

**Direction des Statistiques Démographiques et Sociales**

**F1808**

**La construction d'un indice des prix des  
logements neufs**

**Thomas Balcone**

**Document de travail**



**Institut National de la Statistique et des Études Économiques**



**INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES**

Série des Documents de Travail  
de la  
DIRECTION DES STATISTIQUES DÉMOGRAPHIQUES ET SOCIALES

**F1808**

**La construction d'un indice des prix des logements neufs**

THOMAS BALCONE \*

**Document de travail**

**décembre 2018**

---

\* Division des prix à la consommation au moment du lancement de ce travail

Ces documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.  
Working-papers do not reflect the position of INSEE but only their authors' views.



# **LA CONSTRUCTION D'UN INDICE DES PRIX DES LOGEMENTS NEUFS**

## **THE DRAFTING OF A NEW DWELLINGS INDEX**

Décembre 2018

Thomas BALCONE



## Résumé

Dans le cadre d'un projet européen (le projet Owner-Occupied Housing - OOH), des travaux ont été menés au cours des années 2000 afin de construire un indice des prix des logements (HPI en anglais pour *House Price Index*). Cet indice comprend deux composantes : une pour les logements anciens et l'autre pour les logements neufs. En France, un indice couvrait déjà le champ de l'ancien (l'indice Notaires-Insee des prix des logements anciens) alors qu'il n'existait pas d'indice pour le neuf. Ainsi, l'Insee a dû créer un indice de l'évolution des prix des logements neufs. Les travaux menés au sein de la Division des prix à la consommation ont permis la publication d'un indice du neuf en janvier 2013 portant sur la période 2006T1-2012T3, conformément aux directives européennes.

Nous avons, tout d'abord, mis au point un modèle hédonique en utilisant les données issues de l'enquête de Commercialisation des logements neufs (ECLN) du Service de l'Observation et des Statistiques (SOeS) du Ministère de l'Écologie, du Développement durable, des Transports et du Logement. On obtient alors un indice de prix expurgé des effets qualité en utilisant la méthode dite sur périodes adjacentes : le modèle est enrichi d'une indicatrice temporelle puis il est estimé chaque trimestre en utilisant les données ECLN des deux derniers trimestres.

Après avoir constaté des différences d'évolution entre l'indice du neuf et l'indice Notaires-Insee des prix des logements anciens sur la période 2006T1-2012T3, deux études méthodologiques complémentaires ont été menées. Ces travaux nous ont permis de nous assurer que ces divergences ne correspondaient pas à un biais de méthode (1<sup>ère</sup> étude) et ne provenaient pas d'une mauvaise prise en compte de la géographie dans le modèle sur périodes adjacentes (2<sup>nde</sup> étude). De fait, la grande majorité d'entre elles est imputable à la différence de nature des logements.

Suite à la première diffusion de l'indice du neuf en janvier 2013, nous avons poursuivi les analyses pour essayer d'améliorer la qualité du modèle hédonique. Cela nous a conduits à légèrement réviser sa spécification, et à utiliser cinq trimestres glissants plutôt que deux. Le nouvel indice du neuf ainsi obtenu a été publié à compter du premier trimestre 2017.

Mots-clés : indices, modèles hédoniques, simulation, demande de logements

## Abstract

As part of a European project (the Owner-Occupied Housing project - OOH), works were carried out in the 2000s to draft a House Price Index (HPI). This index has two components: one for existing dwellings and the other for new dwellings. In France, an index already covered the former ones (the *Notaires-Insee* index of prices of existing dwellings) while there was no index for the latter. Thus, INSEE had to create an index of the evolution of new housing prices. The works carried out within the Consumer Prices Division led to the publication of an index for new dwellings in January 2013 covering the period 2006Q1-2012Q3, in accordance with European directives.

A hedonic model was first developed, using data from the survey on the commercialisation of new dwellings (in French *Enquête sur la commercialisation des logements neufs*, ECLN) conducted by the Observation and Statistics Service (SOeS) of the Ministry of Ecology, Sustainable Development, Transport and Housing. A price index purged of quality effects is then obtained by using the so-called adjacent period method: the model is enriched with a time dummy and then estimated quarterly using ECLN data from the last two quarters.

After noting differences in the evolution between the new homes index and the *Notaires-Insee* index of existing homes' prices over the period 2006Q1-2012Q3, two additional methodological studies were carried out. These works allowed us to ensure that these divergences did not correspond to a method bias (1<sup>st</sup> study) and did not result from not taking geography well enough into account in the adjacent period model (2<sup>nd</sup> study). Indeed the vast majority of the divergences are due to the difference in the nature of the dwellings themselves.

Following the first release of the new dwellings index in January 2013, the analyses were continued to try to improve the quality of the hedonic model. This led us to slightly revise its specification, and to use five rolling quarters instead of two. The renewed new dwellings index thus obtained was published from the first quarter of 2017 onwards.

Keywords : indices, hedonic models, simulation, housing demand



## **Remerciements**

Je remercie très sincèrement Anne Laferrère d'avoir accepté d'encadrer mes travaux. Son expertise sur les indices des prix des logements a été précieuse.

Je tiens également à adresser mes remerciements à Patrick Sillard, Axelle Chauvet-Peyrard, Catherine Rougerie, Hélène Thélot, Cédric Cailly et Ismaël Haffoud pour leurs remarques toujours pertinentes.



## Table des matières

<b>PREAMBULE</b> .....	<b>8</b>
<b>INTRODUCTION</b> .....	<b>9</b>
<b>1. L'INDICE DES PRIX DES LOGEMENTS (IPL) ET L'INDICE DES PRIX DES LOGEMENTS NEUFS (IPLN) PUBLIES EN JANVIER 2013</b> .....	<b>10</b>
1.1. LA GENESE : LE PROJET OOH (OWNER-OCCUPIED HOUSING).....	10
1.2. L'INDICE DES PRIX DES LOGEMENTS (IPL) .....	11
1.2.1. <i>Champ et définition</i> .....	11
1.2.2. <i>Méthode de calcul</i> .....	12
1.3. L'INDICE DES PRIX DES LOGEMENTS NEUFS (IPLN) .....	13
1.3.1. <i>La source utilisée : l'Enquête sur la commercialisation des logements neufs (ECLN)</i> .....	13
1.3.2. <i>Le champ de l'IPLN</i> .....	16
1.3.3. <i>La méthode de calcul de l'IPLN</i> .....	16
<b>2. ÉTUDES COMPLEMENTAIRES SUR LES INDICES DES PRIX DES LOGEMENTS NEUFS ET ANCIENS</b> .....	<b>22</b>
2.1. PREMIERE ETUDE : TEST DE SENSIBILITE DE L'INDICE DES PRIX DES LOGEMENTS ANCIENS AU CHOIX DE METHODES DE CALCUL.....	24
2.1.1. <i>Les données sur les logements anciens utilisées</i> .....	24
2.1.2. <i>La méthode de calcul utilisée : la méthode sur périodes adjacentes</i> .....	25
2.1.3. <i>Les résultats : comparaison des indices expérimentaux « 1<sup>ère</sup> simulation » avec les indices Notaires-Insee</i> .....	27
2.1.4. <i>Conclusions de l'étude</i> .....	31
2.2. SECONDE ETUDE : TEST DE SENSIBILITE DE L'INDICE DES PRIX DES LOGEMENTS ANCIENS A LA DISTRIBUTION GEOGRAPHIQUE DES LOGEMENTS.....	31
2.2.1. <i>Les données utilisées</i> .....	32
2.2.2. <i>Constitution de la population des « clones » de logements neufs</i> .....	32
2.2.3. <i>Calcul d'un indice expérimental annuel des prix des logements anciens pour chaque type de construction à partir de la population des « clones »</i> .....	33
2.2.4. <i>Les résultats : comparaison de l'indice « clone » avec l'IPLN et l'indice expérimental « 1<sup>ère</sup> simulation » pour chaque type de construction</i> .....	34
2.2.5. <i>Conclusions de l'étude</i> .....	37
<b>3. LA POURSUITE DE LA REFLEXION SUR LE MODELE DE REGRESSION A LA BASE DU CALCUL DE L'IPLN SUR LA PERIODE 2006T1-2012T3</b> .....	<b>39</b>
3.1. DIAGNOSTICS .....	39
3.1.1. <i>Rappel : une régression pondérée</i> .....	40
3.1.2. <i>(H1) : la non colinéarité des régresseurs</i> .....	41
3.1.3. <i>(H2)-(H5) : analyse graphique des résidus</i> .....	43
3.1.4. <i>Traitement des observations aberrantes et influentes</i> .....	43
3.1.5. <i>Étude des estimateurs des paramètres du modèle</i> .....	44
3.1.6. <i>Qualité d'ajustement du modèle</i> .....	44
3.2. VERS UNE NOUVELLE SPECIFICATION DU MODELE (1).....	44
3.2.1. <i>Un examen des variables explicatives</i> .....	44
3.2.2. <i>Des modèles alternatifs</i> .....	47
<b>4. CONCLUSION ET PRECONISATIONS</b> .....	<b>50</b>
<b>ANNEXES</b> .....	<b>51</b>
<b>BIBLIOGRAPHIE</b> .....	<b>142</b>

## Préambule

Le présent document de travail a été rédigé à partir d'un rapport de stage de master de statistique publique (méthodologie de la statistique publique) réalisé entre avril et août 2013, sur la mise en place d'un indice des prix des logements neufs à l'Insee. Il présente donc, sauf mention contraire, les choix réalisés à cette époque pour assurer la réponse à une demande européenne de production d'une série de 2006 à 2012. La première série d'indices des prix des logements neufs a été publiée en janvier 2013.

L'indice des prix des logements neufs a continué d'être publié à méthodologie constante, mais des aménagements ponctuels ont eu lieu (notamment sur les choix des variables prises en compte dans la régression en raison de la refonte programmée de l'enquête dont sont issues les données). Ce document de travail ne présente donc pas exactement la méthode effectivement utilisée jusqu'en 2016.

Par ailleurs, début 2017, une refonte de l'Enquête sur la commercialisation des logements neufs produite par le Service de l'Observation et des Statistiques (SOeS) du Ministère de l'Écologie, du Développement durable, des Transports et du Logement<sup>1</sup> a atterri. Cette enquête est utilisée par l'Insee pour le calcul de l'indice de prix des logements neufs (*cf. infra*), celui-ci a donc été mis à jour en conséquence (*cf. annexe 0*) Le champ de l'indice a été légèrement modifié (avec notamment l'ajout de la Corse). Les données sont désormais disponibles au niveau logement, alors qu'elles étaient renseignées en moyenne par tranches de commercialisation des programmes immobiliers avant la refonte. Au-delà de ce gain en précision, une actualisation du modèle hédonique utilisé a également été effectuée pour l'occasion, pour suivre les préconisations du présent document de travail : actualisation des variables utilisées, estimation du modèle sur cinq trimestres au lieu de deux. Ces modifications ont eu un impact assez faible, l'indice n'a donc pas été rétropolé.

---

<sup>1</sup> Il s'agit depuis juin 2017 du service de la donnée et des études statistiques (SDES) du ministère de la Transition écologique et solidaire. On l'appelle SOeS dans l'ensemble de ce document.

## Introduction

Dans le cadre du projet OOH (Owner-Occupied Housing) conduit par Eurostat, des travaux ont été menés au cours des années 2000 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 (OOH index en anglais pour *Owner-Occupied Housing index*), et un indice des prix des logements (HPI en anglais pour *House Price Index*). Nous nous concentrons ici uniquement sur la construction de ce second indice.

L'indice des prix des logements comprend deux composantes : une pour les logements anciens et l'autre pour les logements neufs. En France, un indice couvrait déjà le champ de l'ancien (l'indice Notaires-Insee des prix des logements anciens) alors qu'il n'existait pas d'indice pour le neuf. Ainsi, la Division des prix à la consommation de l'Insee a mis en place un indice mesurant l'évolution des prix des logements neufs.

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. Pour tenir compte de cet effet qualité, une modélisation des prix des logements est nécessaire. Ainsi, nous avons établi un modèle hédonique expliquant le prix d'un logement neuf en fonction de ses caractéristiques physiques (surface habitable, nombre de pièces, ...) et des caractéristiques géographiques de la commune de résidence (taille de l'unité urbaine, zone d'études et d'aménagement du territoire (ZEAT), ...). Pour mettre au point ce modèle, nous avons utilisé les données issues de l'Enquête sur la commercialisation des logements neufs (ECLN).

Afin de s'assurer de la qualité de l'indice hédonique des prix de transaction de logements neufs ainsi obtenu, nous avons ensuite mené plusieurs travaux méthodologiques. Nous avons notamment cherché à vérifier que les différences d'évolution constatées entre l'indice du neuf et de l'ancien ne constituaient pas des artefacts de méthode, les indices du neuf et de l'ancien ne reposant pas sur la même méthode de calcul, ni sur la même source de données<sup>2</sup>.

Le premier chapitre rappelle le contexte dans lequel s'est inscrite la création d'un indice des prix des logements aux niveaux européen et national. Il donne également une présentation de cet indice pour la France et de l'indice des prix des logements neufs publiés depuis janvier 2013.

Le deuxième chapitre revient sur les deux études méthodologiques qui ont permis la publication de ces deux indices à partir du 1<sup>er</sup> trimestre de l'année 2006.

Nous étudions dans le troisième chapitre la qualité du modèle de régression à la base du calcul de l'indice des prix des logements neufs sur la période allant du 1<sup>er</sup> trimestre de l'année 2006 au troisième trimestre de l'année 2012. Ceci nous amène à proposer un nouvel indice du neuf basé sur un modèle hédonique amélioré par rapport au modèle initial. De fait, l'Insee profitera des mises à jour nécessaires au calcul de l'IPLN suite à la refonte de l'ECLN en 2017 pour suivre les préconisations méthodologiques de ce rapport.

Enfin, le quatrième et dernier chapitre fait le bilan de notre étude et recense nos préconisations pour l'avenir.

---

<sup>2</sup> L'indice Notaires-Insee des prix des logements anciens est calculé à partir des bases notariales selon la méthode V3 (cf. *bibliographie*).

## **1. L'indice des prix des logements (IPL) et l'indice des prix des logements neufs (IPLN) publiés en janvier 2013**

Le marché du logement est suivi attentivement par de nombreux agents économiques, comme par exemple les ménages, les investisseurs, les promoteurs immobiliers, ou encore les décideurs politiques. Jusqu'en 2013, aucun indicateur unifié portant sur l'évolution des prix sur ce marché dans l'Union européenne n'était publié. Chaque pays publiait les données statistiques qu'il souhaitait. En France, l'Insee produisait, en collaboration avec les Notaires, les indices Notaires-Insee des prix des logements anciens. Quant au marché du logement neuf, il était couvert par la diffusion par le SOeS du nombre de logements mis en chantier, du nombre de logements neufs réservés (i.e. vendus) et du prix de vente moyen<sup>3</sup>.

Afin de disposer d'indicateurs validés au niveau européen, Eurostat a initié le projet OOH, dont l'un des objectifs principaux était l'élaboration d'un indice des prix des logements pour l'Union européenne et la zone euro. Chaque pays a été invité à développer un indice des prix des logements. La France disposant déjà d'indices de qualité couvrant le marché de l'ancien, les indices Notaires-Insee, seule l'élaboration d'un indice des prix des logements neufs a été nécessaire.

### **1.1. La genèse : le 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 frais liés à l'habitation pour les propriétaires occupant leur logement dans l'indice des prix à la consommation a été mis en place. L'objectif de ce projet est double. D'une part, il s'agit 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 (OOH index en anglais - indice OOH ; cf. annexe 1). D'autre part, la création d'un indice des prix des logements doit permettre de suivre l'évolution des prix immobiliers et de pouvoir ainsi anticiper une éventuelle bulle immobilière (HPI en anglais - IPL). Il est important de rappeler que l'indice OOH et l'indice des prix des logements ayant été élaborés pour répondre à deux besoins distincts, ils ne sont pas directement comparables (cf. annexe 2).

L'indice des prix des logements déflaté par l'indice des prix à la consommation constitue l'un des indicateurs du tableau de bord défini dans le cadre de la Procédure de Déséquilibre Macroéconomique (PDM). Ce tableau de bord, suivi par la Commission européenne, se compose de dix indicateurs économiques, financiers et structurels pertinents pour la détection précoce de déséquilibres macro-économiques apparaissant dans les États membres.

Dans la suite de ce document, nous nous intéressons uniquement à l'indice des prix des logements.

---

<sup>3</sup> En euros/m<sup>2</sup> pour les appartements.

## 1.2. L'indice des prix des logements (IPL)

### 1.2.1. Champ et définition

L'indice des prix des logements (IPL) est un indice de prix de transaction trimestriel mesurant l'évolution pure<sup>4</sup> des prix des logements achetés par les ménages. L'IPL couvre l'ensemble des types de logements neufs ou anciens (appartements, maisons individuelles, ...) hors logements atypiques (hôtels particuliers, ateliers, ...), à usage d'habitation (quel qu'il soit : résidence principale, résidence secondaire, investissement locatif).

Seuls les prix de transaction sont retenus. Les logements construits à l'initiative de l'acquéreur (*self-built dwellings* en anglais) sont, par conséquent, exclus du champ, car dans le cas où le ménage est à la fois le conducteur des travaux et l'utilisateur final du logement, il n'y a pas de prix de transaction<sup>5</sup>. Cette exclusion n'a cependant qu'un faible impact sur l'évolution de l'indice des prix des logements neufs (voir Balcone & Laferrère 2015).

Par ailleurs, le prix retenu inclut le prix du terrain.

Au champ de l'indice des prix des logements est associée la nomenclature suivante :

**Tableau 1.1 : Champ de l'indice des prix des logements**

Code	Libellé
H.1.	ACHATS DE LOGEMENTS
H.1.1.	Achats de logements nouveaux <sup>6</sup>
H.1.2.	Achats de logements existants

Le champ géographique de l'IPL français est la France métropolitaine<sup>7</sup>. La sous-section H.1.2. est couverte par l'indice Notaires-Insee des prix des logements anciens. Il a fallu construire un indice des prix des logements neufs, l'IPLN, pour couvrir l'autre sous-section. L'élaboration de cet indice est l'objet du présent document.

---

<sup>4</sup> i.e. hors changement de qualité des biens acquis.

<sup>5</sup> Les maisons neuves achetées « clés en main » par des ménages, c'est-à-dire les maisons choisies sur plan ou dans un état de construction intermédiaire, peuvent être considérées comme des exceptions. En effet, pour ce type d'acquisitions, un prix de transaction est observé.

<sup>6</sup> Voir la définition précise des logements nouveaux (ou neufs) et existants (ou anciens) en annexe 3.

<sup>7</sup> La Corse était également exclue jusqu'à la publication de l'indice du 1<sup>er</sup> trimestre 2017, avec un impact très faible sur l'indice, voir note 17.

## 1.2.2. Méthode de calcul

L'indice des prix des logements français est un indice trimestriel chaîné annuellement. Pour un trimestre donné, il est obtenu comme la moyenne pondérée des deux indices constituant ses deux composantes :

- l'indice des prix des logements neufs (IPLN) pour les logements nouveaux ;
- l'indice Notaires-Insee des prix des logements anciens (IPLA) pour les logements existants<sup>8</sup>.

Plus précisément, l'indice des prix des logements du trimestre T de l'année A est calculé en deux étapes :

- **1)** on calcule tout d'abord la moyenne pondérée des deux indices exprimés en base 100 au trimestre 4 de l'année A-1 :

$$IPL_{A,T}^{A-1,T4} = w_{IPLN,A} IPLN_{A,T}^{A-1,T4} + w_{IPLA,A} IPLA_{A,T}^{A-1,T4}$$

où  $w_{IPLN,A}$  et  $w_{IPLA,A}$  sont respectivement les poids du neuf et de l'ancien pour l'année A ;

puis

- **2)** on procède au chaînage de l'indice obtenu pour aboutir à un indice exprimé en base 100, moyenne annuelle 2010<sup>9</sup> :

$$IPL_{A,T}^{MA,2010} = \frac{IPL_{A,T}^{A-1,T4} IPL_{A-1,T4}^{MA,2010}}{100}$$

Les poids des deux secteurs pour l'année A,  $w_{IPLN,A}$  et  $w_{IPLA,A}$ , correspondent à la part respective du neuf et de l'ancien dans les dépenses consacrées par les ménages à l'acquisition de logements au cours de l'année A-2. Ces dépenses sont issues des comptes satellites du logement publiés par le SOeS (cf. annexe 4).

Le poids de l'ancien est beaucoup plus important que celui du neuf :

**Tableau 1.2 : Poids (en %) du neuf et de l'ancien sur la période 2006-2012**

Code	Libellé	Pondération (en %)						
		2006	2007	2008	2009	2010	2011	2012
H.1	Achats de logements	100	100	100	100	100	100	100
H.1.1.	Achats de logements nouveaux	11,2	11,6	11,8	12,7	13,3	14,0	13,2
H.1.2.	Achats de logements existants	88,8	88,4	88,2	87,3	86,7	86,0	86,8

source : comptes satellites du logement - SOeS

<sup>8</sup> Série brute (i.e. non corrigée des variations saisonnières).

<sup>9</sup> On précise que la base est modifiée en 2017 pour passer à la base 100 en moyenne en 2015.



### 1.3. L'indice des prix des logements neufs (IPLN)

Pour construire un indice des prix des logements neufs, il nous a fallu choisir une base de données contenant un maximum d'informations sur les logements neufs achetés par les ménages, avec un souci de production trimestrielle (les indices de prix calculés doivent représenter le marché au plus proche temporellement de la date de calcul). Notre choix s'est porté sur l'Enquête sur la commercialisation des logements neufs (ECLN) du SOeS.

#### 1.3.1. La source utilisée : l'Enquête sur la commercialisation des logements neufs (ECLN)

##### 1.3.1.1. Présentation de l'enquête

L'ECLN est une enquête **trimestrielle** assurant le suivi de la commercialisation des logements neufs destinés à la vente aux particuliers depuis 1985<sup>10</sup>. Le champ de l'enquête est constitué des permis de construire de 5 logements et plus destinés à la vente aux particuliers et situés en France métropolitaine. 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. Sont exclus en revanche les permis de construire de 1 à 4 logements, qui représentent une part négligeable du marché (<5% en termes de logements commencés chaque trimestre, entre 2006 et 2012, cf. annexe 7).

Le questionnaire associé à l'enquête est composé de deux parties<sup>11</sup> : le questionnaire de base et les fiches tranches (cf. annexe 6). Le questionnaire de base décrit le programme de construction. Il est rempli pour tout nouveau programme suivi par l'enquête. Il permet notamment de connaître l'adresse du programme, son niveau de standing (courant ou ordinaire, confortable, très confortable, luxe) ou encore ses éléments de confort (présence d'un dispositif de sécurité, présence d'une piscine dans la résidence, ...).

La deuxième partie du questionnaire, les fiches tranches, 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. En outre, elle ne comprend que des logements de même type (individuel groupé ou collectif) et de même nature (logements ordinaires, résidences avec services, résidences de vacances ou logements sociaux).

Chaque trimestre, une fiche tranche donne notamment, **suivant le nombre de pièces des logements** :

- **le nombre total de logements réservés**, i.e. le nombre de logements ayant fait l'objet d'une réservation avec dépôt d'arrhes ou vendus sans réservation préalable ;
- le nombre de logements réservés pour investissement locatif ;
- la surface moyenne des terrains, balcons et terrasses en m<sup>2</sup> des logements réservés ;
- **la surface habitable<sup>12</sup> moyenne** en m<sup>2</sup> des logements réservés ;
- **le prix moyen** en euros des logements réservés.

<sup>10</sup> Voir la fiche méthodologique de l'enquête en annexe 5.

<sup>11</sup> Une refonte de l'ECLN a été mise en oeuvre en 2017 et modifie entre autres son questionnaire.

<sup>12</sup> i.e. la surface de plancher construite, après déduction des surfaces occupées par les murs, cloisons, marches et cages d'escaliers, gaines, embrasures de portes et de fenêtres (cf. article R111-2 du Code de la construction).

**Le prix au sens d'ECLN** correspond au prix à la réservation, i.e. à la signature du premier contrat. La TVA est incluse dans ce prix, alors que les frais de notaires ne le sont pas. 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.

### 1.3.1.2. *Les raisons de ce choix*

Une autre source possible pour établir l'indice des prix des logements neufs aurait été, comme pour les indices Notaires-Insee des prix des logements anciens, les bases notariales. La principale motivation aurait résidé dans le fait que ces bases ne présentent pas les deux défauts majeurs de l'ECLN. D'une part, elles couvrent les logements neufs correspondant à des programmes de construction de 1 à 4 logements et d'autre part, elles enregistrent les prix unitaires des transactions<sup>13</sup>.

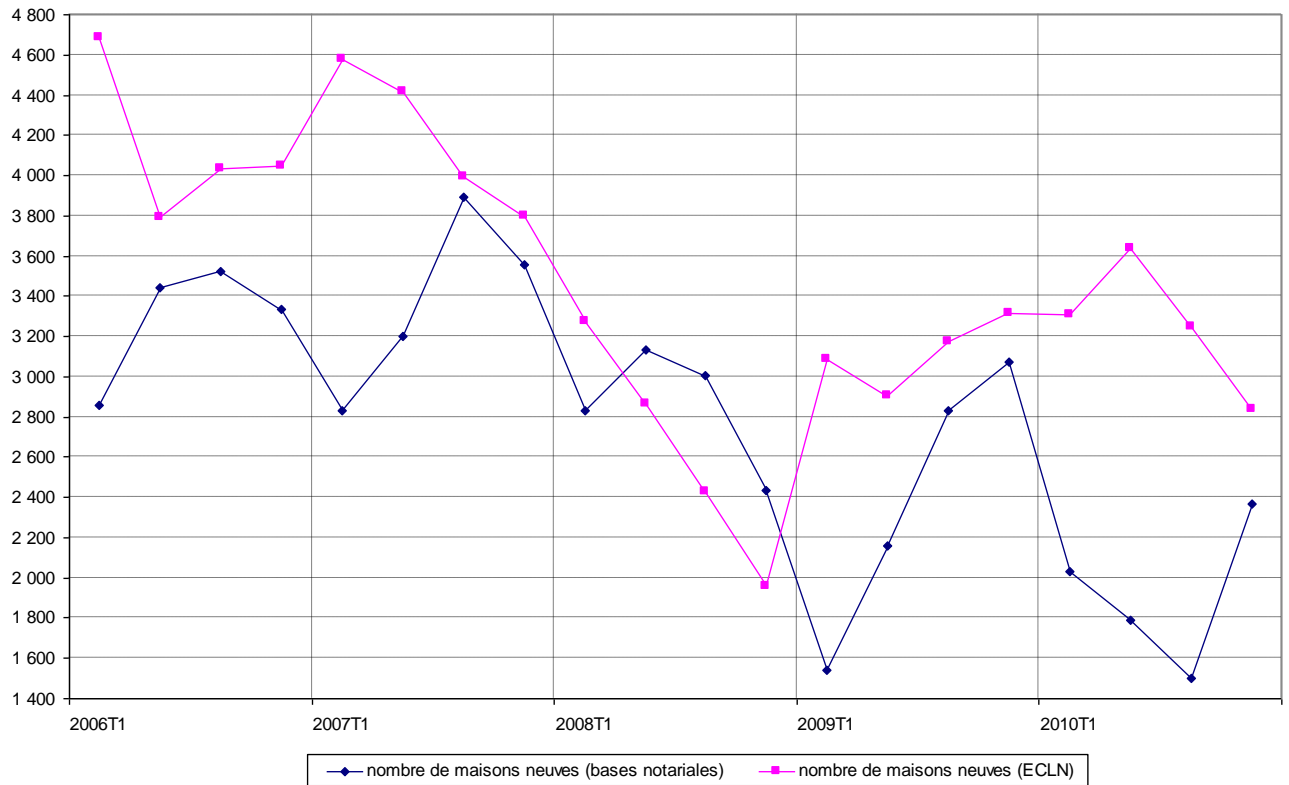
Deux raisons nous ont poussés à privilégier l'ECLN. D'abord, l'indice des prix des logements doit représenter les prix du marché au plus près de la date de publication. Or les bases notariales sont constituées à partir des actes de ventes ; la signature de l'acte pour un logement neuf a lieu plusieurs mois après la date de réservation (en général dans l'année). Ainsi, les données issues des bases notariales sont moins fraîches conjoncturellement que les données à la réservation. Par ailleurs, le taux de couverture des bases notariales n'est pas de 100 %, les notaires envoyant les actes sur la base du volontariat<sup>14</sup>. Or lorsqu'on compare le nombre de transactions de logements neufs contenues dans les deux sources de données, l'ECLN et les bases notariales, sur la période 2006-2010, on constate que l'ECLN répertorie en moyenne un plus grand nombre de ventes que les bases notariales. Par exemple, le nombre trimestriel de ventes de maisons neuves s'établit à 3 470 vs 2 767, soit un écart de 25,4% (cf. graphique 1.1). Le constat est le même pour les appartements neufs, l'écart étant même beaucoup plus conséquent : 23 824 vs 10 932, soit un écart de 117,9% (cf. graphique 1.2). Ces écarts sont cependant à relativiser dans la mesure où ils reposent sur des définitions des logements neufs ne coïncidant pas exactement (cf. annexe 14).

---

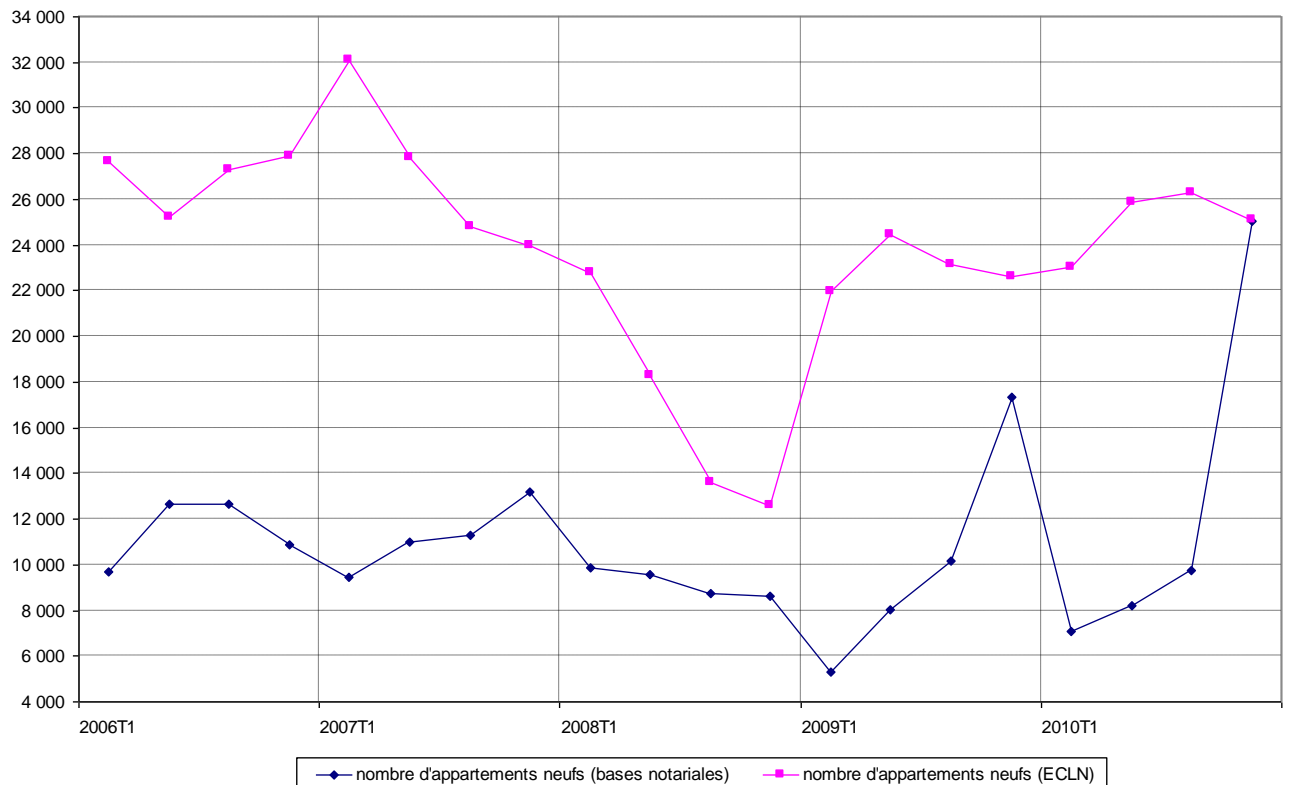
<sup>13</sup> Depuis la refonte de l'enquête en 2017, les prix unitaires des transactions sont aussi disponibles dans l'ECLN.

<sup>14</sup> Cependant, à compter de 2017, les notaires ont l'obligation d'alimenter les bases. Ceci pourrait entraîner une forte augmentation des taux de couverture et pourrait justifier de mener ultérieurement des travaux de comparaison entre des indices calculés avec les deux sources.

**Graphique 1.1 : Le nombre de maisons neuves contenues dans les bases notariales et dans l'ECLN (2006-2010)**



**Graphique 1.2 : Le nombre d'appartements neufs contenus dans les bases notariales et dans l'ECLN (2006-2010)**



### 1.3.2. Le champ de l'IPLN

Le champ géographique retenu pour l'indice des prix des logements neufs est le même que celui de l'IPLA, i.e. la France métropolitaine hors Corse<sup>15</sup>.

Concernant le champ sectoriel, eu égard au champ de l'indice des prix des logements (cf. §1.2.1.), tous les logements réservés figurant dans l'ECLN appartiennent au champ de l'IPLN. En particulier, les logements neufs réservés pour investissement locatif sont couverts par l'IPLN. Cependant, les logements sociaux sont exclus du champ car un logement social neuf ne peut pas être acquis par un ménage. En effet, seuls peuvent être vendus à un ménage les logements sociaux construits ou acquis depuis plus de 10 ans par un organisme HLM<sup>16</sup>.

Après avoir choisi une source de données (l'ECLN) et défini le champ de l'IPLN, nous avons pu mettre au point la méthode de calcul de l'indice des prix des logements neufs publié en janvier 2013.

### 1.3.3. La méthode de calcul de l'IPLN

#### 1.3.3.1. *La nécessité d'une modélisation hédonique*

Eu égard à la définition de l'indice des prix des logements (cf. §1.2.1.), on souhaite construire un indice de prix de transaction qui mesure l'évolution pure, c'est-à-dire hors effet qualité, des prix des logements neufs vendus entre deux trimestres consécutifs, T et T-1. Ainsi, définir l'indice des prix des logements neufs comme le rapport des prix moyens des logements nouveaux aux deux trimestres considérés n'est pas pertinent. En effet, un tel indice n'est un indice de prix pur, i.e. expurgé des effets qualité, que si les logements considérés entre deux trimestres consécutifs ont les mêmes caractéristiques (physiques et géographiques), i.e. ont la même qualité. Or, il est difficile de trouver entre deux trimestres consécutifs des logements neufs réservés ayant exactement les mêmes caractéristiques, le marché du neuf étant en constante évolution. C'est pourquoi, pour construire un indice de prix des logements neufs expurgé des effets qualité, la modélisation du prix d'un logement en fonction de ses caractéristiques physiques et géographiques est nécessaire.

Ainsi, on a construit un modèle économétrique, dit « hédonique », permettant de séparer les facteurs explicatifs du prix du logement relevant de la qualité du bien de ceux correspondant à l'évolution pure des prix.

#### 1.3.3.2. *Le modèle hédonique*

Le prix au m<sup>2</sup> est beaucoup plus élevé à Paris qu'en milieu rural. Ainsi, on ne peut pas comparer directement le prix d'un logement construit à Paris avec celui d'un logement ayant les mêmes caractéristiques physiques construit en milieu rural. C'est pourquoi, on a tout d'abord cherché à prendre en compte les effets de la localisation dans notre modélisation (voir Balcone, 2012). Pour cela, nous avons construit des zones géographiques homogènes en termes de prix au m<sup>2</sup> pour les logements neufs.

---

<sup>15</sup> La Corse sera couverte par l'indice des prix des logements neufs à partir du 2<sup>ème</sup> trimestre 2017 et par l'indice Notaires-Insee des prix des logements anciens à partir du 2<sup>ème</sup> trimestre 2018. L'introduction de ce département dans l'IPLN ne change presque rien.

<sup>16</sup> Habitation à loyer modéré.

## **La prise en compte de l'effet géographique**

Pour construire de telles zones, nous effectuons une **classification ascendante hiérarchique (CAH)** sur les couples (ZEAT<sup>17</sup> ; TUU<sup>18</sup>) en utilisant :

- la **méthode de Ward**, cette dernière reposant sur le principe de minimiser l'inertie intra-classe à chaque étape de la classification ;
- les données ECLN correspondant à la période 2006T1-2011T2 ;
- les unités urbaines définies à partir du 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<sup>er</sup> janvier 2010.

On dénombre huit ZEAT et neuf tailles d'unité urbaine (cf. annexe 8). Le nuage de points est ainsi constitué de 63 couples (ZEAT ; taille d'unité urbaine) et non pas 72, certains croisements n'existant pas.

À chacun de ces éléments  $j$  est associé :

- **une masse**  $m_j$  qui est la part du nombre de logements réservés dans la zone géographique définie par le couple  $j$  dans l'ensemble des logements réservés au cours de la période 2006T1-2011T2 ;
- **le prix moyen en euros**  $\text{prix\_moy}_j$  des logements réservés dans la zone géographique définie par le couple  $j$  au cours de la période 2006T1-2011T2 ;
- **la surface habitable moyenne en m<sup>2</sup>**  $\text{shab\_moy}_j$  des logements réservés dans la zone géographique définie par le couple  $j$  au cours de la période 2006T1-2011T2.

Le nuage de points considéré est constitué de 63 points  $j$  dans  $\mathbf{R}^2$ , chaque point ayant pour coordonnées  $(\text{prix\_moy}_j ; \text{shab\_moy}_j)$  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 horizontale afin d'obtenir une partition de notre nuage en classes homogènes en termes de prix moyen et de surface habitable moyenne. On décide arbitrairement de considérer 15 classes<sup>19</sup> afin de ne pas introduire un trop grand nombre de variables dans notre modèle (voir infra).

Afin de compléter ce découpage géographique, on va également considérer d'autres caractéristiques 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 multicommunale 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 multicommunale, elle est seule ville-centre. Sinon,

<sup>17</sup> Zone d'étude et d'aménagement du territoire.

<sup>18</sup> Taille d'unité urbaine ; on appelle unité urbaine une commune ou un ensemble de communes présentant une zone de bâti continu (i.e. une absence de coupure de plus de 200 mètres entre deux constructions) qui compte au moins 2 000 habitants.

<sup>19</sup> Voir la composition des 15 classes géographiques en annexe 9.

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 multicommunale qui ne sont pas villes-centres constituent la banlieue. Enfin, les communes ne rentrant pas dans la constitution d'une unité urbaine correspondent aux **communes rurales**.

Ensuite, on repère **les communes littorales** et les communes situées dans **l'arrière-pays littoral**<sup>20</sup> : on dénombre 885 communes littorales et l'arrière-pays littoral est constitué de 1178 communes. Enfin, on distingue **les stations de sport d'hiver et d'alpinisme**<sup>21</sup> : les communes correspondant à ces stations sont au nombre de 62.

## **Le modèle**

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 des indices Notaires-Insee des prix des logements anciens. Cependant, en raison du faible nombre d'observations issues chaque trimestre de l'ECLN (cf. annexe 10), la séparation appartements/maisons faite dans la méthodologie Notaires-Insee n'a pas semblé 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 un seul modèle plutôt que d'estimer un modèle par zone comme pour les logements anciens. 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.

