEMENTS NEUFS... 3	
2.1. LA SOURCE UTILISÉE : L'ENQUÊTE SUR LA COMMERCIALISATION DES LOGEMENTS NEUFS (ECLN)	3
2.1.1. <i>Le choix de la source</i>	3
2.1.2. <i>Présentation de l'enquête</i>	3
2.1.3. <i>Questionnaire associé à l'enquête</i>	4
2.2. LE CHAMP DE L'INDICE	5
3. LES INDICES « NAÏFS ».....	6
4. VERS UN INDICE DES PRIX DES LOGEMENTS NEUFS À QUALITÉ CONSTANTE	7
4.1. LA NÉCESSITÉ D'UN DÉCOUPAGE GÉOGRAPHIQUE	7
4.1.1. <i>Étude des prix moyens au m² pour un découpage géographique donné</i>	7
4.1.2. <i>Étude des prix moyens au m² par région</i>	8
4.1.3. <i>Première tentative de découpage géographique</i>	9
4.1.4. <i>Découpage de la France métropolitaine hors Corse en zones de prix au m² homogènes en utilisant la méthode de Ward</i>	11
4.1.5. <i>Autres caractéristiques géographiques pertinentes</i>	11
4.2. LE MODÈLE HÉDONIQUE.....	12
4.2.1. <i>La méthode de calcul l'indice Notaires-Insee des prix des logements anciens</i>	12
4.2.2. <i>La modélisation du prix des logements neufs</i>	12
4.2.3. <i>Premier modèle : le modèle à indicatrices temporelles</i>	13
4.2.4. <i>Deuxième méthode : le modèle à « période de référence »</i>	15
4.2.5. <i>Troisième méthode : le modèle sur périodes adjacentes</i>	16
4.2.6. <i>Le modèle retenu</i>	18
5. L'INDICE DES PRIX DES LOGEMENTS (HPI)	21
5.1. DÉFINITION	21
5.2. CHAMP THÉORIQUE.....	21
5.3. CHAMP EFFECTIF.....	22
5.4. MÉTHODE DE CALCUL.....	22
5.5. ÉVOLUTION DE L'HPI.....	23
CONCLUSION.....	25
ANNEXES	26
BIBLIOGRAPHIE.....	35

1. Présentation du projet OOH (Owner Occupied Housing)

Le projet OOH a été lancé par Eurostat au début des années 2000. Un groupe de travail sur la prise en compte des propriétaires occupant leur logement dans l'indice des prix à la consommation a ainsi été mis en place (projet OOH pour « owner occupied housing »). L'objectif de ce projet est double. Le premier est d'essayer de rendre compte du coût du logement de manière plus exhaustive dans l'indice des prix, en incluant l'évolution des dépenses liées à l'acquisition et à l'utilisation de leur logement par les propriétaires occupants. Le deuxième est de créer un indice des prix des logements afin de suivre l'évolution des prix sur le marché immobilier et de pouvoir ainsi anticiper une éventuelle bulle immobilière. Cet indice expérimental fait d'ores et déjà partie des 22 indicateurs du tableau de bord sur les déséquilibres macroéconomiques suivi par la Commission européenne.

Le projet OOH en est désormais à sa phase terminale, la phase 4. Un projet de règlement OOH proposé par Eurostat a été présenté pour la dernière fois au groupe de travail « Harmonisation of Consumer Price Indices » au mois d'octobre dernier. Il sera proposé pour adoption au comité des directeurs généraux en février 2012. Eurostat envisage de publier un indice des prix des logements neufs non expérimental courant 2012 et un indice des prix des propriétaires-occupants non expérimental courant 2014. La France (division des prix à la consommation) a participé aux travaux depuis la phase 2.

Dans le cadre de cette étude, nous nous intéressons uniquement à l'indice des prix des logements.

2. La source utilisée et le champ de l'indice des prix des logements neufs

2.1. La source utilisée : l'Enquête sur la Commercialisation des Logements Neufs (ECLN)

2.1.1. Le choix de la source

Pour construire un indice des prix des logements neufs, il nous a fallu choisir une base de données contenant l'ensemble des ventes de logements neufs destinés aux particuliers.

On a tout d'abord pensé à utiliser les mêmes bases de données que pour l'indice Notaires-INSEE des prix des logements anciens, à savoir les bases notariales, mais ces bases se sont avérées présenter des taux de couverture insuffisants comparativement aux statistiques de la construction neuve du SOeS¹. Par exemple, en 2007, les taux de couverture en Ile de France étaient de l'ordre de 50 à 60% pour les appartements et 40% pour les maisons.

On a ainsi cherché une autre source de données. Notre choix s'est porté sur l'Enquête de Commercialisation des Logements Neufs (ECLN) du SOeS

2.1.2. Présentation de l'enquête

L'enquête ECLN est une enquête trimestrielle assurant le suivi de la commercialisation des logements neufs destinés à la vente aux particuliers depuis 1985. L'enquête est inscrite au programme des enquêtes d'intérêt général à caractère obligatoire arrêté par le CNIS. Elle est soumise au visa du CNIS et à une déclaration à la CNIL.

La base de l'enquête est la base Sitadel des permis de construire déposés par les pétitionnaires auprès des mairies et des Directions départementales de l'équipement (DDE) et exploités par les Directions régionales de l'équipement (DRE). Cette base est mise à jour chaque trimestre par l'introduction des nouveaux programmes autorisés et la suppression des opérations dont la commercialisation est achevée.

¹ service statistique du ministère du développement durable

Le champ de l'enquête est constitué des permis de 5 logements et plus destinés à la vente aux particuliers. Par conséquent cette enquête ne couvre pas :

- le secteur locatif (permis de construire ne comprenant que des logements destinés à la location)
- les constructions individuelles réalisées par des particuliers en vue de leur occupation personnelle (permis de construire d'un logement)
- les logements construits par l'État, les collectivités locales et les sociétés nationalisées
- l'ensemble des logements de fonction
- les programmes de réhabilitation

L'enquête est exhaustive sur son champ. Ainsi, tous les promoteurs ayant déposé un permis de construire d'au moins cinq logements destinés à la vente aux particuliers sont interrogés. Cette restriction du champ de l'enquête aux permis de construire de cinq logements ou plus destinés à la vente aux particuliers n'est pas vraiment pénalisante pour l'analyse statistique. En effet, d'après le SoeS, la part des programmes de moins de 5 logements destinés à la vente aux particuliers et relevant du champ de la promotion immobilière (champ de l'enquête ECLN) est de l'ordre de 1% pour les appartements et d'environ 5% pour les maisons individuelles groupées².

L'unité statistique de l'enquête est le programme de construction et les tranches de commercialisation qui lui sont rattachées. L'unité enquêtée est le promoteur ou l'entreprise effectuant la commercialisation d'un projet ayant fait l'objet d'un permis de construire.

L'enquête est réalisée en France métropolitaine. La collecte des données est faite par des enquêteurs et/ou par voie postale.

2.1.3. Questionnaire associé à l'enquête

Le questionnaire est composé de deux parties, à savoir le questionnaire de base et les fiches tranches. Le questionnaire de base décrit le programme de construction. Il est rempli la première fois pour tout nouveau programme suivi par l'enquête. Il permet notamment :

- d'identifier le programme de construction et de connaître ses caractéristiques (numéro de permis de construire, nom et adresse du programme, ...)
- depuis la nouvelle version de l'enquête datant de 2006, de recueillir des éléments portant sur la qualité du programme comme sa catégorie (courant ou ordinaire, confortable, très confortable, luxe) ou ses éléments de confort (présence d'un dispositif de sécurité, présence d'une piscine dans la résidence, proximité des transports collectifs, présence d'un système de climatisation)

La deuxième partie du questionnaire, la fiche tranche, assure le suivi de la commercialisation du programme. Une fiche tranche est ouverte dès qu'une partie du programme de construction est offerte à la vente et que sa commercialisation est en cours lors du trimestre enquêté. Elle ne comprend que des logements :

- de même type (individuel groupé ou collectif)
- de même nature (logements ordinaires, résidences avec services, résidences de vacances, logements sociaux)
- et au même stade d'avancement des travaux pour le collectif (en projet, en cours, achevé)

Chaque trimestre, une fiche tranche donne notamment **suivant le nombre de pièces des logements** (1 pièce, 2 pièces, 3 pièces, 4 pièces, 5 pièces, 6 pièces ou plus) :

- le nombre total de logements réservés : ce sont les logements ayant fait l'objet d'une réservation avec dépôt d'arrhes ou vendus sans réservation préalable au cours du trimestre écoulé
- le nombre de logements réservés pour investissement locatif : ces logements sont achetés pour les louer et permettent ainsi à leur acheteur de se constituer un patrimoine et percevoir des revenus complémentaires
- la surface moyenne des terrains, ou balcons et terrasses en m² des logements vendus
- la surface habitable moyenne en m² des logements réservés, i.e. la surface des pièces principales d'habitation, de service (cuisine, WC, salle de bains) et de circulation (hall, couloir). Cette surface ne tient pas compte de la superficie des combles non aménagés,

² Des travaux complémentaires seront menés afin de s'assurer que ces logements n'ont pas une évolution de prix atypique

caves, sous-sols, remises, garages, terrasses, loggias, balcons, séchoirs extérieurs au logement, vérandas, volumes vitrés, locaux communs et autres dépendances des logements, ni des parties de locaux d'une hauteur inférieure à 1,80 m selon l'article R. 111-2 du Code de la construction

- le prix moyen en euros des logements réservés : il s'agit du prix à la réservation, i.e. à la signature du premier contrat. Les frais de notaires ne sont pas inclus dans ce prix. De plus, pour le collectif, il s'agit du prix du logement, garage ou parking exclu, alors que pour les maisons individuelles groupées, il s'agit du prix de la maison, de ses dépendances et du terrain.

2.2. Le champ de l'indice

Le champ géographique retenu pour l'indice des prix des logements neufs est la France métropolitaine hors Corse. Ainsi, les logements neufs réservés en Corse ou dans les départements et territoires d'Outre-mer ne sont pas couverts par l'indice des prix des logements neufs que nous allons établir. On a choisi ce champ géographique dans un souci de cohérence avec l'indice Notaires-Insee des prix des logements anciens qui est un indice « France hors Corse et territoires d'Outre-mer ».

Comme la priorité d'Eurostat est de publier un indice des prix des logements (HPI) en 2012, on va retenir tous les logements réservés figurant dans l'enquête ECLN, quelle que soit l'utilisation faite de ce logement par le ménage acquéreur, pour calculer un indice des prix des logements neufs. En particulier, les logements réservés pour investissement locatif vont être couverts par notre indice expérimental. Cependant, nous excluons les logements réservés correspondant à des logements sociaux car un logement social neuf ne peut pas être acquis par un ménage. En effet, seuls peuvent être vendus les logements sociaux à usage locatif construits ou acquis depuis plus de 10 ans par un organisme d'HLM³.

Une nouvelle version de l'enquête ECLN a vu le jour au 1^{er} trimestre 2006. Cependant, cette dernière a cohabité avec l'ancienne version pendant un trimestre. Ainsi, pour construire notre indice des prix des logements neufs, nous allons diviser les observations contenues dans l'enquête ECLN en deux ensembles de données :

- le premier correspond à la période allant du 1^{er} trimestre 2000 au 1^{er} trimestre 2006 (ancienne enquête)
- le deuxième correspond à la période allant du 1^{er} trimestre 2006 au 2^{ème} trimestre 2011

Une observation issue de l'enquête ECLN est caractérisée par un identifiant de fiche tranche, un nombre de pièces et un numéro de trimestre. Ainsi, par exemple, si pour un trimestre et un programme de construction donnés, 3 appartements de 2 pièces et 4 appartements de 5 pièces correspondant à des logements ordinaires ont été réservés au cours du trimestre écoulé, on considérera que deux observations et sept logements réservés sont associés à la fiche tranche répertoriant les appartements ordinaires du programme de construction considéré.