Nous avons opté pour un modèle classique pour ce genre de modélisation : nous avons choisi un modèle log-log dans lequel la variable dépendante est le logarithme du prix du logement vendu, les variables explicatives continues sont également en logarithme et les variables qualitatives sont transformées en indicatrices. Ce modèle est un cas particulier du modèle général Box-Cox et est du même type que 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
- nb\_obs(t) est le nombre d'observations issues de l'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'ECLN au trimestre t
- shab\_moy<sub>i,t</sub> est la surface habitable moyenne en m<sup>2</sup> des logements réservés correspondant à l'observation i issue de l'ECLN au trimestre t

<sup>20</sup> 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.

<sup>21</sup> 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.

- $(I_{i,t,1}, \dots, I_{i,t,k}, \dots, I_{i,t,K})$  est un vecteur de K indicatrices correspondant aux caractéristiques physiques et géographiques des logements associé à l'observation i issue de l'ECLN au trimestre t
- $\alpha, \beta_{shab}, \beta_1, \dots, \beta_K, \dots, \beta_K$  sont les paramètres (réels) du modèle à estimer
- $\varepsilon_{i,t}$  est le terme d'erreur associé à l'observation i issue de l'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<sup>22</sup> 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 disponibles dans l'ECLN et introduites dans le modèle :
  - 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
 (la modalité de référence est donc l'individuel groupé)
  - o le niveau de standing 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, une première possibilité est de définir un parc fixe de logements et d'estimer la valeur de ce parc à différentes dates (c'est la méthodologie des indices Notaires-Insee). Dans la mesure où le marché du neuf est en constante évolution, il apparaît difficile de définir un parc fixe de logements neufs. On estime donc **un indice de transactions**.

L'indice des prix expurgé des effets qualité est obtenu en introduisant des indicatrices temporelles dans le modèle général : la date de réservation du logement devient une variable explicative de son prix.

---

<sup>22</sup> Voir annexe 11.1 pour une description détaillée.

Ensuite se pose la question des transactions prises en compte. Comme l'influence d'une caractéristique physique ou géographique du logement sur son prix de vente, toutes choses égales par ailleurs, varie au cours du temps (cf. annexe 12), le modèle est estimé, chaque trimestre, à partir des données ECLN correspondant aux deux derniers trimestres, et non pas à partir de l'ensemble des données ECLN disponibles. C'est la méthode dite sur périodes adjacentes. On estime alors le modèle suivant (appelé **modèle sur périodes adjacentes**) :

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

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

où :

- $D_{t_1+1}$  est l'indicatrice temporelle du trimestre  $t_1+1$  définie de la manière suivante :
$$\forall t = t_1, t_1 + 1 \text{ et } \forall i = 1, \dots, \text{nb\_obs}(t), D_{i,t,t_1+1} = \begin{cases} 1 & \text{si } t = t_1 + 1 \\ 0 & \text{sin on} \end{cases}$$
- $\delta_{t_1+1}$  est le paramètre (réel) associé à l'indicatrice temporelle du trimestre  $t_1+1$

### 1.3.3.3. Du modèle hédonique à un indice de prix expurgé des effets qualité

Pour obtenir un indice de prix expurgé des effets qualité à partir du modèle sur périodes adjacentes, il suffit de repartir de l'indice « naïf », défini comme le rapport des prix moyens des logements neufs vendus aux deux trimestres considérés, et de remplacer les prix de vente réellement observés par leurs estimateurs des moindres carrés ordinaires (MCO).

L'indice, noté  $I_{t_1+1/t_1}$ , se réécrit alors sous la forme suivante :

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

où :

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

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



et

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

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

$\hat{\alpha}, \hat{\beta}_{\text{shab}}, \hat{\beta}_1, \dots, \hat{\beta}_k, \dots, \hat{\beta}_K, \hat{\delta}_{t_1+1}$  étant les estimateurs des MCO des paramètres  $\alpha, \beta_{\text{shab}}, \beta_1, \dots, \beta_k, \dots, \beta_K, \delta_{t_1+1}$  obtenus à partir du modèle **(1)** en pondérant chaque observation  $i$  d'un trimestre  $t$  par le nombre de logements réservés correspondant  $n_{i,t}$ .

$\hat{P}_{i,t_1}$  est un estimateur biaisé de  $\bar{P}_{i,t_1}$ . Cependant, ce biais peut être considéré comme négligeable (cf. annexe 11.2).

En remplaçant les estimateurs des prix de vente par leurs expressions respectives, l'indice de prix moyens  $I_{t_1+1/t_1}$  s'écrit alors sous la forme d'un produit d'un **indice de prix à qualité constante**  $I_{t_1+1/t_1}^*$  et d'un indice de qualité à prix constants  $I_{t_1+1/t_1}^Q$  :

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

où :

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

et

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

avec  $\forall k = 1, \dots, K$  et  $\forall t = t_1, t_1 + 1$ ,

$$\bar{I}_{t,k} = \sum_{i=1}^{\text{nb\_obs}(t)} \frac{n_{i,t}}{n_t} I_{i,t,k}$$

L'indice des prix des logements neufs expurgé des effets qualité d'un trimestre  $T_1$ , exprimé en base 100 au trimestre  $T_{\text{base}}$  (avec  $T_{\text{base}} < T_1$ ), est alors obtenu par **chaînage trimestriel** :

$$\text{IPLN}_{T_1}^{T_{\text{base}}} = \prod_{t=T_{\text{base}}}^{T_1-1} I_{t+1/t}^* \times 100$$

$$\text{i.e. } \text{IPLN}_{T_1}^{T_{\text{base}}} = \exp\left(\sum_{t=T_{\text{base}}}^{T_1-1} \hat{\delta}_{t+1}\right) \times 100$$

Les premiers travaux méthodologiques ont ainsi permis d'aboutir à un indice des prix des logements neufs hédonique. Cet indice, expurgé des effets qualité, a été validé par Eurostat. Cependant, les différences d'évolution observées entre les indices du neuf et de l'ancien (cf. graphique 2.1) ont motivé l'Insee à mener deux études méthodologiques complémentaires.

Ces études, encouragées par Eurostat, nous ont permis de nous assurer que ces différences ne correspondaient pas à un biais de méthode et ne provenaient pas d'une mauvaise prise en compte de la géographie dans le modèle sur périodes adjacentes.

## 2. Études complémentaires sur les indices des prix des logements neufs et anciens

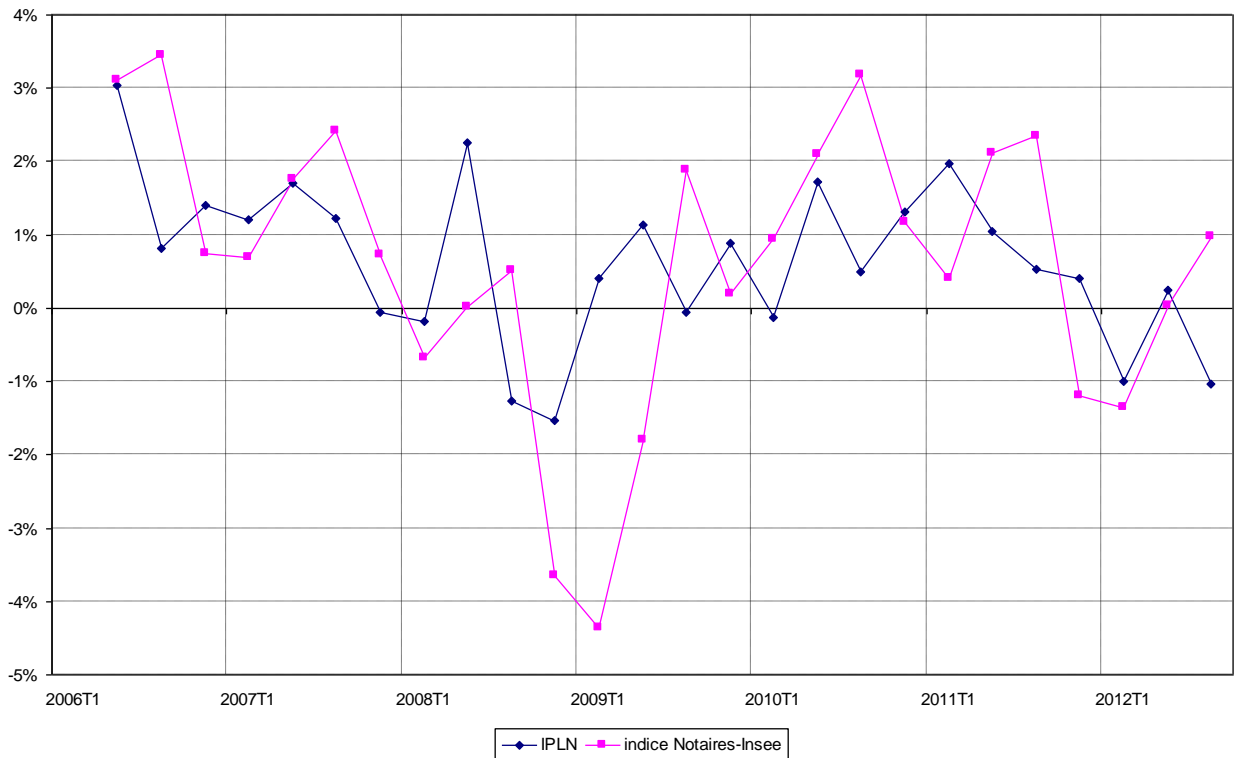
Sur la période 2006T1-2012T3, l'indice des prix des logements est mécaniquement très proche de l'indice des prix des logements anciens, compte tenu de la prépondérance de l'ancien dans les dépenses consacrées par les ménages à l'acquisition de logements (cf. graphique 2.1).

On constate cependant que l'indice des prix des logements neufs et l'indice Notaires-Insee des prix des logements anciens évoluent différemment (cf. graphiques 2.1, 2.2 et 2.3) : en glissement trimestriel et en glissement annuel, les écarts en valeur absolue entre les taux de croissance des deux indices sont relativement importants (respectivement 1,4% et 2,5%). C'est surtout visible pendant la période de crise (du 4<sup>ème</sup> trimestre 2008 au 1<sup>er</sup> trimestre 2010). Le tout est de savoir si ces différences sont liées à des biais de méthode ou à des phénomènes réels (moins de sensibilité des prix du neuf aux chocs par exemple, ou ajustements sur les volumes plutôt que sur les prix).

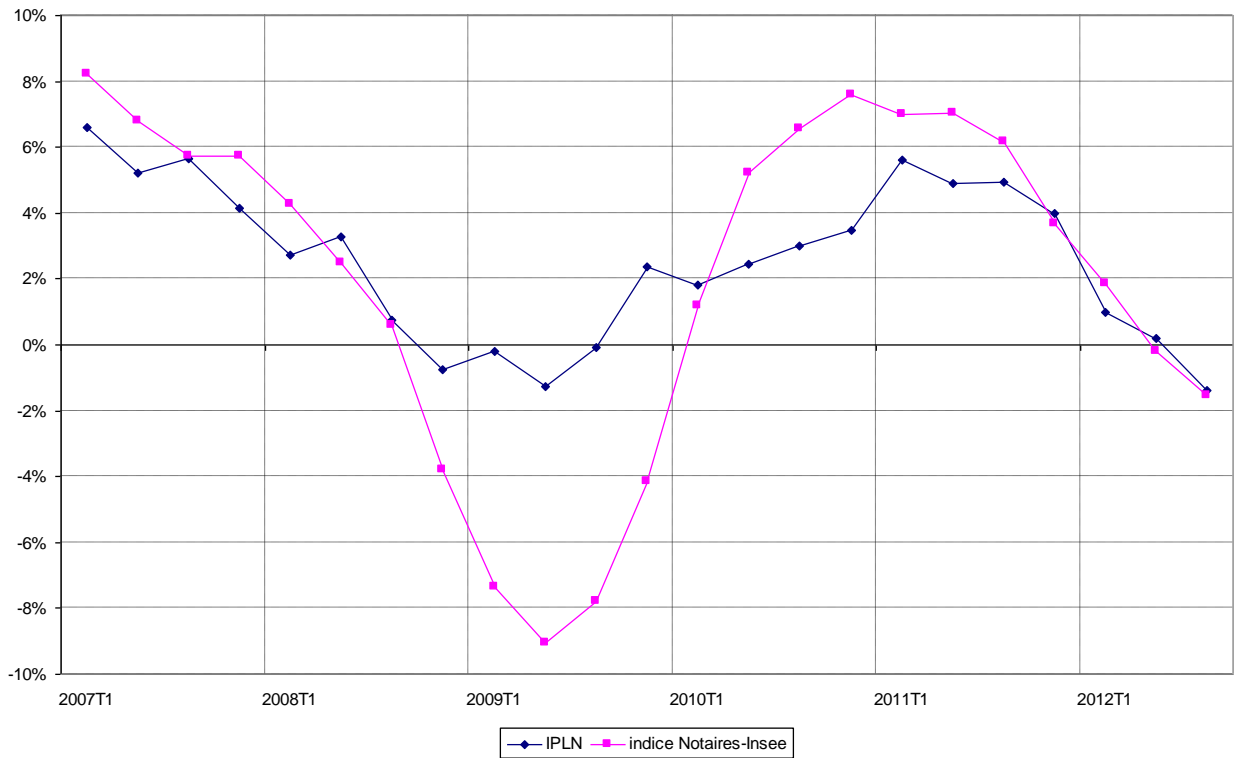
**Graphique 2.1 : L'indice des prix des logements, l'indice des prix des logements neufs (IPLN) et l'indice Notaires-Insee des prix des logements anciens (2006T1-2012T3 – base 100, moyenne annuelle 2010)**



**Graphique 2.2 : Taux de croissance en glissement trimestriel de l'indice des prix des logements neufs (IPLN) et de l'indice Notaires-Insee des prix des logements anciens (2006T1-2012T3)**



**Graphique 2.3 : Taux de croissance en glissement annuel de l'indice des prix des logements neufs (IPLN) et de l'indice Notaires-Insee des prix des logements anciens (2006T1-2012T3)**



Suite à ce constat, il a été décidé de mener **deux études méthodologiques** destinées à bien comprendre l'origine des différences, ou tout du moins, à vérifier que les différences observées ne provenaient pas d'artefacts de méthode. En effet, l'indice des prix des logements neufs est fondé sur une méthodologie de calcul d'indice et de prise en compte de la géographie différente de celle de l'indice Notaires-Insee. Par ailleurs, les logements neufs ne sont généralement pas construits aux mêmes endroits que les logements anciens, souvent en périphérie des villes. De plus, d'importants effets de grappe caractérisent la distribution géographique des logements neufs vendus chaque trimestre : elle peut ainsi différer sensiblement d'un trimestre à l'autre. Le traitement de la géographie pourrait donc lui aussi expliquer une dérive entre les deux indices.

Dans la première étude, nous nous concentrons sur la question de la méthode de calcul, indépendamment des questions de distribution géographique des logements. Ce n'est que dans la seconde étude que nous nous focaliserons sur la localisation.

## **2.1. Première étude : test de sensibilité de l'indice des prix des logements anciens au choix de méthodes de calcul**

Pour tester la sensibilité de l'indice des prix des logements anciens à un changement de méthode de calcul, nous travaillons sur des échantillons rigoureusement identiques à ceux utilisés pour calculer les indices Notaires-Insee et nous procédons, sur ces échantillons, au calcul d'un indice avec la méthode sur périodes adjacentes utilisée pour le calcul de l'indice des prix des logements neufs (en utilisant la même modélisation de la géographie et de dépendance aux caractéristiques des biens que celle utilisée pour l'IPLN – cf. §1.3.3.). Deux indices sont calculés sur :

- les maisons anciennes contenues dans les bases notariales mutées entre 2006 et 2010
- les appartements anciens contenus dans les bases notariales mutés entre 2006 et 2010

Les indices obtenus sont comparés aux indices Notaires-Insee publiés. Ainsi, la seule source de différence entre les indices Notaires-Insee publiés et les indices calculés avec la méthodologie IPLN est la méthodologie elle-même. L'écart entre les indices nous permet ainsi d'**évaluer le « biais de méthode » de calcul.**

### **2.1.1. Les données sur les logements anciens utilisées**

Les données sur les logements anciens proviennent des deux bases notariales utilisées pour le calcul des indices Notaires-Insee. La base « BIEN »<sup>23</sup> couvre l'Île-de-France alors que la base « Perval »<sup>24</sup> couvre la province. Ces bases enregistrent les transactions de tous les types de biens immobiliers<sup>25</sup>. Parmi eux, on ne garde que les logements anciens<sup>26</sup> situés en France métropolitaine hors Corse, mutés entre 2006 et 2010 et respectant l'ensemble des restrictions appliquées pour le calcul des indices Notaires-Insee (cf. annexe 14).

On travaille alors sur les deux lots de données suivants :

- le premier lot de données : les maisons anciennes
- le second lot de données : les appartements anciens

---

<sup>23</sup> Gérée par Paris Notaires Services.

<sup>24</sup> Gérée par Min.not.

<sup>25</sup> Appartements, maisons, immeubles, locaux d'activité, terrains, garages, vignobles et autres biens agricoles.

<sup>26</sup> Au sens fiscal, c'est-à-dire de plus de 5 ans ou connaissant une deuxième mutation.

## 2.1.2. La méthode de calcul utilisée : la méthode sur périodes adjacentes

La méthode sur périodes adjacentes est la méthode retenue pour le calcul de l'indice des prix des logements neufs (cf. §1.3.3.). Il s'agit ici d'implémenter cette méthode à l'identique sur les données Notaires-Insee. Le calcul procède en deux étapes :

- 1) ajustement d'un modèle hédonique sur périodes adjacentes
- 2) calcul d'un indice à partir des éléments fournis par le modèle hédonique

### 2.1.2.1. Le modèle hédonique pour les maisons anciennes (1<sup>er</sup> lot de données)

Comme il n'existe pas une correspondance exacte entre les variables issues de l'ECLN et les variables contenues dans les bases notariales, il convient d'adapter le modèle **(1)** présenté au §1.3.3.2. afin de pouvoir calculer un indice fondé sur les données Notaires-Insee.

Pour les maisons anciennes, le modèle hédonique considéré s'écrit sous la forme suivante :

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

$$\ln(p_{i,t}) = \alpha^{\text{mais}} + \beta_{\text{shab}}^{\text{mais}} \ln(\text{shab}_{i,t}) + \sum_{k=1}^K \beta_k^{\text{mais}} I_{i,t,k} + \delta_{t_1+1}^{\text{mais}} D_{i,t,t_1+1} + \varepsilon_{i,t}^{\text{mais}} \quad (2)$$

où :

- nb\_mais(t) est le nombre d'observations contenues dans les bases notariales correspondant à des maisons anciennes mutées au trimestre t
- $\forall i = 1, \dots, \text{nb\_mais}(t)$ ,
  - o  $p_{i,t}$  est le prix net vendeur en euros du logement i muté au trimestre t
  - o  $\text{shab}_{i,t}$  est la surface habitable en m<sup>2</sup> du logement i muté au trimestre t

### 2.1.2.2. Le modèle hédonique pour les appartements anciens (2<sup>nd</sup> lot de données)

Pour les appartements anciens, on utilise le modèle suivant :

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

$$\ln(p_{i,t}) = \alpha^{\text{appart}} + \beta_{\text{shab}}^{\text{appart}} \ln(\text{shab}_{i,t}) + \sum_{k=1}^K \beta_k^{\text{appart}} I_{i,t,k} + \delta_{t_1+1}^{\text{appart}} D_{i,t,t_1+1} + \varepsilon_{i,t}^{\text{appart}} \quad (3)$$

où nb\_appart(t) est le nombre d'observations contenues dans les bases notariales correspondant à des appartements anciens mutés au trimestre t.

### 2.1.2.3. Les indicatrices correspondant aux caractéristiques physiques et géographiques des logements utilisées dans les modèles (2) et (3)

Les indicatrices correspondant aux caractéristiques géographiques des logements figurant dans les modèles (2) et (3) sont exactement les mêmes que celles apparaissant dans le modèle (1). Celles correspondant aux caractéristiques physiques sont uniquement définies à partir du nombre de pièces des logements (2 indicatrices) :

- logement d'une pièce
- logement de 2 à 4 pièces
- logement de plus de 5 pièces (modalité de référence)

En outre, une indicatrice signalant la présence d'un balcon ou d'une terrasse figure dans le modèle hédonique pour les appartements anciens (modèle (3)).

### 2.1.2.4. Des modèles hédoniques à des indices de prix des logements anciens expérimentaux

En appliquant la méthode sur périodes adjacentes au premier lot de données, i.e. aux maisons anciennes, il vient que l'indice expérimental « 1<sup>ère</sup> simulation » des prix des maisons anciennes d'un trimestre  $T_1$ , exprimé en base 100 au trimestre  $T_{base}$  (avec  $T_{base} < T_1$ ), est donné par la relation suivante :

$$IPLA\_prem\_simul_{T_1}^{mais, T_{base}} = \exp \left( \sum_{t=T_{base}}^{T_1-1} \hat{\delta}_{t+1}^{mais} \right) \times 100$$

où  $\forall t = T_{base}, \dots, T_1 - 1$ ,  $\hat{\delta}_{t+1}^{mais}$  est l'estimateur des MCO du paramètre  $\delta_{t+1}^{mais}$  obtenu à partir du modèle (2).

De même, il vient que l'indice expérimental « 1<sup>ère</sup> simulation » des prix des appartements anciens d'un trimestre  $T_1$ , exprimé en base 100 au trimestre  $T_{base}$  (avec  $T_{base} < T_1$ ), est donné par la relation suivante :

$$IPLA\_prem\_simul_{T_1}^{appart, T_{base}} = \exp \left( \sum_{t=T_{base}}^{T_1-1} \hat{\delta}_{t+1}^{appart} \right) \times 100$$

où  $\forall t = T_{base}, \dots, T_1 - 1$ ,  $\hat{\delta}_{t+1}^{appart}$  est l'estimateur des MCO du paramètre  $\delta_{t+1}^{appart}$  obtenu à partir du modèle (3).

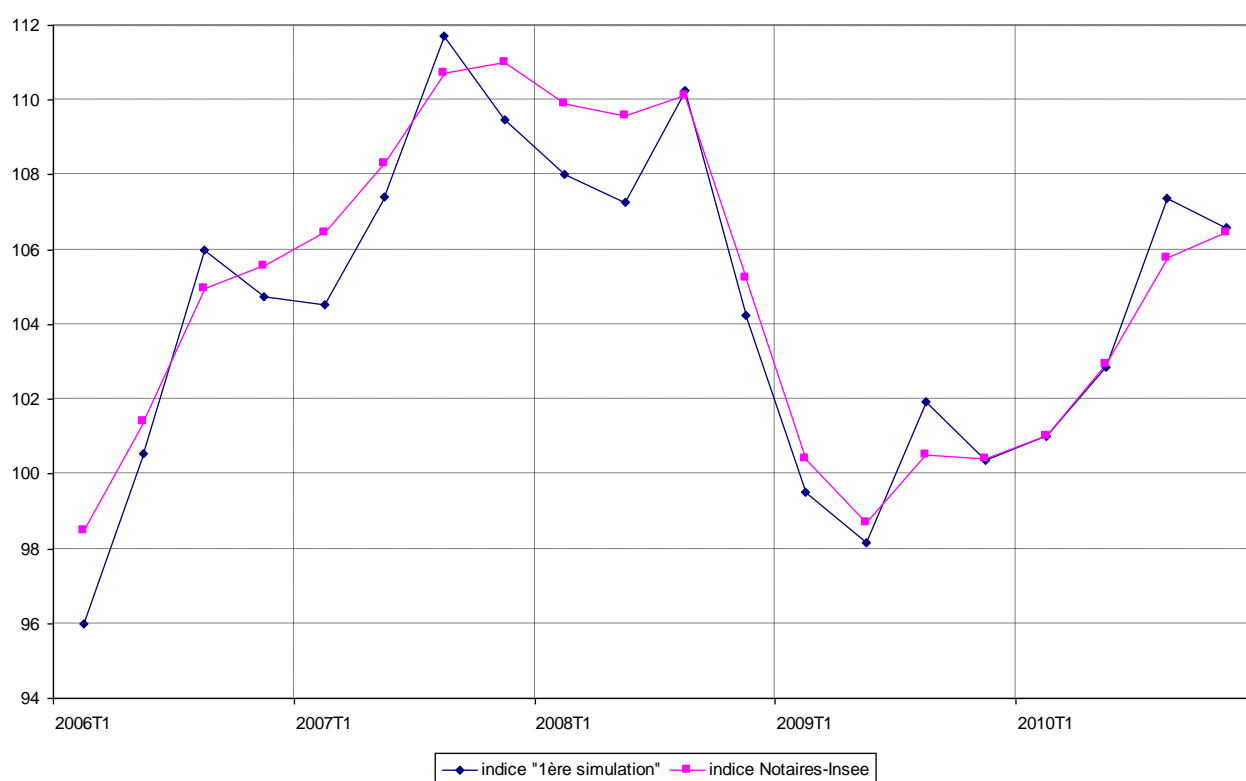
Il ne reste ensuite plus qu'à comparer ces deux indices expérimentaux avec les indices Notaires-Insee correspondants afin de mettre en évidence un éventuel « biais de méthode » de calcul.

### 2.1.3. Les résultats : comparaison des indices expérimentaux « 1<sup>ère</sup> simulation » avec les indices Notaires-Insee

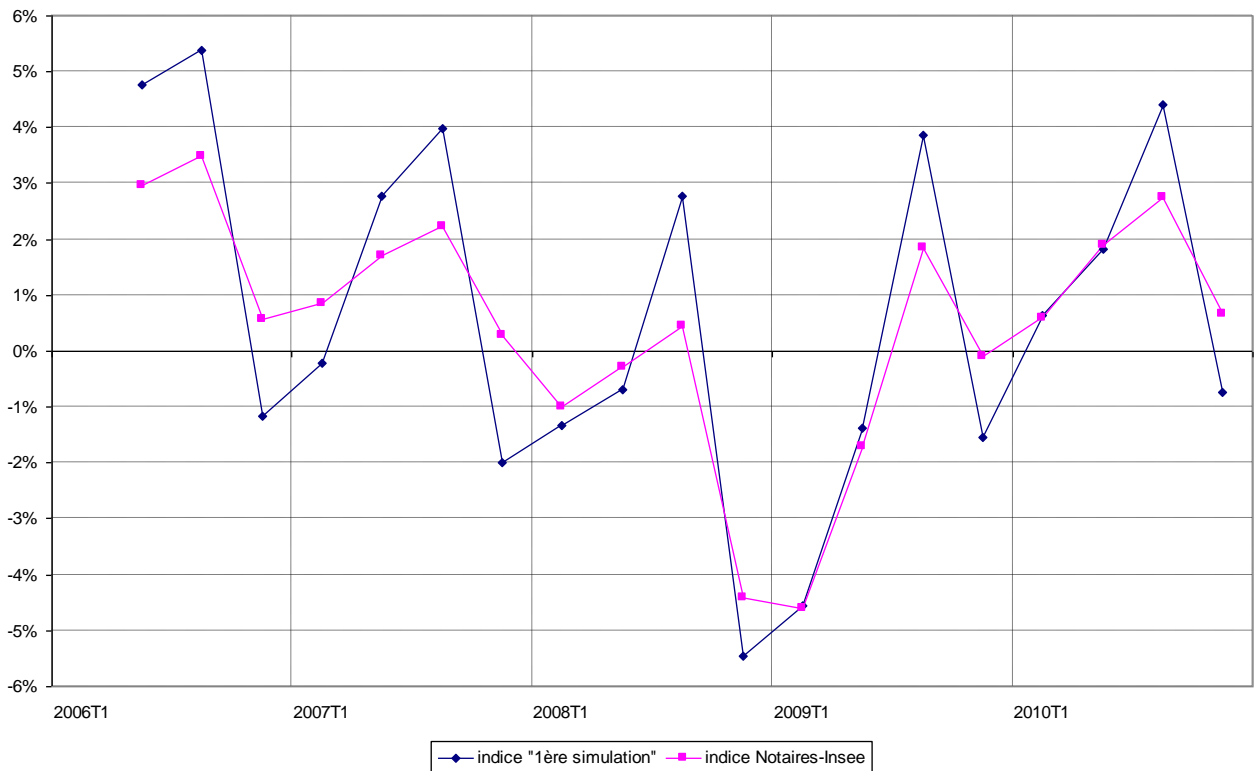
#### 2.1.3.1. Le premier lot de données : les maisons anciennes

L'écart entre l'indice expérimental « 1<sup>ère</sup> simulation » et l'indice Notaires-Insee des prix des maisons anciennes est faible : l'écart en valeur absolue entre les deux indices n'excède pas les 2,6 points d'indice (cf. graphique 2.4). De plus, les évolutions de ces deux indices sont similaires (cf. graphique 2.6), même si les variations de l'indice « 1<sup>ère</sup> simulation » sont un peu plus prononcées (cf. graphique 2.5).

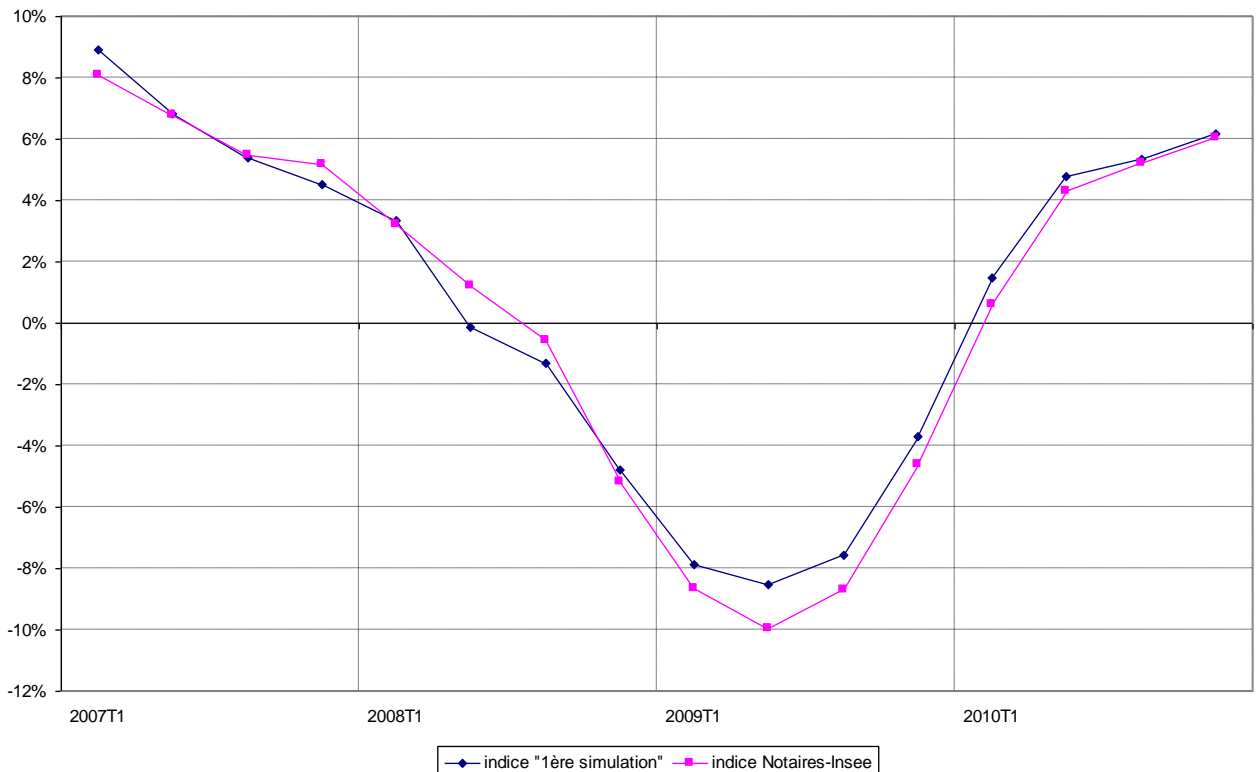
**Graphique 2.4 : L'indice Notaires-Insee et l'indice expérimental « 1<sup>ère</sup> simulation » des prix des maisons anciennes (2006T1-2010T4 – base 100, moyenne annuelle 2009)**



**Graphique 2.5 : Taux de croissance en glissement trimestriel de l'indice Notaires-Insee et de l'indice expérimental « 1<sup>ère</sup> simulation » des prix des maisons anciennes (2006T1-2010T4)**



**Graphique 2.6 : Taux de croissance en glissement annuel de l'indice Notaires-Insee et de l'indice expérimental « 1<sup>ère</sup> simulation » des prix des maisons anciennes (2006T1-2010T4)**

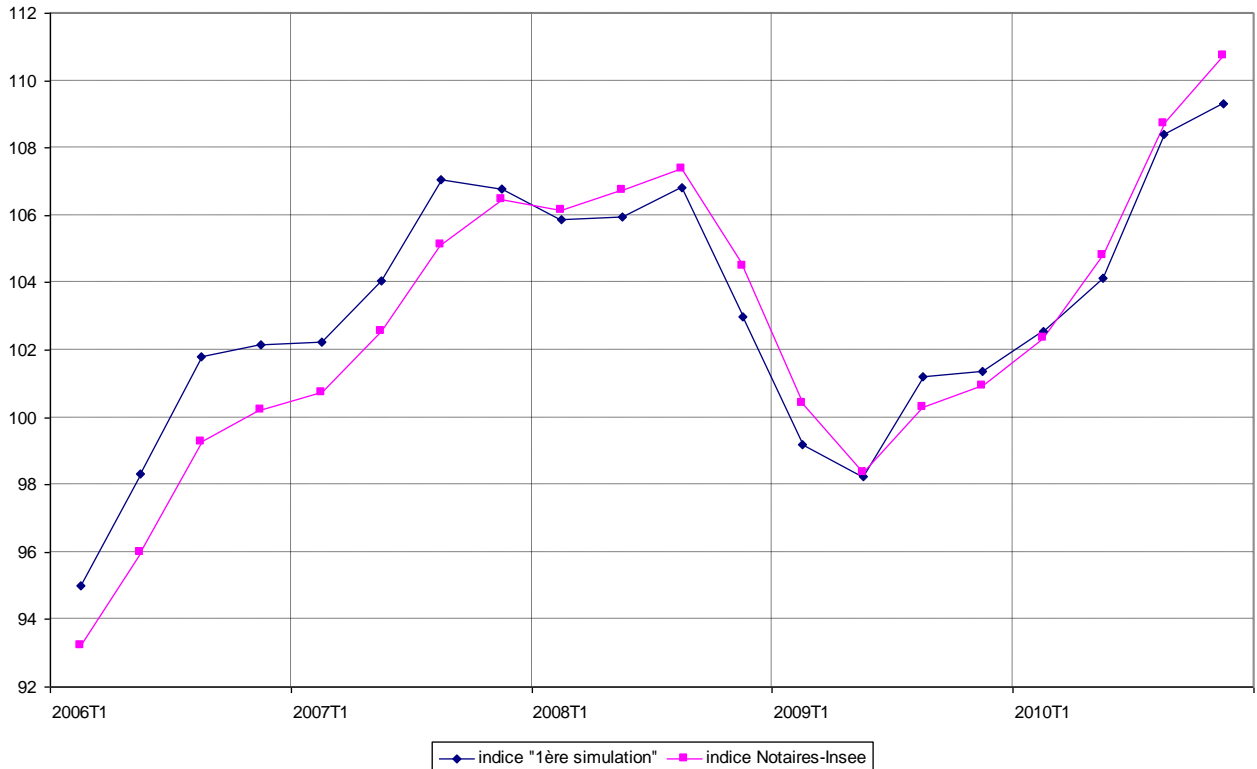




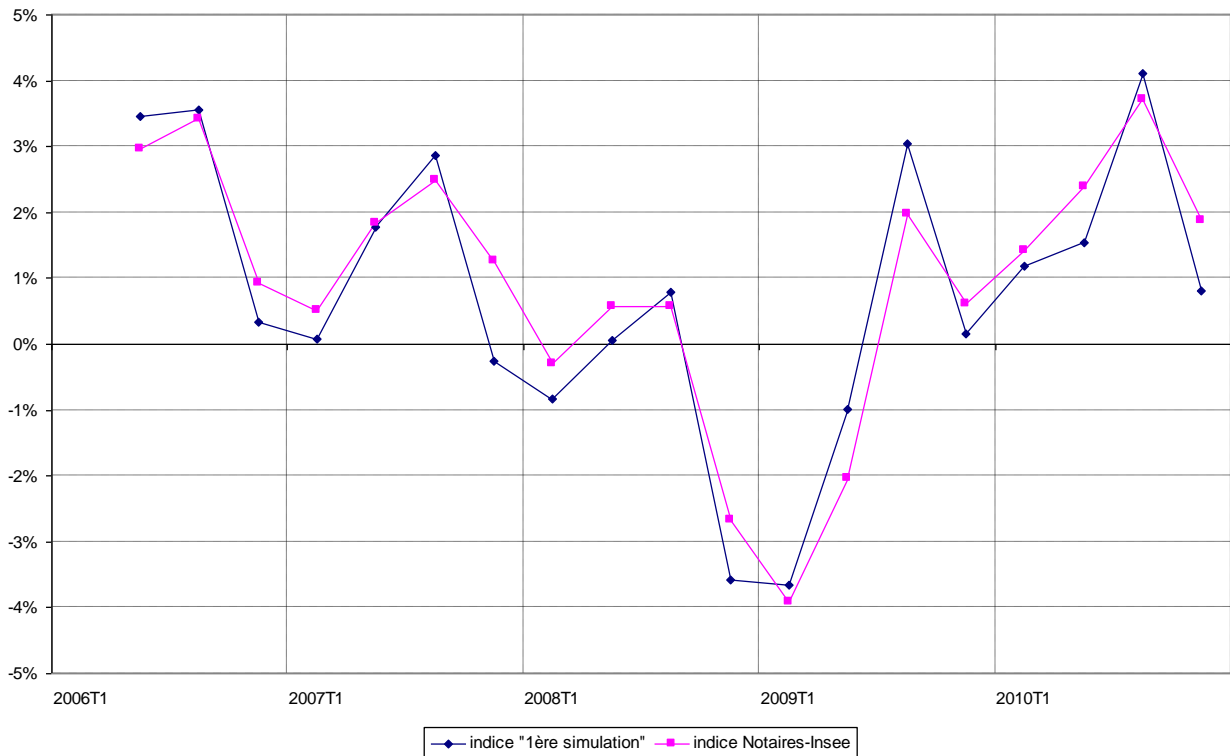
### 2.1.3.2. Le second lot de données : les appartements anciens

L'écart entre l'indice expérimental « 1<sup>ère</sup> simulation » et l'indice Notaires-Insee des prix des appartements anciens est faible : l'écart en valeur absolue entre les deux indices n'excède pas les 2,6 points d'indice (cf. graphique 2.7). De plus, les évolutions de ces deux indices sont très proches (cf. graphiques 2.8 et 2.9).

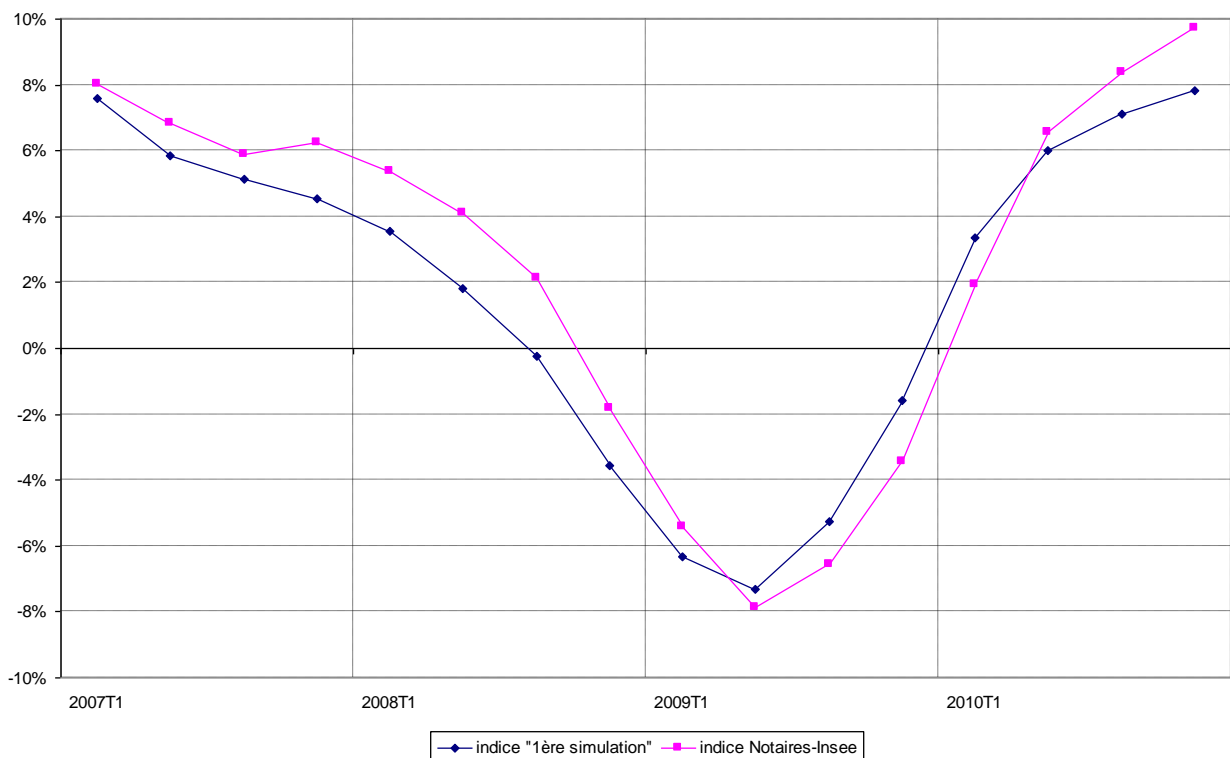
**Graphique 2.7 : L'indice Notaires-Insee et l'indice expérimental « 1<sup>ère</sup> simulation » des prix des appartements anciens (2006T1-2010T4 – base 100, moyenne annuelle 2009)**



**Graphique 2.8 : Taux de croissance en glissement trimestriel de l'indice Notaires-Insee et de l'indice expérimental « 1<sup>ère</sup> simulation » des prix des appartements anciens (2006T1-2010T4)**



**Graphique 2.9 : Taux de croissance en glissement annuel de l'indice Notaires-Insee et de l'indice expérimental « 1<sup>ère</sup> simulation » des prix des appartements anciens (2006T1-2010T4)**



#### 2.1.4. Conclusions de l'étude

Cette première étude avait pour but de mettre en évidence un éventuel biais de méthode susceptible d'expliquer en partie l'écart constaté entre l'indice des prix des logements neufs et l'indice Notaires-INSEE des prix des logements anciens. Les résultats obtenus pour les deux lots de données, à savoir les maisons anciennes et les appartements anciens, montrent que **le biais de méthode est négligeable en tendance**. Des écarts saisonniers subsistent.

### 2.2. Seconde étude : test de sensibilité de l'indice des prix des logements anciens à la distribution géographique des logements

Dans le cadre de cette seconde étude méthodologique, nous nous concentrons sur la question de la distribution géographique des logements neufs et anciens. En effet, la localisation des logements neufs évolue dans le temps de manière plus contrainte que celle des logements anciens. Ainsi, on peut se demander si la différence entre la distribution géographique des logements neufs et celle des logements anciens pourrait être à l'origine de la différence d'évolution des deux indices.

Afin d'établir l'impact de la différence de géographie qui existe (et évolue au cours du temps) entre les logements neufs et les logements anciens, nous allons constituer un sous-ensemble de logements anciens, « clones géographiques » des logements neufs, dans lequel la répartition géographique des logements sera la même que celle des logements neufs. Si l'indice « clone » se rapproche de l'indice des logements neufs, on pourra considérer que la distribution géographique des logements explique une partie des écarts. Si, au contraire, il reste proche de l'indice des logements anciens, on pourra considérer que les différences géographiques expliquent peu les écarts, et donc qu'il y a bien deux marchés au comportement différent : l'un du neuf, l'autre de l'ancien.

En pratique, la base de clones est construite de la manière suivante : on établit, pour chaque triplet (année ; type de construction<sup>27</sup> ; commune), le nombre de logements neufs vendus dans ECLN et le nombre de logements anciens vendus dans les bases notariales en considérant la période allant du 1<sup>er</sup> trimestre 2006 au 4<sup>ème</sup> trimestre 2010. Puis, pour chacun de ces triplets, on tire des « clones » des logements neufs parmi les logements anciens en utilisant un sondage aléatoire simple avec remise, la taille de l'échantillon étant égale au nombre de logements neufs du triplet considéré. Cette population de « clones » ainsi constituée aura alors la même distribution géographique que l'ensemble des logements neufs contenus dans ECLN. Enfin, nous procéderons sur cette population de « clones » au calcul d'un indice pour chaque type de construction avec la méthode sur périodes adjacentes utilisée pour le calcul de l'indice des prix des logements neufs (en utilisant la même modélisation de la géographie et de dépendance aux caractéristiques des biens que celle utilisée pour l'IPLN – cf. §1.3.3.).

Deux indices sont ainsi calculés sur :

- les maisons anciennes « clones » contenues dans les bases notariales mutées entre 2006 et 2010
- les appartements anciens « clones » contenus dans les bases notariales mutés entre 2006 et 2010

et les indices obtenus sont comparés aux indices des prix des logements neufs et aux indices expérimentaux « 1<sup>ère</sup> simulation » des prix des logements anciens correspondants (cf. §2.1.).

---

<sup>27</sup> Individuel (individuel groupé dans ECLN) et collectif.

### 2.2.1. Les données utilisées

Dans le cadre de cette simulation, les logements neufs considérés sont ceux appartenant au champ de l'IPLN (cf. §1.3.2.) dont l'année de vente est comprise entre 2006 et 2010. Quant aux logements anciens retenus, ce sont les mêmes que ceux utilisés dans la première étude (cf. §2.1.1.).

### 2.2.2. Constitution de la population des « clones » de logements neufs

Dans le cadre de cette simulation, on souhaite calculer, pour chaque type de construction, un indice de prix sur des « clones » de logements neufs sélectionnés parmi les logements anciens échangés, de sorte qu'ils aient la même distribution géographique que les logements neufs contenus dans la base ECLN. On détermine tout d'abord la distribution géographique des logements anciens contenus dans les bases notariales et celle des logements neufs contenus dans la base ECLN pour chaque type de construction. Pour les tirages de clones, la maille élémentaire de description de la géographie est le niveau communal<sup>28</sup>. Celle du temps pourrait être le trimestre. Mais pour avoir un nombre suffisant d'observations (cf. annexe 15), nous avons finalement opté pour un grain de description du temps annuel.

Pour constituer la population des « clones » des logements neufs, on considère ainsi les triplets (type de construction ; commune ; année) pour lesquels il y a existence de ventes de logements neufs et existence de ventes de logements anciens. Pour chacun de ces triplets, on tire avec remise (sondage aléatoire simple) un échantillon de  $n$  logements parmi les  $N$  logements anciens correspondant où  $n$  est le nombre de logements neufs associé.

L'ensemble de ces échantillons de logements anciens ainsi tirés constitue la population des « clones » des logements neufs. Par construction, cette population a la même distribution géographique annuelle que la population des logements neufs contenus dans la base ECLN. En d'autres termes, pour chaque triplet (type de construction ; commune ; année), le nombre de « clones » est égal au nombre de logements neufs vendus.

Par exemple, considérons le triplet (collectif ; Amiens ; 2008) présenté au tableau 2.1 :

**Tableau 2.1 : Exemple de tirage de « clones » pour un triplet à Amiens**

Type de construction	Commune	Année	Nombre de logements neufs vendus (n)	Nombre de logements anciens mutés (N)
Collectif	Amiens	2008	220	320

Les « clones » associés à ce triplet correspondent à un échantillon de taille  $n(=220)$  tiré dans la population des logements anciens mutés de taille  $N(=320)$  selon un sondage aléatoire simple avec remise. La population des « clones » correspondant à ce triplet sera ainsi constituée de 220 logements anciens (cf. tableau 2.2).

**Tableau 2.2 : Exemple de tirage de « clones » pour un triplet à Amiens (suite)**

Type de construction	Commune	Année	Nombre de logements neufs vendus (n)	Nombre de logements anciens mutés (N)	Nombre de « clones »
Collectif	Amiens	2008	220	320	<b>220</b>

<sup>28</sup> Niveau arrondissement pour Paris, Marseille et Lyon.

### 2.2.3. Calcul d'un indice expérimental annuel des prix des logements anciens pour chaque type de construction à partir de la population des « clones »

Pour chaque type de construction, on calcule un indice expérimental trimestriel des prix des logements anciens à partir de la population des « clones » en suivant la même méthodologie que celle qui nous a permis d'obtenir les indices expérimentaux « 1<sup>ère</sup> simulation » (cf. §2.1.2.). On appelle indice « clone » des prix des maisons anciennes et indice « clone » des prix des appartements anciens les indices trimestriels ainsi calculés.

Cependant, pour les raisons évoquées en annexe 15, la granularité temporelle retenue dans cette seconde étude est l'année. Nous devons ainsi calculer un indice « clone » annuel pour chaque type de construction. En notant :

- $IPLA\_clone_T^{mais, T_{base}}$  l'indice « clone » des prix des maisons anciennes d'un trimestre T, exprimé en base 100 au trimestre  $T_{base}$ ,
- et
- $IPLA\_clone_A^{mais, A_{base}}$  l'indice « clone » des prix des maisons anciennes d'une année A, exprimé en base 100 moyenne annuelle  $A_{base}$ ,

on a, par définition, la relation suivante :

$$IPLA\_clone_A^{mais, A_{base}} = \frac{\sum_{T=AT1}^{AT4} IPLA\_clone_T^{mais, T_{base}}}{A_{base} T4} \times 100$$

où ATt est le t-ième trimestre de l'année A

De même, on a :

$$IPLA\_clone_A^{appart, A_{base}} = \frac{\sum_{T=AT1}^{AT4} IPLA\_clone_T^{appart, T_{base}}}{A_{base} T4} \times 100$$

## 2.2.4. Les résultats : comparaison de l'indice « clone » avec l'IPLN et l'indice expérimental « 1<sup>ère</sup> simulation » pour chaque type de construction

### 2.2.4.1. Comment interpréter les résultats ?

Le but de cette seconde étude est d'établir l'impact de la différence de géographie qui existe entre les logements neufs et les logements anciens sur la différence d'évolution des indices correspondants. Pour atteindre cet objectif, on va considérer, pour chaque type de construction, le taux de croissance en glissement annuel des trois indices suivants obtenus avec la même méthode de calcul, à savoir la méthode sur périodes adjacentes :

- l'indice « clone » des prix des logements anciens
- l'indice « 1<sup>ère</sup> simulation » des prix des logements anciens
- l'IPLN

**Tableau 2.3 : Les méthodes de calcul et les échantillons utilisés pour calculer les trois indices**

Indice	Méthode de calcul	Échantillon de logements utilisé
<i>Indice « clone »</i>	Méthode sur périodes adjacentes	Logements anciens <sup>29</sup> « clones » des logements neufs <sup>30</sup> (cf. §2.2.2.)
<i>Indice « 1<sup>ère</sup> simulation »</i>	Méthode sur périodes adjacentes	Logements anciens utilisés pour calculer les indices Notaires-Insee
<i>IPLN</i>	Méthode sur périodes adjacentes	Logements neufs contenus dans la base ECLN et appartenant au champ de l'IPLN

Pour avoir une meilleure estimation du taux de croissance en glissement annuel de l'indice « clone », nous tirons **50 échantillons différents de logements « clones »**. Ainsi, nous considérons dans la suite, la moyenne des taux de croissance en glissement annuel des 50 indices « clone » et l'écart-type associé. Pour simplifier l'écriture, on parlera du taux de croissance en glissement annuel de l'indice « clone ».

Par construction des trois indices (cf. tableau 2.3), plus les évolutions de l'indice « clone » seront proches de celles de l'IPLN, plus la différence de géographie des logements échangés expliquera l'écart d'évolution entre les prix du neuf et ceux de l'ancien diffusés. Au contraire, plus les évolutions « clone » seront proches des évolutions « 1<sup>ère</sup> simulation », moins la géographie expliquera les différences observées. C'est cette grille de lecture que nous allons utiliser pour interpréter les résultats obtenus pour les maisons et les appartements.

<sup>29</sup> Contenus dans les bases notariales.

<sup>30</sup> Contenus dans la base ECLN et appartenant au champ de l'IPLN.

#### 2.2.4.2. *Le premier lot de données : les maisons anciennes « clones »*

Dans ce paragraphe, on étudie les écarts existant entre les taux de croissance en glissement annuel des 3 indices suivants :

- l'indice « clone »
- l'indice « 1<sup>ère</sup> simulation »

et

- l'IPLN-maisons neuves<sup>31</sup>

Pour cela, nous représentons sur un même graphique ces trois taux ainsi que l'intervalle de confiance à deux écarts-types associé à celui de l'indice « clone » (cf. graphique 2.10). En première analyse, on constate que l'écart entre les prix du neuf et ceux de l'ancien « à géographie constante »<sup>32</sup> est près de 1,6 fois plus faible en moyenne que l'écart existant sans se ramener à géographie équivalente<sup>33</sup> sur la période 2006-2010 (2,22 points de % contre 3,65 points de %, cf. graphique 2.10). Ainsi, la différence de géographie existant entre les maisons neuves et les maisons anciennes semble expliquer en moyenne une part importante (près de 40 %) de l'écart existant entre le taux de croissance de l'indice « 1<sup>ère</sup> simulation » et celui de l'IPLN-maisons neuves sur la période 2006-2010.

Cependant, ce résultat est à relativiser, car sur cette période, le taux moyen de l'IPLN-maisons neuves (1,86%) appartient à l'intervalle de confiance à deux écarts-types du taux moyen de l'indice « clone » ([0,81% ; 1,95%]) et celui de l'indice « 1<sup>ère</sup> simulation » est très proche de la borne inférieure (0,78%).

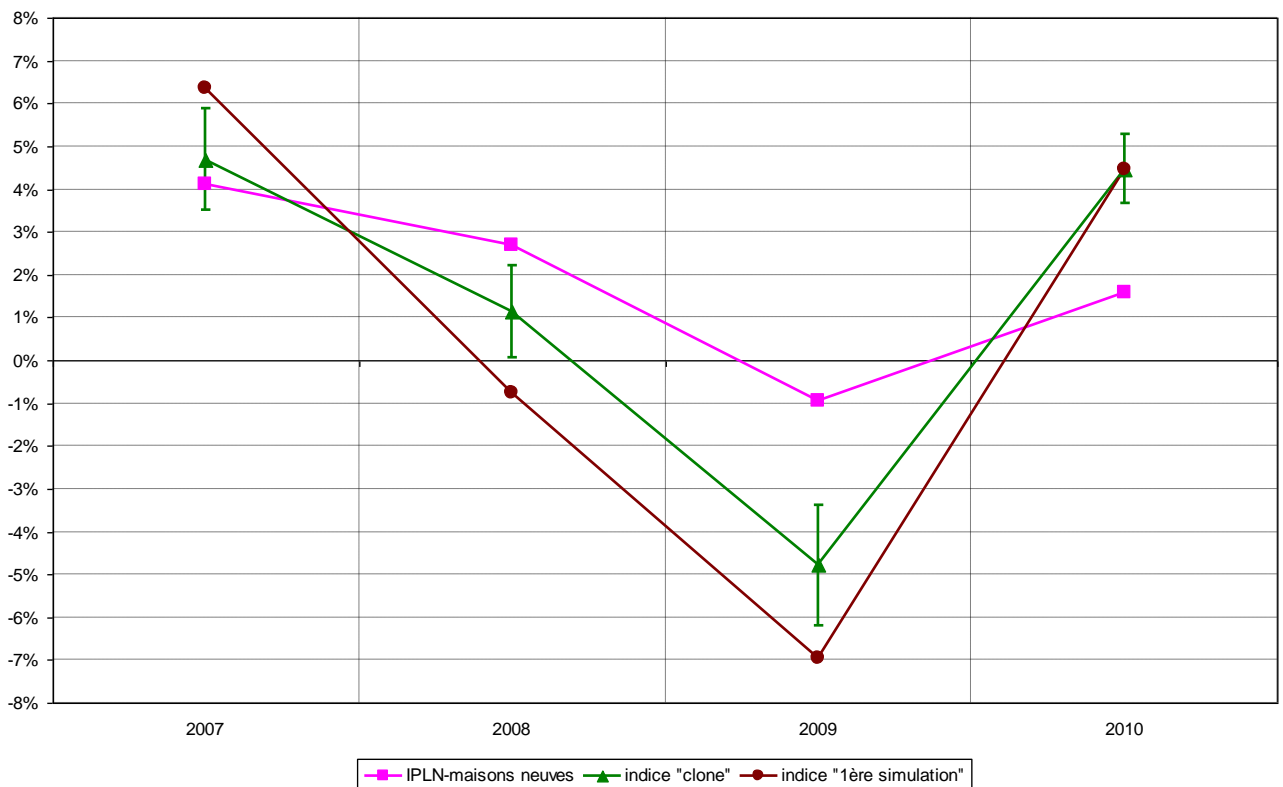
---

<sup>31</sup> Voir la méthode de calcul en annexe 16.

<sup>32</sup> Écart en valeur absolue entre le taux de l'indice « clone » et celui de l'IPLN-maisons neuves.

<sup>33</sup> Écart en valeur absolue entre le taux de l'indice « 1<sup>ère</sup> simulation » et celui de l'IPLN-maisons neuves.

**Graphique 2.10 : Taux de croissance en glissement annuel de l'IPLN-maisons neuves, de l'indice « clone » et de l'indice « 1<sup>ère</sup> simulation » des prix des maisons anciennes (2006-2010)**



**Note :** Pour l'indice « clone », les barres d'erreur correspondent à deux écarts-types

**Lecture :** Plus le taux de croissance de l'indice « clone » est proche de celui de l'IPLN-maisons neuves, plus la différence de répartition géographique entre maisons neuves et maisons anciennes est de nature à expliquer l'écart constaté entre l'indice des maisons anciennes (indice « 1<sup>ère</sup> simulation ») et l'indice des maisons neuves (IPLN-maisons neuves)

### 2.2.4.3. Le second lot de données : les appartements anciens « clones »

Dans ce paragraphe, on étudie les écarts existant entre les taux de croissance en glissement annuel des 3 indices suivants :

- l'indice « clone »
- l'indice « 1<sup>ère</sup> simulation »

et

- l'IPLN-appartements neufs<sup>34</sup>

Pour cela, nous représentons sur un même graphique ces trois taux ainsi que l'intervalle de confiance à deux écarts-types associé au taux de croissance en glissement annuel de l'indice « clone » (cf. graphique 2.11). On constate que le taux de croissance de l'indice « clone » est très proche de celui de l'indice « 1<sup>ère</sup> simulation ». En effet, l'écart en valeur absolue entre les taux de croissance de ces deux indices est en moyenne inférieur à 0,70 points de % sur la période 2006-2010

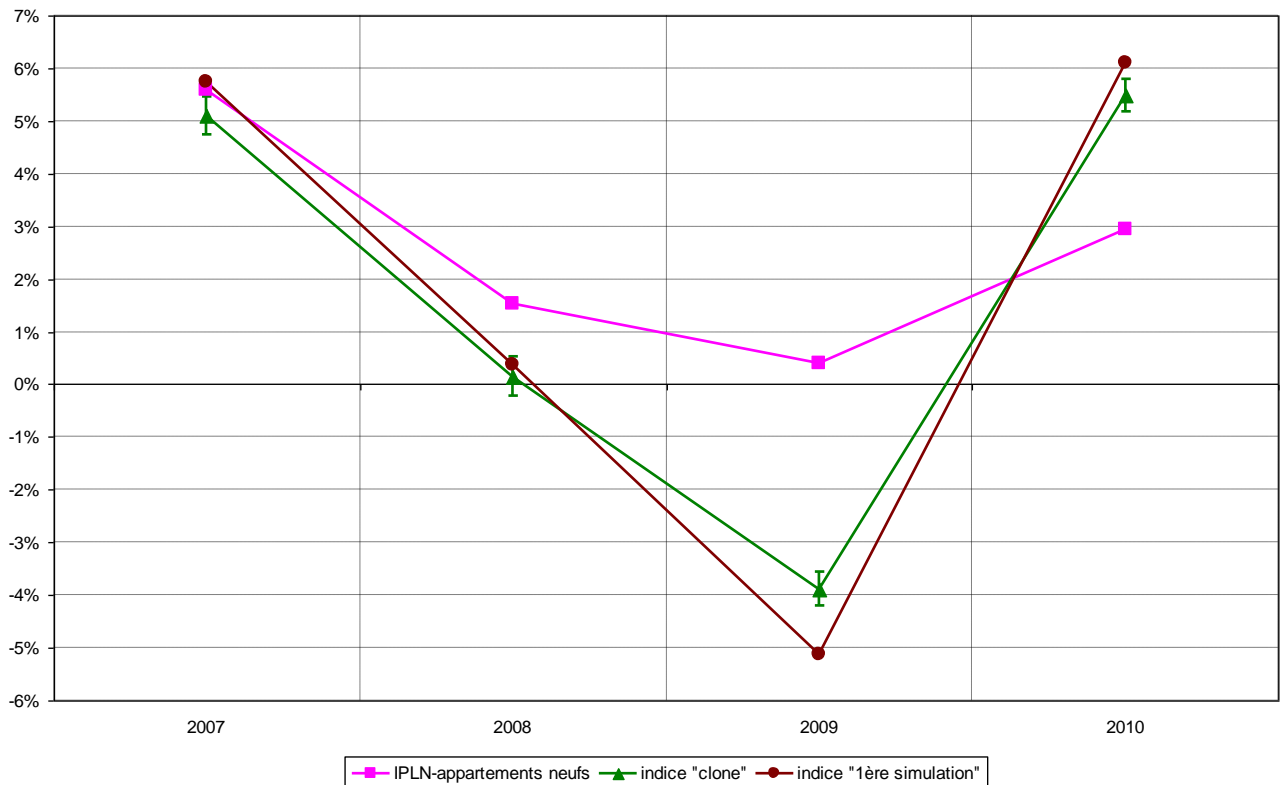
<sup>34</sup> Voir la méthode de calcul en annexe 16.



(0,68 points de % - cf. graphique 2.11). Si on exclut l'année 2009 de notre analyse, cet écart passe sous la barre des 0,5 points de % (0,49 points de %). Ainsi, il s'avère que la différence de distribution géographique n'explique qu'une très faible part des écarts entre le taux de croissance de l'IPLN-appartements neufs et celui de l'indice « 1<sup>ère</sup> simulation ».

Par conséquent, un alignement de la géographie pour les appartements neufs et anciens conduirait à un très léger rapprochement des deux indices. Un écart résiduel conséquent demeurerait, témoignant du fait que la dynamique des prix est différente pour ces deux types d'appartements.

**Graphique 2.11 : Taux de croissance en glissement annuel de l'IPLN-appartements neufs, de l'indice « clone » et de l'indice « 1<sup>ère</sup> simulation » des prix des appartements anciens (2006-2010)**



**Note :** Pour l'indice « clone », les barres d'erreur correspondent à deux écarts-types

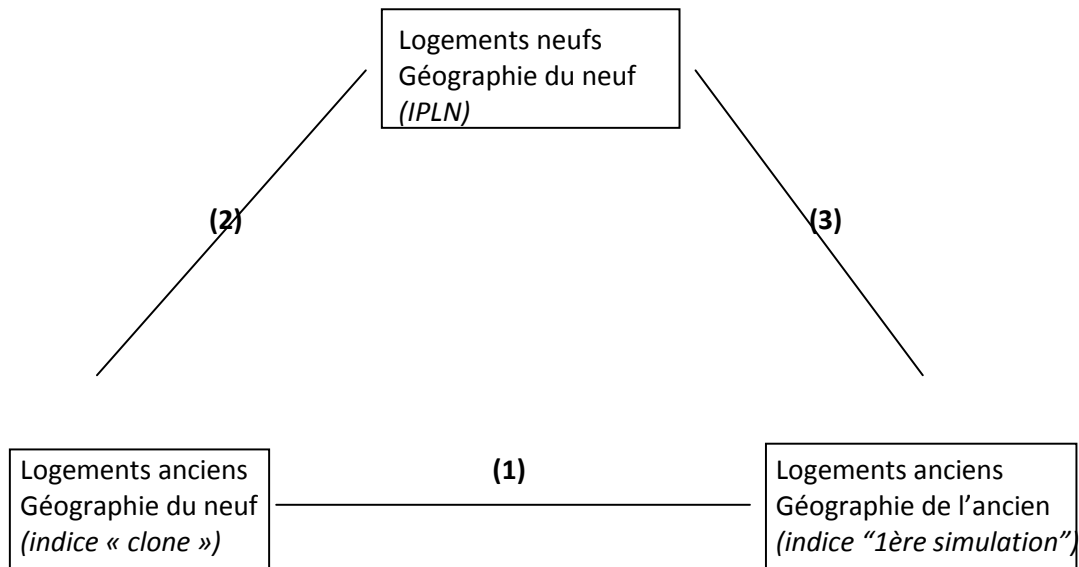
**Lecture :** Plus le taux de croissance de l'indice « clone » est proche de celui de l'IPLN-appartements neufs, plus la différence de répartition géographique entre appartements neufs et appartements anciens est de nature à expliquer l'écart constaté entre l'indice des appartements anciens (indice « 1<sup>ère</sup> simulation ») et l'indice des appartements neufs (IPLN-appartements neufs)

### 2.2.5. Conclusions de l'étude

Conceptuellement, la géographie est prise en compte, aussi bien dans les traitements réalisés sur les logements neufs (cf. modèle hédonique **(1)** utilisé pour l'IPLN - §1.3.3.2.) que dans ceux portant sur les logements anciens (cf. modèles hédoniques **(2)** et **(3)** utilisés pour l'indice « 1<sup>ère</sup> simulation » - §2.1.2.1 et §2.1.2.2.). Si les spécifications économétriques sont très proches, les paramètres liés à la géographie sont estimés séparément pour les deux types de logements, neufs et anciens. Afin de bien isoler l'impact de ces estimations différentes sur les taux de croissance des indices hédoniques obtenus, nous avons tout d'abord cloné les logements neufs par des logements

anciens, puis comparé les indices fondés sur les données de logements anciens « clones » et d'origine pour évaluer l'impact de la configuration géographique de l'échantillon de logements sur l'estimation des coefficients géographiques et in fine des indices.

**Figure 2.1 : Schéma récapitulatif représentant les calculs menés**



La configuration des trois calculs d'indices est présentée à la figure 2.1. On cherche à comprendre la source des différences entre les deux couples (logements neufs ;géographie du neuf) et (logements anciens ;géographie de l'ancien) (relation **(3)** de la figure 2.1). Pour cela, on décompose ces différences en deux :

- **(1)** : une différence de géographie (i.e. une différence qui joue in fine sur les paramètres des modèles liés à la géographie)
- et
- **(2)** : une différence de nature de logements, à géographie donnée

Nos résultats établissent que s'agissant des maisons, les écarts constatés entre l'indice de l'ancien et l'indice du neuf peuvent être considérés comme étant dans le bruit de mesure. S'agissant des appartements, la situation est plus lisible, puisque les barres d'erreurs associées aux indices « clones » sont beaucoup plus petites que pour les maisons. Sur la période 2006-2010, les écarts observés entre la croissance annuelle des indices du neuf et de l'ancien imputables à une différence de géographie sont négligeables dans le bilan, ces écarts correspondant principalement à une différence d'évolution liée à la différence de nature des deux types de biens.

On peut donc penser que la spécification économétrique employée pour l'IPLN (modèle **(1)**) permet d'ajuster convenablement les effets qualité liés à la géographie : **les écarts constatés entre l'indice des logements anciens et l'indice des logements neufs tiennent principalement à des effets de marché et ne sont pas des artefacts de traitement.**

Compte tenu de ces éléments, l'Insee, avec l'aval du conseil scientifique Notaires-Insee, a décidé de publier l'IPLN et l'IPL français le 31 janvier 2013 à partir de 2006, conformément à la réglementation européenne.

La suite de ce document de travail approfondit la réflexion sur la modélisation de cet indice des prix des logements neufs. Cela aboutit à une nouvelle méthodologie qui a été mise en œuvre au moment de l'atterrissage de la refonte de l'ECLN en 2017 (cf. partie 4 sur les préconisations).

### 3. La poursuite de la réflexion sur le modèle de régression à la base du calcul de l'IPLN sur la période 2006T1-2012T3

La priorité jusqu'en janvier 2013 était de répondre à la demande européenne. Après avoir assuré la diffusion des indices IPLN et IPL, nous avons approfondi la réflexion sur le modèle **(1)** qui est pour rappel :

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

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

Dans un premier temps, nous avons cherché à valider les régressions à la base du calcul de l'IPLN. Puis, dans un deuxième temps, nous avons considéré des modèles alternatifs, notamment avec l'étude des variables de l'ECLN.

#### 3.1. Diagnostics

Nous avons tâché de nous assurer que le modèle **(1)** vérifiait les hypothèses de base sur lesquelles repose le modèle de régression linéaire gaussien, à savoir :

- **(H1)** : la non colinéarité des régresseurs (i.e. les régresseurs sont linéairement indépendants)
- **(H2)** : les erreurs sont d'espérance nulle
- **(H3)** : l'homoscédasticité des erreurs
- **(H4)** : l'indépendance des erreurs
- **(H5)** : la normalité des erreurs

Si le modèle **(1)** viole l'une de ces hypothèses, nous reverrons alors la spécification du modèle à la base du calcul de l'IPLN.

### 3.1.1. Rappel : une régression pondérée

Sous forme matricielle, le modèle **(1)** peut se réécrire sous la forme suivante :

$$Y = X\beta + \varepsilon$$

où

$$- Y = \left[ \ln(\bar{p}_{i,t}) \right]_{i=1, \dots, \text{nb\_obs}(t), t=t_1, t_1+1} = \left[ y_{i,t} \right]_{i=1, \dots, \text{nb\_obs}(t), t=t_1, t_1+1}$$

$$- X = \left[ x_{i,t} \right]_{i=1, \dots, \text{nb\_obs}(t), t=t_1, t_1+1}$$

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

$$x_{i,t} = \left[ 1, \ln(\text{shab\_moy}_{i,t}), I_{i,t,1}, \dots, I_{i,t,k}, \dots, I_{i,t,K}, D_{i,t,t_1+1} \right]$$

$$- \beta = \left( \alpha, \beta_{\text{shab}}, \beta_1, \dots, \beta_k, \dots, \beta_K, \delta_{t_1+1} \right)'$$

$$- \varepsilon = \left( \varepsilon_{i,t} \right)_{i=1, \dots, \text{nb\_obs}(t), t=t_1, t_1+1}$$

Comme précisé au §1.3.3.3., l'estimateur  $\hat{\beta}$  utilisé dans le calcul de l'indice des prix des logements neufs d'un trimestre donné est l'estimateur des moindres carrés ordinaires (MCO) obtenu à partir du modèle **(1)** en pondérant chaque observation  $i$  d'un trimestre  $t$  par le nombre de logements réservés correspondant  $n_{i,t}$ .  $\hat{\beta}$  est ainsi la solution du problème de minimisation suivant :

$$\min_{\beta} \left\| \sqrt{W}(Y - X\beta) \right\|^2$$

où :

$$W = \text{diag} \left[ \left( n_{i,t} \right)_{i=1, \dots, \text{nb\_obs}(t), t=t_1, t_1+1} \right]$$

L'estimateur  $\hat{\beta}$  s'écrit ainsi :

$$\hat{\beta} = (X'WX)^{-1} X'WY$$

i.e.

$$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{Y}$$

où :

$$- \tilde{Y} = \sqrt{W}Y = \left[ \sqrt{n_{i,t}} \ln(\bar{p}_{i,t}) \right]_{i=1, \dots, \text{nb\_obs}(t), t=t_1, t_1+1} = \left[ \tilde{y}_{i,t} \right]_{i=1, \dots, \text{nb\_obs}(t), t=t_1, t_1+1}$$

et

$$- \tilde{X} = \sqrt{W}X = \left[ \tilde{1}, \tilde{X}_1, \dots, \tilde{X}_j, \dots, \tilde{X}_p \right] = \left[ \tilde{1}, \tilde{X} \right]$$

$\hat{\beta}$  est ainsi l'estimateur des moindres carrés ordinaires correspondant au modèle de régression linéaire suivant :

$$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon} \quad (4)$$

où  $\tilde{\varepsilon} = \sqrt{W}\varepsilon$

C'est ce modèle de régression que nous allons chercher à valider. Par conséquent, nous allons vérifier que les hypothèses suivantes sont satisfaites :

- **(H1)** : la matrice  $\tilde{X}$  est de plein rang (i.e. de rang  $p+1$ )
- **(H2)** :  $E(\tilde{\varepsilon} / \tilde{X}) = 0$
- **(H3+H4)** :  $Cov(\tilde{\varepsilon} / \tilde{X}) = \sigma^2 I_n$   
Où  $Cov(\tilde{\varepsilon} / \tilde{X})$  est la matrice de variance-covariance de  $\tilde{\varepsilon}$   
et  $n = nb\_obs(t_1) + nb\_obs(t_1 + 1)$
- **(H5)** :  $\tilde{\varepsilon} / \tilde{X} \sim N(0, \sigma^2)$

### 3.1.2. (H1) : la non colinéarité des régresseurs

Il est nécessaire que le modèle **(4)** vérifie cette hypothèse pour que l'estimateur des MCO puisse être défini. De plus, la principale source d'instabilité dans l'estimation de  $\beta$  est la multicollinéarité (cas où les régresseurs sont très corrélés entre eux) car la matrice de variance-covariance de l'estimateur des MCO de  $\beta$ ,  $\hat{\beta}$ , est donné par la relation suivante (sous réserve que les hypothèses **(H3)** et **(H4)** soient vérifiées) :

$$Cov(\hat{\beta} / \tilde{X}) = \sigma^2 (\tilde{X}'\tilde{X})^{-1}$$

Il est ainsi fondamental de s'assurer que les variables explicatives ne sont pas très corrélées entre elles.

Pour mesurer la colinéarité entre les régresseurs, nous allons considérer **le facteur d'inflation de la variance (VIF pour Variance Inflation Factor)** associé à chaque variable explicative et **les indices de conditionnement (CI pour Condition Index)** associés à chacune des valeurs propres de la matrice  $\tilde{X}'\tilde{X}$ .

#### 3.1.2.1. Le facteur d'inflation de la variance (VIF)

Pour rappel, le facteur d'inflation de la variance associé à la variable explicative  $\tilde{X}_j$  noté  $VIF_j$  est donné par la relation suivante :

$$VIF_j = \frac{1}{1 - R_j^2}$$

où  $R_j^2$  est le carré du coefficient de corrélation multiple de  $\tilde{X}_j$  avec les  $p-1$  autres variables explicatives.

La loi du paramètre  $VIF_j$  étant inconnue, Besley, Kuh et Welsch ont défini empiriquement un seuil limite égal à 10. Or, sur la période 2006T1-2012T3, on constate que les VIF associés aux classes géographiques et, à un degré moindre, au statut de la commune sont supérieurs à 10 (cf. annexe 17). Ainsi, les régresseurs de ce modèle présentent **un problème de colinéarité**.

Pour confirmer cette première analyse, nous allons considérer les indices de conditionnement associés à chacune des valeurs propres de la matrice  $\tilde{X}'\tilde{X}$ .

### 3.1.2.2. L'indice de conditionnement (CI)

On note  $\lambda_1, \dots, \lambda_p$  avec  $\lambda_1 \geq \dots \geq \lambda_p$  les  $p$  valeurs propres de la matrice  $\tilde{X}'\tilde{X}$  préalablement transformée (les variables explicatives  $\tilde{X}_1, \dots, \tilde{X}_j, \dots, \tilde{X}_p$  sont centrées et réduites). On définit alors l'indice de conditionnement associée à la valeur propre  $k$  ( $k \in \{1, \dots, p\}$ ) noté  $CI_k$  de la manière suivante :

$$CI_k = \sqrt{\frac{\lambda_1}{\lambda_k}}$$

À chaque indice de conditionnement  $CI_k$  sont associées des « proportions de variance » (ou « variance proportions ») pour chacune des variables explicatives. Ces « proportions » permettent d'identifier les variables responsables de la colinéarité révélée par un indice de conditionnement élevé.

Comme dans le cas du VIF, la loi du paramètre  $CI_k$  n'est pas connue. Besley et al. ont également défini de façon empirique un seuil limite pour ce paramètre : ils considèrent qu'un indice de conditionnement supérieur à 30 révèle un problème de colinéarité entre certaines variables explicatives du modèle. Pour les identifier, ils précisent qu'elles sont caractérisées par une proportion de variance supérieure à 0,5.

L'examen des indices de conditionnement sur la période 2006T1-2012T3 confirme l'existence d'un problème de colinéarité dont les classes géographiques semblent être la source (cf. annexe 18). Cette analyse est confirmée par l'étude des estimateurs des paramètres associés à ces variables et les écarts-types correspondants (cf. annexe 12) qui révèle une grande instabilité des estimations. Pour résoudre ce problème, nous avons décidé de réduire le nombre de classes géographiques retenues en regroupant certaines d'entre elles (cf. §3.2.1.).

### 3.1.3. (H2)-(H5) : analyse graphique des résidus

Les hypothèses **(H2)**, **(H3)**, **(H4)** et **(H5)** font intervenir le vecteur des termes d'erreur  $\tilde{\varepsilon}$ . Ces termes étant inobservées, nous allons vérifier ces hypothèses en considérant les résidus  $\tilde{u}$  qui sont donnés par la relation suivante :

$$\tilde{u} = \tilde{Y} - \hat{Y}$$

où  $\hat{Y} = \tilde{X}\hat{\beta}$

En effet, les résidus  $\tilde{u}$  peuvent être vus comme des estimateurs des erreurs  $\tilde{\varepsilon}$ .

#### 3.1.3.1. Vérification des hypothèses (H2), (H3) et (H4)

Pour vérifier que les résidus sont d'espérance nulle (i.e.  $E(\tilde{u}/\tilde{X}) = 0$ ) et respectent l'hypothèse d'homoscédasticité, nous traçons les résidus  $\tilde{u}_{i,t}$  en fonction des valeurs estimées  $\hat{Y}_{i,t}$  (cf. annexe 19). On constate que pour chacune des régressions, les résidus sont répartis uniformément dans une bande horizontale centrée sur l'axe des abscisses. On en conclut ainsi que les hypothèses **(H2)** et **(H3)** ne peuvent pas être rejetées.

En outre, on considère que l'hypothèse d'indépendance des résidus (hypothèse **(H4)**) est vérifiée, la variable dépendante n'étant pas une série temporelle.

#### 3.1.3.2. Vérification de l'hypothèse (H5)

Afin de vérifier l'hypothèse de normalité des résidus (hypothèse **(H5)**), nous traçons le QQ-plot des résidus pour chaque régression. Nous représentons ainsi sur un même graphique la droite d'équation  $y = x$  et le « nuage de points »  $(q_i; \tilde{u}_{(i)})_{i=1, \dots, 100}$  où  $q_i$  est le  $i$ ème centile de la loi normale  $N(0; \sigma^2)$ ,  $\sigma^2$  étant la variance des résidus, et où  $\tilde{u}_{(i)}$  correspond au  $i$ ème centile de la distribution des résidus.