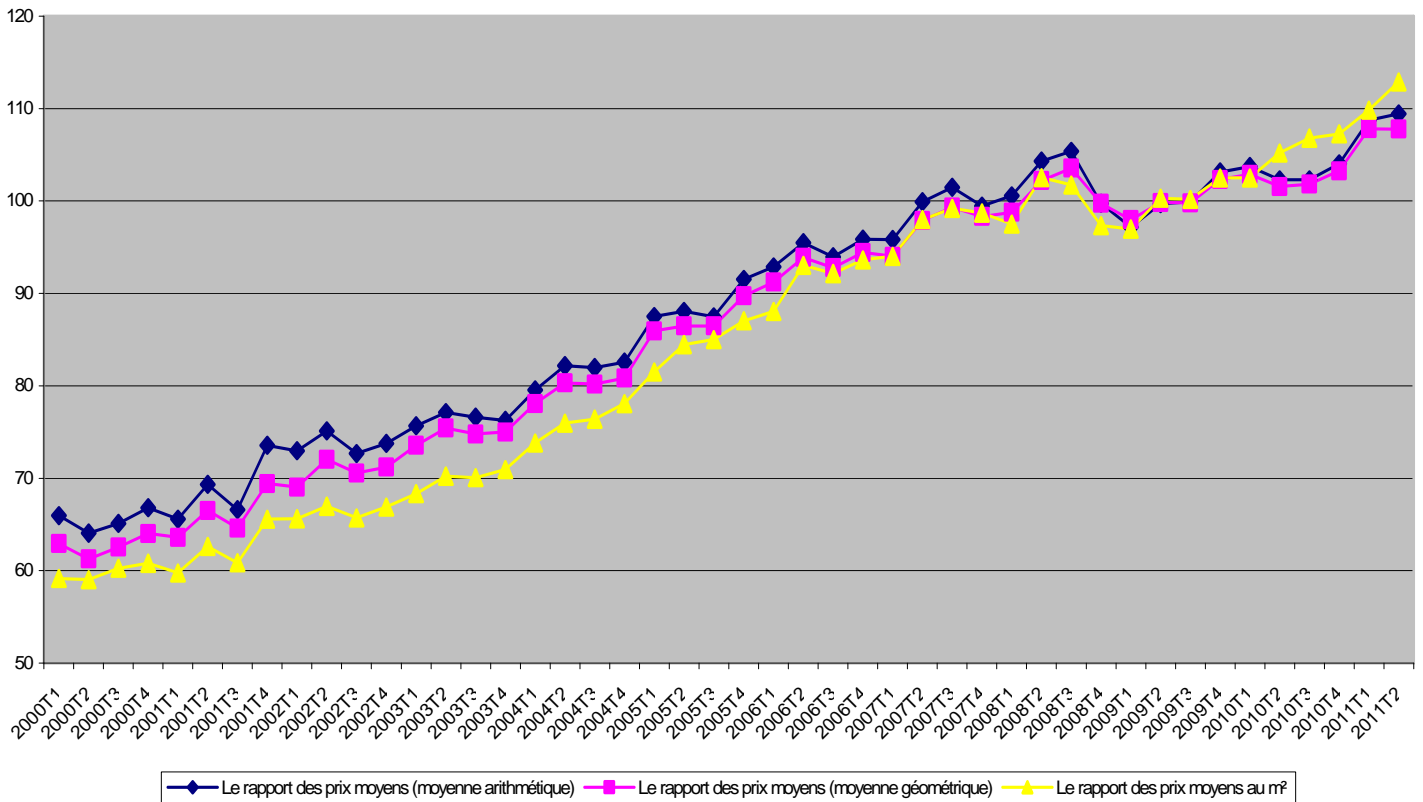
Sur la période allant du 1^{er} trimestre 2000 au 2^{ème} trimestre 2011, chaque trimestre, l'enquête ECLN contient en moyenne environ 8 000 observations qui correspondent en moyenne à environ 25 500 logements réservés.

³ Habitation à Loyer Modéré

3. Les indices « naïfs »

On souhaite construire un indice qui mesure l'évolution des prix des logements neufs vendus entre deux trimestres t et $t+1$. Une première possibilité est de définir notre indice comme le rapport des prix moyens des logements neufs vendus aux deux trimestres considérés, les prix moyens étant calculés en utilisant la moyenne arithmétique. Une deuxième idée est de construire l'indice de la même manière mais en utilisant la moyenne géométrique. Enfin, une dernière idée revient à s'inspirer de la méthodologie de calcul de l'indice des loyers, i.e. à calculer notre indice comme le rapport des prix moyens au m^2 des logements neufs vendus aux deux trimestres considérés.

En chaînant les indices ainsi construits et en choisissant comme base 100 la moyenne annuelle en 2009, on obtient les 3 indices suivants :



Graphique 1 : comparaison de l'évolution des indices naïfs (2000T1-2011T2 - base 100, moyenne annuelle 2009)

Ces trois indices sont relativement proches et présentent tous le même défaut : ils ne prennent pas en compte l'**effet qualité**. En effet, de tels indices sont des indices de prix pur uniquement si les logements considérés entre deux trimestres consécutifs ont les mêmes caractéristiques (physiques et géographiques), i.e. ont la même qualité. Or, le marché du neuf est en constante évolution : un logement neuf vendu au trimestre t n'apparaîtra plus sur le marché du neuf au trimestre $t+1$. Ainsi, il est difficile de trouver entre deux trimestres consécutifs des logements neufs vendus ayant les mêmes caractéristiques.

Il nous faut ainsi essayer de contrôler cet effet qualité. C'est la raison pour laquelle nous allons essayer de modéliser le prix d'un logement neuf en fonction de ses caractéristiques physiques (surface habitable, type de construction, nombre de pièces, ...) et géographiques. On va ainsi essayer de construire un **modèle hédonique** afin de calculer un **indice de prix à qualité constante**.

4. Vers un indice des prix des logements neufs à qualité constante

4.1. La nécessité d'un découpage géographique

Le prix au m² pour un logement neuf est beaucoup plus élevé à Paris qu'en milieu rural. Ainsi, on ne peut pas comparer directement le prix d'un logement neuf construit à Paris avec celui d'un logement neuf ayant les mêmes caractéristiques physiques construit en milieu rural. Ainsi, on va tout d'abord devoir essayer de prendre en compte les effets de localisation géographique dans notre modélisation en construisant des zones géographiques homogènes en termes de prix au m² pour les logements neufs.

4.1.1. Étude des prix moyens au m² pour un découpage géographique donné

Pour une zone géographique z donnée, on va définir un prix moyen au m² à partir des données de l'enquête ECLN de la manière suivante :

$$\text{prix_m2}_{z,t_0-t_1} = \frac{\text{prix_moy}_{z,t_0-t_1}}{\text{shab_moy}_{z,t_0-t_1}} = \frac{\frac{1}{\sum_{t=t_0}^{t_1} n_t(z)} \times \sum_{t=t_0}^{t_1} \sum_{i=1}^{\text{nb_obs}(t)} n_{i,t} \bar{p}_{i,t} I_z(i)}{\frac{1}{\sum_{t=t_0}^{t_1} n_t(z)} \times \sum_{t=t_0}^{t_1} \sum_{i=1}^{\text{nb_obs}(t)} n_{i,t} \text{shab_moy}_{i,t} I_z(i)}$$

où :

- $\text{prix_m2}_{z,t_0-t_1}$ est le prix moyen au m² des logements vendus dans la zone géographique z entre les trimestres t_0 et t_1 (en euros)
- $\text{prix_moy}_{z,t_0-t_1}$ est le prix moyen des logements vendus dans la zone géographique z entre les trimestres t_0 et t_1 (en euros)
- $\text{shab_moy}_{z,t_0-t_1}$ est la surface habitable moyenne des logements vendus dans la zone géographique z entre les trimestres t_0 et t_1 (en m²)
- pour chaque observation i de l'enquête ECLN,
 $I_z(i) = 1$ si l'observation i correspond à des logements vendus dans la zone géographique z
 $I_z(i) = 0$ sinon
- $n_t(z) = \sum_{i=1}^{\text{nb_obs}(t)} n_{i,t} I_z(i)$ est le nombre de logements réservés au cours du trimestre t situés dans la zone géographique z
- $\text{nb_obs}(t)$ est le nombre d'observations issues de l'enquête ECLN au trimestre t
- $n_{i,t}$ est le nombre de logements réservés correspondant à l'observation i issue de l'enquête ECLN au trimestre t
- $\bar{p}_{i,t}$ est le prix moyen en euros des logements réservés correspondant à l'observation i issue de l'enquête ECLN au trimestre t
- $\text{shab_moy}_{i,t}$ est la surface habitable moyenne en m² des logements réservés correspondant à l'observation i issue de l'enquête ECLN au trimestre t

On peut alors définir un écart-type pour le prix moyen au m² pour une zone géographique z donnée de la manière suivante :

$$\text{ecart_type}(\text{prix_m}2_{z,t_0-t_1}) = \sqrt{\frac{1}{\sum_{t=t_0}^{t_1} n_t(z)} \times \sum_{t=t_0}^{t_1} \sum_{i=1}^{\text{nb_obs}(t)} n_{i,t} \left(\frac{\bar{p}_{i,t}}{\text{shab_moy}_{i,t}} - \text{prix_m}2_{z,t_0-t_1} \right)^2} I_z(i)}$$

Cet écart-type nous donne une mesure de la volatilité du prix au m² des logements neufs vendus dans la zone géographique considérée entre deux trimestres donnés.

4.1.2. Étude des prix moyens au m² par région

Une première idée est de choisir le découpage administratif de la France métropolitaine, hors Corse, en régions. Si on étudie la volatilité des prix au m² pour les logements neufs à l'intérieur de chaque région, on obtient les résultats suivants :

Région	nombre moyen de logements réservés par trimestre	prix de vente moyen (en €)	surface habitable moyenne des logements vendus (en m ²)	prix de vente moyen au m ² (en €)	écart-type du prix de vente moyen au m ² (en €)
ILE-DE-FRANCE	5 125	204 824,95	69,35	2 953,67	1 230,65
PROVENCE-ALPES-COTE D'AZUR	2 753	179 855,12	63,95	2 812,32	1 113,67
PICARDIE	176	140 968,88	61,33	2 298,71	832,75
RHONE-ALPES	3 598	157 074,71	68,71	2 286,08	922,85
AQUITAINE	1 587	129 619,42	59,81	2 167,05	639,82
LANGUEDOC-ROUSSILLON	1 530	132 510,41	61,37	2 159,25	679,41
PAYS DE LA LOIRE	1 279	132 559,80	63,72	2 080,38	632,21
MIDI-PYRENEES	1 792	125 574,24	60,77	2 066,29	578,13
HAUTE-NORMANDIE	364	129 140,44	63,48	2 034,26	626,29
BRETAGNE	1 334	124 774,04	62,49	1 996,79	485,27
POITOU-CHARENTES	525	120 786,80	60,53	1 995,64	560,35
AUVERGNE	247	128 359,73	64,97	1 975,62	539,18
BASSE-NORMANDIE	375	127 882,13	65,10	1 964,35	460,78
ALSACE	952	140 392,31	71,79	1 955,67	373,59
NORD-PAS-DE-CALAIS	716	135 270,33	70,67	1 914,19	502,43
LIMOUSIN	207	115 749,15	61,20	1 891,24	477,70
CENTRE	622	128 353,42	68,11	1 884,46	455,24
CHAMPAGNE-ARDENNE	115	134 057,52	73,50	1 823,96	478,82
BOURGOGNE	277	127 455,70	70,10	1 818,08	378,44
FRANCHE-COMTE	239	127 628,63	70,47	1 811,23	460,30
LORRAINE	441	128 361,36	71,70	1 790,21	473,81

Tableau 1 : étude de la volatilité du prix moyen au m² des logements neufs vendus par région (2000T1-2006T1 - ancienne version de l'enquête)

Région	nombre moyen de logements réservés par trimestre	prix de vente moyen (en €)	surface habitable moyenne des logements vendus (en m ²)	prix de vente moyen au m ² (en €)	écart-type du prix de vente moyen au m ² (en €)
ILE-DE-FRANCE	4 771	259 475,66	62,96	4 121,47	1 508,01
PROVENCE-ALPES-COTE D'AZUR	3 483	229 027,48	58,97	3 883,72	1 318,92
RHONE-ALPES	4 328	205 921,11	61,89	3 327,39	1 156,64
LANGUEDOC-ROUSSILLON	1 976	176 188,20	57,05	3 088,57	736,02
AQUITAINE	1 584	177 189,68	59,15	2 995,54	806,46
MIDI-PYRENEES	1 578	172 329,44	59,25	2 908,60	635,16
BASSE-NORMANDIE	433	175 540,65	60,40	2 906,42	1 009,32
PAYS DE LA LOIRE	1 510	167 665,61	58,87	2 848,26	774,31
PICARDIE	353	177 278,72	63,62	2 786,46	555,07
HAUTE-NORMANDIE	428	172 079,22	62,43	2 756,51	571,56
POITOU-CHARENTES	388	164 023,56	60,90	2 693,28	883,53
NORD-PAS-DE-CALAIS	887	181 264,99	67,39	2 689,86	771,53
BRETAGNE	1 650	156 935,64	58,60	2 678,22	810,74
ALSACE	921	180 219,96	68,40	2 634,79	440,09
CHAMPAGNE-ARDENNE	196	163 149,65	62,09	2 627,48	622,94
CENTRE	731	163 677,06	63,39	2 581,94	643,82
BOURGOGNE	304	158 327,25	63,27	2 502,57	518,48
AUVERGNE	306	149 417,61	60,33	2 476,70	530,86
FRANCHE-COMTE	288	158 934,71	65,58	2 423,63	481,82
LIMOUSIN	155	144 795,02	59,86	2 418,92	509,32
LORRAINE	564	156 978,58	68,32	2 297,78	425,31

Tableau 2 : étude de la volatilité du prix moyen au m² des logements neufs vendus par région (2006T1-2011T2 - nouvelle version de l'enquête)

Au vu des écarts-types obtenus, on peut espérer construire des zones géographiques dans lesquelles les prix au m² des logements vendus soient plus homogènes en utilisant par exemple des caractéristiques des villes où ils sont construits qui pourraient avoir une influence sur le prix de vente, comme le nombre d'habitants.