En considérant les différents QQ-plots (cf. annexe 20), on constate que les points du nuage sont à peu près alignés sur la droite d'équation  $y = x$ . On en conclut ainsi que l'hypothèse de normalité des résidus ne peut pas être rejetée.

### 3.1.4. Traitement des observations aberrantes et influentes

Chaque trimestre, nous effectuons deux régressions successives afin de calculer l'indice des prix des logements neufs. Après la première régression, nous détectons les observations aberrantes et les observations influentes en considérant que :

- une observation est aberrante si le résidu studentisé associé est supérieur à 2 en valeur absolue
- une observation est influente si la distance de Cook associée est supérieure à 1 en valeur absolue

Les observations aberrantes sont relativement peu nombreuses (4,82 % des observations, soit 14,21 % des logements en moyenne chaque trimestre – cf. annexe 21) et les observations influentes sont extrêmement rares. Ceci nous a poussés à les exclure et, ainsi, à ré-estimer le modèle

à partir des observations restantes. On procède ainsi de la même manière que pour les logements anciens.

### 3.1.5. Étude des estimateurs des paramètres du modèle

Certains estimateurs des coefficients du modèle **(1)** présentent des évolutions suspectes ou une certaine instabilité sur la période 2006T1-2012T3. En particulier, comme précisé au §3.1.2., les estimations des paramètres associés aux classes géographiques sont instables, les écarts-types associés étant conséquents (cf. annexe 12). De plus, tous ces coefficients présentent une tendance haussière et un « creux » en 2010T2-2010T3 suspects. Ce constat nous a poussés à examiner attentivement la composition des 15 classes géographiques retenues (cf. §3.2.1.).

D'autre part, on note une grande volatilité des coefficients correspondant à la catégorie et aux éléments de confort du programme de construction<sup>35</sup>. Certaines estimations<sup>36</sup> sont, de surcroît, instables. Cette analyse nous a poussés à nous assurer de la bonne qualité de ces variables (cf. §3.2.1.).

### 3.1.6. Qualité d'ajustement du modèle

Pour évaluer la qualité d'ajustement du modèle de régression **(1)** sur la période 2006T1-2012T3, nous considérons les coefficients de détermination  $R^2$  associés à chaque régression au cours de cette période<sup>37</sup>. Le  $R^2$  moyen correspondant est de 0,78, le  $R^2$  minimum étant de 0,75 et le  $R^2$  maximum s'élevant à 0,82 (cf. graphique 3.1.). On peut ainsi considérer que l'ajustement du modèle **(1)** est de bonne qualité sur la période 2006T1-2012T3 malgré les problèmes de colinéarité, d'instabilité et de volatilité évoqués précédemment<sup>38</sup>.

## 3.2. Vers une nouvelle spécification du modèle (1)

### 3.2.1. Un examen des variables explicatives

Pour résoudre le problème de colinéarité entre les classes géographiques révélé au §3.1.2., nous avons examiné attentivement la composition des 15 classes géographiques retenues (cf. annexe 9). Nous avons ainsi constaté que les deux dernières classes n'étaient composées que d'un seul couple (ZEAT ; TUU) et que les deux couples composant la 13<sup>ème</sup> classe pouvaient être logiquement réaffectés à deux autres regroupements<sup>39</sup>. C'est pourquoi, nous avons proposé de ne considérer plus que 12 classes géographiques<sup>40</sup> (soit 11 indicatrices géographiques dans le modèle). En analysant les VIF et les indices de conditionnement obtenus à partir de ce nouveau modèle, on constate que les régresseurs peuvent être maintenant considérés comme linéairement indépendants

---

<sup>35</sup> La certification QUALITEL, la présence d'un dispositif de sécurité, la présence d'une piscine dans la résidence, la proximité d'un transport collectif par voie ferrée, la présence d'un système de climatisation et la présence d'un balcon ou d'une terrasse pour les appartements.

<sup>36</sup> La certification QUALITEL et la présence d'un système de climatisation.

<sup>37</sup> Considérer le  $R^2$  ajusté est inutile ici, le nombre de régresseurs étant le même pour chaque régression.

<sup>38</sup> Les  $R^2$  ont été recalculés sur la période 2013T1-2016T3 et sont en moyenne de 0,75 en considérant le même modèle.

<sup>39</sup> Le couple (Région parisienne ; 5000 à 9999 habitants) à la 11<sup>ème</sup> classe et le couple (Centre-est ; commune rurale) à la 8<sup>ème</sup>.

<sup>40</sup> La 14<sup>ème</sup> classe étant affectée à la 11<sup>ème</sup> et la 15<sup>ème</sup> à la 6<sup>ème</sup> – voir en annexe 22 la composition des 12 nouvelles classes.



(aucune variable explicative pour laquelle  $VIF \geq 10$  et aucune valeur propre pour laquelle  $CI \geq 30$ , le VIF maximal sur la période 2006T1-2012T3 étant de 9,30 et le CI maximal de 7,19).

Pour les raisons évoquées au §3.1.5., nous avons étudié les valeurs prises par les variables associées à la catégorie et aux éléments de confort du programme de construction. Cette analyse a révélé une mauvaise qualité de ces données (cf. annexe 23). Nous nous sommes donc interrogés sur la pertinence de la présence de ces variables dans le modèle. Le coefficient associé à la modalité « luxe » étant très significatif, nous avons décidé de ne maintenir que la variable « catégorie du programme ». Cependant, pour limiter les effets indésirables de la non-réponse sur les estimations, on a décidé de regrouper les modalités « confortable » et « très confortable » en une modalité « intermédiaire » (cf. annexe 24).

Le régresseur « présence d'un balcon ou d'une terrasse pour les appartements » est pour sa part construit à partir de la variable de l'enquête « surface extérieure » correspondant à :

- La surface du balcon et/ou de la terrasse pour les appartements
- La taille du terrain pour les maisons

Pour les appartements, cette variable est mal renseignée jusque fin 2007 et prend des valeurs suspectes à partir de 2008. Pour les maisons, elle est de bonne qualité à partir de 2008 (cf. annexe 25). Cette variable étant importante dans la formation du prix des maisons, on l'utilise pour distinguer les maisons bâties sur un grand terrain<sup>41</sup> à partir de 2008. On décide ainsi de remplacer à partir de cette date l'ancien régresseur « présence d'un balcon pour les appartements » par une indicatrice « maisons bâties sur un grand terrain » dans le modèle.

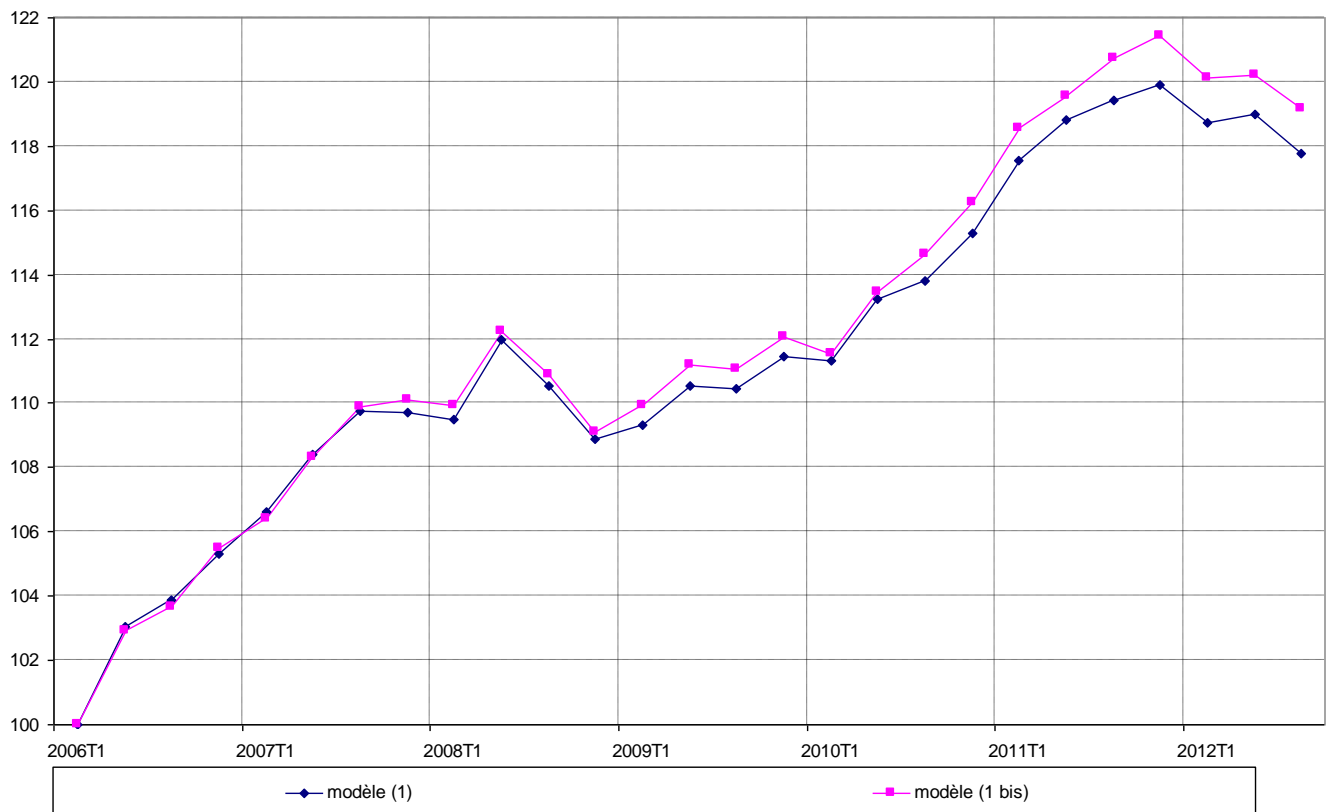
Les coefficients associés aux nouveaux régresseurs sont significatifs au niveau 5% ( $p\text{-value} < 0,05$  – cf. annexe 13) et de signe constant sur la période 2006T1-2012T3 (cf. annexe 12). L'indice obtenu<sup>42</sup> à partir du nouveau modèle (appelé modèle **(1bis)**) s'éloigne progressivement de celui calculé en utilisant le modèle **(1)** à partir du 4<sup>ème</sup> trimestre 2007, l'écart final au 3<sup>ème</sup> trimestre 2012 étant de 1,38 points d'indice (cf. graphique 3.1).

---

<sup>41</sup> Terrain correspondant à une surface supérieure à 200 m<sup>2</sup>.

<sup>42</sup> Exprimé en base 100 = 2006T1.

**Graphique 3.1 : Les indices du neuf calculés à partir des modèles (1) et (1bis)**



Cependant, en étudiant les taux de croissance en glissement trimestriel, on constate que les deux indices ont une évolution similaire : l'écart moyen est quasi nul (0,05% - cf. annexe 25) même si l'écart est plus marqué au 2010T3 (0,52%) et au 2011T3 (0,47%). En termes de glissement annuel, l'écart moyen est plus important mais reste faible (0,22%).

Cet examen sur les variables explicatives nous a incités à **préférer le nouveau modèle de régression (1 bis) pour les travaux de comparaison de méthodes réalisés dans la suite de ce document.**

## 3.2.2. Des modèles alternatifs

### 3.2.2.1. Le modèle global avec indicatrices temporelles

Comme précisé au §1.3.3.2., le modèle **(1)** est un modèle sur périodes adjacentes. Nous considérons ainsi les données ECLN correspondant aux trimestres T et T-1 pour calculer l'indice des prix des logements neufs du trimestre T. Pour d'autres indices, comme par exemple l'Indice du coût de la construction (ICC<sup>43</sup>), l'ensemble des données disponibles sont utilisées pour calculer l'indice. Le pendant de ce modèle pour l'indice du neuf, appelé **modèle global**<sup>44</sup> dans la suite, est le suivant :

$$\forall t = 2006T1, \dots, T \text{ et } \forall i = 1, \dots, \text{nb\_obs}(t),$$
$$\ln(\bar{p}_{i,t}) = \alpha + \beta_{\text{shab}} \ln(\text{shab\_moy}_{i,t}) + \sum_{k=1}^K \beta_k I_{i,t,k} + \sum_{j=2006T2}^T \delta_j D_{i,t,j} + \varepsilon_{i,t} \quad (5)$$

On obtient directement l'indice en base 100 2006T1 pour un trimestre T (avec T>2006T1) :

$$I_{\text{global}}^{2006T1} = \exp(\hat{\delta}_T) \times 100$$

Les estimations des coefficients du modèle sont beaucoup moins instables et présentent une volatilité beaucoup moins marquée que précédemment (cf. annexe 27). Ces résultats étaient attendus dans la mesure où chaque trimestre, nous considérons un plus grand nombre d'observations que dans le cas du modèle sur périodes adjacentes. Nous appelons dans la suite l'indice basé sur le modèle **(5)**  $I_{\text{global}}$ .

Cependant, ce modèle repose sur une **hypothèse forte** : l'effet d'une caractéristique physique ou géographique du logement sur son prix de vente est constant au cours du temps. Cette hypothèse est difficilement acceptable sur un intervalle de temps trop long dans le cas des constructions neuves. La volatilité des coefficients estimés à partir du modèle sur périodes adjacentes (modèle **(1)**) le prouve (cf. annexe 12). C'est pourquoi, nous nous sommes orientés vers un modèle reposant sur une hypothèse de stabilité moins forte.

### 3.2.2.2. Le modèle sur cinq trimestres glissants

Le modèle utilisé jusqu'à la refonte de l'ECLN en 2017 était estimé sur les deux trimestres T et T-1.

Pour corriger ses défauts tout en gardant ses avantages, nous considérons un modèle basé sur une hypothèse de stabilité des coefficients sur une période plus longue (5 trimestres) mais sans pour autant retenir la méthode globale présentée en §3.2.2.1. Nous utilisons ainsi les données ECLN correspondant aux cinq trimestres compris entre AT et (A-1)T pour calculer l'indice des prix des

<sup>43</sup> Voir la fiche méthodologique de cet indice en annexe 26

<sup>44</sup> L'indicatrice « maisons bâties sur un grand terrain », non disponible sur toute la période, n'est pas utilisée dans ce modèle.

logements neufs du trimestre T de l'année A<sup>45</sup>. Le modèle considéré<sup>46</sup>, appelé **modèle sur cinq trimestres glissants** dans la suite, est alors le suivant :

$$\forall t = (A-1)T, \dots, AT \text{ et } \forall i = 1, \dots, \text{nb\_obs}(t),$$

$$\ln(\bar{p}_{i,t}) = \alpha + \beta_{\text{shab}} \ln(\text{shab\_moy}_{i,t}) + \sum_{k=1}^K \beta_k I_{i,t,k} + \sum_{j \in \{(A-1)T; AT\}} \delta_j D_{i,t,j} + \varepsilon_{i,t} \quad (6)$$

On obtient directement l'indice en base 100 (A-1)T<sup>47</sup> pour un trimestre T de l'année A :

$$\text{IPLN\_new}_{AT}^{(A-1)T4} = \exp(\hat{\delta}_{AT}) \times 100$$

Pour obtenir l'indice en base 100 2006T1, un **chaînage** est donc nécessaire<sup>48</sup> :

$$\text{IPLN\_new}_{AT}^{2006T1} = \frac{\text{IPLN\_new}_{AT}^{(A-1)T4} \times \text{IPLN\_new}_{(A-1)T4}^{2006T1}}{100}$$

Les estimations des coefficients du modèle sont certes plus instables et présentent une volatilité plus marquée qu'en utilisant le modèle global mais elles sont plus précises et évoluent de manière moins heurtée que dans le cas du modèle sur périodes adjacentes (cf. annexe 27). Nous appelons dans la suite l'indice basé sur le modèle **(6)** IPLN\_new.

### 3.2.2.3. Comparaison des trois indices

Lorsqu'on compare les indices<sup>49</sup> obtenus avec les 3 modèles (modèle **(1 bis)**, **(5)** et **(6)**), on constate que les trois indices sont très proches jusqu'au 1<sup>er</sup> trimestre 2008 (cf. graphique 3.2.). Puis, à partir du deuxième trimestre de cette même année, l'IPLN\_new (modèle **(6)**) s'éloigne des deux autres indices avant de s'en rapprocher de nouveau à partir du troisième trimestre 2010. La période de crise est donc perçue de façon plus « violente » par l'IPLN\_new que par les deux autres. L'écart moyen entre l'IPLN\_new et l'indice obtenu à partir du modèle **(1 bis)** est ainsi plus élevé que celui entre l'I\_global et ce même indice sur la période 2006T1-2012T3 (0,22 points d'indice contre -0,09). Cependant, l'étude des taux de croissance en glissement trimestriel révèle que les 3 indices ont une évolution similaire : l'écart moyen entre les deux nouveaux indices et l'indice « modèle **(1 bis)** » est quasi nul (0,01% pour l'I\_global et 0,00% pour l'IPLN\_new, l'écart entre ces deux indices étant de -0,01% – cf. annexe 28). L'étude des taux de croissance en glissement annuel révèle que l'écart moyen entre les deux nouveaux indices est très faible (-0,01%) tout comme ceux entre ces deux

<sup>45</sup> Excepté pour l'année 2006 pour laquelle on adapte le modèle aux données disponibles

<sup>46</sup> L'indicatrice « maisons bâties sur un grand terrain », retenue à partir de début 2008, est donc introduite dans le modèle à partir du 2009T1.

<sup>47</sup> Base 100 2006T1 pour les trimestres de l'année 2006

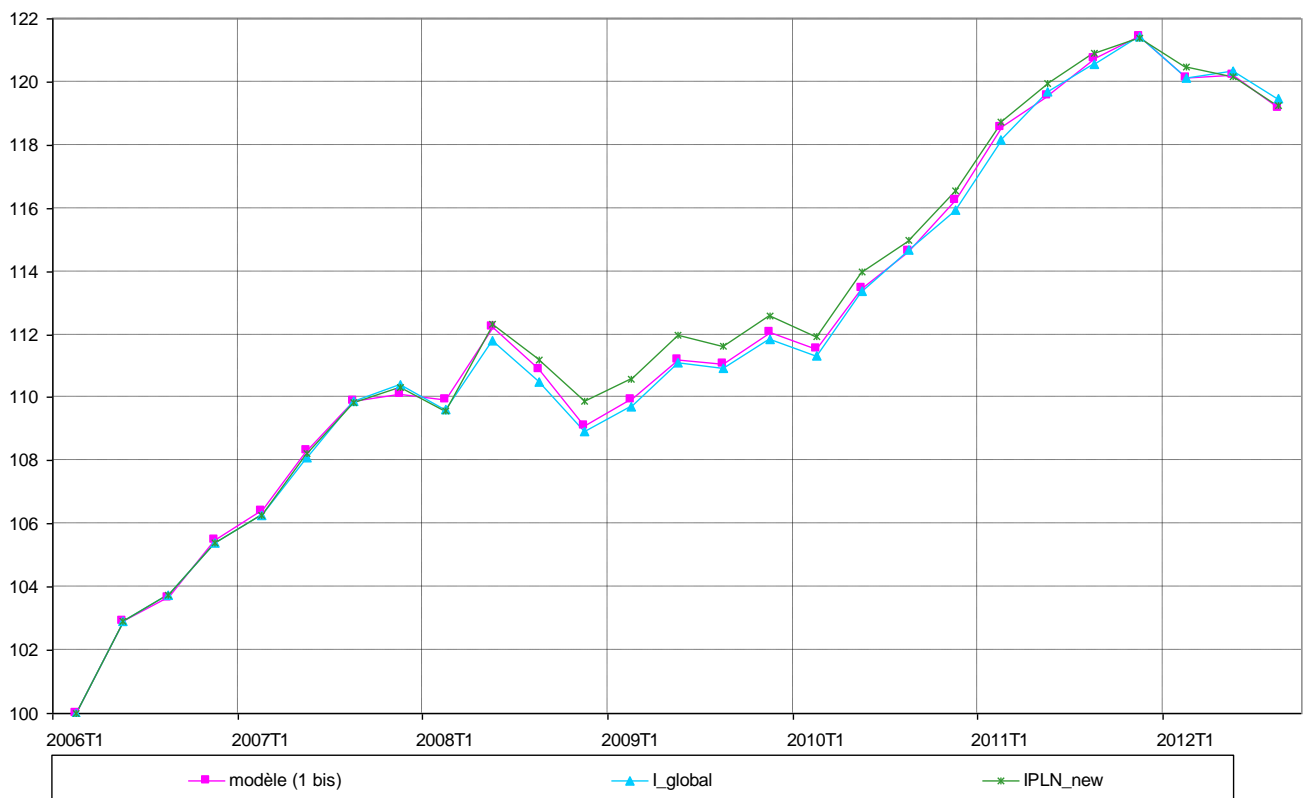
<sup>48</sup> Cette méthode a été retenue en production à partir de 2017, avec un chaînage trimestriel (et pas annuel). La référence en production est donc T-1 et non pas (A-1)T4 comme indiqué ici.

<sup>49</sup> Exprimés en base 100 2006T1

indices et l'indice « modèle (1 bis) » (0,02% pour l'I\_global et 0,01% pour le second). Cependant, un écart conséquent entre le taux de croissance de l'IPLN\_new et ceux des deux autres indices est à noter sur la période 2008T4-2009T2 en glissement annuel.

Afin d'éviter une trop grande volatilité des estimations, tout en n'imposant pas une trop forte hypothèse de stabilité des coefficients, nous privilégions le modèle (6) et donc l'indice IPLN\_new. Il permet de traiter les problèmes de colinéarité et aboutit à des estimateurs plus précis marqués par des évolutions moins heurtées.

**Graphique 3.2 : Les indices du neuf calculés à partir des modèles (1 bis), (5) et (6)**



## 4. Conclusion et préconisations

Le premier objectif de ce travail était la mise au point d'un indice du neuf de bonne qualité. Il a été atteint par l'élaboration d'un modèle de régression sur la période 2006T1-2012T3 (modèle **(1)**). Le deuxième était de s'assurer que les écarts observés entre les indices du neuf et de l'ancien n'étaient pas dus à la différence de méthodologie. Les deux études méthodologiques menées ont montré qu'ils étaient imputables à des différences de nature des logements (le marché de l'ancien est différent de celui du neuf).

Cependant, après examen du modèle initial (modèle **(1)**), il est apparu que certains coefficients étaient très volatils en raison d'une mauvaise qualité de certaines variables de l'ECLN. On a décidé d'exclure les variables correspondantes, à savoir les éléments de confort du programme de construction. Ce choix nécessaire n'est pas définitif. Dès que ces variables présenteront un taux de réponse satisfaisant, on pourra envisager leur réintroduction dans le modèle. Cela permettra d'améliorer la qualité de l'indice.

Cette médiocre qualité de ces variables s'explique par les objectifs actuels de l'ECLN. Elle a pour objectif principal la publication du nombre de logements neufs réservés au cours du trimestre écoulé et des prix moyens correspondants selon le type de construction (individuel groupé/collectif). Un nouvel objectif est donc assigné à cette enquête : la construction d'un indice des prix des logements neufs. Cela permettra d'insister sur l'importance pour le calcul de l'indice de certaines variables de l'enquête non utilisées jusqu'à présent et, in fine, une amélioration du modèle à la base de l'indice. De fait, le projet de refonte de cette enquête, qui a atterri en 2017, prend en compte les besoins de l'Insee pour le calcul de son indice.

Un enseignement de notre étude est que les indices obtenus sont assez robustes aux changements de méthodes. Par exemple, appliquer la méthode sur périodes adjacentes ne modifie pas l'indice Notaires-Insee basé sur la méthode de valorisation d'un panier de logements ; ou encore, pour les prix des logements neufs, appliquer la méthode du modèle global avec indicatrices temporelles conduit à un indice qui n'est pas si différent de ceux obtenus à partir du modèle sur périodes adjacentes ou sur 5 trimestres glissants. Ceci est relativement rassurant sur la validité des indices publiés. Ce qui fait la qualité de l'indice, ce sont d'abord les données utilisées. Cependant, ce document de travail préconise d'utiliser la méthode sur cinq trimestres glissants (modèle (6) - IPLN\_new), qui permet de limiter les effets de volatilité liés à l'utilisation de seulement deux trimestres de données chaque trimestre, et sans pour autant imposer une hypothèse trop forte de stabilité des coefficients dans le temps comme le fait le modèle global. C'est ce qui a été fait par l'Insee au moment de la mise à jour du calcul de l'IPLN suite à la refonte de l'ECLN en 2017.

Différentes pistes d'étude sont encore possibles. Par exemple, la méthode de calcul de l'indice du neuf pourra être revue. En effet, la spécification d'un modèle propre à chaque type de construction, individuel et collectif, paraît souhaitable. L'indice du neuf serait alors obtenu par agrégation de ces deux sous-indices. De plus, la possibilité de l'élaboration d'un modèle par zone géographique pourra de nouveau être étudiée si le nombre d'observations devient suffisant. Cela permettra la publication d'un indice du neuf pour chacune de ces zones. Avant de définir la finesse du découpage géographique, il faudra au préalable s'assurer qu'on dispose des pondérations permettant d'agréger les sous-indices.

## ANNEXES

<b>ANNEXE 0 : MISE A JOUR DE L'INDICE DES PRIX DES LOGEMENTS NEUFS APRES ATTERRISSAGE DE LA REFONTE DE L'ECLN (T1 2017) – SYNTHÈSE .....</b>	<b>53</b>
<b>L'INDICE DES PROPRIETAIRES-OCCUPANTS (INDICE OOH – OOH INDEX).....</b>	<b>54</b>
ANNEXE 1 : L'INDICE DES PROPRIETAIRES-OCCUPANTS (INDICE OOH) .....	54
ANNEXE 2 : L'INDICE DES PRIX DES LOGEMENTS (IPL) VS L'INDICE DES PROPRIETAIRES-OCCUPANTS (INDICE OOH).....	56
<b>L'INDICE DES PRIX DES LOGEMENTS (IPL) .....</b>	<b>59</b>
ANNEXE 3 : LA DEFINITION DES LOGEMENTS NOUVEAUX (OU NEUFS) ET DES LOGEMENTS EXISTANTS (OU ANCIENS) .....	59
ANNEXE 4 : LE CALCUL DES POIDS DU NEUF ET DE L'ANCIEN .....	61
<b>L'ENQUETE SUR LA COMMERCIALISATION DES LOGEMENTS NEUFS (ECLN) .....</b>	<b>63</b>
ANNEXE 5 : LA FICHE METHODOLOGIQUE DE L'ECLN AVANT 2017.....	63
ANNEXE 6 : LES QUESTIONNAIRES ASSOCIES A L'ECLN AVANT LA REFONTE DE 2017.....	66
<b>L'INDICE DES PRIX DES LOGEMENTS NEUFS (IPLN) .....</b>	<b>69</b>
ANNEXE 7 : LE TAUX DE COUVERTURE DE L'IPLN .....	69
ANNEXE 8 : LES ZONES D'ETUDE ET D'AMENAGEMENTS DU TERRITOIRE (ZEAT) ET LES TAILLES D'UNITE URBAINE (TUU).....	70
ANNEXE 9 : LA COMPOSITION DES 15 CLASSES GEOGRAPHIQUES RETENUES.....	71
ANNEXE 10 : LE NOMBRE D'OBSERVATIONS ET LE NOMBRE DE LOGEMENTS RESERVES TRIMESTRIELS ISSUS DE L'ECLN (2006T1-2012T3) .....	73
ANNEXE 11.1 : DEFINITION DES VARIABLES INDICATRICES INTERVENANT DANS LE MODELE (1).....	75
ANNEXE 11.2 : DES RESIDUS D'ESPERANCE NULLE .....	76
ANNEXE 12 : LES VALEURS DES ESTIMATEURS DES PARAMETRES ET LES INTERVALLES DE CONFIANCE ASSOCIES AU NIVEAU 5% POUR LES MODELES (1) ET (1 BIS) .....	77
ANNEXE 13 : LES P-VALUE CORRESPONDANT AU TEST DE NULLITE DU PARAMETRE POUR LES MODELES (1) ET (1 BIS) .....	97
<b>PREMIERE ETUDE : TEST DE SENSIBILITE DE L'INDICE DES PRIX DES LOGEMENTS ANCIENS AU CHOIX DE METHODES DE CALCUL.....</b>	<b>100</b>
ANNEXE 14 : LES LOGEMENTS ANCIENS RETENUS POUR LE CALCUL DES INDICES NOTAIRES-INSEE .....	100
<b>SECONDE ETUDE : TEST DE SENSIBILITE DE L'INDICE DES PRIX DES LOGEMENTS ANCIENS A LA DISTRIBUTION GEOGRAPHIQUE DES LOGEMENTS .....</b>	<b>102</b>
ANNEXE 15 : LA LIMITE DES DISTRIBUTIONS GEOGRAPHIQUES TRIMESTRIELLES .....	102
ANNEXE 16 : L'IPLN-MAISONS NEUVES ET L'IPLN-APPARTEMENTS NEUFS .....	105
<b>DIAGNOSTICS.....</b>	<b>107</b>
ANNEXE 17 : LES VARIABLES EXPLICATIVES DU MODELE (4) PRESENTANT UN PROBLEME DE COLINEARITE AU SENS DU VIF SUR LA PERIODE 2006T1-2012T3 .....	107
ANNEXE 18 : LES INDICES DE CONDITIONNEMENT ET LES PROPORTIONS DE VARIANCE ASSOCIEES CORRESPONDANT AU MODELE (4) ET REVELANT UN PROBLEME DE COLINEARITE (2006T1-2012T3).....	108
ANNEXE 19 : GRAPHIQUES DES RESIDUS EN FONCTION DES VALEURS ESTIMEES POUR LE MODELE (4) (2006T1- 2012T3).....	109
ANNEXE 20 : LES QQ-PLOTS DES RESIDUS POUR LE MODELE (4) (2006T1-2012T3).....	110
ANNEXE 21 : LE NOMBRE D'OBSERVATIONS ABERRANTES ET INFLUENTES POUR CHAQUE REGRESSION POUR LE MODELE (1) (2006T1-2012T3) .....	111
ANNEXE 22 : LA COMPOSITION DES 12 NOUVELLES CLASSES GEOGRAPHIQUES .....	112
ANNEXE 23 : LES VARIABLES DE CONFORT DU MODELE (1) .....	114
ANNEXE 24 : LES NOUVELLES VARIABLES APPARAISSANT DANS LE MODELE (1 BIS) .....	118
ANNEXE 25 : LES TAUX DE CROISSANCE EN GLISSEMENT TRIMESTRIEL ET EN GLISSEMENT ANNUEL DES INDICES OBTENUS A PARTIR DES MODELES (1) ET (1 BIS) .....	120
ANNEXE 26 : LA FICHE METHODOLOGIQUE DE L'INDICE DU COUT DE LA CONSTRUCTION (ICC).....	122

ANNEXE 27 : LES VALEURS DES ESTIMATEURS DES PARAMETRES ET LES INTERVALLES DE CONFIANCE ASSOCIES AU NIVEAU 5% POUR LES MODELES (1 BIS), (5) ET (6).....	126
ANNEXE 28 : LES TAUX DE CROISSANCE EN GLISSEMENT TRIMESTRIEL ET EN GLISSEMENT ANNUEL DES INDICES OBTENUS A PARTIR DES MODELES (1 BIS), (5) ET (6) .....	140



## **Annexe 0 : Mise à jour de l'indice des prix des logements neufs après atterrissage de la refonte de l'ECLN (T1 2017) – Synthèse**

L'enquête sur le prix des logements neufs (ECLN) produite par le SOeS fait l'objet d'une refonte qui a atterri pour les données du T1 2017. Les données sont dorénavant disponibles au logement, alors qu'elles étaient auparavant renseignées en moyenne par tranches de commercialisation des programmes immobiliers, ce qui constitue un gain de précision en matière de collecte.

L'ECLN est utilisée par l'Insee pour le calcul de l'indice de prix des logements neufs, celui-ci a donc été mis à jour en conséquence. Le champ de l'indice a été légèrement modifié avec l'ajout de la Corse. Une actualisation du modèle hédonique utilisé a été effectuée à cette occasion : certaines variables ont été supprimées et d'autres ont été ajoutées, le modèle est estimé sur cinq trimestres au lieu de deux.

Les modifications ont un impact assez faible, l'indice n'a pas été rétrapolé ; les nouvelles évolutions ont été chaînées à l'indice existant.

## L'indice des propriétaires-occupants (indice OOH – OOH index)

### Annexe 1 : L'indice des propriétaires-occupants (indice OOH)

L'adoption d'une monnaie commune a fait naître le besoin d'une mesure harmonisée de l'inflation dans la zone euro. Ainsi, Eurostat a développé, en coopération avec les instituts statistiques nationaux, des indices des prix à la consommation harmonisés (IPCH), établis selon une méthodologie commune dans tous les pays de l'Union européenne. Les indices nationaux, qui sont encore publiés parallèlement aux IPCH, sont très hétérogènes car ils reposent sur des conceptions sensiblement différentes de ce qui doit être mesuré d'un pays à l'autre. En effet, le champ d'un indice des prix à la consommation peut être plus ou moins étendu, selon le choix de la frontière entre consommation et investissement.

- En particulier, les services de logement sont intégrés dans l'IPC français, a minima, à travers :
- les loyers d'habitation versés par les locataires (regroupement 04.1 de la COICOP),
  - le prix des biens et services d'entretien et réparation courante du logement, à l'exclusion des gros travaux (regroupement 04.3).
  - le prix de la fourniture d'eau et d'autres services liés au logement (enlèvement des ordures ménagères, assainissement, ...) (regroupement 04.4)

Cependant, certains instituts statistiques prennent également en compte les services de logement fournis aux propriétaires occupant leur logement dans leur indice des prix à la consommation. Cette inclusion est délicate car les dépenses correspondant à ces services se situent à la limite de la consommation et de l'investissement. En outre, elles ne sont pas directement observables, et leur mesure fait l'objet de conventions statistiques dans les pays qui ont choisi de les intégrer dans leur indice des prix à la consommation national. Il en résulte, par conséquent, des différences majeures dans le traitement de ces dépenses par les instituts statistiques nationaux. En France, un indice des prix à la consommation complémentaire incluant les loyers imputés est calculé : il est obtenu en ajoutant les dépenses correspondant aux loyers imputés à la pondération associée au poste IPC « Loyers des résidences principales » (04.1.1.1.1).

Dans un souci d'harmonisation de la prise en compte dans les IPC nationaux des dépenses des propriétaires occupant leur logement liées à l'acquisition et à l'entretien de leur logement, Eurostat a lancé le projet OOH. L'un des objectifs de ce projet est, en effet, la création d'un indice des propriétaires-occupants (indice OOH). Cet indice a pour but de mesurer les variations des prix de transaction des logements nouveaux dans le secteur des ménages (**concept des acquisitions<sup>50</sup> nettes<sup>51</sup>**) ainsi que d'autres biens et services que les ménages acquièrent en qualité de propriétaires-occupants. Il pourrait éventuellement être intégré à l'IPCH à la fin du projet.

---

<sup>50</sup> Ce concept définit la valeur totale de l'acquisition d'un bien comme étant celle au moment de l'achat, peu importe la date à laquelle le bien commencera à être consommé et la manière dont il sera payé (par opposition au concept d'utilisation et au concept des paiements)

<sup>51</sup> On considère uniquement les acquisitions nouvelles pour le champ macroéconomique des ménages. Les ventes entre ménages sont ainsi exclues.

Le règlement OOH, entré en vigueur le 22 février 2013, contient une nomenclature définissant les différentes catégories de dépense que devra couvrir le futur indice des propriétaires-occupants :

**Tableau 6.1 : Le champ de l'indice des propriétaires-occupants (indice OOH)**

Code	Libellé
O.1	DEPENSES EN MATIERE DE LOGEMENT DES PROPRIETAIRES-OCCUPANTS
O.1.1.	ACQUISITIONS DE LOGEMENTS
O.1.1.1.	Logements nouveaux
O.1.1.1.1.	<i>Achats de logements nouveaux</i>
O.1.1.1.2.	<i>Logements construits par les propriétaires eux-mêmes et gros travaux de rénovations</i>
O.1.1.2.	Logements existants, nouveaux dans le secteur des ménages
O.1.1.3.	Autres services liés à l'acquisition de logements
O.1.2.	PROPRIETE DES LOGEMENTS
O.1.2.1.	Grosses réparations et entretien
O.1.2.2.	Assurances liées aux logements
O.1.2.3.	Autres services liés à la propriété des logements

La cinquième phase du projet OOH qui se terminera en mars 2014 a pour but de finaliser la construction d'un indice des propriétaires-occupants dans chaque pays membre. Cette phase aura notamment pour but de résoudre les derniers problèmes méthodologiques auxquels sont confrontés certains pays. En particulier, la cinquième phase sera l'occasion de finaliser la construction d'un indice des prix des logements neufs ne prenant pas en compte le prix du terrain, i.e. « expurgé » de la composante « investissement » du prix d'un logement. En effet, on peut considérer, en première approximation, que le prix du terrain correspond à la composante « investissement » dans l'achat d'un logement, et que le prix du logement hors terrain correspond à la composante « consommation ». Dans la mesure où l'indice des propriétaires-occupants a pour vocation d'intégrer éventuellement l'indice des prix à la consommation, cette exclusion apparaît indispensable. Devant les difficultés rencontrées par certains États membres à réaliser cette exclusion, Eurostat tolère pour le moment un indice des prix des logements neufs incluant le prix du terrain.

## Annexe 2 : L'indice des prix des logements (IPL) vs l'indice des propriétaires-occupants (indice OOH)

Comme précisé au §1.1., il est important de garder à l'esprit que **l'indice des prix des logements (IPL) et l'indice des propriétaires-occupants (indice OOH) sont deux indices différents qui ne doivent pas être comparés directement.**

En effet, l'indice des prix des logements mesure l'évolution des **prix de transaction** de toutes les propriétés résidentielles acquises par les ménages. Contrairement à l'indice des propriétaires-occupants, l'IPL n'a pas été élaboré selon le concept des acquisitions nettes mais selon le concept des acquisitions brutes, et le prix du terrain est inclus dans les prix des logements et dans les pondérations.

Par conséquent,

- puisque l'accent est mis sur les prix de transactions des logements, les prix ne correspondant à aucune transaction sont exclus du champ de l'IPL ; un exemple d'un tel prix serait le prix final obtenu pour un logement qui a été construit petit à petit à l'initiative de l'acquéreur ;
- Cependant, les maisons neuves achetées « clés en main » peuvent être considérées comme ayant elles ont un prix de marché ; mais Eurostat laisse pour le moment la possibilité aux États membres de les exclure du champ de l'IPL ;
- comme l'objectif de l'IPL est de mesurer l'évolution des prix de **tous les logements acquis par les ménages**, l'usage qu'en font les ménages n'intervient pas dans la définition du champ de l'IPL ; ainsi, les logements acquis par des ménages ne souhaitant pas y habiter sont couverts par l'IPL (exemple : investissement locatif) ;
- comme tous les achats de logements doivent être couverts par l'IPL, toutes les acquisitions de logements anciens appartiennent au champ de l'IPL, y compris ceux correspondant à une transaction entre ménages ;
- dans la mesure où dans la définition du champ de l'IPL, l'accent est uniquement mis sur le coût d'acquisition du logement, et non pas sur le coût total correspondant à son acquisition, à sa propriété et à son entretien, les « autres coûts liés à l'acquisition du logement » (frais et taxes) et les grosses réparations sont exclus du champ de l'IPL.