4.1.3. Première tentative de découpage géographique

On a essayé de construire de telles zones homogènes en utilisant les deux caractéristiques géographiques des communes suivantes:

- la **ZEAT** (Zone d'Étude et d'Aménagement du Territoire)
- la **taille de l'unité urbaine**

Le découpage de la France métropolitaine en ZEAT date de 1967. Il a été établi par l'Insee en collaboration avec la DATAR (Délégation à l'Aménagement du Territoire et à l'Action Régionale) lors

du grand courant de la planification territoriale. Il découpe le territoire en huit grandes zones d'études et d'aménagement définis de la manière suivante :

ZEAT	REGIONS
Région parisienne	Ile de France
Bassin parisien	Bourgogne, Centre, Champagne-Ardenne, Basse et Haute Normandie, Picardie
Nord	Nord Pas-de-Calais
Est	Alsace, Franche-Comté, Lorraine
Ouest	Bretagne, Pays de la Loire, Poitou-Charentes
Sud-ouest	Aquitaine, Limousin, Midi-Pyrénées
Centre-est	Auvergne, Rhône-Alpes
Méditerranée	Languedoc-Roussillon, Provence-Alpes-Côte d'Azur, Corse

Tableau 3 : Composition des huit ZEAT

Au niveau européen, le découpage en ZEAT correspond au niveau 1 de la nomenclature des unités territoriales statistiques (NUTS1).

La notion d'unité urbaine repose, quant à elle, sur la continuité du bâti et le nombre d'habitants. On appelle unité urbaine une commune ou un ensemble de communes présentant une zone de bâti continu (pas de coupure de plus de 200 mètres entre deux constructions) qui compte au moins 2 000 habitants. Les communes qui ne rentrent pas dans la constitution d'une unité urbaine sont considérées comme rurales.

La taille d'une unité urbaine correspond au troisième caractère de son code d'identification qui en comprend cinq. Ce code précise la taille de l'unité urbaine à laquelle appartient la commune au dernier recensement de la population. Il est calculé à partir de la population municipale de l'ensemble de l'unité urbaine. Dans le cas d'unités urbaines internationales, seules sont prises en compte les communes situées en France. Il y a 9 tailles d'unité urbaine différentes :

Taille d'unité urbaine	Commune
0	Commune rurale
1	Commune appartenant à une unité urbaine de 2 000 à 4 999 habitants
2	Commune appartenant à une unité urbaine de 5 000 à 9 999 habitants
3	Commune appartenant à une unité urbaine de 10 000 à 19 999 habitants
4	Commune appartenant à une unité urbaine de 20 000 à 49 999 habitants
5	Commune appartenant à une unité urbaine de 50 000 à 99 999 habitants
6	Commune appartenant à une unité urbaine de 100 000 à 199 999 habitants
7	Commune appartenant à une unité urbaine de 200 000 à 1 999 999 habitants
8	Commune appartenant à l'unité urbaine de Paris

Tableau 4 : les différentes tailles d'unité urbaine

Les unités urbaines sont redéfinies périodiquement. C'est pourquoi, pour la nouvelle enquête (données allant du 1^{er} trimestre 2006 au 2^{ème} trimestre 2011), nous allons utiliser le zonage daté de 2010 qui a été établi en référence à la population connue au recensement de 2007 et sur la géographie du territoire au 1^{er} janvier 2010. Pour l'ancienne enquête, nous utiliserons les tailles d'unité urbaine définies à partir du recensement de 1999.

Il apparaît pertinent d'utiliser ces deux variables pour définir des zones géographiques homogènes en termes de prix au m² des logements vendus, dans la mesure où d'après les tableaux 4 et 5, les prix de vente moyens au m² sont plus élevés dans les régions fortement peuplées (Ile de France) que dans les régions où la densité de population est moins importante (Limousin).

On va ainsi croiser les variables ZEAT (Zone d'Étude et d'Aménagement du Territoire) et taille d'unité urbaine. On obtient ainsi un découpage de la France métropolitaine en 63 classes (et non pas 72, certains croisements n'existant pas). Nous pouvons alors étudier la volatilité du prix moyen au m² des logements vendus par zone géographique ainsi définie de la même manière que pour les régions (cf. **annexe 1** et **annexe 2**).

On constate que pour certains croisements, on obtient des écarts-type plus faibles que ceux obtenus avec un découpage en régions. Cependant, certains croisements présentent un écart-type conséquent. Ce découpage n'est donc pas satisfaisant.

Comme on ne peut pas obtenir directement des zones géographiques homogènes en termes de prix de vente moyen au m², on va essayer de les construire, dans une première approche, en effectuant une **classification ascendante hiérarchique** (CAH) sur les couples (ZEAT ; taille d'unité urbaine) en utilisant la **méthode de Ward**, cette dernière reposant sur le principe de minimiser l'inertie intra-classe à chaque étape de la classification. Cette première approche ne nous permettra certainement pas d'obtenir des classes beaucoup plus homogènes et sera prolongée par des travaux complémentaires.

4.1.4. Découpage de la France métropolitaine hors Corse en zones de prix au m² homogènes en utilisant la méthode de Ward

Dans le cadre de notre étude, notre nuage de points est constitué des 63 couples (ZEAT ; taille d'unité urbaine). A chacun de ces éléments j est associé :

- **une masse** m_j qui est la part des logements réservés dans la zone géographique définie par le couple j dans l'ensemble des logements réservés au cours de la période considérée
- **le prix moyen en euros** $\text{prix}_{\text{moy}_j}$ des logements réservés dans la zone géographique définie par le couple j au cours de la période considérée
- **la surface habitable moyenne en m²** $\text{shab}_{\text{moy}_j}$ des logements réservés dans la zone géographique définie par le couple j au cours de la période considérée

Ainsi, le nuage de points considéré est constitué de 63 points j dans \mathbf{R}^2 , chaque point ayant pour coordonnées $(\text{prix}_{\text{moy}_j} ; \text{shab}_{\text{moy}_j})$ et pour masse m_j . Après avoir réalisé la classification ascendante hiérarchique sur ce nuage de points, il suffit de couper l'arbre de classification obtenu par une droite afin d'obtenir une partition de notre nuage en classes homogènes en terme de prix moyen au m² et de surface habitable moyenne. On décide de considérer 15 classes.

Si on étudie la volatilité des prix au m² à l'intérieur de chacune de ces 15 classes, on constate que l'écart-type du prix de vente moyen au m² reste parfois conséquent (cf. **annexe 3** et **annexe 4**). Ainsi, ce découpage n'est pas entièrement satisfaisant. Cependant, on va retenir pour le moment ce découpage de la France métropolitaine, hors Corse en 15 classes géographiques. Des travaux complémentaires seront menés afin de l'améliorer.

4.1.5. Autres caractéristiques géographiques pertinentes

Afin de compléter ce découpage géographique, on va également considérer d'autres caractéristiques géographiques des communes dans lesquelles sont situées les logements réservés.

On va tout d'abord retenir le statut des communes défini à partir du découpage du territoire en unités urbaines. En effet, une commune peut avoir quatre statuts différents. Lorsqu'une commune constitue à elle seule une unité urbaine, on dit que c'est une ville isolée. Dans le cas où une unité urbaine est constituée de plusieurs communes, on la désigne sous le terme d'agglomération multi communale et les communes qui la composent sont soit ville-centre, soit banlieue. Si une commune représente plus de 50 % de la population de l'agglomération multi communale, elle est seule ville-centre. Sinon, toutes les communes qui ont une population supérieure à 50 % de celle de la commune la plus peuplée, ainsi que cette dernière, sont villes-centres. Les communes d'une agglomération multi communale qui ne sont pas villes-centres constituent la banlieue.

Ensuite, on va repérer les communes littorales et les communes situées dans l'arrière-pays littoral. C'est la loi « Littoral » n°86-2 du 3 janvier 1986 qui établit ce classement. Une commune littorale (ou maritime) est une commune située directement en bord de mer, d'océans ou d'étangs salés. L'arrière-pays littoral est, quant à lui, défini comme l'ensemble des communes non littorales des cantons littoraux, un canton littoral étant un canton ayant au moins une commune littorale. On dénombre 885 communes littorales et l'arrière-pays littoral est constitué de 1178 communes.

Enfin, on va également repérer les stations de sport d'hiver et d'alpinisme à partir de la liste des stations classées établie au mois de février 2009 avant l'entrée en vigueur de la réforme du classement des communes touristiques et des stations classées de tourisme introduite par la loi du 14 avril 2006. On dénombre ainsi 62 communes correspondant à des stations de sport d'hiver et d'alpinisme.

4.2. Le modèle hédonique

Pour mettre au point la modélisation des prix des logements neufs en fonction de leurs caractéristiques, nous nous sommes inspirés des modèles utilisés pour le calcul de l'indice Notaires-INSEE des prix des logements anciens.

4.2.1. La méthode de calcul l'indice Notaires-Insee des prix des logements anciens

La méthodologie de calcul de l'indice Notaires-Insee repose sur des modèles décomposant les prix des logements anciens en fonction des contributions de leurs caractéristiques. Pour prendre en compte l'effet géographique, le territoire métropolitain hors Corse est découpé, pour chaque type de bien (appartements et maisons) en zones élémentaires (environ 300) homogènes en termes de prix. A l'aide de modèles économétriques, des coefficients mesurant l'impact de chaque caractéristique physique sont estimés au sein de chacune des ces zones. Ils servent à estimer, à partir des transactions observées, les prix de biens de référence dont les caractéristiques sont fixes au cours du temps. Les évolutions de prix de ces biens de référence sont ensuite utilisées pour le calcul d'indices de prix élémentaires. Les indices agrégés sont calculés à partir des évolutions des indices élémentaires en tenant compte du poids de chaque indice élémentaire au sein du parc de référence, ce dernier étant actualisé tous les deux ans.

4.2.2. La modélisation du prix des logements neufs

En raison du faible nombre d'observations issues chaque trimestre de l'enquête ECLN, la séparation appartements/maisons faite dans la méthodologie Notaires-Insee n'est pas applicable dans le cadre de la modélisation des prix des logements neufs. Pour la même raison, nous avons été contraints d'inclure le zonage géographique dans notre modèle. Ainsi, les prix des logements neufs ont été modélisés à partir d'un seul modèle hédonique qui contient les caractéristiques physiques et géographiques des logements. Par analogie avec l'indice Notaires-INSEE des prix des appartements anciens, nous avons tout d'abord essayé de modéliser le prix au m² mais le faible nombre de variables descriptives du logement contenues dans l'enquête ECLN ne nous permettait pas d'obtenir une modélisation satisfaisante.

Nous avons alors opté pour un modèle multiplicatif afin de prendre en compte les interactions entre les variables descriptives du logement. Nous avons ainsi choisi un modèle log-log dans lequel la variable dépendante est le logarithme du prix du logement vendu, les variables explicatives continues sont également passées en logarithme et les variables qualitatives sont transformées en indicatrices. Ce modèle qui est un cas particulier du modèle général Box-Cox est celui utilisé pour la modélisation des prix des maisons anciennes.