Le tableau suivant donne une vue synthétique des différences existant entre le champ de l'indice OOH et le champ de l'IPL :

**Tableau 6.2 : Le champ de l'IPL vs le champ de l'indice OOH**

Code OOH	Libellé OOH	Dans le champ de l'indice OOH ?	Dans le champ de l'IPL ?	Code IPL	Libellé IPL
O.1	DEPENSES EN MATIERE DE LOGEMENT DES PROPRIETAIRES- OCCUPANTS			H.1.	ACHATS DE LOGEMENTS
O.1.1.	ACQUISITIONS DE LOGEMENTS				
O.1.1.1.	Logements nouveaux				
<b>O.1.1.1.1.</b>	<b>Achats de logements nouveaux (=neufs)</b>	<b>* uniquement les logements neufs acquis à des fins d'occupation par le propriétaire * le prix du terrain doit être, en théorie, exclu du prix de vente du logement</b>	<b>* tous les logements neufs acquis par des ménages quel que soit l'usage qu'en font les ménages * le prix du terrain est inclus dans le prix de vente du logement</b>	<b>H.1.1.</b>	<b>Achats de logements nouveaux</b>
O.1.1.1.2.	<i>Logements construits par les propriétaires eux-mêmes et gros travaux de rénovations</i>	<i>oui</i>	<i>non</i>		
<b>O.1.1.2.</b>	<b>Logements existants, nouveaux dans le secteur des ménages</b>	<b>* oui * le prix du terrain doit être, en théorie, exclu du prix de vente du logement</b>	<b>* oui + logements existants, dont l'ancien propriétaire est un ménage * le prix du terrain est inclus dans le prix de vente du logement</b>	<b>H.1.2.</b>	<b>Achats de logements existants</b>
O.1.1.3.	Autres services liés à l'acquisition de logements	oui	non		
O.1.2.	PROPRIETE DES LOGEMENTS				
O.1.2.1.	Grosses réparations et entretien	oui	non		
O.1.2.2.	Assurances liées aux logements	oui	non		
O.1.2.3.	Autres services liés à la propriété des logements	oui	non		

En résumé, les principales différences entre le champ de l'indice OOH et le champ de l'IPL sont les suivantes :

- le prix du terrain est inclus dans le prix retenu pour calculer l'IPL alors qu'il est en principe exclu dans le cadre de l'indice OOH ;
- l'IPL couvre tous les logements neufs acquis par un ménage quelle que soit l'utilisation qu'en fait le ménage alors que l'indice OOH ne couvre que les logements neufs acquis par un ménage à des fins d'occupation personnelle ;
- l'IPL couvre tous les logements anciens acquis par un ménage, y compris ceux faisant l'objet d'une transaction entre ménages alors que ces derniers sont exclus du champ de l'indice OOH.

## L'indice des prix des logements (IPL)

### Annexe 3 : La définition des logements nouveaux (ou neufs) et des logements existants (ou anciens)

Au niveau européen

#### *Les achats de logements nouveaux*

Dans le guide technique sur l'OOH, il est précisé que la sous-section « H.1.1. Achats de logements nouveaux » couvre l'acquisition de logements nouvellement construits. Ces logements sont de différents types : appartements, maisons individuelles (y compris celles « clés en main » construites par des promoteurs), ...

La sous-section H.1.1. peut être divisée en deux grandes catégories :

- les appartements neufs (logements constituant une unité d'habitation individuelle dans un immeuble neuf) ;
- les maisons neuves (tous les types de maisons : maisons individuelles, maisons mitoyennes, etc.).

Par ailleurs, il est précisé dans le guide qu'un logement construit à un trimestre T mais vendu quelques trimestres plus tard doit être considéré comme un logement neuf au trimestre durant lequel il est vendu.

#### *Les achats de logements existants*

Selon le guide technique sur l'OOH, les achats de logements anciens couverts par la sous-section H.1.2. sont les acquisitions de logements qui ont déjà fait l'objet d'au moins une transaction ou qui ont été construits à l'initiative de l'acquéreur puis occupés par ce dernier.

En théorie, la nomenclature européenne forme une partition de l'ensemble des logements. Cependant, en pratique, Eurostat tolère la non-couverture des maisons individuelles vendues « clés en main ».

Au niveau français

#### *Les achats de logements nouveaux*

Les achats de logements nouveaux couverts par l'IPL français sont exclusivement ceux appartenant au champ de l'Enquête de commercialisation des logements neufs (ECLN). Ils correspondent ainsi à l'acquisition d'un logement neuf faisant partie d'un programme de construction de 5 logements et plus destinés à la vente aux particuliers et situés en France métropolitaine. De fait, les acquisitions de logements neufs par des ménages correspondant à des programmes de construction de 1 à 4 logements ne sont pas couverts par l'indice des prix des logements neufs français.

Par conséquent, l'indice des prix des logements neufs français présente un défaut de couverture. Cependant, ce dernier est minime (cf. annexe 7).

### *Les achats de logements existants*

Les achats de logements existants couverts par l'indice des prix des logements anciens sont les mêmes que ceux appartenant au champ de l'indice Notaires-Insee. L'indice Notaires-Insee porte sur les logements anciens au sens fiscal<sup>52</sup>, i.e. de plus de 5 ans ou connaissant une deuxième mutation. Afin d'éviter les transactions atypiques, les biens non standards tels que les chambres, lofts, ateliers ou châteaux et les acquisitions effectuées par des professionnels de l'immobilier sont exclus. De plus, le champ couvert par l'indice Notaires-Insee n'est constitué que de biens :

- libres d'occupation au moment de la vente (ou occupés par le vendeur)
- acquis en pleine propriété par une vente de gré à gré
- destinés à un usage strict d'habitation

Par conséquent, on peut considérer que l'indice des prix des logements anciens français ne présente pas de défaut de couverture.

Les achats de logements neufs et les acquisitions de logements existants ne provenant pas de la même source et l'indice des prix des logements neufs français présentant un défaut de couverture, les sous-sections H.1.1. et H.1.2. ne forment pas exactement une partition de l'ensemble des logements au niveau français.

---

<sup>52</sup> Cette définition peut donc changer au cours du temps au gré des changements de fiscalité



## **Annexe 4 : Le calcul des poids du neuf et de l'ancien**

Pour déterminer les pondérations annuelles intervenant dans le calcul de l'indice des prix des logements français, on utilise les données relatives aux acquisitions de logements et aux travaux figurant dans les comptes satellites du logement établis par le SOeS.

On dispose du montant de chacun des types d'activité immobilière (acquisitions de logements neufs, de logements d'occasion, travaux sur logements existants) pour chaque catégorie de propriétaires (propriétaires occupants accédants, propriétaires occupants non accédants, bailleurs personnes physiques, organismes HLM, autres bailleurs sociaux, autres personnes morales, propriétaires de résidences secondaires). Pour les acquisitions de logements neuf et d'occasion, on distingue :

- le coût des bâtiments hors terrain,
- le coût des terrains,
- les frais induits,
- les droits de mutation.

Pour les logements neufs, dans les rubriques « bâtiments hors terrain » et « terrains d'assise », on sépare les logements construits à l'initiative de l'acquéreur de ceux achetés à des promoteurs.

Étant donné le champ de l'indice des prix des logements (IPL) et dans la mesure où le prix du terrain doit être pris en compte, les poids du neuf et de l'ancien sont calculés de la manière suivante :

**Tableau 6.3 : Calcul détaillé des pondérations de l'IPL pour l'année 2012**

Code	Libellé	Tableau 23 « Les acquisitions de logements et les travaux »			Montant 2010 (en millions d'euros)		Pondérations 2012 (en %)
H.1.	ACHATS DE LOGEMENTS				155 835,3		100,0
H.1.1.	Achats de logements nouveaux	23.112	Bâtiments hors terrain achetés à des promoteurs	Résidences principales - Ensemble des propriétaires-occupants	8 902,5	20 508,5	13,2
				Résidences principales - Bailleurs personnes physiques	7 100,9		
		23.122	Terrains d'assise des logements vendus par des promoteurs	Résidences principales - Ensemble des propriétaires-occupants	2 633,1		
				Résidences principales - Bailleurs personnes physiques	1 872,0		
H.1.2.	Achats de logements existants	23.31	Montants des acquisitions hors terrains	Résidences principales - Ensemble des propriétaires-occupants	88 376,1	135 326,8	86,8
				Résidences principales - Bailleurs personnes physiques	8 248,6		
		23.32	Montant des terrains dans les acquisitions	Résidences principales - Ensemble des propriétaires-occupants	36 097,3		
				Résidences principales - Bailleurs personnes physiques	2 604,8		

source : comptes satellites du logement 2010 - SOeS

# L'Enquête sur la commercialisation des logements neufs (ECLN)

## Annexe 5 : La fiche méthodologique de l'ECLN avant 2017

### Description de la source

ECLN est une enquête trimestrielle assurant le suivi de la commercialisation des logements neufs destinés à la vente aux particuliers depuis 1985.

Le Comité du label du Cnis lui a attribué le label d'intérêt général et de qualité statistique et a proposé de lui conférer le caractère obligatoire en décembre 2010. Ce label a ainsi été accordé à l'ECLN jusqu'en 2015. Son caractère obligatoire est, quant à lui, indiqué sur le visa ministériel l'autorisant et est rendu public par un arrêté au Journal officiel donnant la liste des enquêtes inscrites au programme annuel<sup>53</sup>. L'ECLN est également soumise à une déclaration à la Cnil.

### Concepts et définitions propres à la source

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.

Le **champ de l'enquête** couvre **les permis de 5 logements et plus destinés à la vente aux particuliers**.

Sont donc exclus :

- le secteur locatif (permis de construire destinés à la location) : il s'agit des programmes de construction dont les logements sont destinés à la location et loués directement par le promoteur (à ne pas confondre avec l'investissement locatif qui correspond à un logement acquis par un ménage pour être loué par la suite),
- les constructions individuelles réalisées par des particuliers en vue de leur occupation personnelle (permis de construire d'un logement). Les maisons individuelles vendues « clés en main » sont des exemples de telles constructions,
- les logements construits par l'État, les collectivités locales et les sociétés nationalisées ; en réalité, seuls les logements destinés à la location sociale ou en accession sociale sont exclus du champ de l'ECLN,
- l'ensemble des logements de fonction,
- les programmes de réhabilitation

L'enquête est exhaustive sur son champ.

---

<sup>53</sup> Voir l'arrêté du 2 octobre 2012 portant approbation du programme d'enquêtes statistiques d'initiative nationale et régionale des services publics pour 2013

L'**unité statistique** 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.

## Questionnaires d'enquête

L'enquête est réalisée en France métropolitaine par les cellules statistiques des Directions régionales de l'équipement.

La collecte des données est faite par des enquêteurs et/ou par voie postale.

Le questionnaire est à deux niveaux :

- 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 :
  - o d'identifier le programme et de connaître ses caractéristiques : numéro de permis de construire, nom et adresse du programme... ;
  - o de recueillir des éléments portant sur sa qualité : catégorie, éléments de confort ;
  - o d'identifier les interlocuteurs : pétitionnaire, promoteur, correspondant.
  
- la **fiche tranche** assure le suivi de la commercialisation du programme suivant le nombre de pièces des logements. Elle est ouverte dès qu'une partie ou la totalité du programme est offerte à la vente et que sa commercialisation est en cours lors du trimestre enquêté. Elle ne comprend que des logements :
  - o de même type (individuel groupé ou collectif),
  - o de même nature (logements ordinaires, résidences avec services pour personnes âgées, résidences de vacances, logements sociaux),
  - o et, pour le collectif, au même stade d'avancement des travaux (en projet, en cours, achevé).

Elle décrit :

- les mises en vente, soit les logements nouvellement offerts à la vente au cours du trimestre enquêté ;
- le nombre de logements vendus : ce sont plus exactement les logements réservés, c'est-à-dire 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 disponibles à la vente à la fin du trimestre. Ils sont obtenus par différence entre les logements mis en vente et les logements vendus, auxquels on rajoute les stocks à la fin du trimestre précédent.

Ils seront éventuellement modifiés sur la fiche tranche du trimestre suivant pour tenir compte d'une part des annulations (renoncement à la vente malgré la réservation), d'autre part des retraits de commercialisation (abandon partiel ou total, mise en location, mise en location-accession, vente en multipropriété ou en globalité, retrait de vente, transformation de logements), ou encore de la modification de la structure par nombre de pièces des logements commercialisés.

Pour les **logements vendus**, sont demandés selon la taille du logement :

- la surface moyenne des terrains, ou balcons et terrasses (en mètres carrés) ;
- le nombre de logements destinés à l'investissement locatif : ce système consiste à acheter un ou plusieurs logements pour le (ou les) louer, et se constituer ainsi un patrimoine et percevoir des revenus complémentaires ;
- la surface habitable moyenne : pièces principales d'habitation, de service (cuisine, WC, salle de bains) et de circulation (hall, couloir) ;
- le prix moyen : il s'agit pour le collectif du prix du logement, garage ou parking exclu, et pour les maisons individuelles groupées du prix global de la maison, de ses dépendances et du terrain.

## Qualité des données et conditions d'utilisation

Il n'y a pas de redressement des données.

Les données sont disponibles au niveau de la France métropolitaine, des régions et de zones plus fines (sous réserve du respect des règles du secret statistique).

## Liste des variables

Les principales variables publiées sont :

- les mises en vente et les réservations,
- les prix de vente,
- les stocks disponibles à la vente,
- les délais d'écoulement des stocks (temps nécessaire pour que les logements disponibles à la vente soient vendus),
- le taux de renouvellement des stocks (rapport entre les mises en vente du trimestre et les stocks en début de trimestre).

## Annexe 6 : Les questionnaires associés à l'ECLN avant la refonte de 2017

### Le questionnaire de base

#### ENQUÊTE COMMERCIALISATION DES LOGEMENTS NEUFS QUESTIONNAIRE DE BASE

Région : .....

Trimestre d'enquête :

200\_\_/\_\_\_

##### IDENTIFICATION DU PROGRAMME

Département	Commune	Année de dépôt	N° du permis	Découpage infra communal	Date avis favorable	Logements autorisés	Type cons.	Trimestre de 1ère enquête
__	____	__	_____	_____	_____	_____	__	200__/___

NOM DU PROGRAMME : .....

Adresse du lieu des travaux : .....

Code postal \_\_\_\_\_ Commune : .....

##### IDENTIFICATION DU PETITIONNAIRE

Nom : ..... N° SIRET : .....

Adresse : .....  
.....

Code postal \_\_\_\_\_ Commune : ..... N° : \_\_\_\_\_  
(Dept. + n° ordre)

Tél : \_\_\_\_\_ Fax : \_\_\_\_\_  
\_\_\_\_\_ Mèl : .....

##### IDENTIFICATION DU PROMOTEUR

Nom : ..... N° SIRET : .....

Adresse : .....  
.....

Code postal \_\_\_\_\_ Commune : ..... N° : \_\_\_\_\_  
(Dept. + n° ordre)

Tél : \_\_\_\_\_ Fax : \_\_\_\_\_  
\_\_\_\_\_ Mèl : .....

##### PERSONNE OU ORGANISME HABILITE A REpondre A L'ENQUETE (le correspondant)

Nom : .....

Adresse : ..... N° : \_\_\_\_\_  
.....

Code postal \_\_\_\_\_ Commune : .....

Tél : \_\_\_\_\_ Fax : \_\_\_\_\_  
\_\_\_\_\_ Mèl : .....

Vu l'avis favorable du Conseil national de l'information statistique, cette enquête, reconnue d'intérêt général et de qualité statistique, est obligatoire.

Visa n°2013T048LO du Ministre de l'écologie, du développement durable et de l'énergie et du Ministre de l'économie et des finances valable pour l'année 2013.

Aux termes de l'article 6 de la loi n° 51-711 du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques, les renseignements transmis en réponse au présent questionnaire ne sauraient en aucun cas être utilisés à des fins de contrôle fiscal ou de répression économique.

L'article 7 de la loi précitée stipule d'autre part que tout défaut de réponse ou une réponse sciemment inexacte peut entraîner l'application d'une amende administrative. Questionnaire confidentiel destiné au Service de l'observation et des Statistiques du commissariat général au développement durable. La loi n° 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, garantit aux personnes physiques concernées un droit d'accès et de et de rectification pour les données les concernant. Ce droit peut être exercé auprès du SoeS – Tour Voltaire – 92055 La Défense cedex.

Conformément à l'avis du comité du secret statistique du conseil national de l'information statistique et après concertation avec la fédération des promoteurs-construc-teurs de France, les règles de diffusion des résultats de cette enquête ont été aménagées, afin de mieux répondre à la demande d'informations du secteur. Il a pour cela été tenu compte du caractère plus ou moins sensible de l'information collectée (données peu sensibles [1], données sensibles [2], repérées sur le questionnaire ci-joint). Les règles mises en place sont précisées sur le site Internet du ministère à l'adresse suivante : <http://www.statistiques.developpement-durable.gouv.fr>, rubrique « Répondre aux enquêtes » - « Enquête sur la commercialisation des logements neufs » - section « Les règles de diffusion ».

## IDENTIFICATION DU PROGRAMME

Département	Commune	Année de dépôt	N° du permis
_ _	_ _ _	_ _	_ _ _ _

## PERSONNE CHARGEE DE LA COLLECTE DE L'INFORMATION

Nom : .....	N° :  _ _ - _ _ _ _
Région où a lieu l'enquête : .....	(Région + n° enquêteur)
Tél :  _ _ _ _ _ _ _ _ _ _	Fax : .....
	Mèl : .....

## QUALITE DU PROGRAMME

### CATEGORIE DU PROGRAMME

En dehors de toute considération et argument promotionnel de vente, et en sachant que votre réponse est strictement confidentielle, quelle est votre opinion sur le niveau de standing de cette opération de construction ?

courant ou ordinaire       confortable       très confortable       luxe

L'opération est-elle certifiée QUALITEL ?       oui       non

### ELEMENTS DE CONFORT DU PROGRAMME

> Présence d'un dispositif de sécurité (gardien, vidéo-surveillance)       oui       non

> Présence d'une piscine dans la résidence       oui       non

> Accès à moins de 10 minutes à pieds d'un transport collectif par voie ferrée (train, tramway, métro, RER)       oui       non

> Climatisation       oui       non

### CARACTERISTIQUES DU PROGRAMME

SI LE PROGRAMME EST DEFINITIVEMENT ABANDONNE, ne pas continuer à remplir le questionnaire.

SINON, parmi les logements autorisés, combien sont destinés :

1. à la VENTE .....	:  _ _ _  (*)	5. à la MULTIPROPRIETE .....	:  _ _ _
2. à la VENTE en globalité (investisseur ou autre) .....	:  _ _ _	6. Autres cas (précisez) .....	:  _ _ _
3. à la LOCATION .....	:  _ _ _	7. à la LOCATION SOCIALE .....	:  _ _ _
4. à la LOCATION ACCESSION .....	:  _ _ _	8. à la LOCATION ACCESSION SOCIALE .....	:  _ _ _

(\*) Pour la première mise en vente :

Si la première mise en vente a eu lieu, pouvez-vous en indiquer l'année et le trimestre      |\_|\_|\_|\_| / |\_|

Sinon, à quelle date prévoyez-vous de la réaliser ? indiquez, si possible, l'année et le trimestre      |\_|\_|\_|\_| / |\_|

### SI LA MISE EN VENTE A DEBUTE, REMPLIR AU MOINS UNE FICHE TRANCHE

Code enquête      |\_|\_|

# La fiche tranche

## ENQUÊTE COMMERCIALISATION DES LOGEMENTS NEUFS FICHE TRANCHE

Région : ..... Trimestre d'enquête : 201|\_| / |\_|  
 N° enquêteur : -  
 Code enquête : |\_|

### PERSONNE OU ORGANISME HABILITE A REpondre A L'ENQUÊTE (le correspondant)

Nom : ..... (Dépt. + n° ..... N° : |\_|\_|\_|\_|\_|\_|  
 Tél : |\_|\_|\_|\_|\_|\_|\_|\_|\_|\_| Mobile : |\_|\_|\_|\_|\_|\_|\_|\_|\_|\_| Mél : .....  
 Adresse : .....  
 Code postal : |\_|\_|\_|\_|\_| Commune : .....

### IDENTIFICATION DE LA TRANCHE

N° SIRET du promoteur : |\_|\_|\_|\_|\_|\_|\_|\_|\_|\_|\_|\_|\_|\_|\_|\_|

Département	Commune	Année de dépôt	N° du permis	N° tranche	Commune du lieu des travaux	Dernier trimestre enquêté
_ _	_ _ _ _	_ _	_ _ _ _ _	_ _	.....	201 _ _  /  _ _
NOM DE LA TRANCHE : .....						(année / trimestre)
Adresse des travaux : .....						

### DESCRIPTION DE LA TRANCHE

<b>TYPE DE CONSTRUCTION :</b>  _  1 - Individuel groupé 2 - Collectif	<b>STADE D'AVANCEMENT DES TRAVAUX :</b>  _  (Pour les individuels : état des vendus du trimestre, pour les collectifs : état de la tranche) 1 - En projet 2 - En cours 3 - Achevé	<b>NATURE DES LOGEMENTS :</b>  _  1 - Logements ordinaires 2 - Résidences avec services 3 - Résidences de vacances 4 - Logements sociaux
Nombre de logements de la tranche déclarés lors de la 1ère enquête :  _ _	<b>NOMBRE DE PARKINGS de la tranche :</b> couverts :  _ _ _ _  non couverts :  _ _ _ _	<b>DATES PREVUES (année / trimestre)</b> pour le début des travaux : 201 _ _  /  _ _  pour la livraison : 201 _ _  /  _ _
LA TRANCHE EST-ELLE EN VENTE ? 2 (Oui=1, Non=2) Si NON : ne pas remplir la suite.		

### SUIVI DE LA TRANCHE (partie à renseigner seulement si la tranche a été déclarée MISE EN VENTE)

Date réelle de mise en vente (année, trimestre) : |\_|\_|\_|\_|\_| / |\_|\_| Nombre de logements mis en vente : |\_|\_|\_|\_|

Depuis la fin du trimestre précédent, si la répartition des logements par type de pièces a été modifiée,  
 1\*) indiquez la raison du changement dans la case suivante (utilisez la nomenclature proposée ci-dessous) : |\_|\_|

2\*) précisez la nouvelle ventilation en mettant à jour les informations pré-remplies de la colonne ci-dessous :

1 - Abandon	6 - Retrait de vente
2 - Mise en location	7 - Transformation de logements
3 - Mise en location accession	8 - Rectificatif
4 - Vente en multi-propriété	9 - Autres cas (précisez) .....
5 - Vente en globalité	

Typologie de la tranche lors de la mise en vente	Nombre de logements disponibles à la fin du trimestre précédent (1)	TYPE DE LOGEMENTS		Surface moyenne des terrains, ou balcons et terrasses (en m2) (2)	Nombre de logements réservés au cours du trimestre (1)	dont pour investissement locatif (2)	Nombre de logements disponibles à la fin du trimestre en cours (1)	Logements réservés au cours du trimestre écoulé	
		Nb de pièces	(**)					Surface habitable moyenne (1)	Prix moyens en EUROS (*) (2)
_ _ _	_ _ _	1	2/3	_ _ _ _	_ _ _	_ _ _	_ _ _	_ _ _	_ _ _ _ _ _ _
_ _ _	_ _ _	2	4	_ _ _ _	_ _ _	_ _ _	_ _ _	_ _ _	_ _ _ _ _ _ _
_ _ _	_ _ _	3	5	_ _ _ _	_ _ _	_ _ _	_ _ _	_ _ _	_ _ _ _ _ _ _
_ _ _	_ _ _	4	6	_ _ _ _	_ _ _	_ _ _	_ _ _	_ _ _	_ _ _ _ _ _ _
_ _ _	_ _ _	5	7	_ _ _ _	_ _ _	_ _ _	_ _ _	_ _ _	_ _ _ _ _ _ _
_ _ _	_ _ _	6 et +	8	_ _ _ _	_ _ _	_ _ _	_ _ _	_ _ _	_ _ _ _ _ _ _
_ _ _	_ _ _	ensemble		_ _ _ _	_ _ _	_ _ _	_ _ _	_ _ _	_ _ _ _ _ _ _

(\*) - individuel : prix moyen des lots (maison + terrain)  
 - collectif : prix moyen des logements (garage ou parking exclu)  
 (\*\*) Nomenclature vacances (couchages)

Nombre d'annulations de réservations enregistrées au cours du trimestre : |\_|\_|\_|\_|

(1) donnée peu sensible ; (2) donnée sensible (cf. lettre d'accompagnement)



# L'indice des prix des logements neufs (IPLN)

## Annexe 7 : Le taux de couverture de l'IPLN

Pour établir le taux de couverture de l'indice des prix des logements neufs, nous avons utilisé la base Sítadel. Nous avons tout d'abord sélectionné les permis correspondant à des logements appartenant au champ de l'IPLN. Nous avons ainsi conservé uniquement ceux tels que :

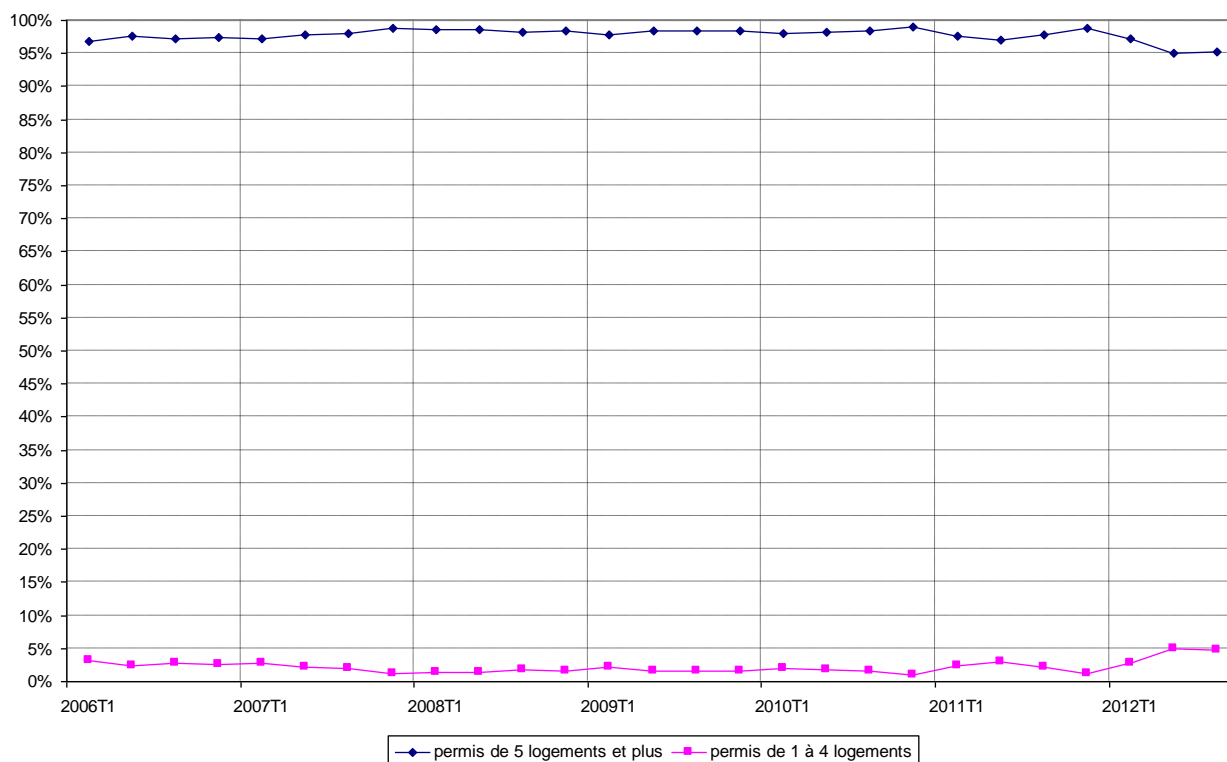
- le maître d'ouvrage est un promoteur
- et
- le mode d'utilisation est « occupation personnelle », « vente » ou « location » (i.e. investissement locatif)

Puis, nous avons établi le nombre trimestriel<sup>54</sup> de mises en chantier (i.e. le nombre de logements commencés) en distinguant :

- les permis de construire de 1 à 4 logements
- et
- les permis de construire de 5 logements et plus (champ effectif de l'IPLN)

Sur la période 2006T1-2012T3, le taux de couverture de l'IPLN est en moyenne proche des 98% (97,8% - cf. graphique 6.1). De plus, ce taux ne passe jamais sous la barre des 95%. On en conclut ainsi que le défaut de couverture de l'IPLN est minime.

**Graphique 6.1 : Le taux de couverture de l'IPLN (2006T1-2012T3)**



<sup>54</sup> En considérant la date réelle (les séries en date réelle rattachent chaque événement - autorisation, mise en chantier, annulation, achèvement - au mois au cours duquel il s'est effectivement produit)

## Annexe 8 : Les Zones d'étude et d'aménagements du territoire (ZEAT) et les tailles d'unité urbaine (TUU)

**Tableau 6.4 : Les ZEAT et leur composition**

	<b>ZEAT</b>	<b>REGIONS</b>
<b>1</b>	<b>Région parisienne</b>	Ile de France
<b>2</b>	<b>Bassin parisien</b>	Bourgogne, Centre, Champagne-Ardenne, Basse et Haute Normandie, Picardie
<b>3</b>	<b>Nord</b>	Nord Pas-de-Calais
<b>4</b>	<b>Est</b>	Alsace, Franche-Comté, Lorraine
<b>5</b>	<b>Ouest</b>	Bretagne, Pays de la Loire, Poitou-Charentes
<b>7</b>	<b>Sud-ouest</b>	Aquitaine, Limousin, Midi-Pyrénées
<b>8</b>	<b>Centre-est</b>	Auvergne, Rhône-Alpes
<b>9</b>	<b>Méditerranée</b>	Languedoc-Roussillon, Provence-Alpes-Côte d'Azur, Corse

**Tableau 6.5 : Les tailles d'unité urbaine**

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

## Annexe 9 : La composition des 15 classes géographiques retenues

**Tableau 6.6 : Les 15 classes géographiques**

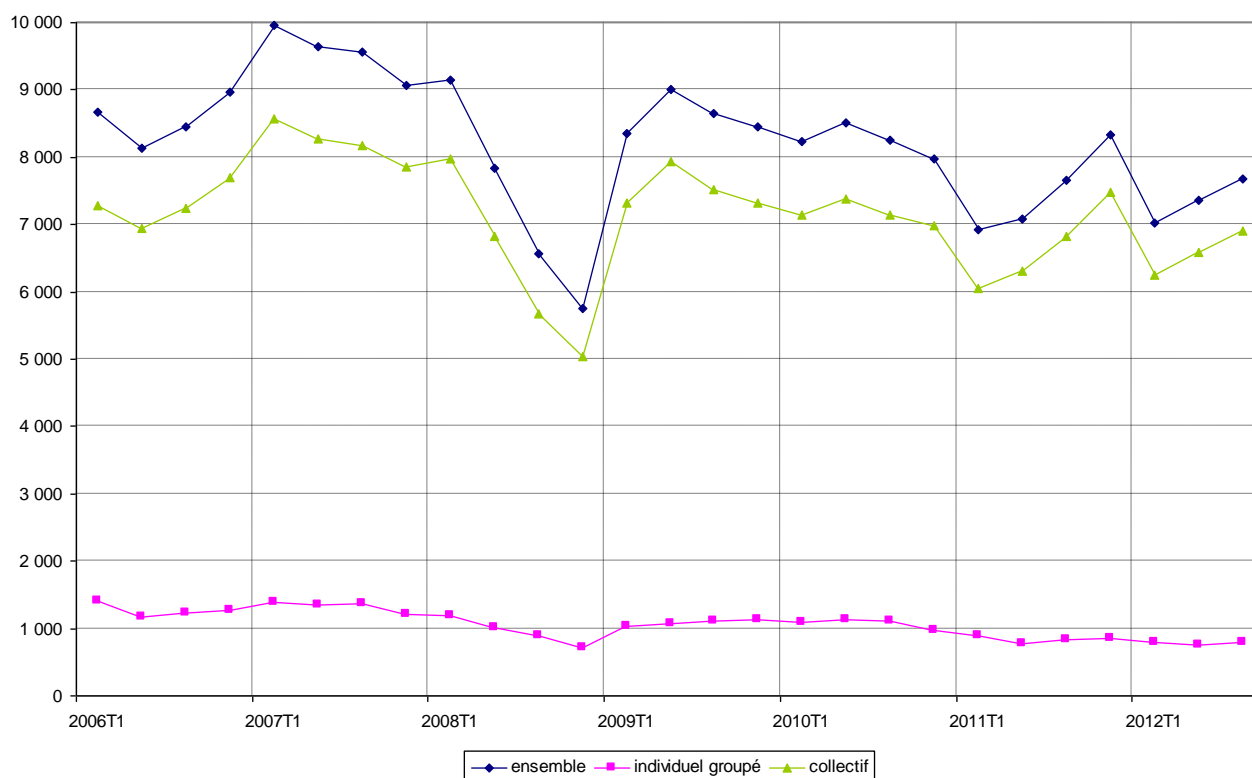
Classe géographique	ZEAT		TUU	
1	2	Bassin parisien	1	2 000 à 4 999 habitants
	4	Est	3	10 000 à 19 999 habitants
	4	Est	4	20 000 à 49 999 habitants
2	1	Région parisienne	4	20 000 à 49 999 habitants
	3	Nord	2	5 000 à 9 999 habitants
	3	Nord	6	100 000 à 199 999 habitants
	8	Centre-est	3	10 000 à 19 999 habitants
	9	Méditerranée	0	commune rurale
	9	Méditerranée	1	2 000 à 4 999 habitants
3	9	Méditerranée	2	5 000 à 9 999 habitants
	2	Bassin parisien	2	5 000 à 9 999 habitants
	2	Bassin parisien	5	50 000 à 99 999 habitants
	3	Nord	5	50 000 à 99 999 habitants
	4	Est	5	50 000 à 99 999 habitants
	5	Ouest	0	commune rurale
	5	Ouest	1	2 000 à 4 999 habitants
	7	Sud-ouest	3	10 000 à 19 999 habitants
	7	Sud-ouest	6	100 000 à 199 999 habitants
4	9	Méditerranée	6	100 000 à 199 999 habitants
	5	Ouest	2	5 000 à 9 999 habitants
	5	Ouest	3	10 000 à 19 999 habitants
	5	Ouest	5	50 000 à 99 999 habitants
	7	Sud-ouest	0	commune rurale
	7	Sud-ouest	1	2 000 à 4 999 habitants
5	7	Sud-ouest	2	5 000 à 9 999 habitants
	2	Bassin parisien	3	10 000 à 19 999 habitants
	4	Est	0	commune rurale
	4	Est	1	2 000 à 4 999 habitants
	4	Est	2	5 000 à 9 999 habitants
	5	Ouest	6	100 000 à 199 999 habitants
6	8	Centre-est	5	50 000 à 99 999 habitants
	2	Bassin parisien	0	commune rurale
	2	Bassin parisien	7	200 000 à 1 999 999 habitants
	3	Nord	1	2 000 à 4 999 habitants
	4	Est	7	200 000 à 1 999 999 habitants
	7	Sud-ouest	5	50 000 à 99 999 habitants
6	9	Méditerranée	4	20 000 à 49 999 habitants

Classe géographique	ZEAT		TUU	
7	2	Bassin parisien	4	20 000 à 49 999 habitants
	2	Bassin parisien	6	100 000 à 199 999 habitants
	4	Est	6	100 000 à 199 999 habitants
	5	Ouest	4	20 000 à 49 999 habitants
	5	Ouest	7	200 000 à 1 999 999 habitants
8	3	Nord	0	commune rurale
	3	Nord	3	10 000 à 19 999 habitants
	8	Centre-est	2	5 000 à 9 999 habitants
	9	Méditerranée	5	50 000 à 99 999 habitants
	9	Méditerranée	7	200 000 à 1 999 999 habitants
9	1	Région parisienne	1	2 000 à 4 999 habitants
	8	Centre-est	1	2 000 à 4 999 habitants
	8	Centre-est	6	100 000 à 199 999 habitants
	9	Méditerranée	3	10 000 à 19 999 habitants
10	1	Région parisienne	5	50 000 à 99 999 habitants
	3	Nord	7	200 000 à 1 999 999 habitants
	8	Centre-est	4	20 000 à 49 999 habitants
11	1	Région parisienne	0	commune rurale
	1	Région parisienne	8	unité urbaine de Paris
12	3	Nord	4	20 000 à 49 999 habitants
	7	Sud-ouest	7	200 000 à 1 999 999 habitants
	8	Centre-est	7	200 000 à 1 999 999 habitants
13	1	Région parisienne	2	5 000 à 9 999 habitants
	8	Centre-est	0	commune rurale
14	1	Région parisienne	3	10 000 à 19 999 habitants
15	7	Sud-ouest	4	20 000 à 49 999 habitants

## Annexe 10 : Le nombre d'observations et le nombre de logements réservés trimestriels issus de l'ECLN<sup>55</sup> (2006T1-2012T3)

Le nombre moyen trimestriel d'observations issus de l'ECLN s'élève à 8 194 sur la période 2006T1-2012T3 : 1 057 pour l'individuel groupé et 7 138 pour le collectif (cf. graphique 6.2).

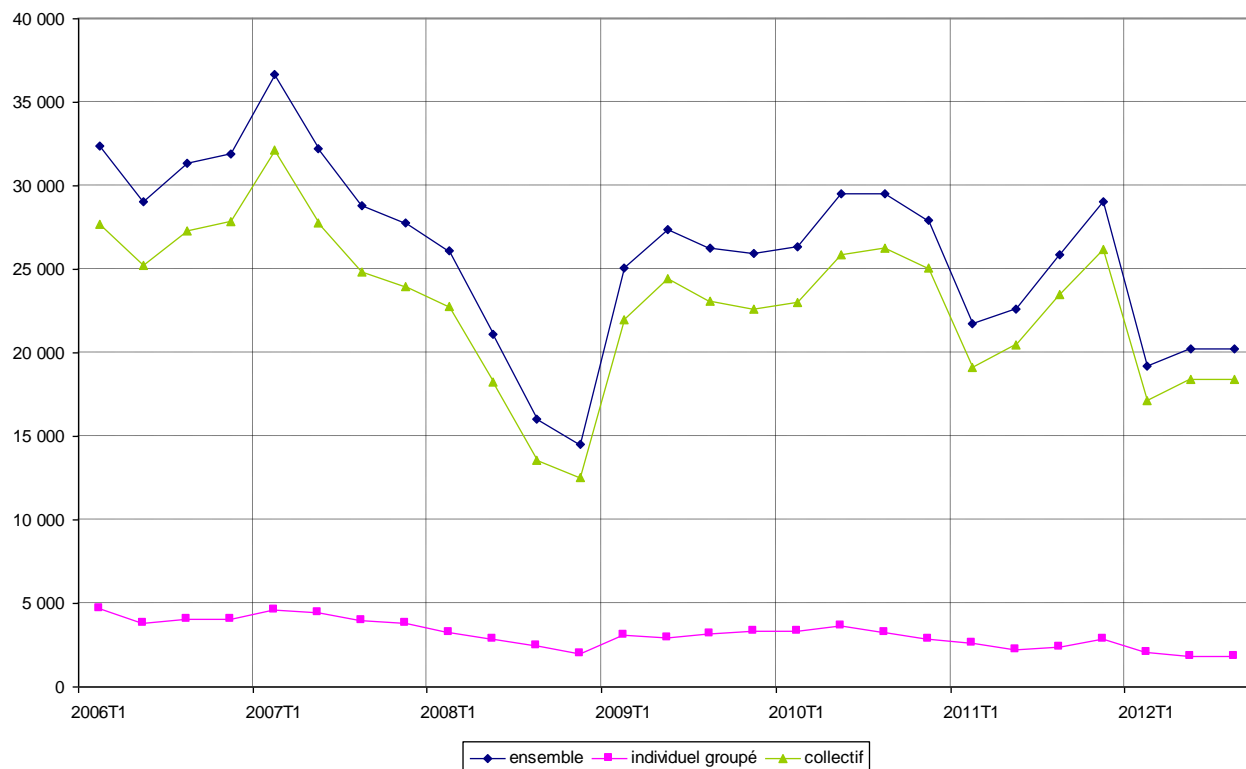
**Graphique 6.2 : Le nombre d'observations trimestriel issus de l'ECLN (2006T1-2012T3)**



<sup>55</sup> Hors logements sociaux

Le nombre moyen trimestriel de logements réservés issus de l'ECLN est de 26 105 sur la période 2006T1-2012T3 : 3 152 pour l'individuel groupé et 22 952 pour le collectif (cf. graphique 6.3).

**Graphique 6.3 : Le nombre de logements réservés trimestriel issus de l'ECLN (2006T1-2012T3)**



## Annexe 11.1 : Définition des variables indicatrices intervenant dans le modèle (1)

**Tableau 6.7 : Définition des variables indicatrices intervenant dans le modèle (1)**

<b>k</b>	<b>Caractéristique géographique ou physique</b>	<b>Définition des variables indicatrices</b>	<b>Modalité de référence</b>
<b>1-14</b>	Le découpage du territoire métropolitain hors Corse en 15 classes géographiques	14 indicatrices correspondant aux 14 premières classes géographiques	La 15 <sup>ème</sup> classe géographique
<b>15-17</b>	Le statut de la commune	Ville-centre	Commune rurale
		Ville de banlieue	
		Ville isolée	
<b>18</b>	Les communes maritimes		
<b>19</b>	L'arrière-pays littoral		
<b>20</b>	Les stations de sport d'hiver et d'alpinisme		
<b>21-23</b>	Le type de construction et le nombre de pièces	Appartement d'une pièce	Individuel groupé (maisons)
		Appartement de 2 à 4 pièces	
		Appartement de plus de 5 pièces	
<b>24-27</b>	Le niveau de standing du programme de construction	Ordinaire	Non-réponse
		Confortable	
		Très confortable	
		Luxe	
<b>28</b>	La certification QUALITEL du programme de construction		Absence de certification ou non-réponse
<b>29</b>	La présence d'un dispositif de sécurité		Absence de dispositif de sécurité ou non-réponse
<b>30</b>	La présence d'une piscine dans la résidence		Absence de piscine ou non-réponse
<b>31</b>	La proximité d'un transport collectif par voie ferrée		Absence de transport collectif par voie ferrée à proximité ou non-réponse
<b>32</b>	La présence d'un système de climatisation dans le programme de construction		Absence de système de climatisation ou non-réponse
<b>33</b>	La présence d'un balcon ou d'une terrasse pour les appartements		(Absence de balcon et de terrasse pour les appartements) ou (maisons)

## Annexe 11.2 : des résidus d'espérance nulle

On peut montrer que :

$$E(\hat{p}_{i,t_1} - p_{i,t_1}) \cong -\exp(X_i\beta) \times \frac{\sigma^2}{2}$$

où  $X_i$  est le vecteur des régresseurs et  $\beta$  le vecteur des coefficients du modèle (1).

$\frac{\sigma^2}{2}$  est de l'ordre de 0,02 sur l'ensemble de la période 2006T1-2012T3 (cf. tableau 6.8). On peut ainsi considérer que les résidus sont d'espérance nulle.

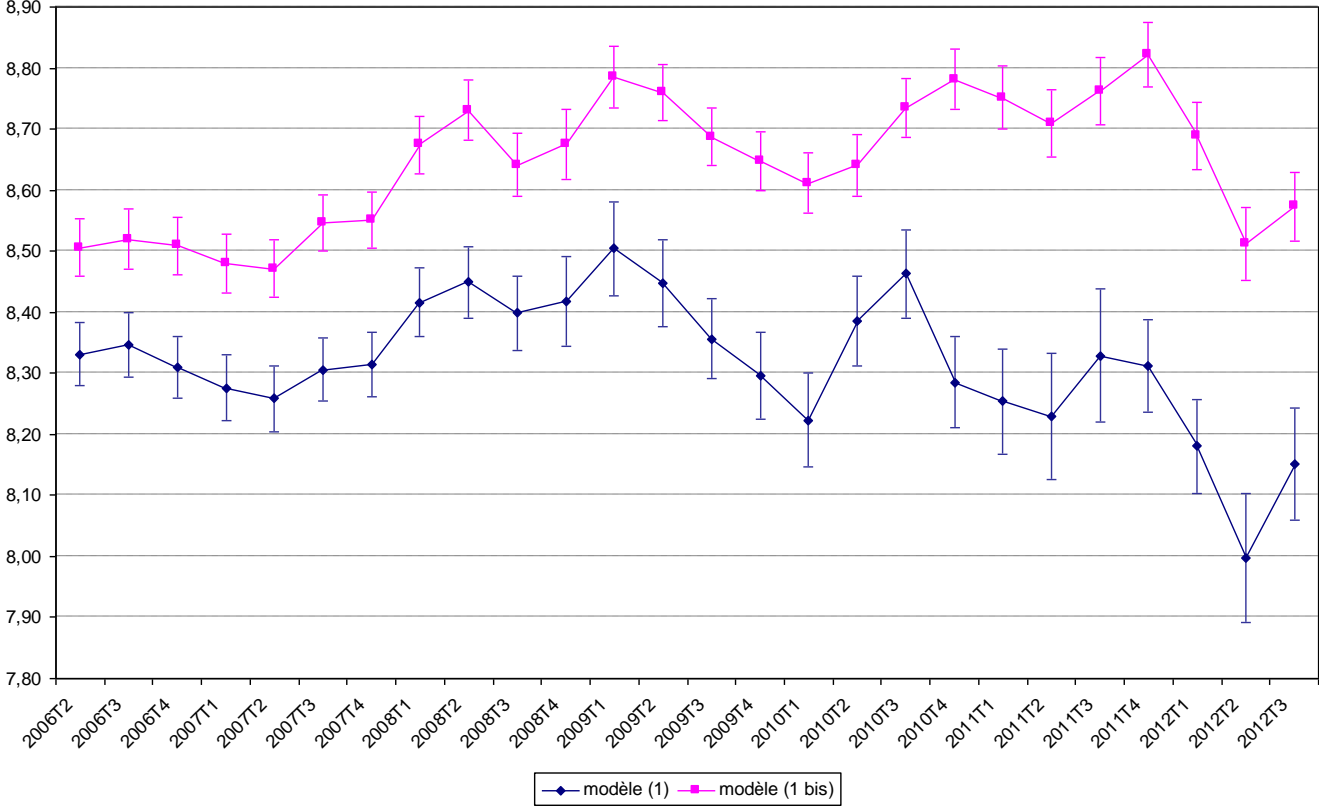
**Tableau 6.8 : Les valeurs de  $\frac{\sigma^2}{2}$  (2006T1-2012T3)**

Trimestre	$\frac{\sigma^2}{2}$
2006T2	0,022
2006T3	0,023
2006T4	0,022
2007T1	0,024
2007T2	0,024
2007T3	0,021
2007T4	0,021
2008T1	0,022
2008T2	0,022
2008T3	0,021
2008T4	0,021
2009T1	0,019
2009T2	0,019
2009T3	0,020
2009T4	0,021
2010T1	0,022
2010T2	0,023
2010T3	0,023
2010T4	0,022
2011T1	0,023
2011T2	0,023
2011T3	0,024
2011T4	0,023
2012T1	0,024
2012T2	0,025
2012T3	0,023
Moyenne	0,022
Minimum	0,019
Maximum	0,025

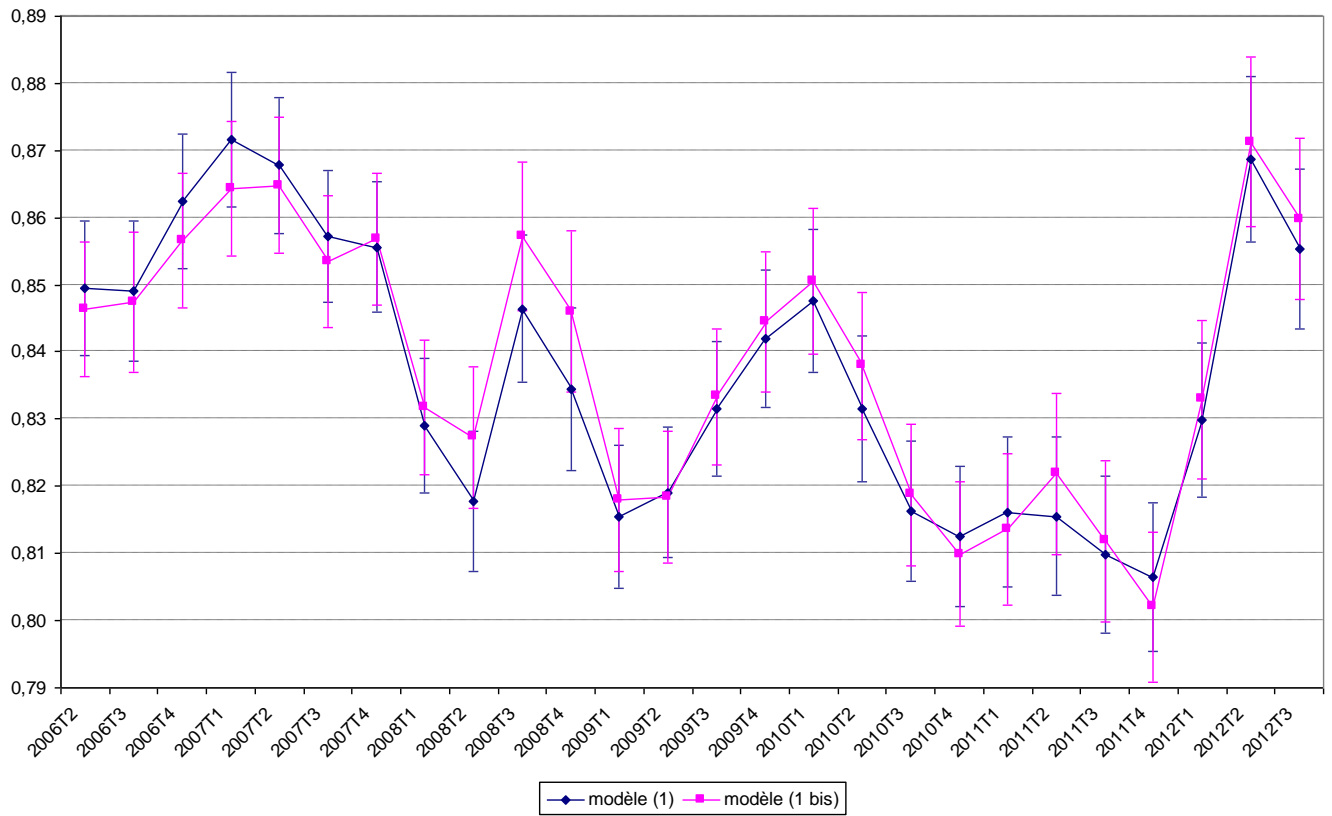


**Annexe 12 : Les valeurs des estimateurs des paramètres et les intervalles de confiance associés au niveau 5% pour les modèles (1) et (1 bis)**

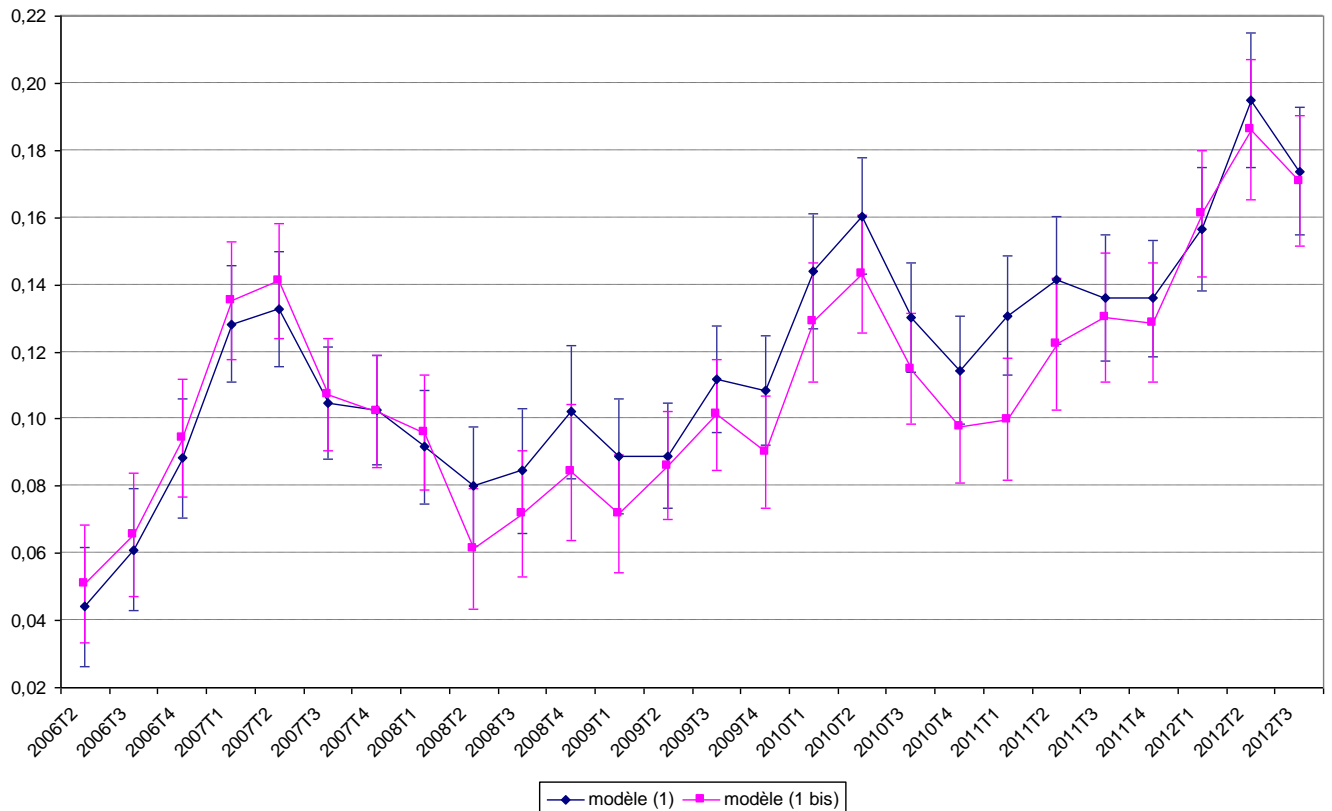
**Graphique 6.4 : La constante**



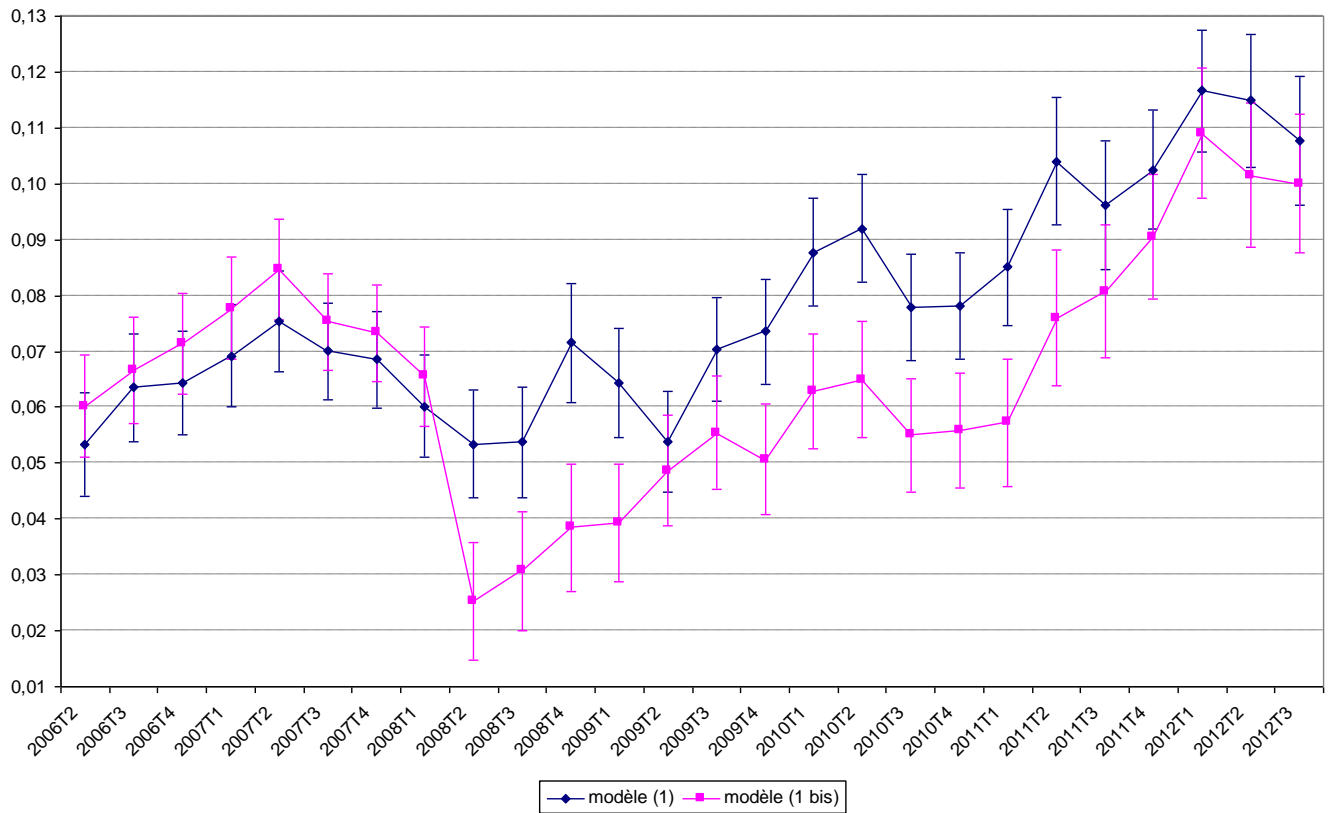
**Graphique 6.5 : La surface habitable moyenne en m<sup>2</sup> du logement**



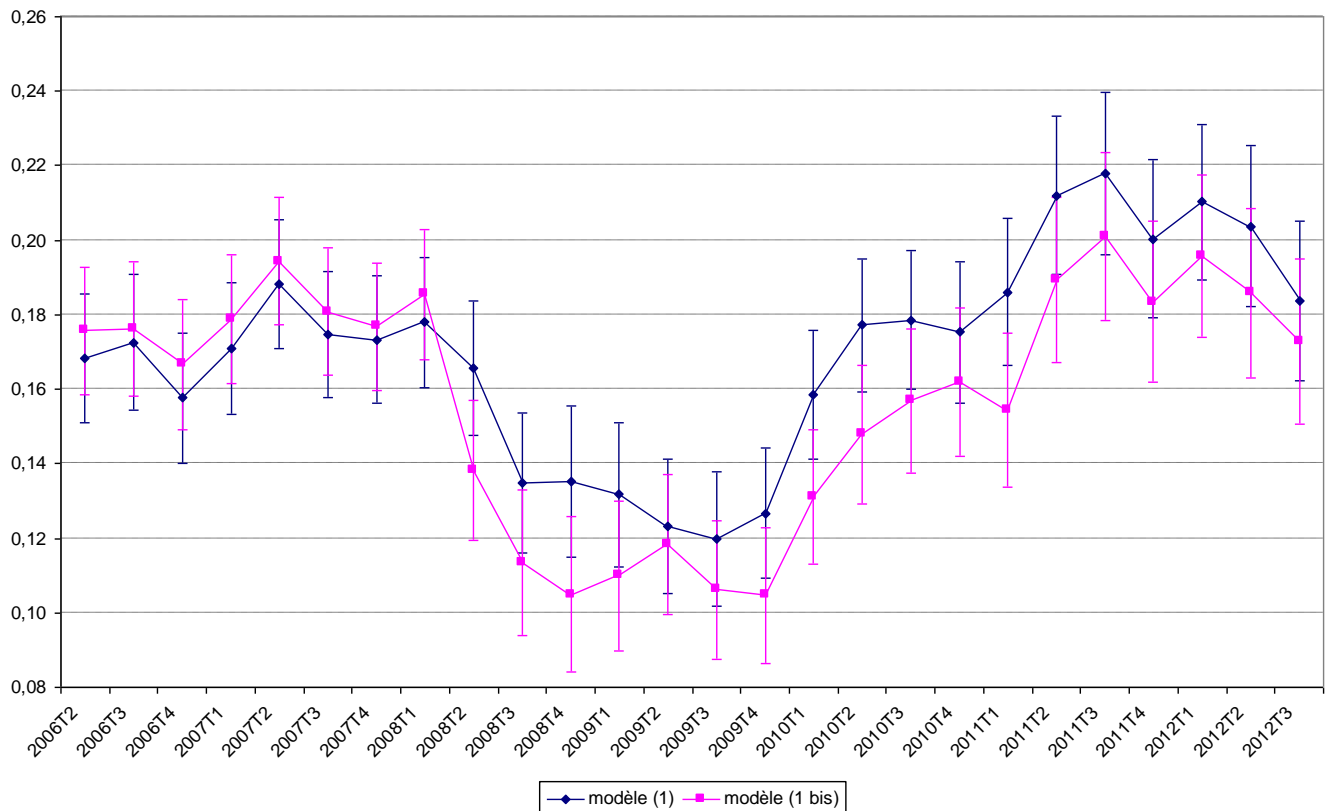
**Graphique 6.6 : Les appartements d'une pièce**



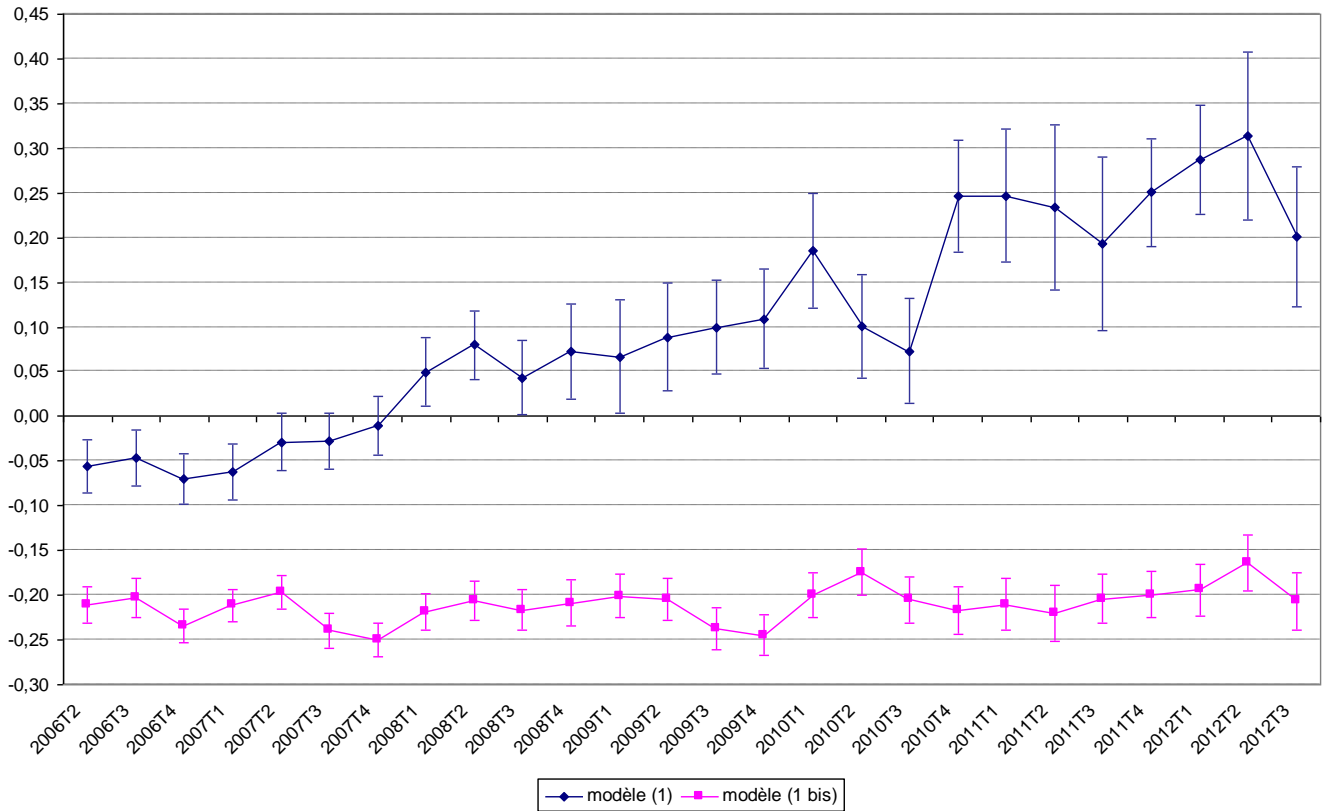
**Graphique 6.7 : Les appartements de 2 à 4 pièces**



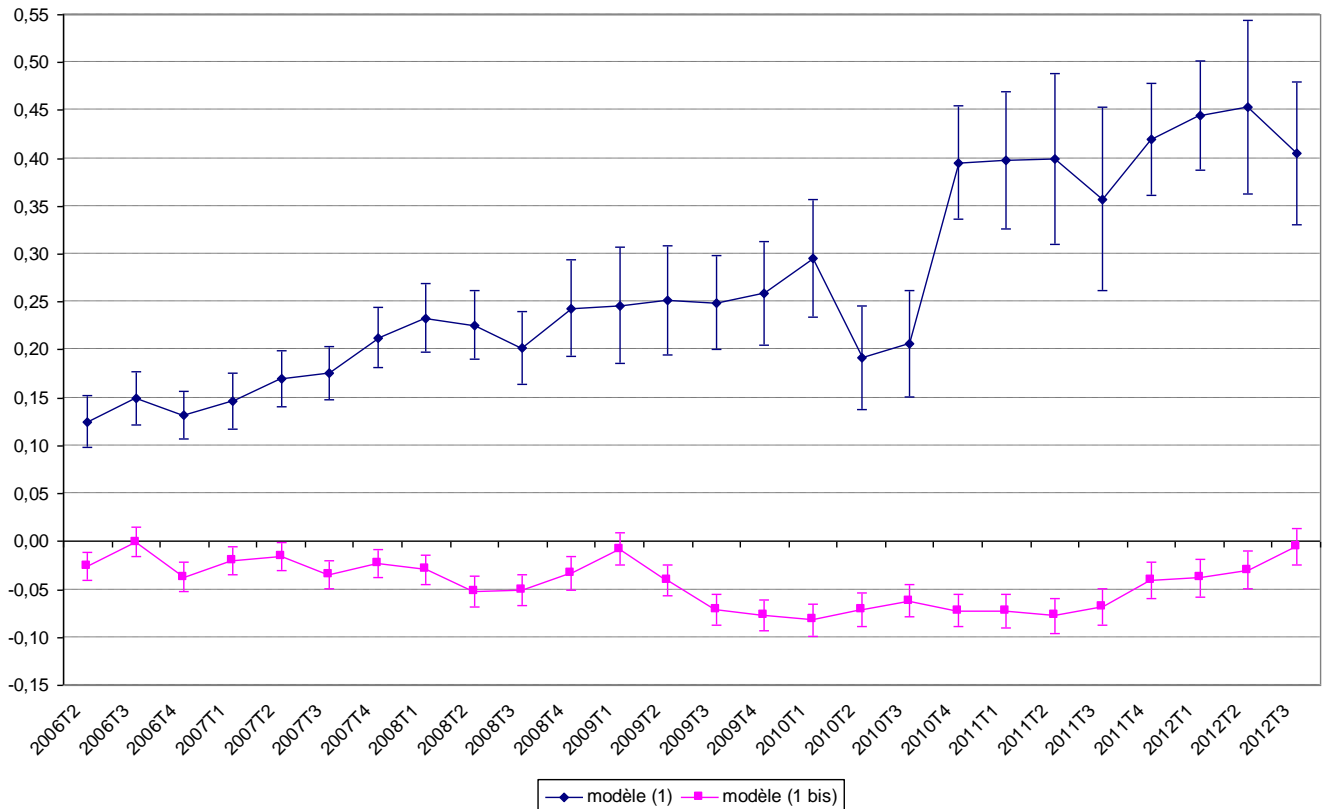
**Graphique 6.8 : Les appartements de plus de 5 pièces**



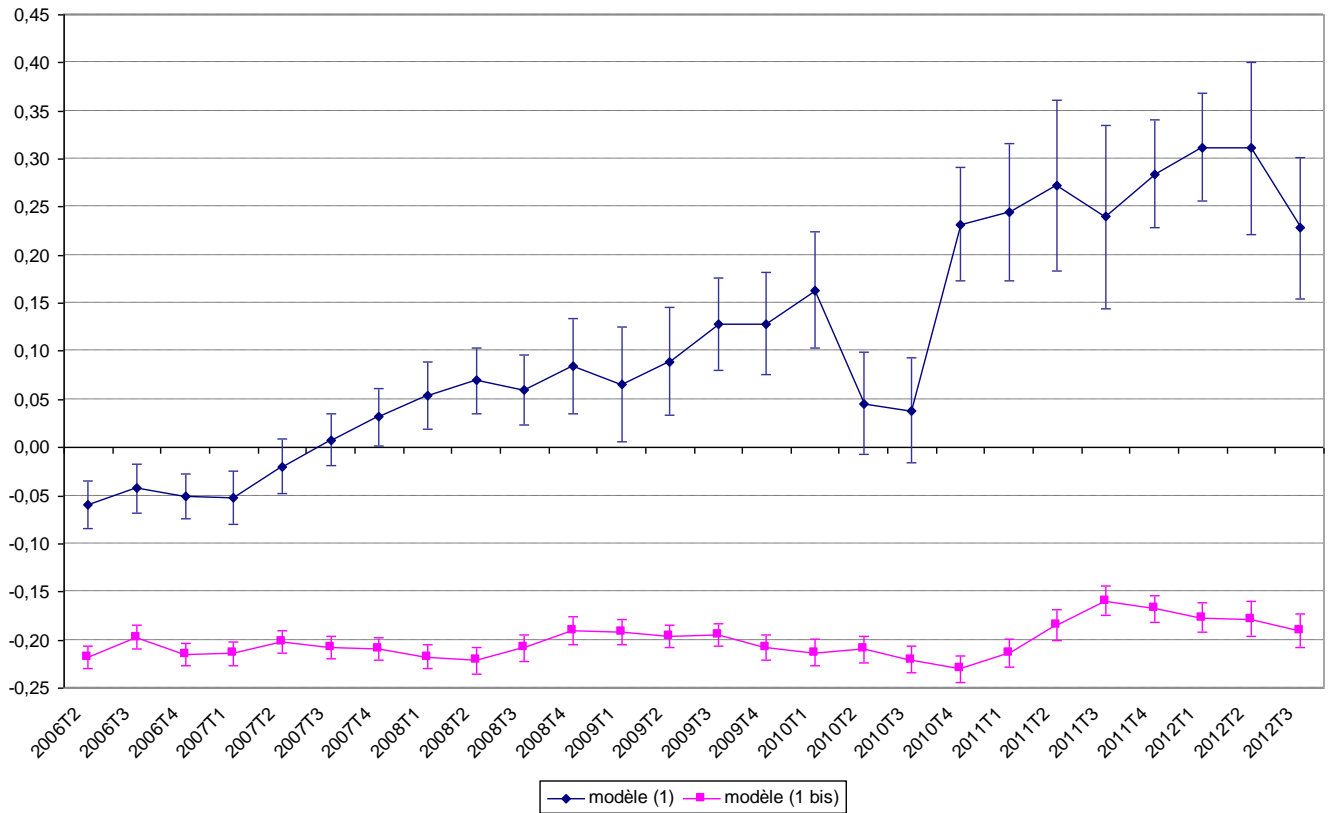
**Graphique 6.9 : Première classe géographique**



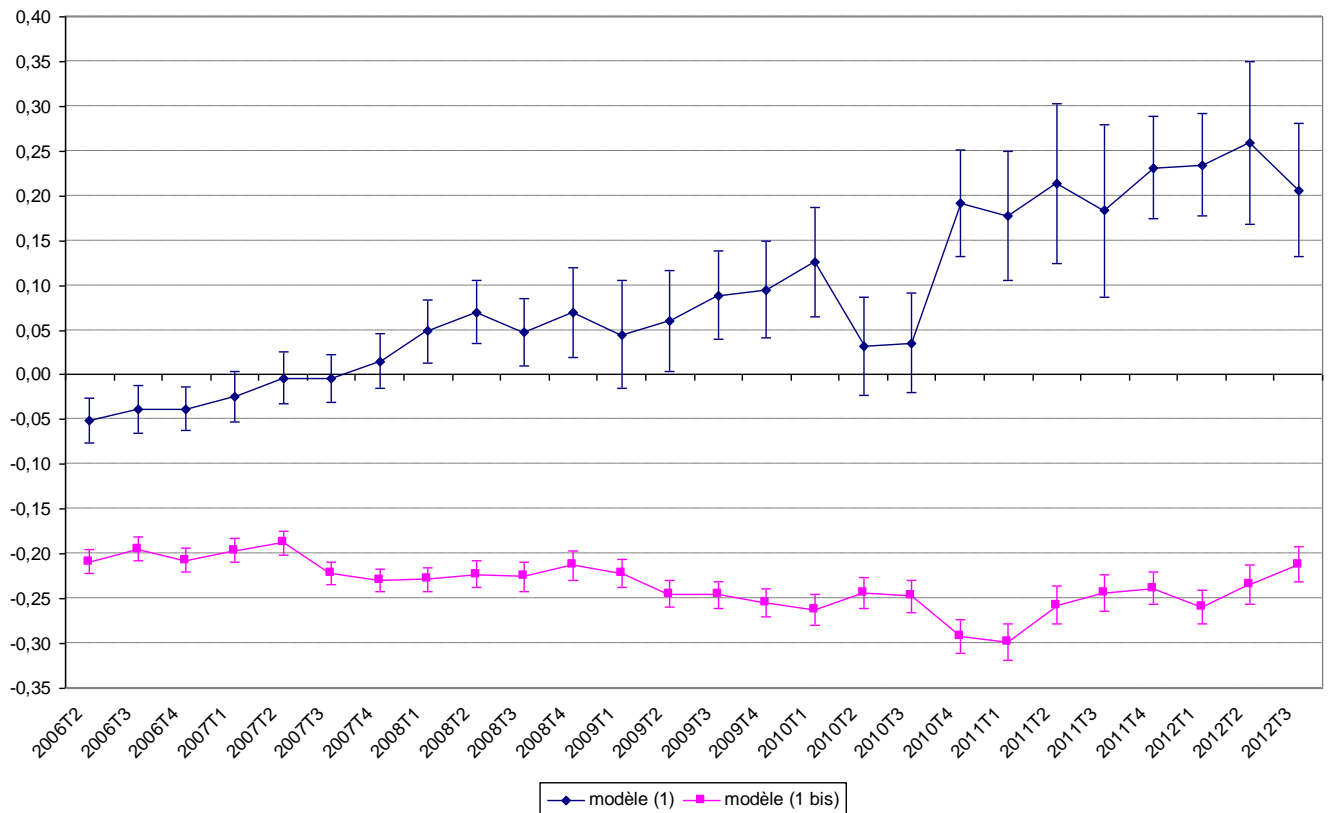
**Graphique 6.10 : Deuxième classe géographique**



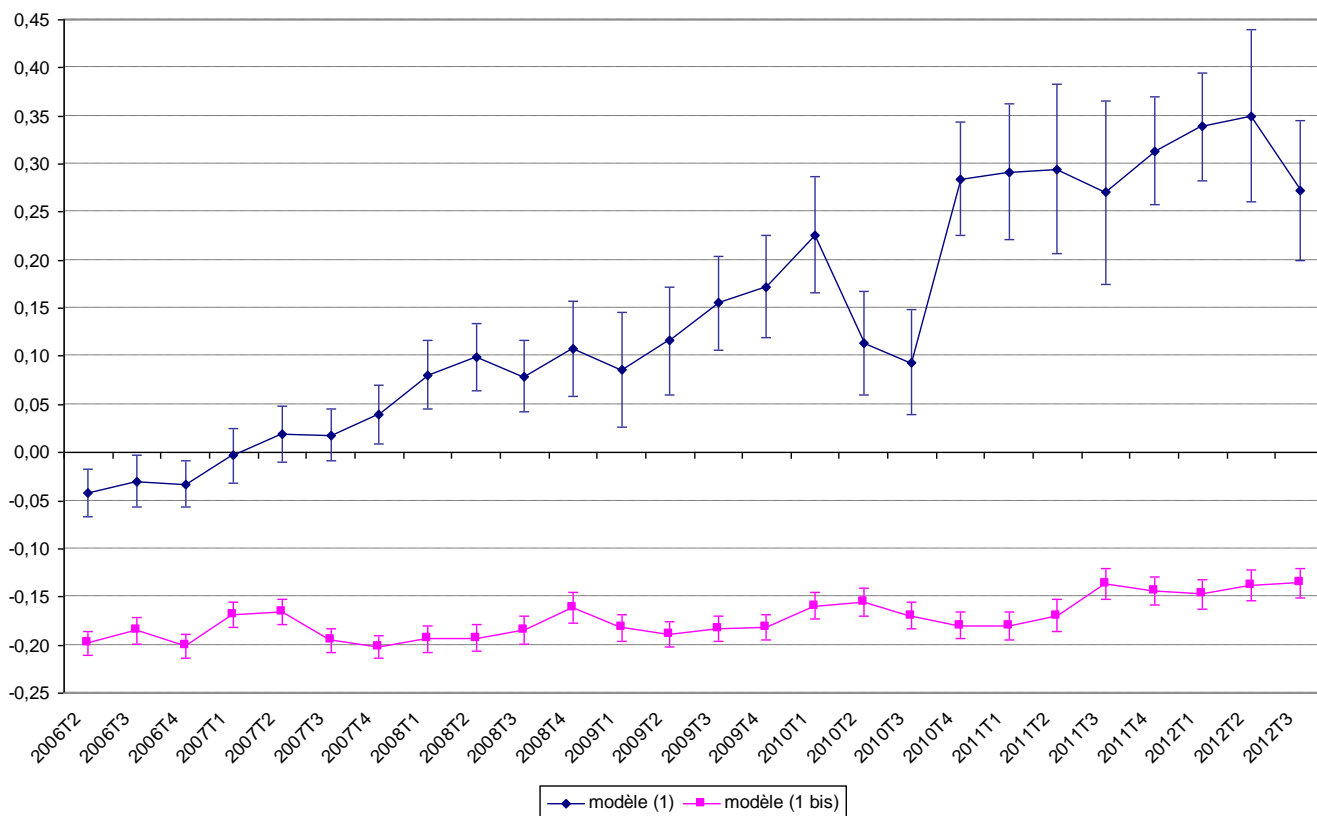
**Graphique 6.11 : Troisième classe géographique**