Le modèle général s'écrit ainsi sous la forme suivante :

$$\forall t = 1, \dots, T \text{ et } \forall i = 1, \dots, \text{nb_obs}(t),$$

$$\ln(\bar{p}_{i,t}) = \alpha + \beta_{\text{shab}} \ln(\text{shab_moy}_{i,t}) + \sum_{k=1}^K \beta_k I_{i,t,k} + \varepsilon_{i,t}$$

où :

- T est le nombre de trimestres considérés
- $(I_{i,t,1}, \dots, I_{i,t,k}, \dots, I_{i,t,K})$ est un vecteur de K indicatrices correspondant aux caractéristiques physiques et géographiques des logements réservés correspondant à l'observation i issue de l'enquête ECLN au trimestre t
- $\alpha, \beta_{\text{shab}}, \beta_1, \dots, \beta_k, \dots, \beta_K$ sont les paramètres (réels) du modèle à estimer
- $\varepsilon_{i,t}$ est le terme d'erreur associé à l'observation i issue de l'enquête ECLN au trimestre t. On suppose que les termes d'erreur $\varepsilon_{i,t}$ sont tels que :

$$\varepsilon_{i,t} \sim N(0, \sigma^2)$$

Les K variables indicatrices correspondent aux caractéristiques suivantes :

- les caractéristiques géographiques :
 - o le découpage du territoire métropolitain hors Corse en 15 classes géographiques (14 indicatrices)
 - o le statut de la commune (3 indicatrices)
 - o les communes maritimes (1 indicatrice)
 - o l'arrière-pays littoral (1 indicatrice)
 - o les stations de sport d'hiver et d'alpinisme (1 indicatrice)
- les caractéristiques physiques:
 - o le type de construction (collectif/individuel groupé) et le nombre de pièces (3 indicatrices) :
 - appartement d'une pièce
 - appartement de 2 à 4 pièces
 - appartement de plus de 5 pièces

Pour les données ECLN issues de la nouvelle enquête (du 1^{er} trimestre 2006 au 2^{ème} trimestre 2011), on introduit des indicatrices supplémentaires correspondant aux caractéristiques physiques suivantes :

- o la catégorie du programme de construction (4 indicatrices)
- o la certification QUALITEL du programme de construction (1 indicatrice)
- o la présence d'un dispositif de sécurité (1 indicatrice)
- o la présence d'une piscine dans la résidence (1 indicatrice)
- o la proximité d'un transport collectif par voie ferrée (1 indicatrice)
- o la présence d'un système de climatisation dans le programme de construction (1 indicatrice)
- o la présence d'un balcon ou d'une terrasse pour les appartements (1 indicatrice)

Pour construire un indice de prix à qualité constante à partir de ce modèle général, on peut soit définir un parc fixe de logements et estimer la valeur de ce parc à différentes dates (cf. la méthodologie de l'indice Notaires-Insee), soit introduire des indicatrices temporelles dans le modèle. Dans la mesure où le marché du neuf est en constante évolution, il apparaît très complexe de définir un parc fixe de logements neufs. On a ainsi décidé de construire un indice des prix à qualité constante en introduisant des indicatrices temporelles dans le modèle général. On a alors considéré les 3 modèles suivants :

- o le modèle à « période de référence »
- o le modèle à indicatrices temporelles
- o le modèle sur périodes adjacentes

4.2.3. Premier modèle : le modèle à indicatrices temporelles

4.2.3.1. Le modèle

Dans un modèle à indicatrices temporelles, les indicatrices temporelles des trimestres où sont réservés les logements sont ajoutées comme variables explicatives dans le modèle de régression général. Dans ce modèle, la date de réservation du logement devient une variable explicative du prix du logement.

Le modèle à indicatrices temporelles s'écrit ainsi de la manière suivante :

$\forall t = 1, \dots, T$ et $\forall i = 1, \dots, nb_obs(t)$,

$$\ln(\bar{p}_{i,t}) = \alpha + \beta_{shab} \ln(shab_moy_{i,t}) + \sum_{k=1}^K \beta_k I_{i,t,k} + \sum_{j=2}^T \delta_j D_{i,t,j} + \varepsilon_{i,t}$$

où $\forall j = 1, \dots, T$,

- D_j est l'indicatrice temporelle du trimestre j définie de la manière suivante :

$$\forall t = 1, \dots, T \text{ et } \forall i = 1, \dots, \text{nb_obs}(t), D_{i,t,j} = \begin{cases} 1 & \text{si } t = j \\ 0 & \text{sin on} \end{cases}$$

- δ_j est le paramètre (réel) associé à l'indicatrice temporelle du trimestre j

4.2.3.2. L'indice de prix à qualité constante

Pour calculer un indice de prix à qualité constante entre deux trimestres t_1 et t_2 noté I_{t_2/t_1}^* , on repart de la définition de l'indice « naïf » comme le rapport des prix moyens des logements neufs vendus aux deux trimestres considérés, les prix moyens étant calculés en utilisant la moyenne géométrique. Mais, pour prendre en compte l'effet qualité, on remplace les prix réellement observés par leur estimateur des moindres carrés ordinaires (MCO) obtenu à partir du modèle à indicatrices temporelles, i.e. $\forall t = 1, \dots, T$ et $\forall i = 1, \dots, \text{nb_obs}(t)$, $\bar{p}_{i,t}$ est remplacé par son estimateur $\hat{\bar{p}}_{i,t}$ qui est donné par la relation suivante :

$$\hat{\bar{p}}_{i,t} = \exp \left(\hat{\alpha} + \hat{\beta}_{\text{shab}} \ln(\text{shab_moy}_{i,t}) + \sum_{k=1}^K \hat{\beta}_k I_{i,t,k} + \hat{\delta}_t \right)$$

où $\hat{\alpha}$, $\hat{\beta}_{\text{shab}}$, $\hat{\beta}_1, \dots, \hat{\beta}_K$, $\hat{\delta}_t$ sont les estimateurs des MCO des paramètres α , β_{shab} , β_1, \dots, β_K , δ_t

Ainsi, l'indice des prix des logements neufs entre deux trimestres t_1 et t_2 , noté I_{t_2/t_1} , se réécrit sous la forme suivante :

$$I_{t_2/t_1} = \frac{\left(\prod_{i=1}^{\text{nb_obs}(t_2)} \hat{\bar{p}}_{i,t_2}^{n_{i,t_2}} \right)^{\frac{1}{n_{t_2}}}}{\left(\prod_{i=1}^{\text{nb_obs}(t_1)} \hat{\bar{p}}_{i,t_1}^{n_{i,t_1}} \right)^{\frac{1}{n_{t_1}}}}$$

En remplaçant les estimateurs $\hat{\bar{p}}_{i,t_1}$ et $\hat{\bar{p}}_{i,t_2}$ par leurs expressions respectives, il vient l'égalité suivante :

$$I_{t_2/t_1} = I_{t_2/t_1}^* \times I_{t_2/t_1}^Q$$

où :

$$I_{t_2/t_1}^* = \exp(\hat{\delta}_{t_2} - \hat{\delta}_{t_1})$$

et

$$I_{t_2/t_1}^Q = \exp \left[\hat{\beta}_{\text{shab}} \left(\sum_{i=1}^{\text{nb_obs}(t_2)} \frac{n_{i,t_2}}{n_{t_2}} \ln(\text{shab_moy}_{i,t_2}) - \sum_{i=1}^{\text{nb_obs}(t_1)} \frac{n_{i,t_1}}{n_{t_1}} \ln(\text{shab_moy}_{i,t_1}) \right) + \sum_{k=1}^K \hat{\beta}_k (\bar{I}_{t_2,k} - \bar{I}_{t_1,k}) \right]$$

avec $\forall k = 1, \dots, K$:

$$\bar{I}_{t_1,k} = \sum_{i=1}^{nb_obs(t_1)} \frac{n_{i,t_1}}{n_{t_1}} I_{i,t_1,k}$$

$$\text{et } \bar{I}_{t_2,k} = \sum_{i=1}^{nb_obs(t_2)} \frac{n_{i,t_2}}{n_{t_2}} I_{i,t_2,k}$$

On a ainsi écrit l'indice de prix moyens, I_{t_2/t_1} , sous la forme d'un produit d'un indice de prix à qualité constante, I_{t_2/t_1}^* , et un indice de qualité à prix constants, I_{t_2/t_1}^Q .

L'indice des prix des logements neufs à qualité constante mesurant l'évolution des prix des logements neufs entre le trimestre 1 et un trimestre t, noté $I_{t/1}^*$ est ainsi donné par la relation suivante :

$$I_{t/1}^* = \exp(\hat{\delta}_t - \hat{\delta}_1)$$

or d'après l'écriture du modèle à indicatrices temporelles, le trimestre 1 est considéré comme étant le trimestre de référence ($\hat{\delta}_1 = 0$).

Ainsi, il vient que :

$$I_{t/1}^* = \exp(\hat{\delta}_t)$$

4.2.4. Deuxième méthode : le modèle à « période de référence »

4.2.4.1. Le modèle

Dans un modèle « à période de référence », on estime les paramètres du modèle à indicatrices temporelles en considérant **uniquement** les observations d'un certain nombre de trimestres. La période couverte par ces observations est la période de référence. On note T_{ref} le nombre de trimestres utilisés. On considère ainsi le modèle de régression suivant :

$\forall t = 1, \dots, T_{ref}$ et $\forall i = 1, \dots, nb_obs(t)$,

$$\ln(\bar{p}_{i,t}) = \alpha + \beta_{shab} \ln(shab_moy_{i,t}) + \sum_{k=1}^K \beta_k I_{i,t,k} + \sum_{j=2}^{T_{ref}} \delta_j D_{i,t,j} + \varepsilon_{i,t}$$

On note $\hat{\alpha}_{ref}$, $\hat{\beta}_{shab,ref}$, $\hat{\beta}_{1,ref}$, ..., $\hat{\beta}_{K,ref}$, $\hat{\delta}_{2,ref}$, ..., $\hat{\delta}_{T_{ref},ref}$ ($\hat{\delta}_{1,ref} = 0$) les estimateurs des MCO des paramètres α , β_{shab} , β_1 , ..., β_K ainsi obtenus.

A partir de ces estimateurs, pour les trimestres t postérieurs à la période de référence ($t > T_{ref}$), on va définir pour chaque observation i **sa valeur implicite**, noté $\bar{p}_{i,t}^*$, de la manière suivante :

$\forall t = T_{ref} + 1, \dots, T$ et $\forall i = 1, \dots, nb_obs(t)$,

$$\bar{p}_{i,t}^* = \frac{\bar{p}_{i,t}}{\bar{p}_{i,t}^{ref}}$$

$$\text{où : } \bar{p}_{i,t}^{ref} = \exp\left(\hat{\alpha}_{ref} + \hat{\beta}_{shab,ref} \ln(shab_moy_{i,t}) + \sum_{k=1}^K \hat{\beta}_{k,ref} I_{i,t,k}\right)$$

On peut définir de la même manière **une valeur implicite** pour les observations correspondant à la période de référence en remplaçant leurs prix observés par leurs estimateurs des MCO obtenus à l'aide du modèle ci-dessus. On a ainsi :

$$\forall t = 1, \dots, T_{\text{ref}} \text{ et } \forall i = 1, \dots, \text{nb_obs}(t),$$

$$\bar{p}_{i,t}^* = \frac{\hat{p}_{i,t}}{\bar{p}_{i,t}^{\text{ref}}}$$

$$\text{i.e. } \bar{p}_{i,t}^* = \frac{\exp\left(\hat{\alpha}_{\text{ref}} + \hat{\beta}_{\text{shab,ref}} \ln(\text{shab_moy}_{i,t}) + \sum_{k=1}^K \hat{\beta}_{k,\text{ref}} I_{i,t,k} + \hat{\delta}_{t,\text{ref}}\right)}{\exp\left(\hat{\alpha}_{\text{ref}} + \hat{\beta}_{\text{shab,ref}} \ln(\text{shab_moy}_{i,t}) + \sum_{k=1}^K \hat{\beta}_{k,\text{ref}} I_{i,t,k}\right)}$$