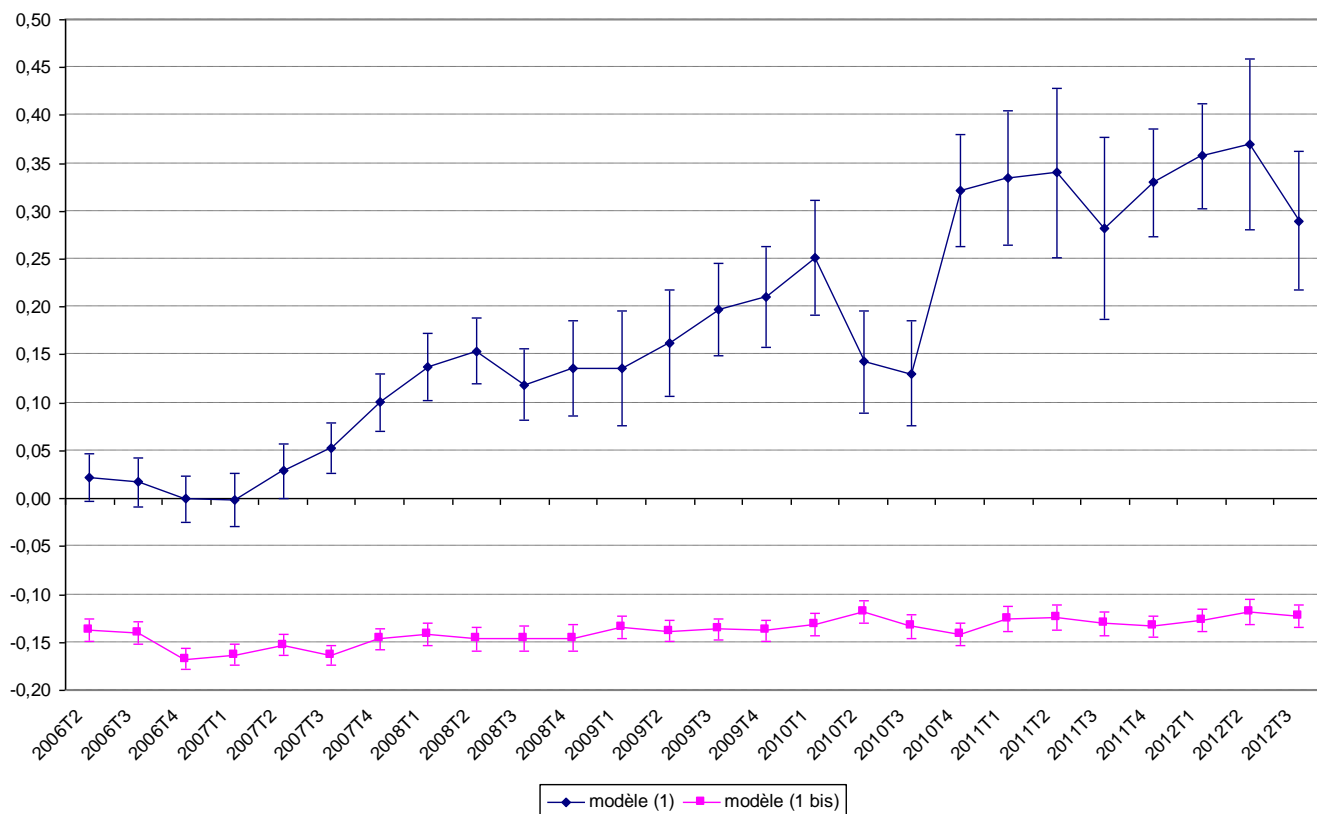
**Graphique 6.12 : Quatrième classe géographique**



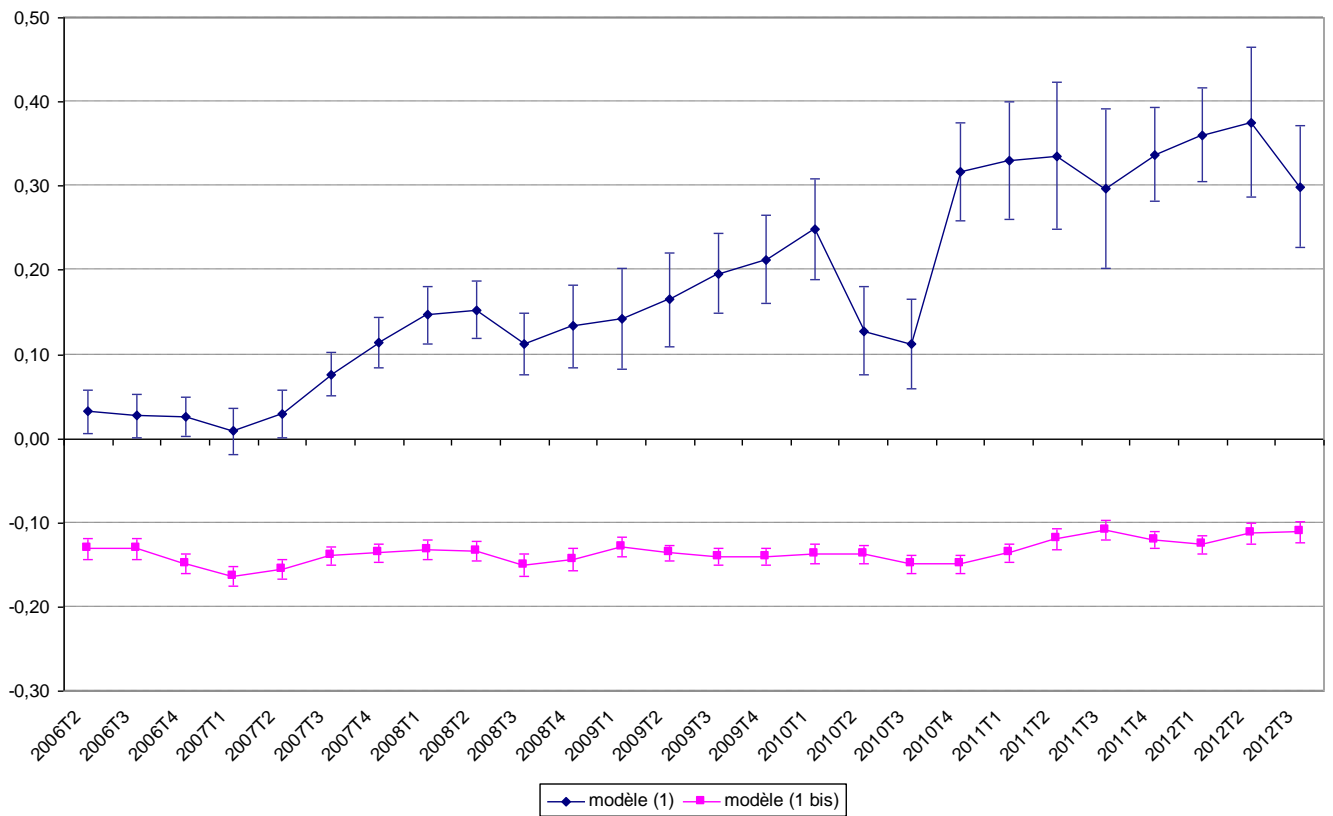
**Graphique 6.13 : Cinquième classe géographique**



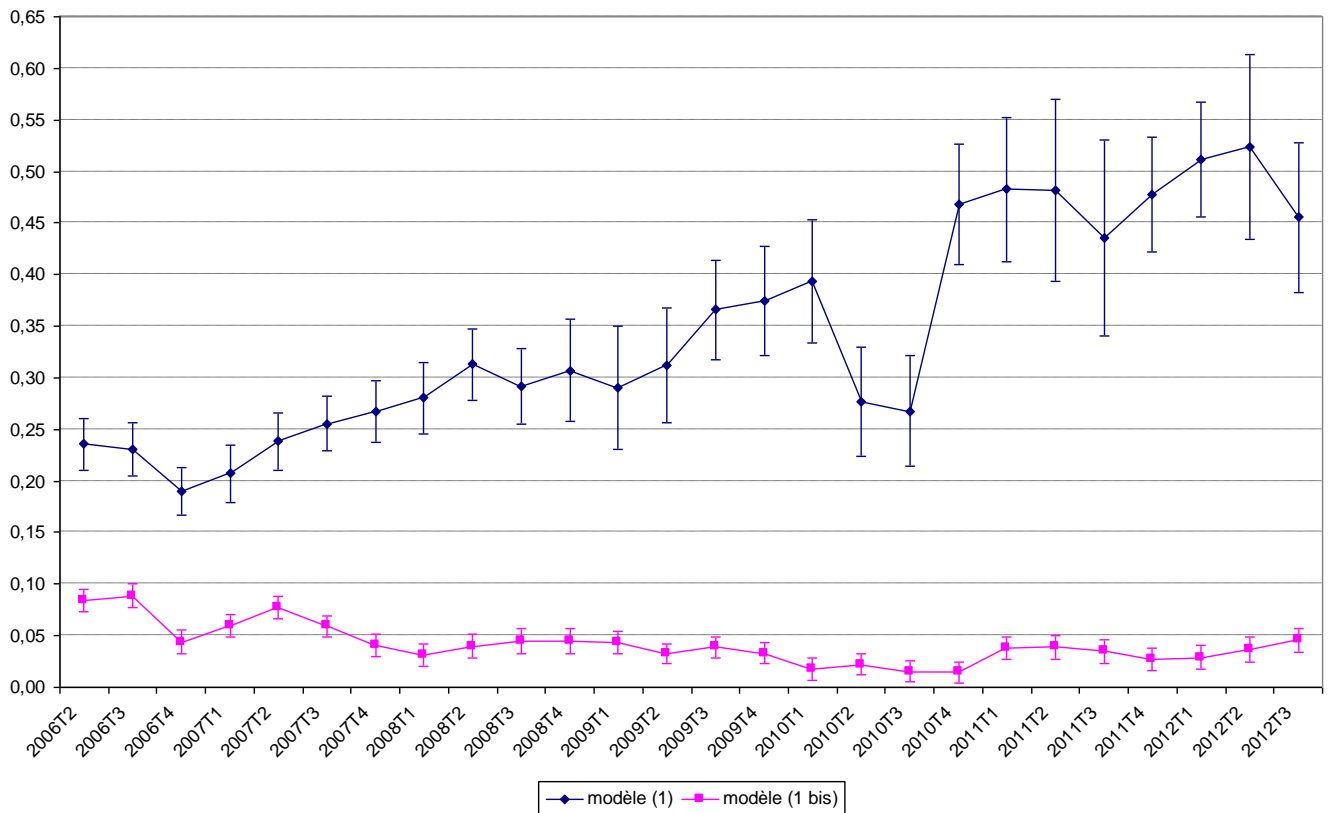
**Graphique 6.14 : Sixième classe géographique**



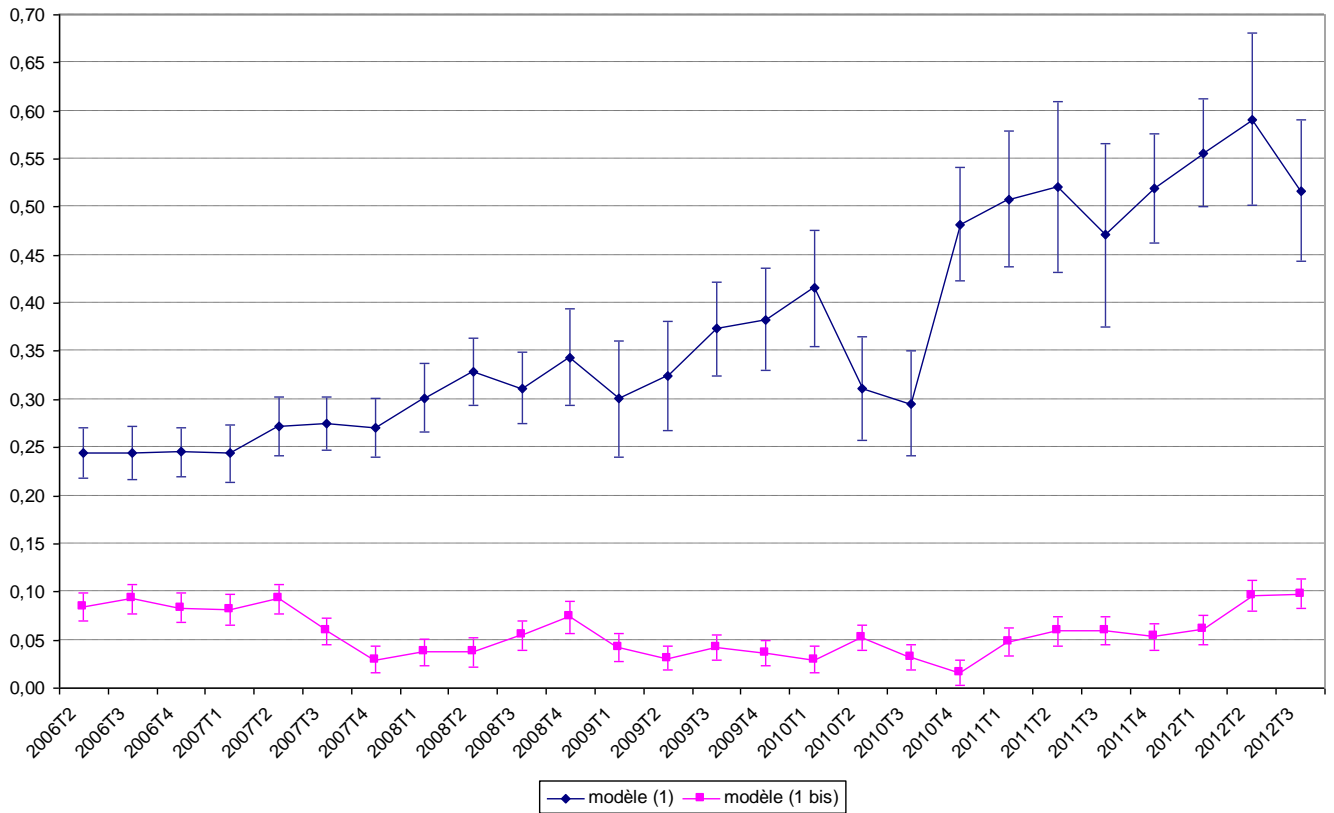
**Graphique 6.15 : Septième classe géographique**



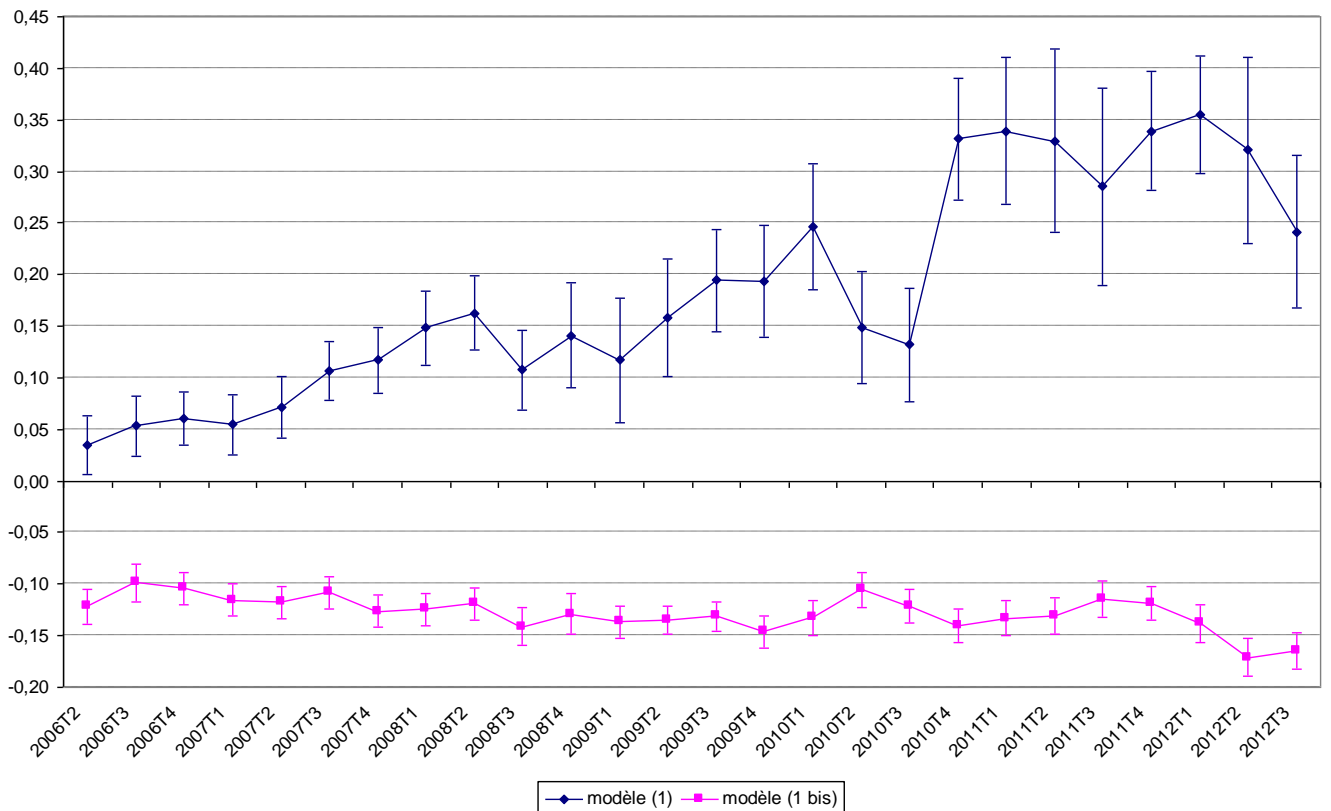
**Graphique 6.16 : Huitième classe géographique**



**Graphique 6.17 : Neuvième classe géographique**

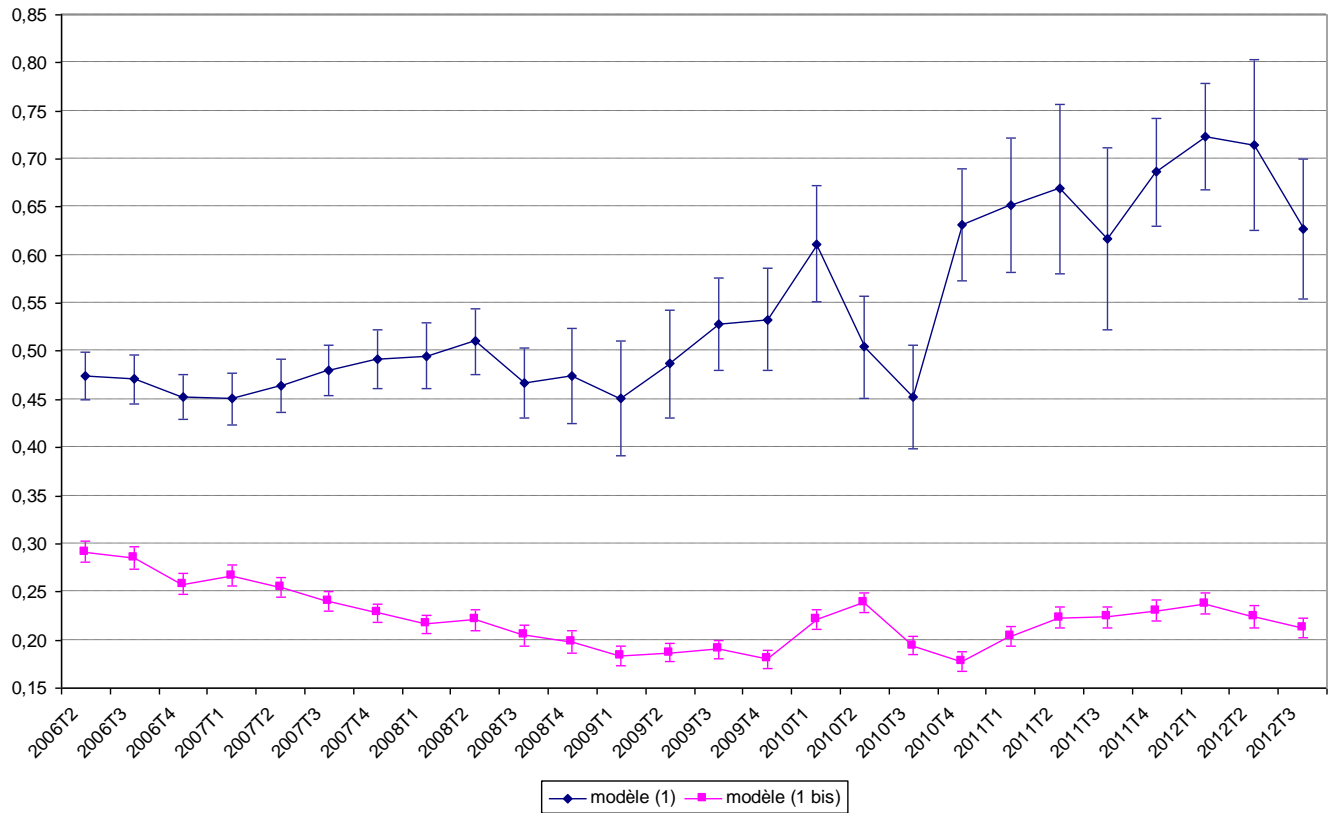


**Graphique 6.18 : Dixième classe géographique**

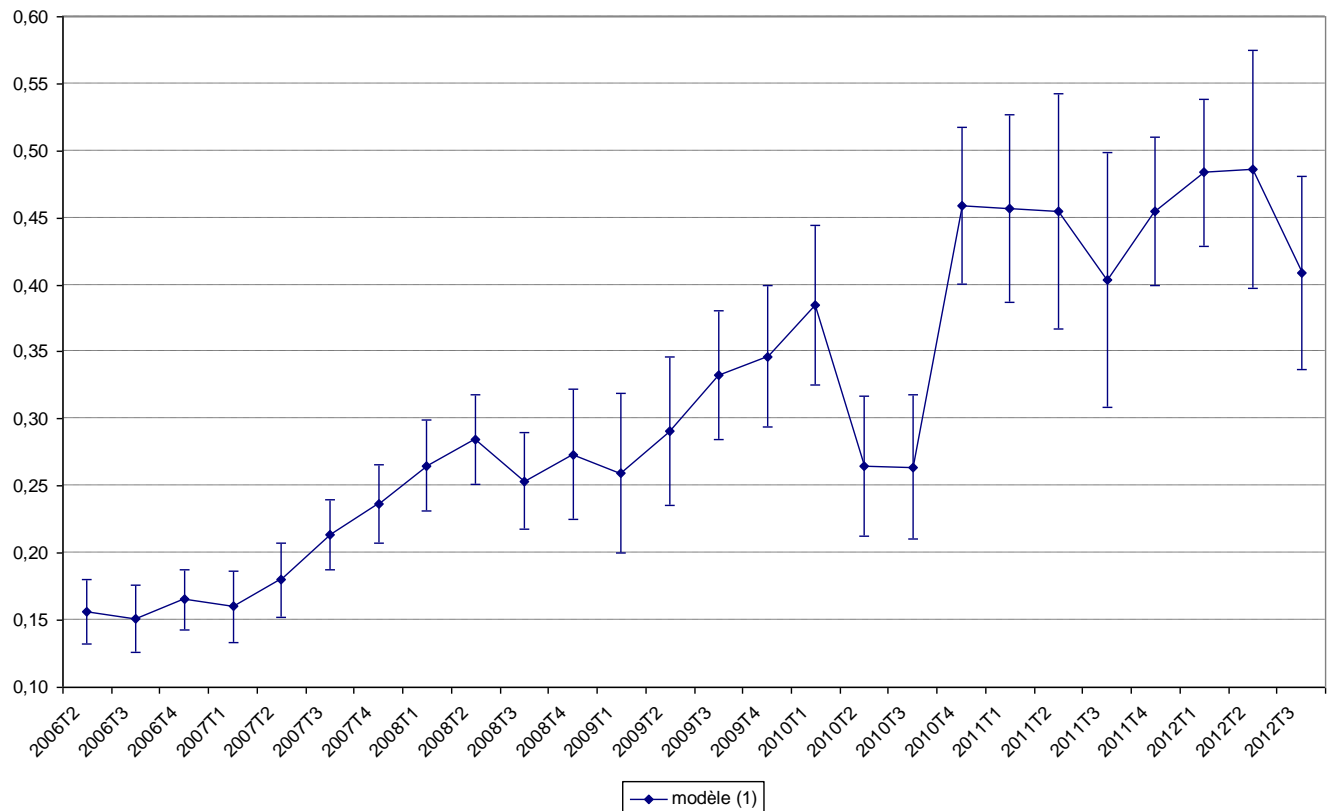




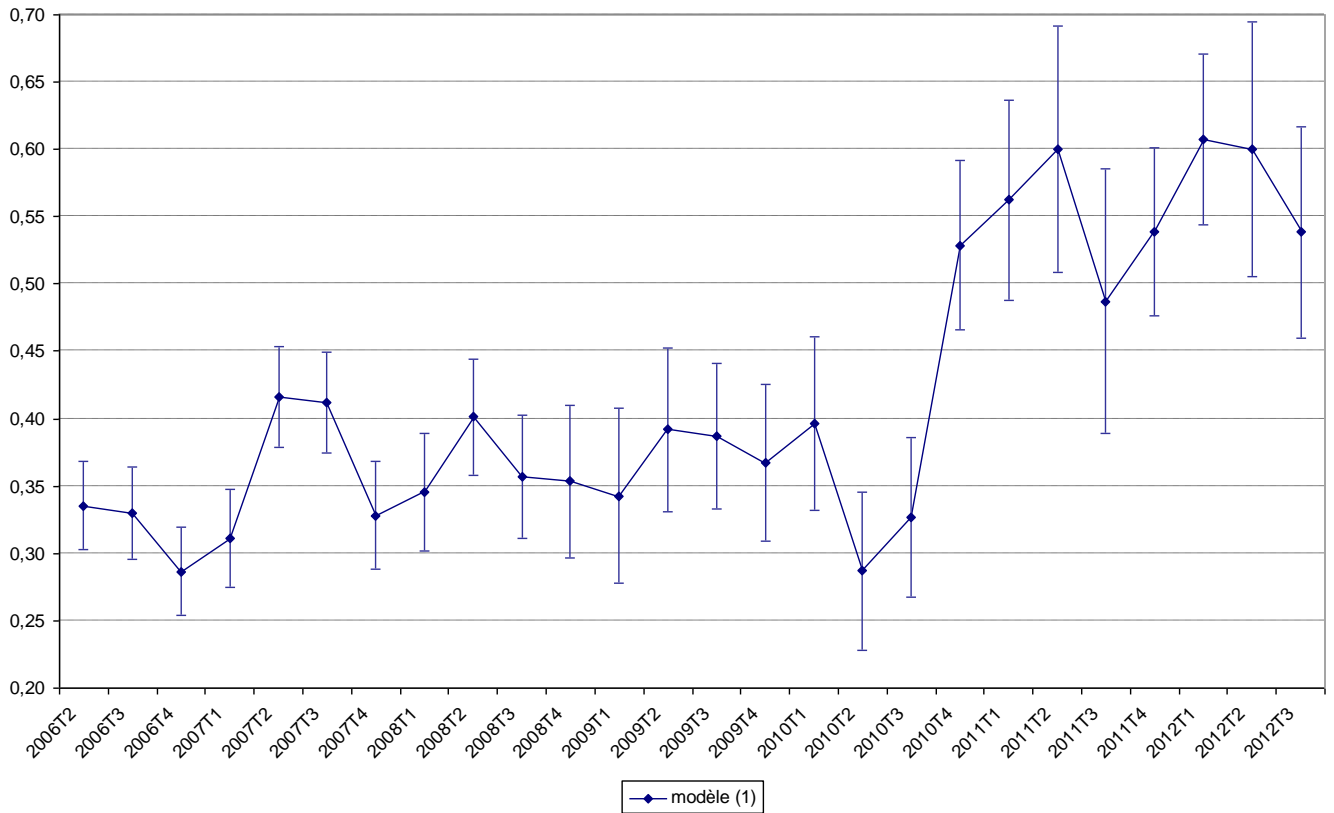
**Graphique 6.19 : Onzième classe géographique**



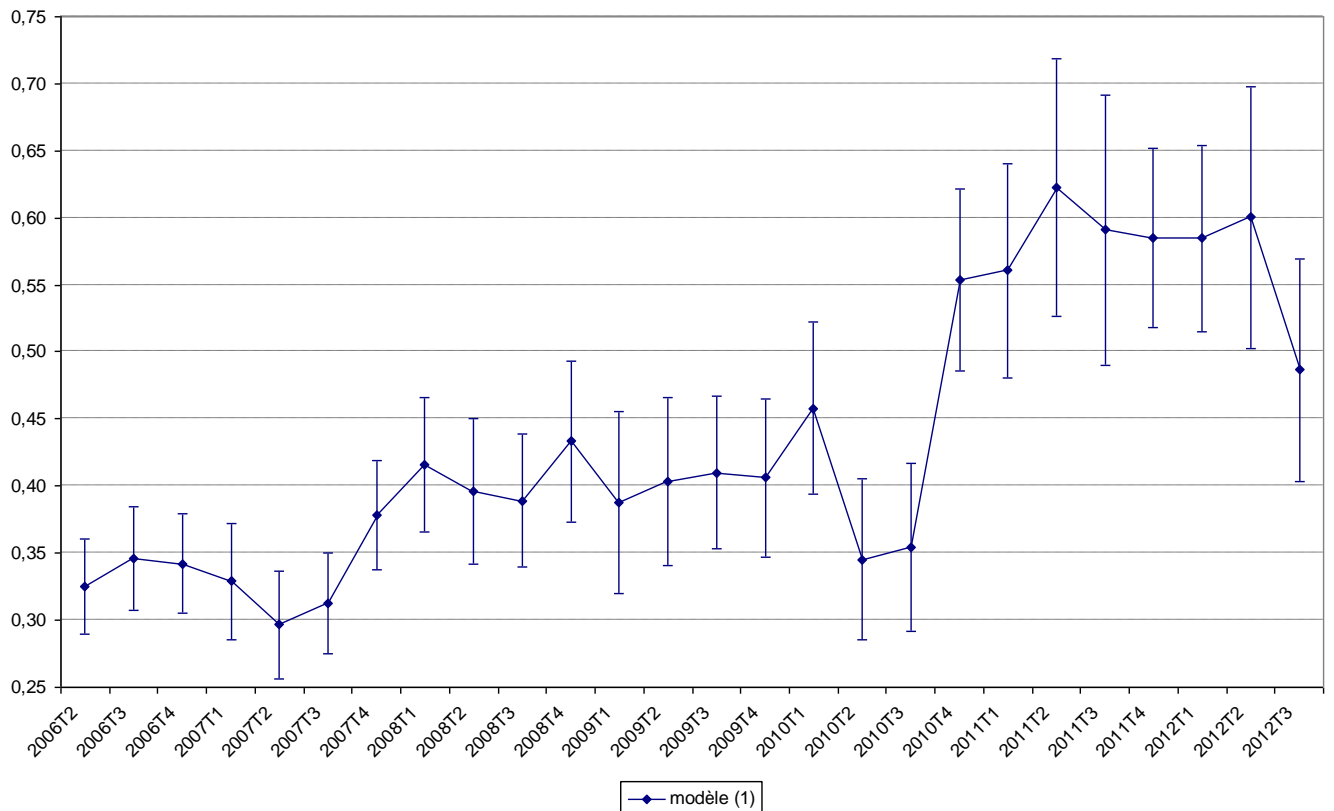
**Graphique 6.20 : Douzième classe géographique**



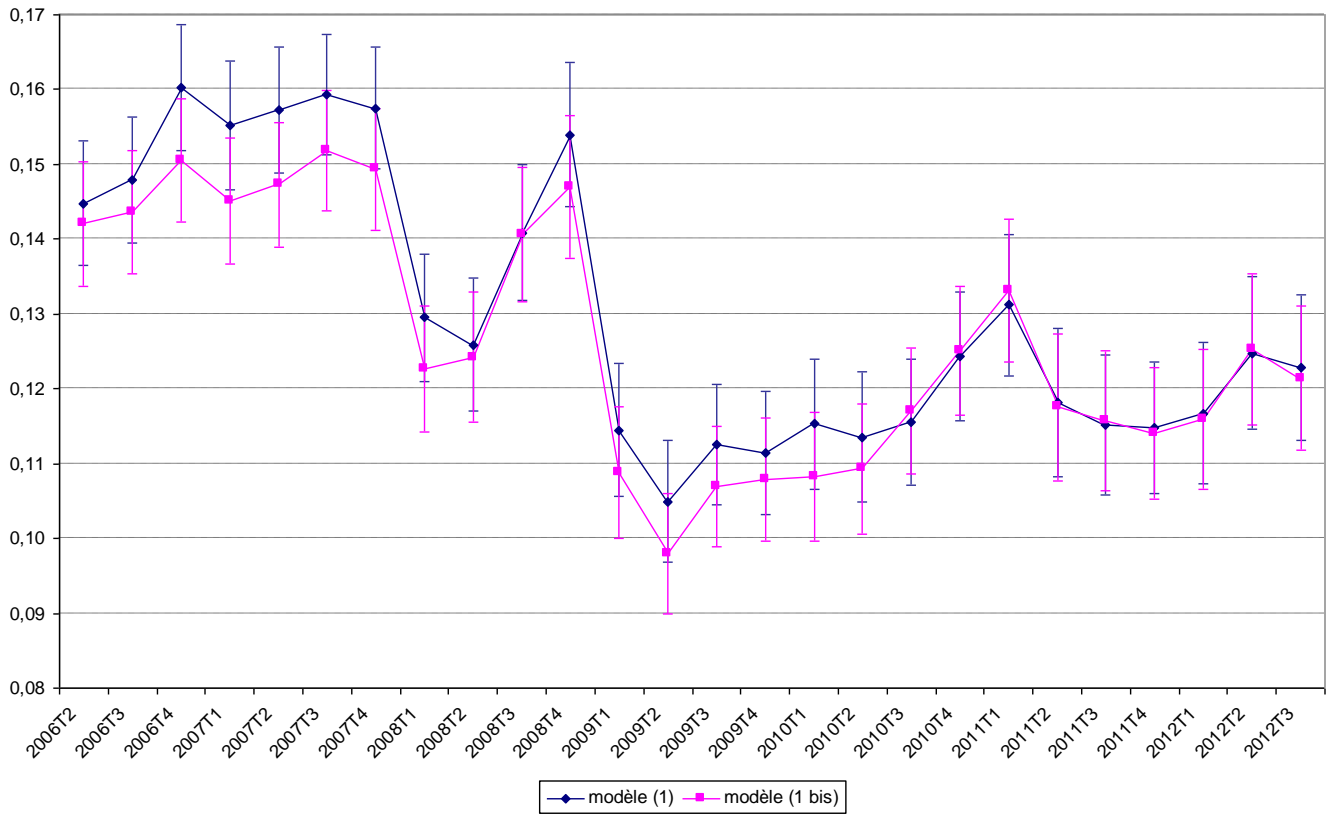
**Graphique 6.21 : Treizième classe géographique**



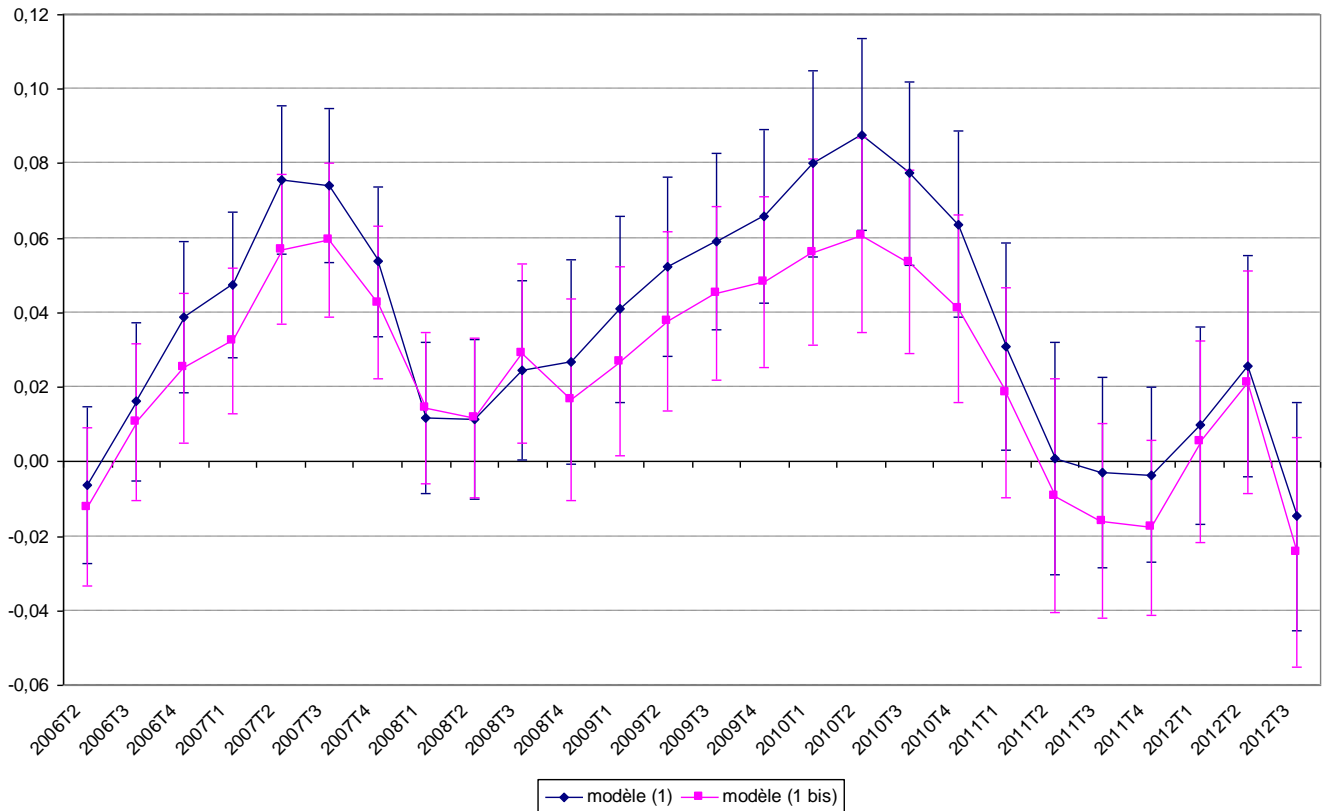
**Graphique 6.22 : Quatorzième classe géographique**



**Graphique 6.23 : Les communes maritimes**



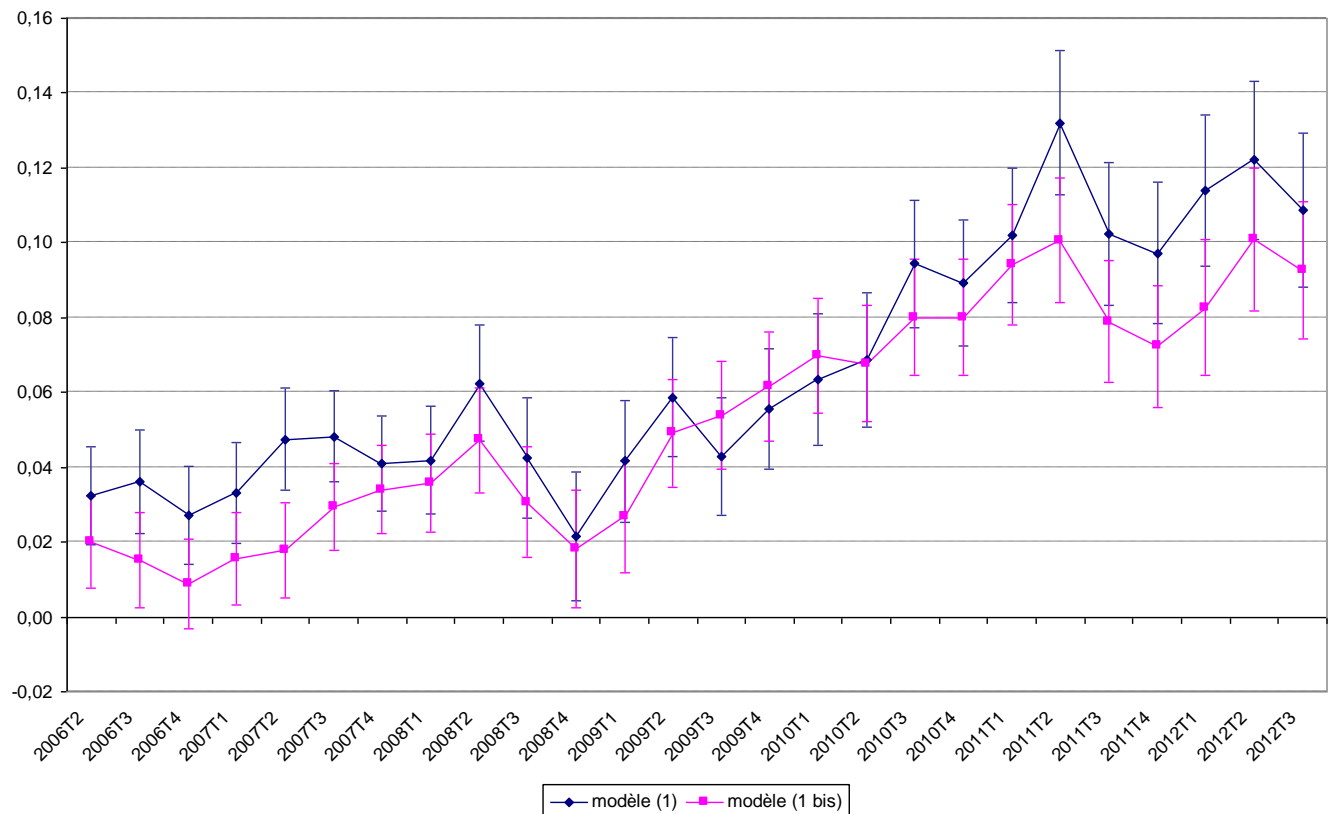
**Graphique 6.24 : L'arrière-pays littoral**