Ainsi, il vient que :

$$\forall t = 1, \dots, T_{\text{ref}} \text{ et } \forall i = 1, \dots, \text{nb_obs}(t),$$

$$\bar{p}_{i,t}^* = \exp(\hat{\delta}_{t,\text{ref}})$$

4.2.4.2. L'indice de prix à qualité constante

De la même manière que pour le modèle à indicatrices temporelles, on en déduit alors un indice de prix à qualité constante entre le 1^{er} trimestre et un trimestre t donné, noté $I_{t/1}^*$, en remplaçant les prix moyens réellement observés par leurs valeurs implicites :

$$I_{t/1}^* = \frac{\left(\prod_{i=1}^{\text{nb_obs}(t)} \bar{p}_{i,t}^* n_{i,t}\right)^{\frac{1}{n_t}}}{\left(\prod_{i=1}^{\text{nb_obs}(1)} \hat{p}_{i,1}^* n_{i,1}\right)^{\frac{1}{n_1}}}$$

Comme $\forall i = 1, \dots, \text{nb_obs}(1)$, $\bar{p}_{i,1}^* = \exp(\hat{\delta}_{1,\text{ref}})$ et que $\hat{\delta}_{1,\text{ref}} = 0$, il vient que :

$$\forall t = 1, \dots, T,$$

$$I_{t/1}^* = \left(\prod_{i=1}^{\text{nb_obs}(t)} \bar{p}_{i,t}^* n_{i,t}\right)^{\frac{1}{n_t}}$$

4.2.5. Troisième méthode : le modèle sur périodes adjacentes

4.2.5.1. Le modèle

Le modèle sur périodes adjacentes est un modèle à indicatrices temporelles basé uniquement sur les observations de deux trimestres consécutifs t_1 et t_1+1 . Dans cette approche, le modèle de régression s'écrit ainsi sous la forme suivante :

$\forall t = t_1, t_1 + 1$ et $\forall i = 1, \dots, \text{nb_obs}(t)$,

$$\ln(\bar{p}_{i,t}) = \alpha + \beta_{\text{shab}} \ln(\text{shab_moy}_{i,t}) + \sum_{k=1}^K \beta_k I_{i,t,k} + \delta_{t_1+1} D_{i,t,t_1+1} + \varepsilon_{i,t}$$

4.2.5.2. L'indice de prix à qualité constante

De la même manière que pour le modèle à indicatrices temporelles, pour calculer un indice de prix à qualité constante entre deux trimestres consécutifs t_1 et t_1+1 noté I_{t_1+1/t_1}^* , on commence par remplacer les prix moyens réellement observés par leur estimateur des moindres carrés ordinaires (MCO) obtenu à partir du modèle de régression ci-dessus, i.e. \bar{p}_{i,t_1+1} est remplacé par son estimateur \hat{p}_{i,t_1+1} et \bar{p}_{i,t_1} est remplacé par son estimateur \hat{p}_{i,t_1} . Ces deux estimateurs sont donnés par les relations suivantes :

$\forall i = 1, \dots, \text{nb_obs}(t_1 + 1)$,

$$\hat{p}_{i,t_1+1} = \exp\left(\hat{\alpha} + \hat{\beta}_{\text{shab}} \ln(\text{shab_moy}_{i,t_1+1}) + \sum_{k=1}^K \hat{\beta}_k I_{i,t_1+1,k} + \hat{\delta}_{t_1+1}\right)$$

et

$\forall i = 1, \dots, \text{nb_obs}(t_1)$,

$$\hat{p}_{i,t_1} = \exp\left(\hat{\alpha} + \hat{\beta}_{\text{shab}} \ln(\text{shab_moy}_{i,t_1}) + \sum_{k=1}^K \hat{\beta}_k I_{i,t_1,k}\right)$$

où $\hat{\alpha}$, $\hat{\beta}_{\text{shab}}$, $\hat{\beta}_1, \dots, \hat{\beta}_K$, $\hat{\delta}_{t_1+1}$ sont les estimateurs des MCO des paramètres α , β_{shab} , β_1, \dots, β_K , δ_{t_1+1}

Ainsi, l'indice des prix des logements neufs entre deux trimestres consécutifs t_1 et t_1+1 , noté I_{t_1+1/t_1} , se réécrit sous la forme suivante :

$$I_{t_1+1/t_1} = \frac{\left(\prod_{i=1}^{\text{nb_obs}(t_1+1)} \hat{p}_{i,t_1+1}^{n_{i,t_1+1}}\right)^{\frac{1}{n_{t_1+1}}}}{\left(\prod_{i=1}^{\text{nb_obs}(t_1)} \hat{p}_{i,t_1}^{n_{i,t_1}}\right)^{\frac{1}{n_{t_1}}}}$$

En remplaçant les estimateurs \hat{p}_{i,t_1} et \hat{p}_{i,t_1+1} par leurs expressions respectives, il vient l'égalité suivante :

$$I_{t_1+1/t_1} = I_{t_1+1/t_1}^* \times I_{t_1+1/t_1}^Q$$

où :

$$I_{t_1+1/t_1}^* = \exp(\hat{\delta}_{t_1+1})$$

et

$$I_{t_1+1/t_1}^Q = \exp \left[\hat{\beta}_{shab} \left(\sum_{i=1}^{nb_obs(t_1+1)} \frac{n_{i,t_1+1}}{n_{t_1+1}} \ln(shab_moy_{i,t_1+1}) - \sum_{i=1}^{nb_obs(t_1)} \frac{n_{i,t_1}}{n_{t_1}} \ln(shab_moy_{i,t_1}) \right) + \sum_{k=1}^K \hat{\beta}_k (\bar{I}_{t_1+1,k} - \bar{I}_{t_1,k}) \right]$$

avec $\forall k = 1, \dots, K$:

$$\bar{I}_{t_1+1,k} = \sum_{i=1}^{nb_obs(t_1+1)} \frac{n_{i,t_1+1}}{n_{t_1+1}} I_{i,t_1+1,k}$$

et
$$\bar{I}_{t_1,k} = \sum_{i=1}^{nb_obs(t_1)} \frac{n_{i,t_1}}{n_{t_1}} I_{i,t_1,k}$$

On a ainsi écrit l'indice de prix moyens, I_{t_1+1/t_1} , sous la forme d'un produit d'un indice de prix à qualité constante, I_{t_1+1/t_1}^* , et un indice de qualité à prix constants, I_{t_1+1/t_1}^Q .

L'indice des prix des logements neufs à qualité constante mesurant l'évolution des prix des logements neufs entre le trimestre 1 et un trimestre t_1 donné, noté $I_{t_1/1}^*$ est alors donné par la relation suivante :

$$I_{t_1/1}^* = \prod_{t=1}^{t_1-1} I_{t+1/t}^*$$

i.e.
$$I_{t_1/1}^* = \exp \left(\sum_{t=1}^{t_1-1} \hat{\delta}_{t+1} \right)$$

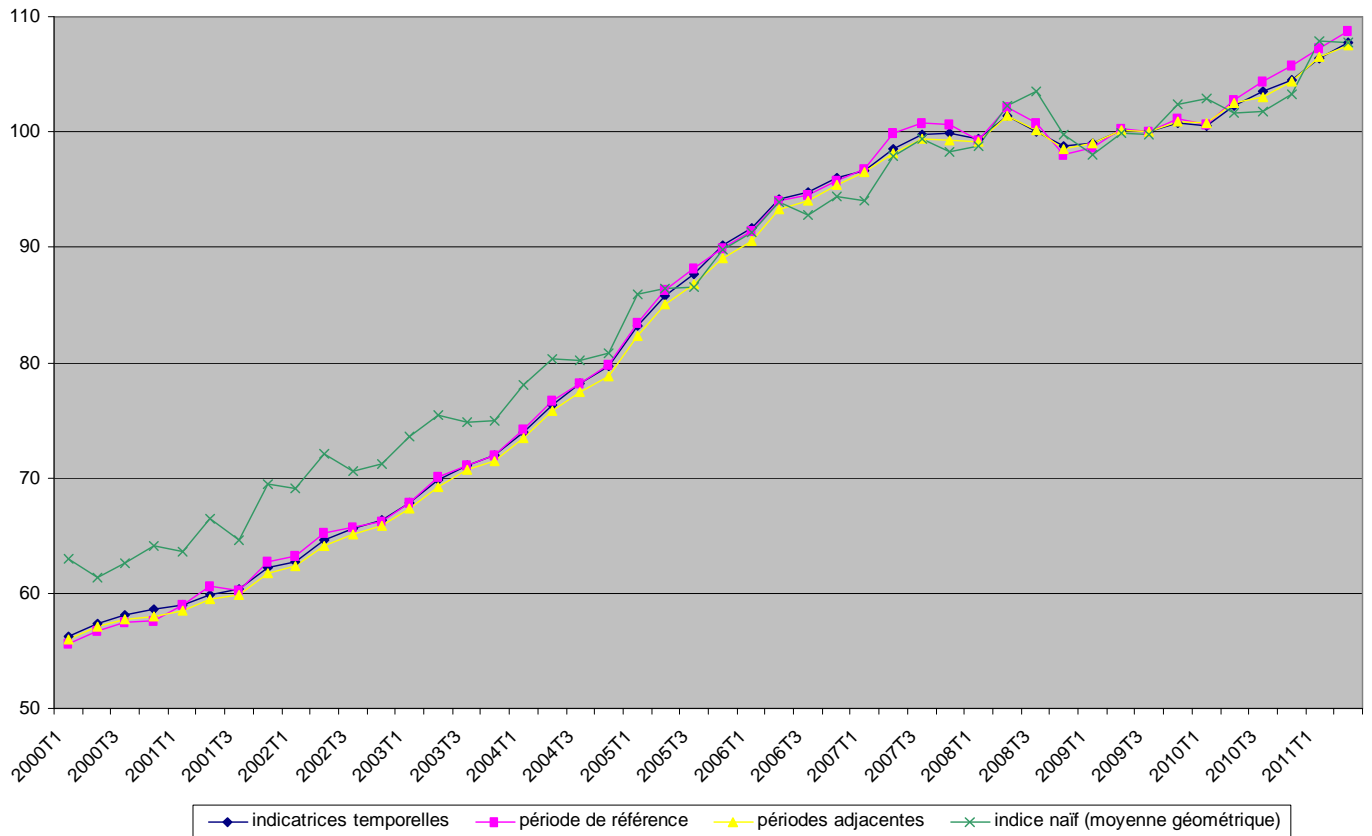
4.2.6. Le modèle retenu

Chacun des modèles présentés ci-dessus a ses avantages et ses inconvénients.

Le modèle à indicatrices temporelles a pour principal avantage d'augmenter sensiblement le nombre des observations qui sont utilisées pour estimer les paramètres de la régression. En effet, chaque trimestre t , on réestime les paramètres du modèle en considérant toutes les observations du premier trimestre jusqu'au trimestre t alors que dans le modèle sur périodes adjacentes, on n'utilise que les observations des trimestres t et $t-1$. Cependant, le modèle à indicatrices temporelles présente un inconvénient majeur : en utilisant un tel modèle, on fait l'hypothèse implicite que les paramètres de la régression associés aux caractéristiques physiques des logements ($\beta_{shab}, \beta_1, \dots, \beta_k, \dots, \beta_K$) sont constants. Or, cette hypothèse n'est en pratique pas vérifiée. Cela traduit le fait que l'influence d'une caractéristique physique du logement sur son prix de vente, toutes choses égales par ailleurs, varie au cours du temps. Le modèle sur périodes adjacentes permet de prendre en compte cette instabilité des paramètres.

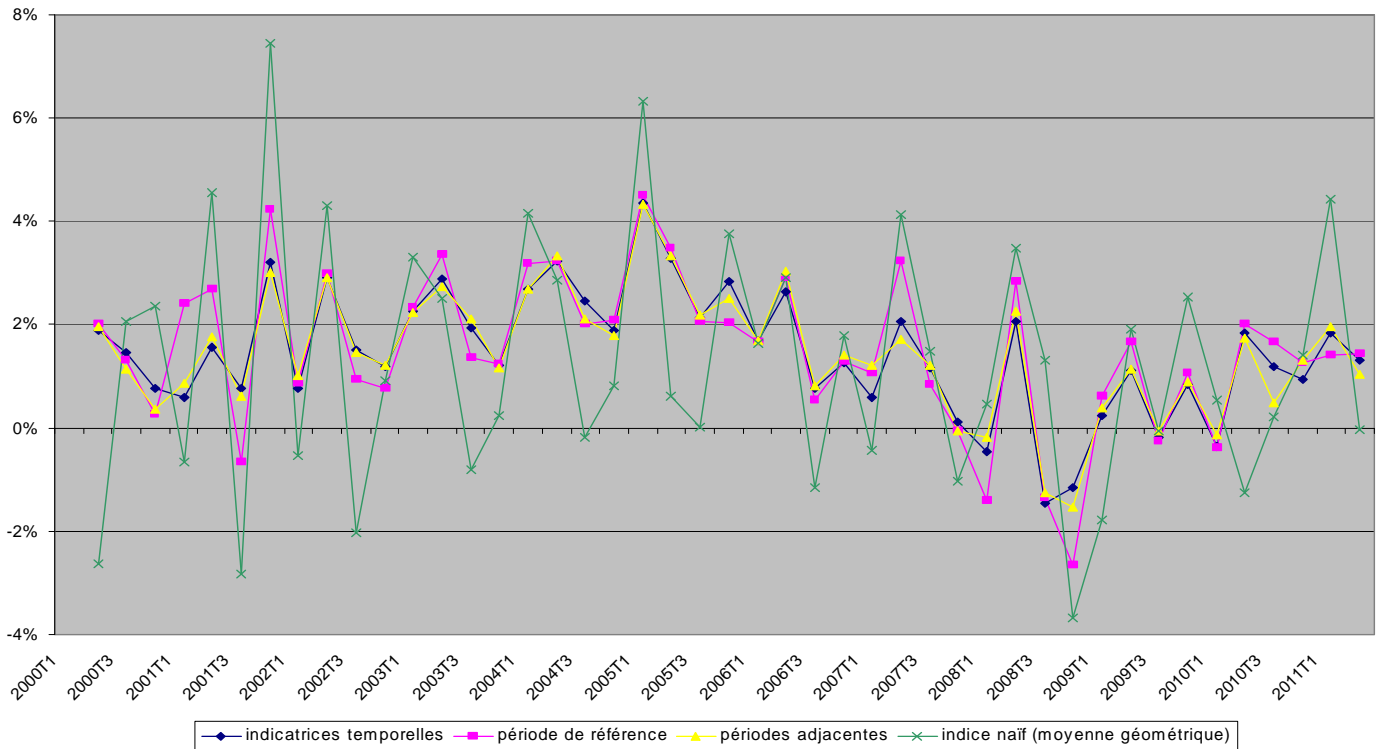
Quant au modèle à « période de référence », la grande difficulté réside dans la détermination de la période de référence et de la durée d'utilisation de ce modèle. De plus, pour utiliser de manière optimale ce modèle, il faudrait se rapprocher de la méthode de calcul de l'indice Notaires-Insee de prix des logements anciens en définissant un parc fixe de logements neufs.

Tout ceci nous a poussé à retenir **le modèle sur périodes adjacentes**.



Graphique 2 : évolution des indices des prix des logements neufs construits à partir des 3 modèles hédoniques considérés et de l'indice naïf « moyenne géométrique » (2000T1-2011T2 - base 100, moyenne annuelle 2009)

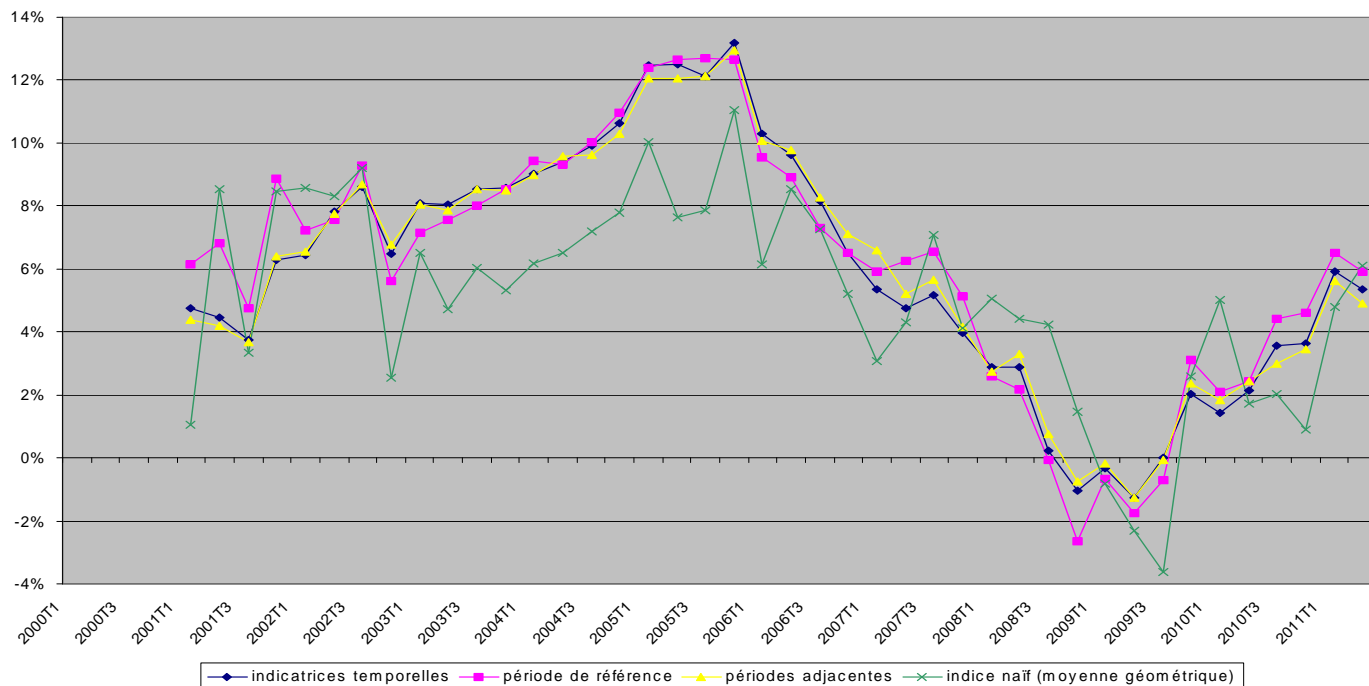
A la lecture de ce graphique, il semblerait que les 3 indices correspondant aux 3 modèles considérés ci-dessus évoluent de la même manière entre le 1^{er} trimestre 2000 et le 2^{ème} trimestre 2011. Cependant, si on considère le taux de croissance en glissement trimestriel, il s'avère que cette première observation est à nuancer :



Graphique 3 : taux de croissance en glissement trimestriel des indices des prix des logements neufs construits à partir des 3 modèles hédoniques considérés et de l'indice naïf « moyenne géométrique » (2000T1-2011T2)

En effet, on constate que le taux de croissance en glissement trimestriel est très souvent plus élevé en valeur absolue pour l'indice construit en utilisant le modèle à « période de référence » que pour les deux autres indices. Ceci montre la nécessité de ré-estimer chaque trimestre les coefficients du modèle hédonique afin de ne pas surestimer les évolutions des prix des logements neufs.

Ce constat est encore vrai si on considère les taux de croissance en glissement annuel :



Graphique 4 : taux de croissance en glissement annuel des indices des prix des logements neufs construits à partir de 3 modèles hédoniques considérés et de l'indice naïf « moyenne géométrique » (2000T1-2011T2)

D'autre part, on constate bien une divergence entre les indices de prix à qualité constante et l'indice naïf. Ceci prouve l'existence d'un effet qualité et la nécessité de le prendre en compte si on veut construire un indice des prix des logements neufs à qualité constante.

5. L'indice des prix des logements (HPI)

Après avoir mis au point un indice des prix des logements neufs, nous avons pu calculer l'indice des prix des logements (HPI) pour la France.

5.1. Définition

L'indice des prix des logements (HPI) a pour but de mesurer l'évolution des prix de marché de tous les logements acquis par les ménages quelle que soit l'utilisation faite de ce logement (résidence principale, résidence secondaire, investissement locatif, ...). Le prix du terrain doit être inclus dans le prix de vente suivi.

5.2. Champ théorique

Le projet de règlement OOH contient une nomenclature partitionnant le champ d'un indice des prix des logements :

Code HPI	Libellé HPI
H.1	ACHATS DE LOGEMENTS
H.1.1.	Achats de logements neufs
H.1.2.	Achats de logements anciens

Tableau 5 : Champ théorique de l'indice des prix des logements

5.3. Champ effectif

Le champ effectif de l'indice des prix des logements coïncide avec son champ théorique :

Code HPI	Libellé HPI	Indice
H.1	ACHATS DE LOGEMENTS	
H.1.1.	Achats de logements neufs	IPLN
H.1.2.	Achats de logements anciens	Indice Notaires-Insee de prix des logements anciens

Tableau 6 : Champ effectif de l'indice des prix des logements

5.4. Méthode de calcul

L'HPI est un indice de Laspeyres chaîné trimestriel. Il est égal à la moyenne pondérée des deux indices qui couvrent le champ effectif de l'HPI, i.e. pour un trimestre t d'une année A donnée, on a :

$$HPI_{A,t} = w_{IPLN,A} IPLN_{A,t} + w_{ILA,A} ILA_{A,t}$$

où :

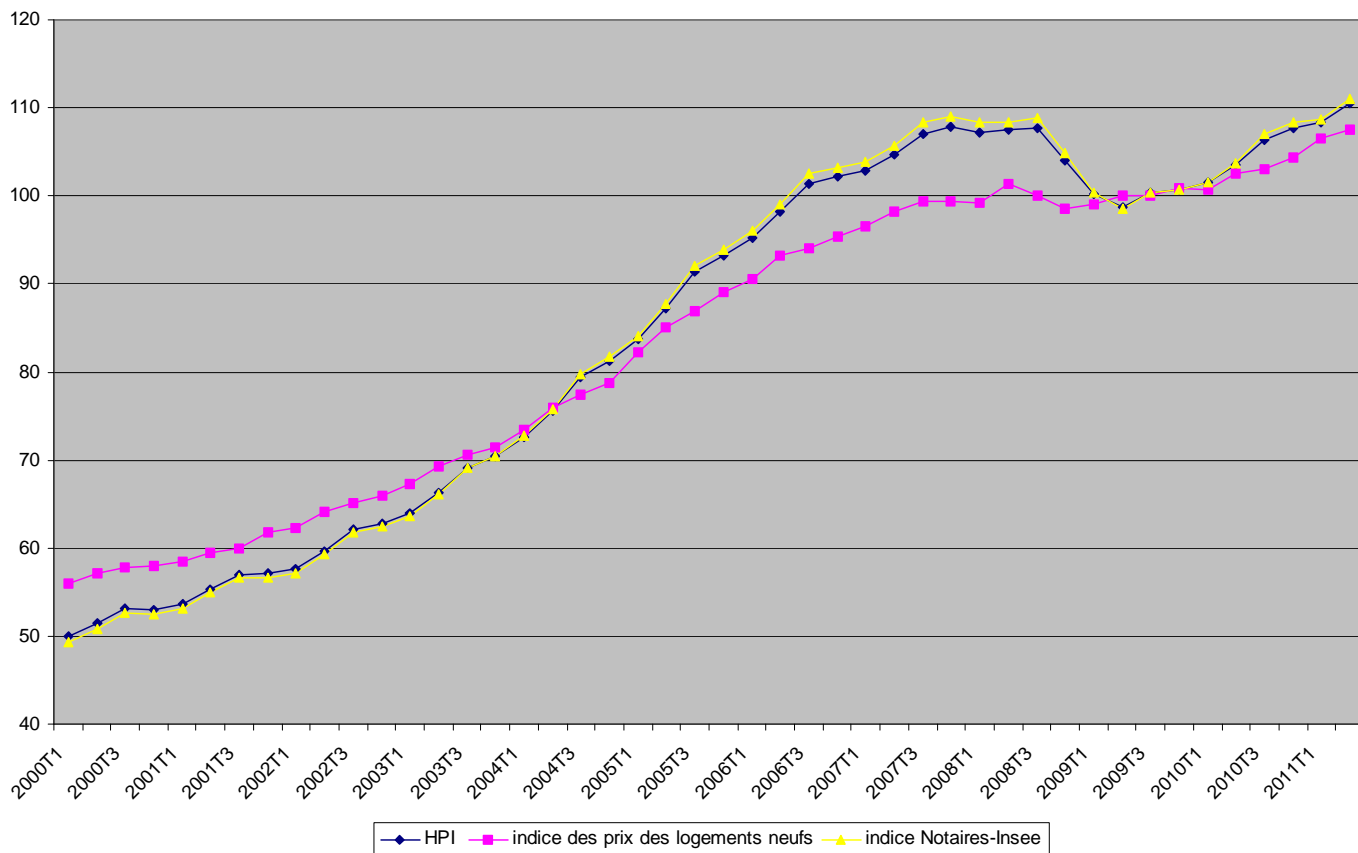
- $HPI_{A,t}$ est l'indice des prix des logements du trimestre t de l'année A considérée
- $IPLN_{A,t}$ est l'indice des prix des logements neufs du trimestre t de l'année A considérée
- $ILA_{A,t}$ est l'indice Notaires-Insee des prix des logements anciens du trimestre t de l'année A considérée
- $w_{IPLN,A}$ et $w_{ILA,A}$ sont les pondérations pour l'année A des deux sous-sections définissant le champ effectif de l'HPI

Les pondérations de l'année A sont calculées à partir des comptes satellites du logement de l'année A-2 (tableau 23 « Les acquisitions de logements et les travaux »). Pour l'année 2010, les pondérations sont les suivantes :

code OOH	Libellé OOH	Pondération 2010
H.1	ACHATS DE LOGEMENTS	
H.1.1.	Achats de logements neufs	13,32%
H.1.2.	Achats de logements anciens	86,68%

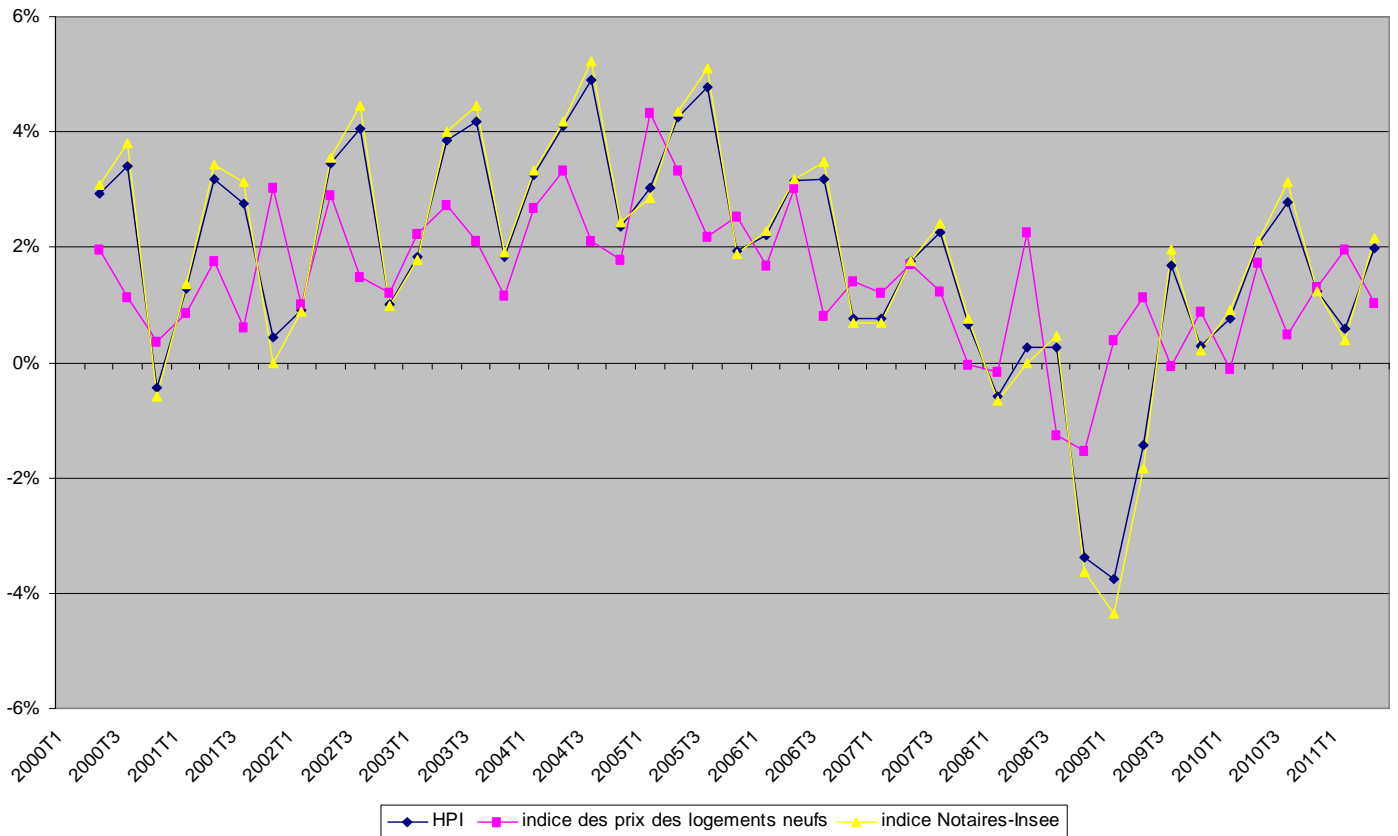
Tableau 7 : pondérations de l'HPI pour l'année 2010 (source : comptes satellites du logement 2008)

5.5. Évolution de l'HPI



Graphique 5 : évolution de l'HPI et de ses 2 composantes (2000T1-2011T2 - base 100, moyenne annuelle 2009)

Étant donné le poids très important des logements anciens, l'indice des prix des logements français évolue de la même manière que l'indice Notaires-Insee des prix des logements anciens. En revanche, ce dernier ne connaît pas la même évolution que l'indice des prix des logements neufs. Ce constat est confirmé par les taux de croissance en glissement trimestriel :



Graphique 6 : taux de croissance en glissement trimestriel de l’HPI et de ses 2 composantes (2000T1-2011T2)

Des travaux complémentaires seront menés afin d’essayer de trouver une explication à cette différence d’évolution. On s’assurera tout d’abord qu’il n’existe pas de biais de méthode, l’indice Notaires-Insee et l’indice des prix des logements neufs n’étant pas calculés selon la même méthodologie. Puis, on se demandera si cette différence n’est pas due à une différence de structure géographique entre le parc des logements neufs et le parc des logements anciens.

Conclusion

La construction d'un indice des prix des logements neufs est une démarche nouvelle. A travers cette étude, nous avons montré que pour rendre compte de l'évolution du marché immobilier neuf en France, on ne pouvait pas se contenter de suivre l'évolution des prix moyens des logements vendus. En effet, pour atteindre un tel objectif, le recours à un modèle hédonique permettant de modéliser le prix des logements neufs en fonction de leurs caractéristiques est inévitable si on veut prendre en compte correctement l'effet qualité.

Cependant, notre indice des prix des logements neufs construit à partir d'un modèle à « périodes adjacentes » est perfectible. En effet, l'influence de la localisation géographique sur le prix des logements pourrait être mieux prise en compte en définissant un meilleur zonage géographique. De plus, la différence d'évolution constatée entre notre indice des prix des logements neufs et l'indice Notaires-Insee des prix des logements anciens ne nous permet pas de valider de manière définitive notre indice.

ANNEXES

Annexe 1 : étude de la volatilité du prix moyen au m² des logements neufs vendus pour chaque couple (ZEAT ;tranche d'unité urbaine) (2000T1-2006T1 - ancienne version de l'enquête)

ZEAT	tranche d'unité urbaine	nombre moyen de logements réservés par trimestre	de prix de vente moyen (en €)	de des logements vendus (en m ²)	surface habitable moyenne des logements vendus (en m ²)	de prix de vente moyen au m ² (en €)	écart-type du prix de vente moyen au m ² (en €)
Région parisienne	unité urbaine de Paris	4 598	207 907,14	68,69	3 026,95	1 260,61	
Centre-est	milieu rural	519	162 165,90	55,03	2 946,79	1 175,74	
Méditerranée	50 000 à 99 999 hab.	469	179 046,69	62,52	2 863,77	1 282,35	
Méditerranée	200 000 à 1 999 999 hab.	2 169	174 918,35	63,91	2 736,75	1 065,49	
Centre-est	5 000 à 9 999 hab.	176	194 114,53	72,42	2 680,25	1 908,71	
Méditerranée	10 000 à 19 999 hab.	361	154 377,05	60,17	2 565,56	913,41	
Méditerranée	5 000 à 9 999 hab.	252	170 749,64	67,21	2 540,63	831,55	
Région parisienne	milieu rural	197	201 169,95	81,29	2 474,62	771,35	
Région parisienne	20 000 à 49 999 hab.	101	156 619,51	63,61	2 462,37	711,94	
Méditerranée	milieu rural	314	136 896,35	56,43	2 426,02	996,54	
Centre-est	100 000 à 199 999 hab.	459	169 067,91	71,94	2 350,04	679,56	
Région parisienne	10 000 à 19 999 hab.	118	172 781,34	75,10	2 300,70	805,21	
Région parisienne	2 000 à 4 999 hab.	33	196 177,77	86,67	2 263,42	504,52	
Centre-est	10 000 à 19 999 hab.	217	160 368,72	71,13	2 254,48	1 048,52	
Bassin parisien	milieu rural	145	151 495,19	68,78	2 202,67	936,96	
Sud-ouest	100 000 à 199 999 hab.	407	132 217,36	60,20	2 196,40	708,75	
Région parisienne	50 000 à 99 999 hab.	21	134 862,48	61,52	2 192,30	480,79	
Sud-ouest	10 000 à 19 999 hab.	220	125 427,71	57,85	2 168,05	575,74	
Sud-ouest	milieu rural	276	118 670,44	55,29	2 146,24	610,49	
Centre-est	200 000 à 1 999 999 hab.	1 692	149 951,92	70,48	2 127,52	577,23	
Ouest	20 000 à 49 999 hab.	326	124 686,71	58,90	2 116,89	532,01	
Sud-ouest	5 000 à 9 999 hab.	237	125 228,27	59,39	2 108,47	556,55	

ZEAT	tranche d'unité urbaine	nombre moyen de logements réservés par trimestre	prix de vente moyen (en €)	surface habitable moyenne des logements vendus (en m²)	prix de vente moyen au m² (en €)	écart-type du prix de vente moyen au m² (en €)
Ouest	200 000 à 1 999 999 hab.	939	136 554,55	64,97	2 101,90	527,23
Sud-ouest	200 000 à 1 999 999 hab.	1 558	132 514,10	63,05	2 101,88	607,44
Sud-ouest	50 000 à 99 999 hab.	415	122 484,97	58,55	2 092,15	517,81
Centre-est	2 000 à 4 999 hab.	143	144 635,03	69,45	2 082,68	742,83
Région parisienne	5 000 à 9 999 hab.	57	150 662,41	72,47	2 078,96	402,59
Centre-est	20 000 à 49 999 hab.	323	144 574,98	69,56	2 078,31	809,48
Méditerranée	100 000 à 199 999 hab.	191	130 049,16	63,02	2 063,45	467,84
Centre-est	50 000 à 99 999 hab.	317	142 549,75	69,11	2 062,59	512,01
Ouest	100 000 à 199 999 hab.	588	128 018,13	62,36	2 052,83	681,69
Sud-ouest	2 000 à 4 999 hab.	198	123 528,27	60,82	2 030,89	634,30
Méditerranée	20 000 à 49 999 hab.	372	129 128,97	63,79	2 024,15	528,19
Méditerranée	2 000 à 4 999 hab.	155	128 249,34	63,58	2 017,01	685,28
Nord	20 000 à 49 999 hab.	88	111 807,15	55,92	1 999,34	475,34
Ouest	5 000 à 9 999 hab.	293	122 993,80	61,78	1 990,95	625,00
Est	200 000 à 1 999 999 hab.	622	137 355,08	69,77	1 968,57	452,99
Bassin parisien	200 000 à 1 999 999 hab.	779	132 814,49	67,59	1 965,01	521,59
Nord	5 000 à 9 999 hab.	19	147 115,04	75,00	1 961,48	665,63
Ouest	10 000 à 19 999 hab.	270	114 625,71	58,60	1 956,14	475,79
Nord	200 000 à 1 999 999 hab.	431	142 158,51	73,12	1 944,13	505,43
Bassin parisien	20 000 à 49 999 hab.	261	125 805,93	64,84	1 940,30	449,71
Ouest	50 000 à 99 999 hab.	262	120 121,40	62,03	1 936,45	499,36
Bassin parisien	10 000 à 19 999 hab.	168	124 196,80	64,26	1 932,75	532,12
Ouest	milieu rural	258	129 775,52	67,33	1 927,52	506,28
Nord	10 000 à 19 999 hab.	29	133 994,19	69,91	1 916,70	558,39
Ouest	2 000 à 4 999 hab.	202	115 448,75	60,45	1 909,74	461,80
Est	50 000 à 99 999 hab.	156	139 175,64	72,89	1 909,29	436,09
Est	10 000 à 19 999 hab.	136	130 938,34	68,69	1 906,28	381,35
Nord	2 000 à 4 999 hab.	29	116 377,24	61,17	1 902,60	504,05
Bassin parisien	100 000 à 199 999 hab.	127	129 470,07	68,32	1 895,13	437,10
Sud-ouest	20 000 à 49 999 hab.	274	105 825,21	55,88	1 893,66	497,56
Bassin parisien	50 000 à 99 999 hab.	260	119 643,17	64,02	1 868,84	386,43
Bassin parisien	2 000 à 4 999 hab.	84	124 361,28	66,98	1 856,75	471,36
Est	2 000 à 4 999 hab.	164	135 994,00	74,43	1 827,17	335,06

ZEAT	tranche d'unité urbaine	nombre moyen de logements réservés par trimestre	prix de vente moyen (en €)	surface habitable moyenne des logements vendus (en m²)	prix de vente moyen au m² (en €)	écart-type du prix de vente moyen au m² (en €)
Est	milieu rural	154	140 241,13	76,78	1 826,54	355,73
Est	5 000 à 9 999 hab.	95	128 910,11	70,82	1 820,34	306,00
Est	20 000 à 49 999 hab.	116	134 275,72	74,21	1 809,48	340,94
Est	100 000 à 199 999 hab.	189	127 446,39	70,50	1 807,80	460,39
Nord	100 000 à 199 999 hab.	41	125 256,56	69,90	1 791,84	380,26
Bassin parisien	5 000 à 9 999 hab.	92	123 340,48	70,38	1 752,53	473,80
Nord	50 000 à 99 999 hab.	66	128 219,15	73,64	1 741,13	468,76
Nord	milieu rural	14	159 201,45	91,62	1 737,60	422,26

Annexe 2 : étude de la volatilité du prix moyen au m² des logements neufs vendus pour chaque couple (ZEAT ; tranche d'unité urbaine) (2006T1-2011T2 - nouvelle version de l'enquête)

ZEAT	tranche d'unité urbaine	nombre moyen de logements réservés par trimestre	prix de vente moyen (en €)	surface habitable moyenne des logements vendus (en m ²)	prix de vente moyen au m ² (en €)	écart-type du prix de vente moyen au m ² (en €)
Région parisienne	unité urbaine de Paris	4 176	263 482,13	61,67	4 272,78	1 552,93
Centre-est	milieu rural	369	247 514,27	60,96	4 060,58	2 023,42
Méditerranée	200 000 à 1 999 999 hab.	3 040	216 554,33	56,34	3 843,46	1 222,50
Méditerranée	milieu rural	312	203 610,21	56,59	3 597,88	1 144,44
Méditerranée	50 000 à 99 999 hab.	744	216 988,79	60,49	3 586,91	1 371,24
Centre-est	2 000 à 4 999 hab.	258	231 827,09	65,85	3 520,34	1 632,67
Centre-est	100 000 à 199 999 hab.	733	229 057,01	66,54	3 442,62	791,37
Méditerranée	10 000 à 19 999 hab.	312	225 610,20	65,93	3 421,74	1 086,99
Région parisienne	5 000 à 9 999 hab.	59	240 211,48	72,20	3 327,19	545,11
Région parisienne	10 000 à 19 999 hab.	185	272 622,49	82,35	3 310,53	549,53
Centre-est	200 000 à 1 999 999 hab.	2 208	189 701,34	58,69	3 232,33	843,10
Centre-est	5 000 à 9 999 hab.	178	215 665,93	67,38	3 200,76	1 514,01
Méditerranée	5 000 à 9 999 hab.	321	201 421,31	63,60	3 167,22	947,97
Région parisienne	milieu rural	47	261 279,36	82,57	3 164,15	471,20
Sud-ouest	200 000 à 1 999 999 hab.	1 973	187 730,64	59,85	3 136,81	633,35
Région parisienne	20 000 à 49 999 hab.	153	198 954,54	63,46	3 135,22	644,53
Région parisienne	50 000 à 99 999 hab.	85	181 302,53	57,91	3 130,99	403,12
Région parisienne	2 000 à 4 999 hab.	67	226 747,02	73,47	3 086,41	448,01

ZEAT	tranche d'unités urbaine	nombre moyen de logements réservés par trimestre	prix de vente moyen (en €)	surface habitable moyenne des logements vendus (en m²)	prix de vente moyen au m² (en €)	écart-type du prix de vente moyen au m² (en €)
Centre-est	10 000 à 19 999 hab.	203	201 084,69	65,50	3 069,78	1 410,96
Méditerranée	2 000 à 4 999 hab.	202	197 452,37	64,75	3 049,32	837,90
Nord	20 000 à 49 999 hab.	46	192 356,04	64,00	3 005,61	1 068,13
Méditerranée	20 000 à 49 999 hab.	261	175 662,89	59,49	2 952,75	657,59
Sud-ouest	50 000 à 99 999 hab.	185	174 046,61	59,79	2 910,81	1 024,93
Ouest	100 000 à 199 999 hab.	407	171 846,27	59,05	2 910,05	881,77
Ouest	200 000 à 1 999 999 hab.	1 379	169 080,91	58,13	2 908,73	749,28
Ouest	20 000 à 49 999 hab.	414	168 273,20	58,03	2 899,77	768,69
Nord	10 000 à 19 999 hab.	22	214 546,43	74,51	2 879,50	1 455,97
Méditerranée	100 000 à 199 999 hab.	268	156 600,73	54,62	2 866,88	464,59
Centre-est	20 000 à 49 999 hab.	227	182 410,74	64,75	2 817,07	1 316,93
Nord	100 000 à 199 999 hab.	23	205 533,58	73,20	2 807,88	806,83
Bassin parisien	100 000 à 199 999 hab.	297	167 331,72	59,69	2 803,16	589,78
Bassin parisien	200 000 à 1 999 999 hab.	926	174 637,69	62,68	2 786,40	601,39
Centre-est	50 000 à 99 999 hab.	458	171 198,53	62,18	2 753,45	643,18
Bassin parisien	20 000 à 49 999 hab.	330	169 263,93	61,55	2 750,22	830,30
Nord	200 000 à 1 999 999 hab.	617	180 324,86	65,65	2 746,72	725,72
Sud-ouest	10 000 à 19 999 hab.	190	156 553,67	57,27	2 733,79	773,96
Est	200 000 à 1 999 999 hab.	569	177 257,33	66,02	2 684,72	517,57
Bassin parisien	10 000 à 19 999 hab.	168	171 465,76	64,37	2 663,73	582,63
Nord	5 000 à 9 999 hab.	20	205 418,42	78,38	2 620,71	762,88
Ouest	50 000 à 99 999 hab.	337	150 060,74	57,38	2 615,17	760,52
Bassin parisien	2 000 à 4 999 hab.	144	163 813,88	63,00	2 600,32	1 117,55
Sud-ouest	milieu rural	251	149 674,45	57,98	2 581,48	766,01
Nord	2 000 à 4 999 hab.	30	176 175,59	68,75	2 562,71	557,54

ZEAT	tranche d'unités urbaine	nombre moyen de logements réservés par trimestre	prix de vente moyen (en €)	surface habitable moyenne des logements vendus (en m²)	prix de vente moyen au m² (en €)	écart-type du prix de vente moyen au m² (en €)
Sud-ouest	5 000 à 9 999 hab.	199	150 686,15	59,17	2 546,62	535,36
Bassin parisien	50 000 à 99 999 hab.	347	154 574,61	60,95	2 536,19	513,43
Sud-ouest	100 000 à 199 999 hab.	180	153 308,13	60,68	2 526,30	555,21
Est	100 000 à 199 999 hab.	211	166 774,93	66,16	2 520,78	458,58
Ouest	5 000 à 9 999 hab.	334	148 221,42	58,94	2 514,60	709,81
Bassin parisien	milieu rural	126	177 667,09	71,23	2 494,20	960,04
Sud-ouest	2 000 à 4 999 hab.	156	148 152,85	59,49	2 490,46	836,83
Ouest	2 000 à 4 999 hab.	269	152 734,56	61,71	2 475,22	978,59
Sud-ouest	20 000 à 49 999 hab.	182	133 397,95	54,15	2 463,42	469,21
Ouest	10 000 à 19 999 hab.	227	146 047,71	59,54	2 453,03	646,08
Est	5 000 à 9 999 hab.	120	172 397,60	70,80	2 435,05	376,11
Est	50 000 à 99 999 hab.	239	156 179,75	64,34	2 427,50	430,18
Est	10 000 à 19 999 hab.	125	162 855,15	67,18	2 424,34	380,33
Ouest	milieu rural	179	158 477,99	65,50	2 419,40	865,05
Est	20 000 à 49 999 hab.	157	163 832,13	68,71	2 384,28	418,64
Est	2 000 à 4 999 hab.	179	171 291,11	72,27	2 370,19	361,51
Nord	50 000 à 99 999 hab.	106	159 952,66	68,18	2 345,93	507,60
Bassin parisien	5 000 à 9 999 hab.	108	152 556,54	65,31	2 335,79	653,12
Nord	milieu rural	24	210 977,51	92,20	2 288,27	713,10
Est	milieu rural	172	170 551,18	74,58	2 286,91	349,75

Annexe 3 : étude de la volatilité du prix moyen au m² des logements neufs vendus par classe géographique issue de la CAH (2000T1-2006T1 - ancienne version de l'enquête)

classes géographiques issues de la CAH	nombre moyen de logements réservés par trimestre	prix de vente moyen (en €)	surface habitable moyenne des logements vendus (en m ²)	prix de vente moyen au m ² (en €)	écart-type du prix de vente moyen au m ² (en €)
10	406	197 704,81	77,88	2 538,70	1 383,88
15	4 598	207 907,14	68,69	3 026,95	1 260,61
5	749	161 591,77	60,36	2 677,30	1 194,30
13	2 638	175 652,47	63,67	2 758,93	1 107,86
12	462	154 866,10	60,92	2 542,06	872,34
11	829	170 108,60	70,95	2 397,45	746,82
7	2 039	136 806,20	65,88	2 076,55	626,52
8	1 895	150 091,82	70,41	2 131,65	609,74
4	2 910	132 654,96	64,37	2 060,89	600,10
3	1 541	142 541,78	71,58	1 991,44	599,56
2	2 177	128 872,83	65,50	1 967,54	560,32
1	2 334	124 144,82	61,30	2 025,36	553,61
9	798	119 463,21	60,35	1 979,65	524,71
14	274	105 825,21	55,88	1 893,66	497,56
6	589	114 573,72	58,96	1 943,19	473,23

Annexe 4 : étude de la volatilité du prix moyen au m² des logements neufs vendus par classe géographique issue de la CAH (2006T1-2011T2 - nouvelle version de l'enquête)

classes géographiques issues de la CAH	nombre moyen de logements réservés par trimestre	prix de vente moyen (en €)	surface habitable moyenne des logements vendus (en m²)	prix de vente moyen au m² (en €)	écart-type du prix de vente moyen au m² (en €)
13	428	246 512,55	62,50	3 944,36	1 909,07
11	4 223	263 457,59	61,90	4 256,30	1 549,63
8	4 008	216 551,39	57,92	3 738,97	1 280,78
2	1 234	201 107,48	62,73	3 206,09	1 072,43
9	1 370	228 681,56	66,61	3 433,21	1 061,34
10	929	180 924,61	64,72	2 795,53	891,21
12	4 226	188 810,04	59,29	3 184,66	760,70
7	2 631	168 594,27	59,36	2 840,08	738,86
1	426	163 538,28	66,33	2 465,36	737,48
4	1 505	148 866,94	58,61	2 540,02	718,71
5	1 503	171 436,15	64,88	2 642,24	687,00
3	1 885	155 440,09	61,28	2 536,53	682,02
6	2 096	175 627,01	63,53	2 764,40	669,65
14	185	272 622,49	82,35	3 310,53	549,53
15	182	133 397,95	54,15	2 463,42	469,21

Bibliographie

- [1] Les indices Notaires-Insee de prix des logements anciens, version 2 des modèles hédoniques, INSEE méthodes n°111
- [2] Gilbert Saporta, « Probabilités, analyse des données, et statistique », *Editions TECHNIP*
- [3] Russell Davidson, James G. MacKinnon, « Estimation and Inference in Econometrics », *Oxford University Press, 1993*
- [4] H.Erkel-Rousse, « Détection de la multi colinéarité dans un modèle linéaire ordinaire :quelques éléments pour un usage averti des indicateurs de Besley, Kuh et Welsh », *Revue de statistique appliquée, tome 43, n°4 (1995), p. 19-42*