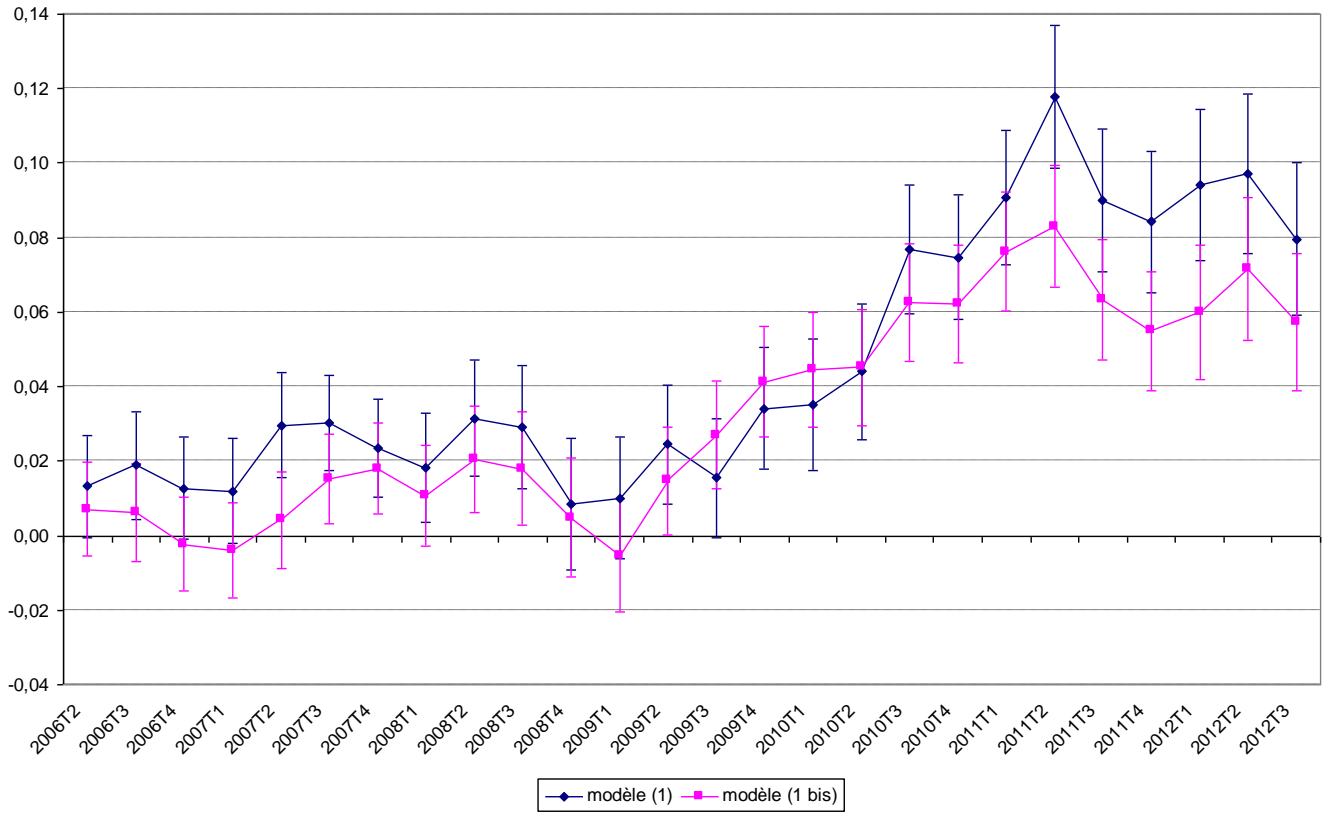
**Graphique 6.25 : Les stations de sport d'hiver et d'alpinisme**



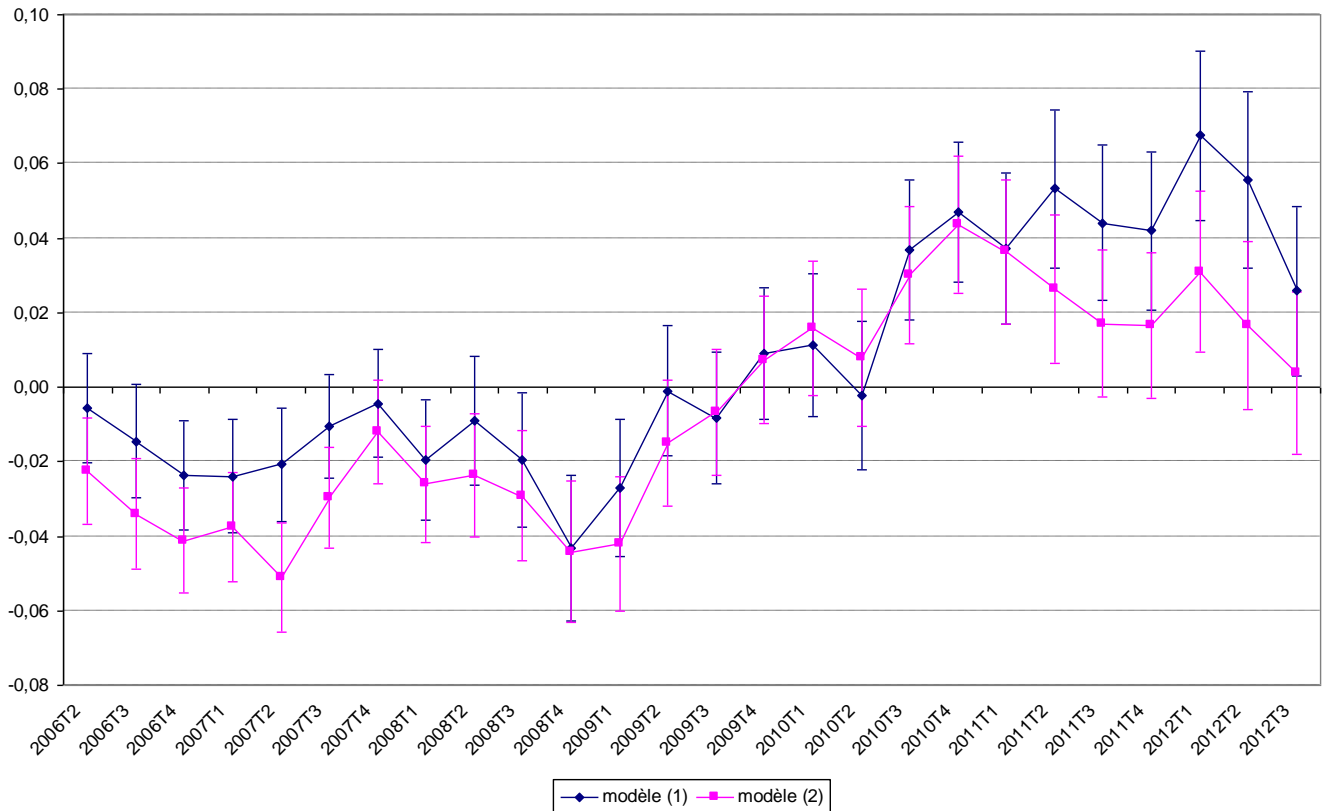
**Graphique 6.26 : Les villes-centres**



**Graphique 6.27 : Les villes de banlieue**



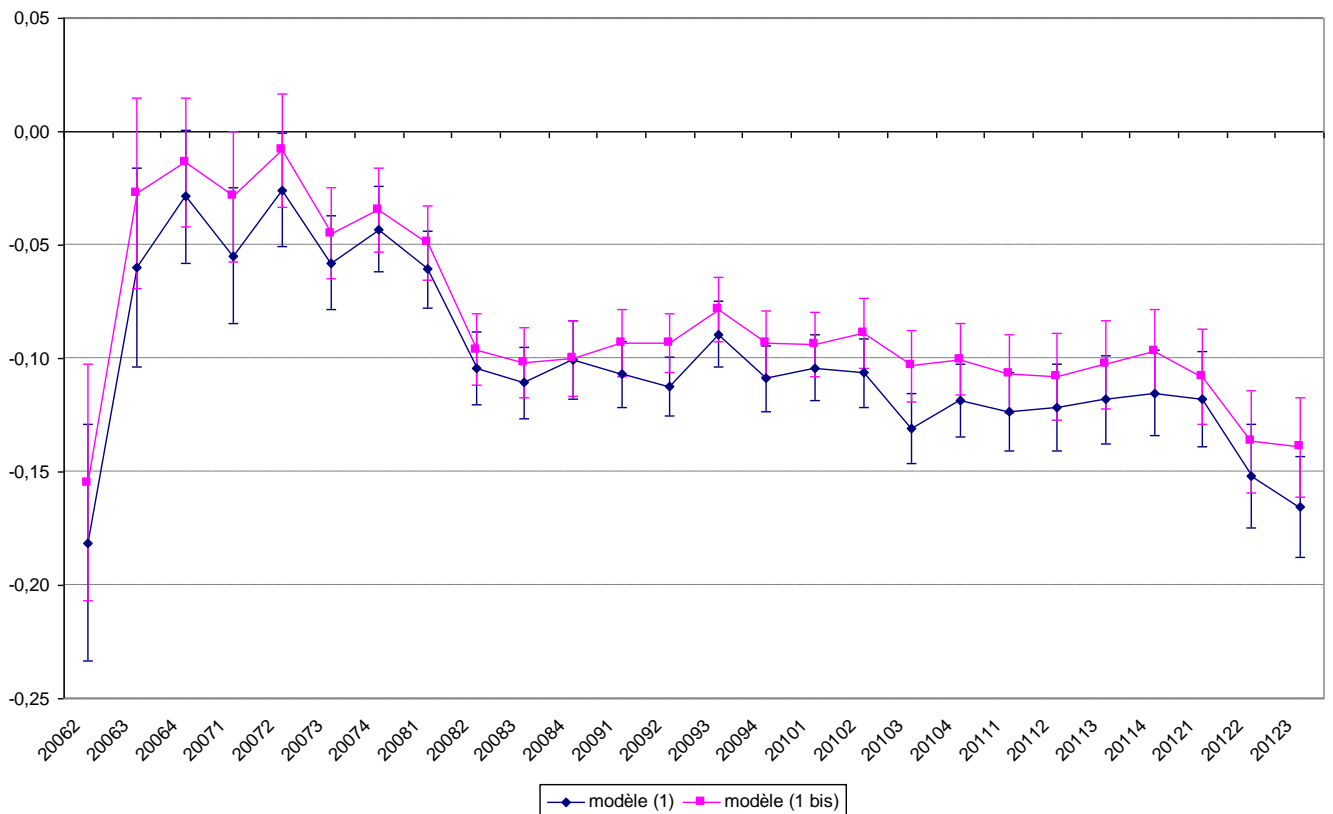
**Graphique 6.28 : Les villes isolées**



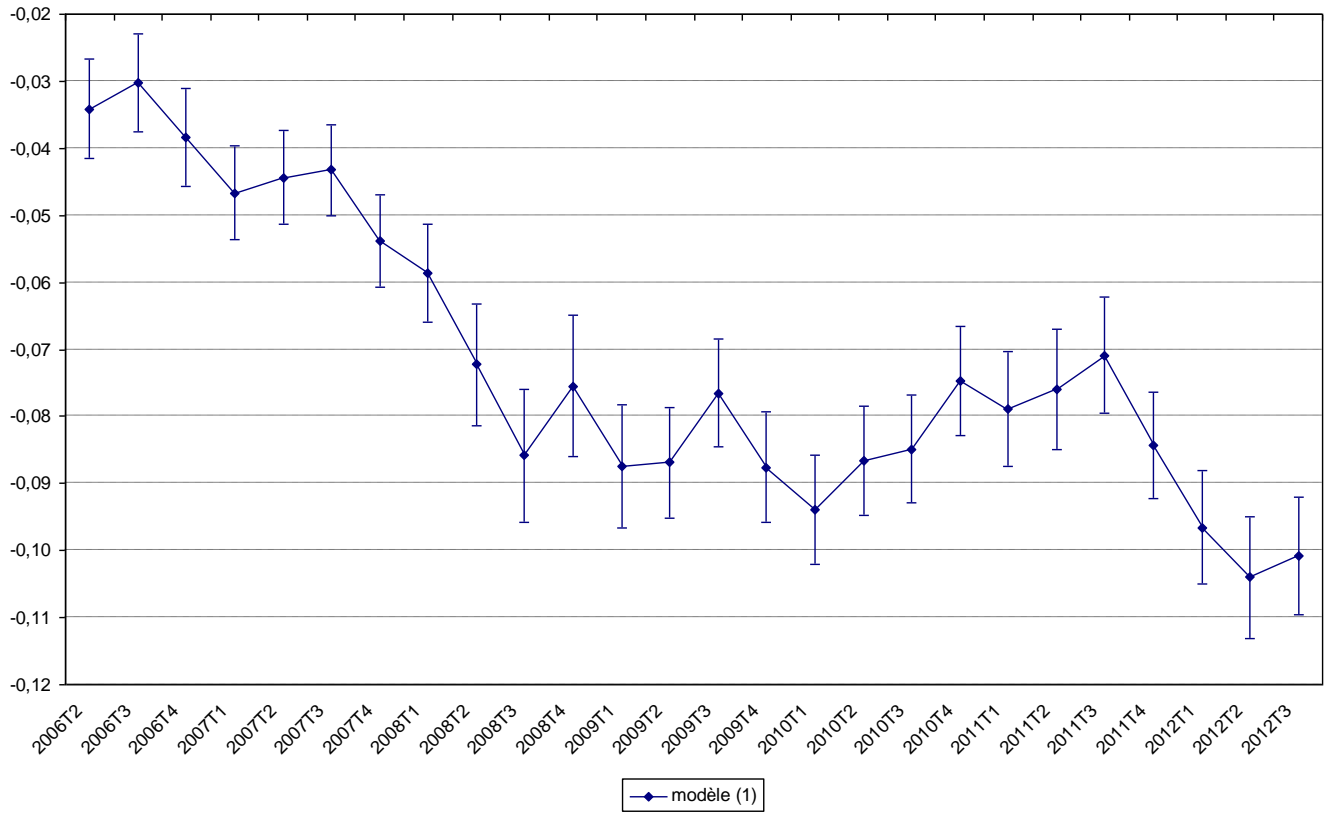
**Graphique 6.29 : L'indicatrice temporelle**



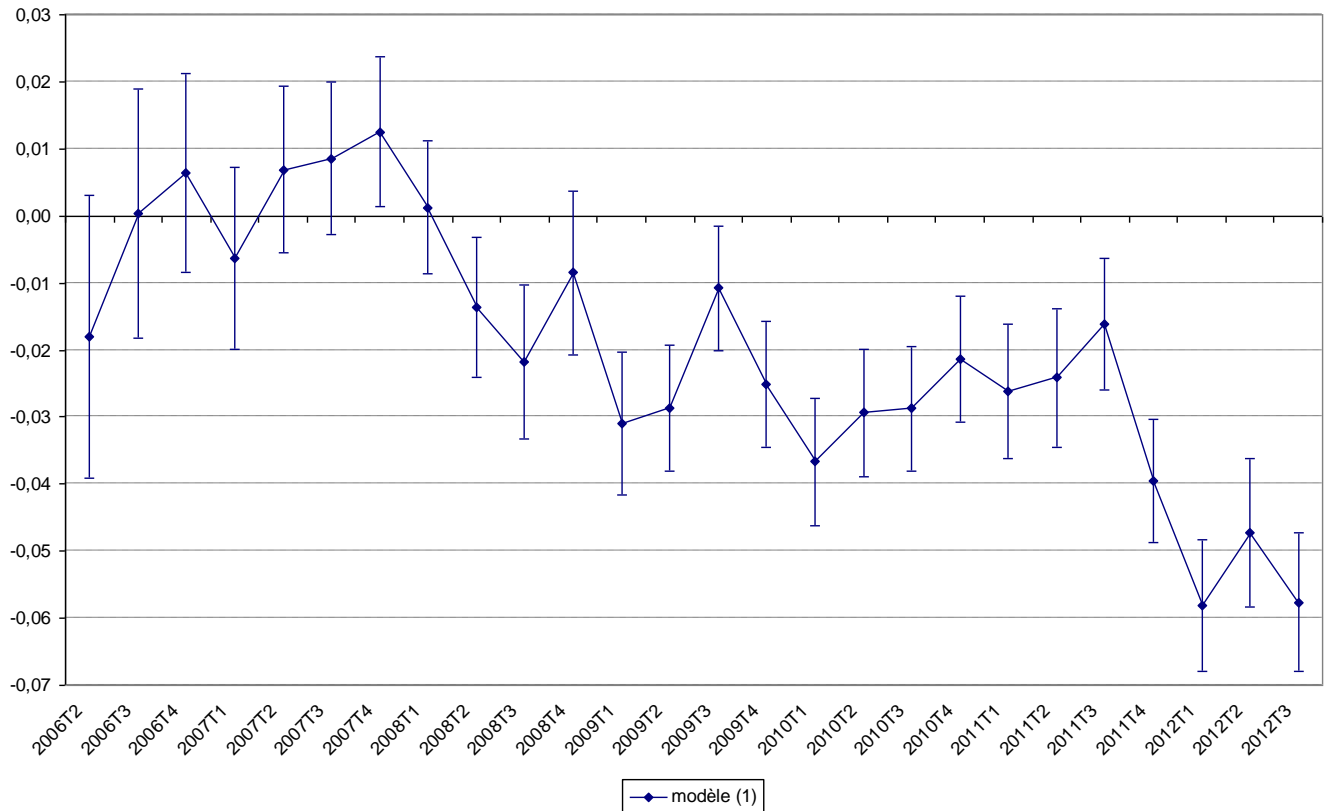
**Graphique 6.30 : La catégorie « ordinaire »**



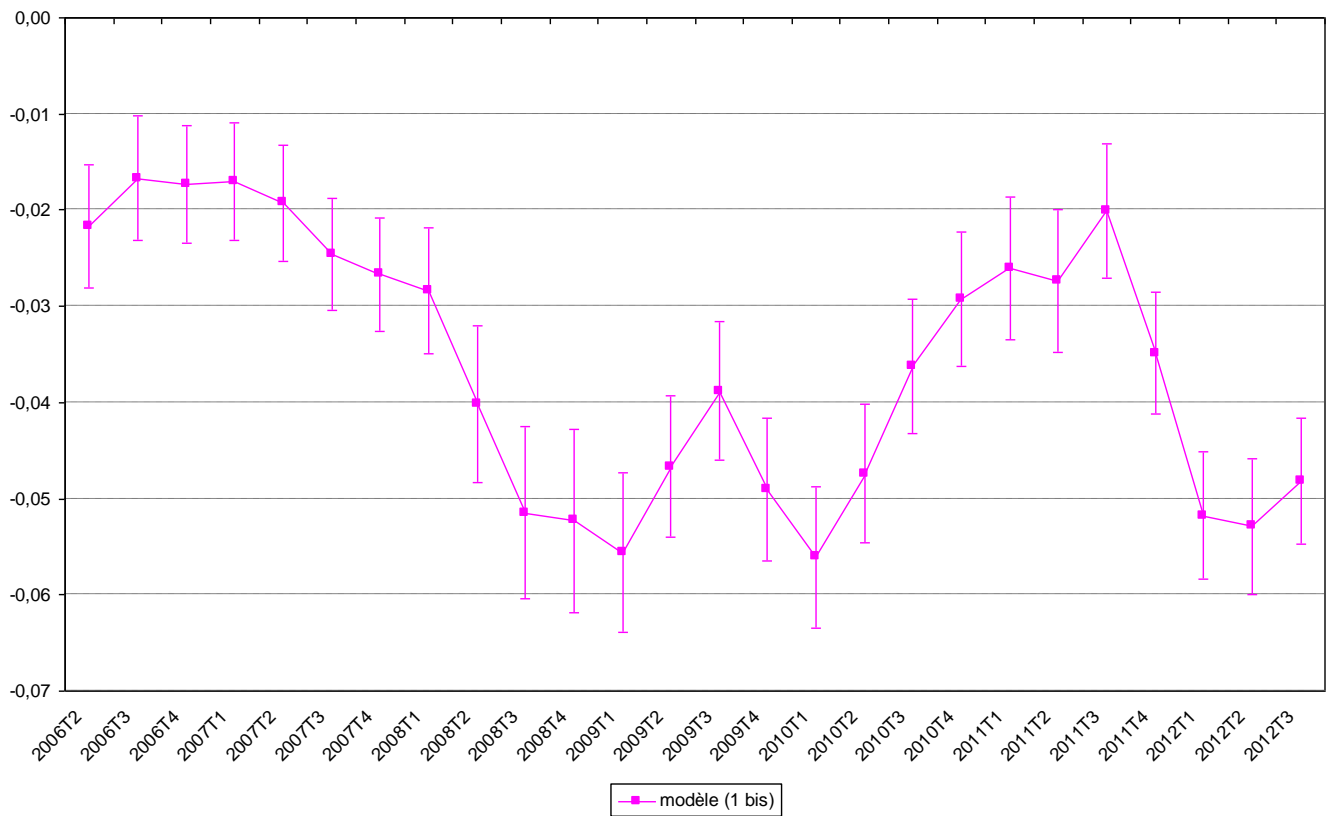
**Graphique 6.31 : La catégorie « confortable »**



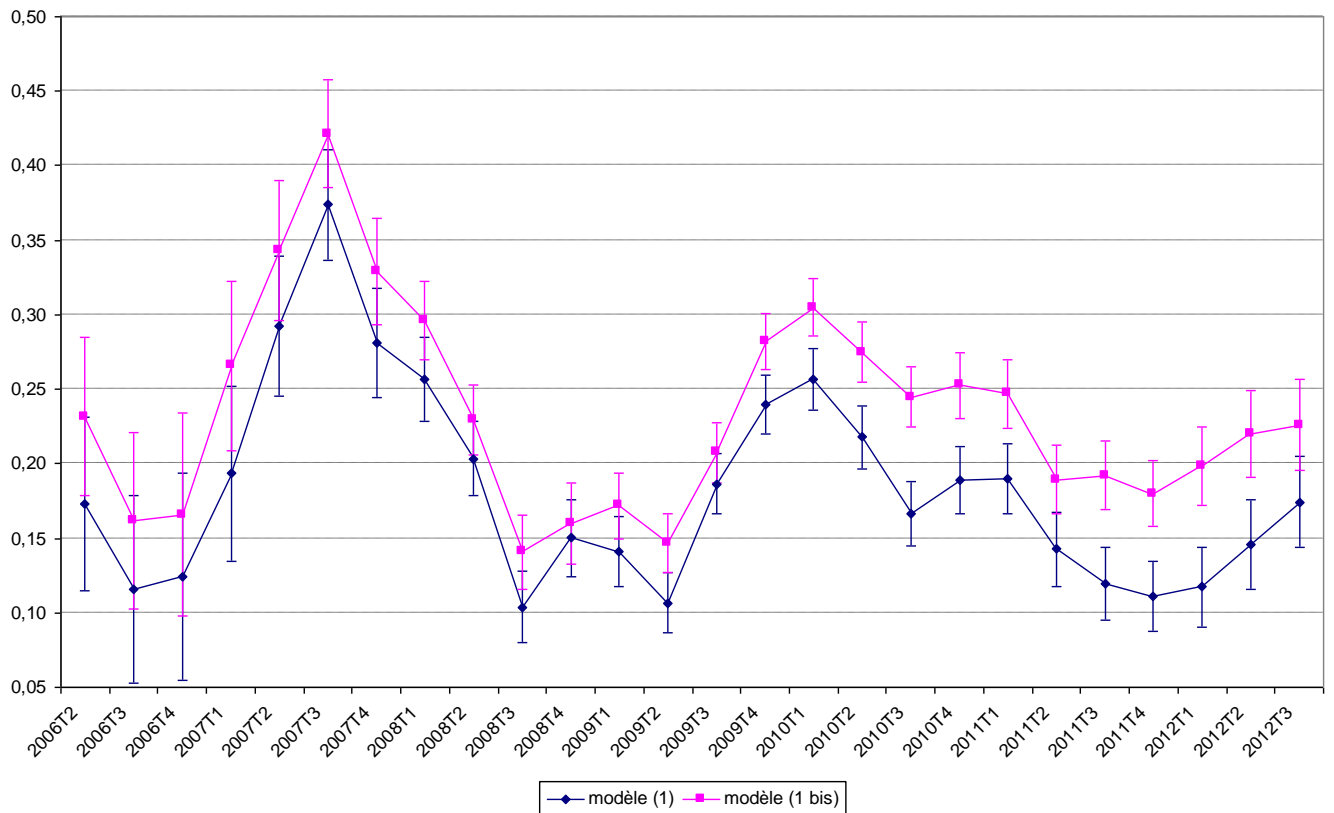
**Graphique 6.32 : La catégorie « très confortable »**



**Graphique 6.33 : La catégorie « intermédiaire »**

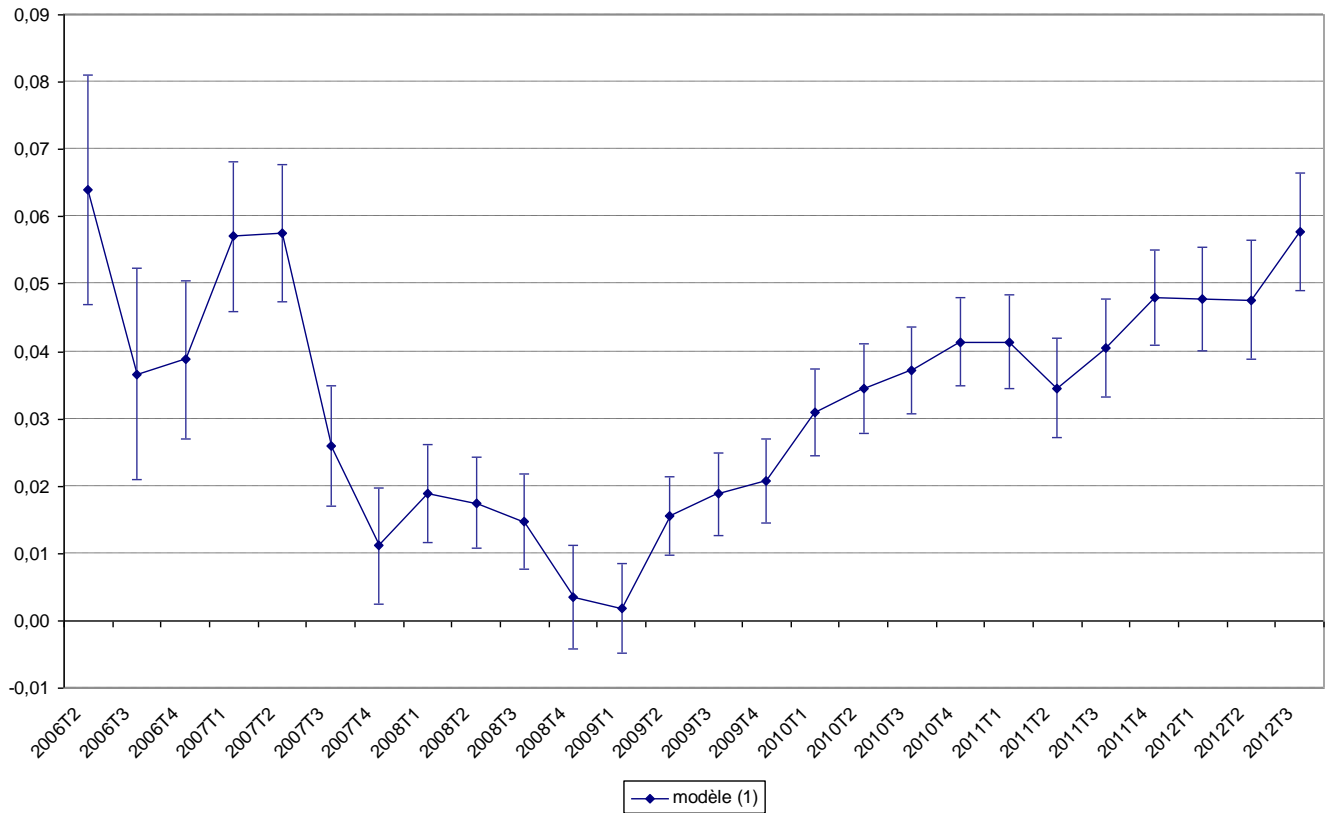


**Graphique 6.34 : La catégorie « luxe »**

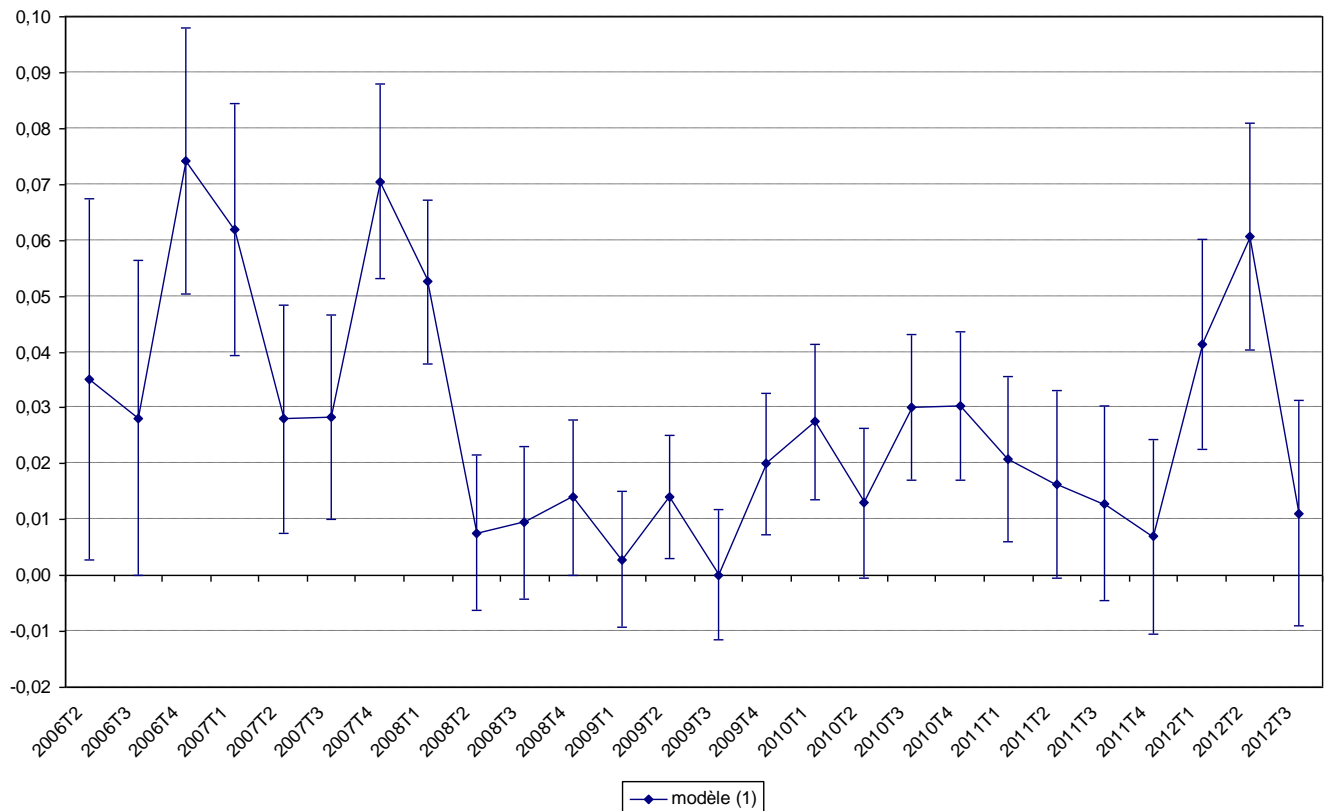




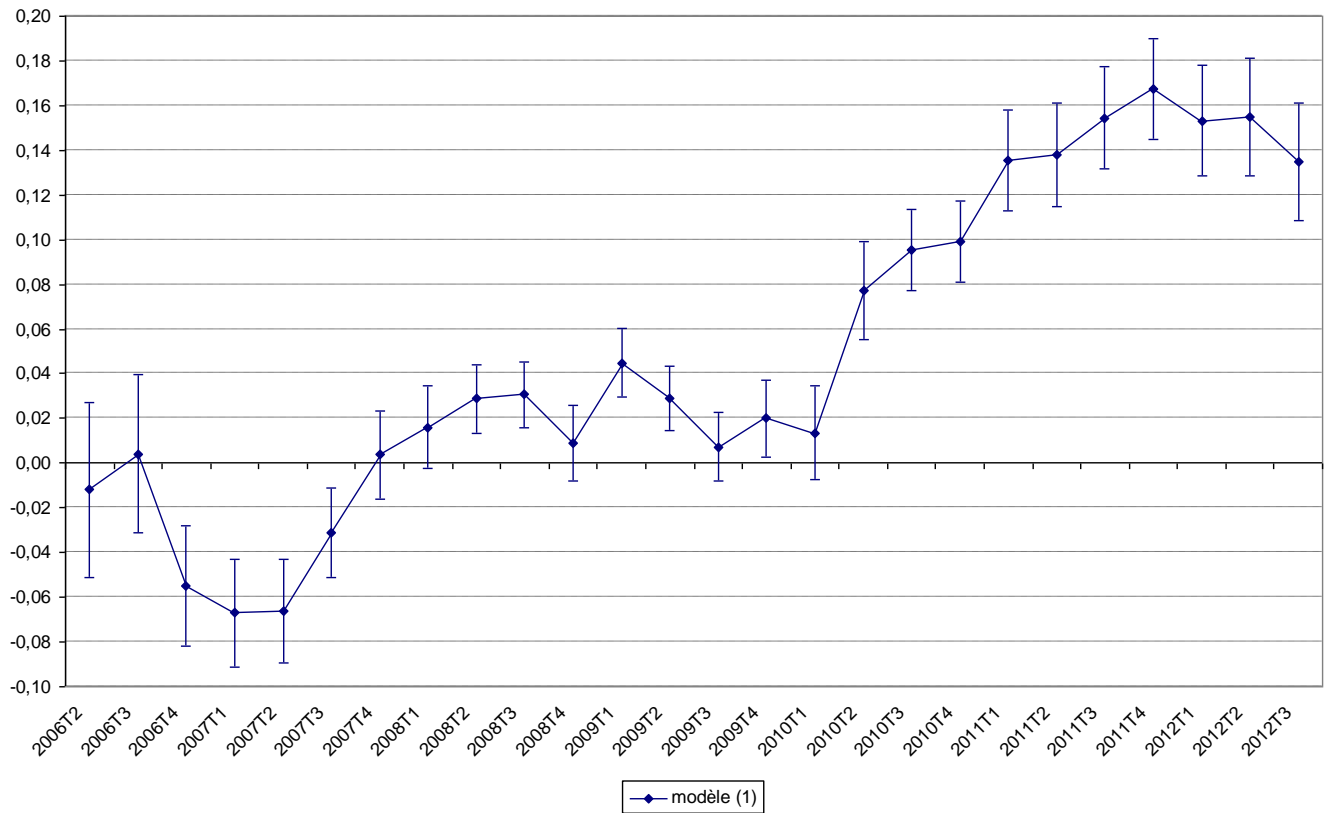
**Graphique 6.35 : La proximité d'un transport collectif par voie ferrée**



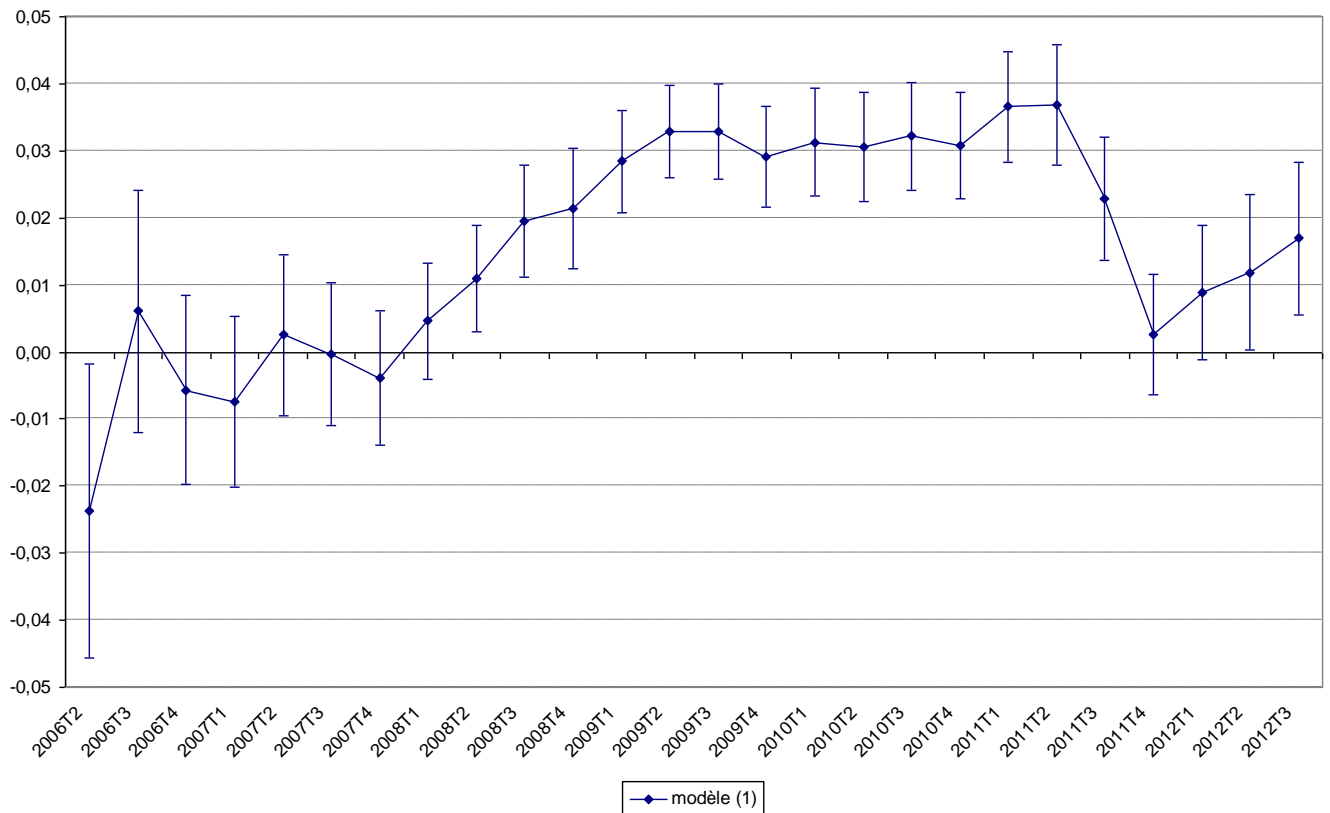
**Graphique 6.36 : La présence d'un système de climatisation dans le programme de construction**



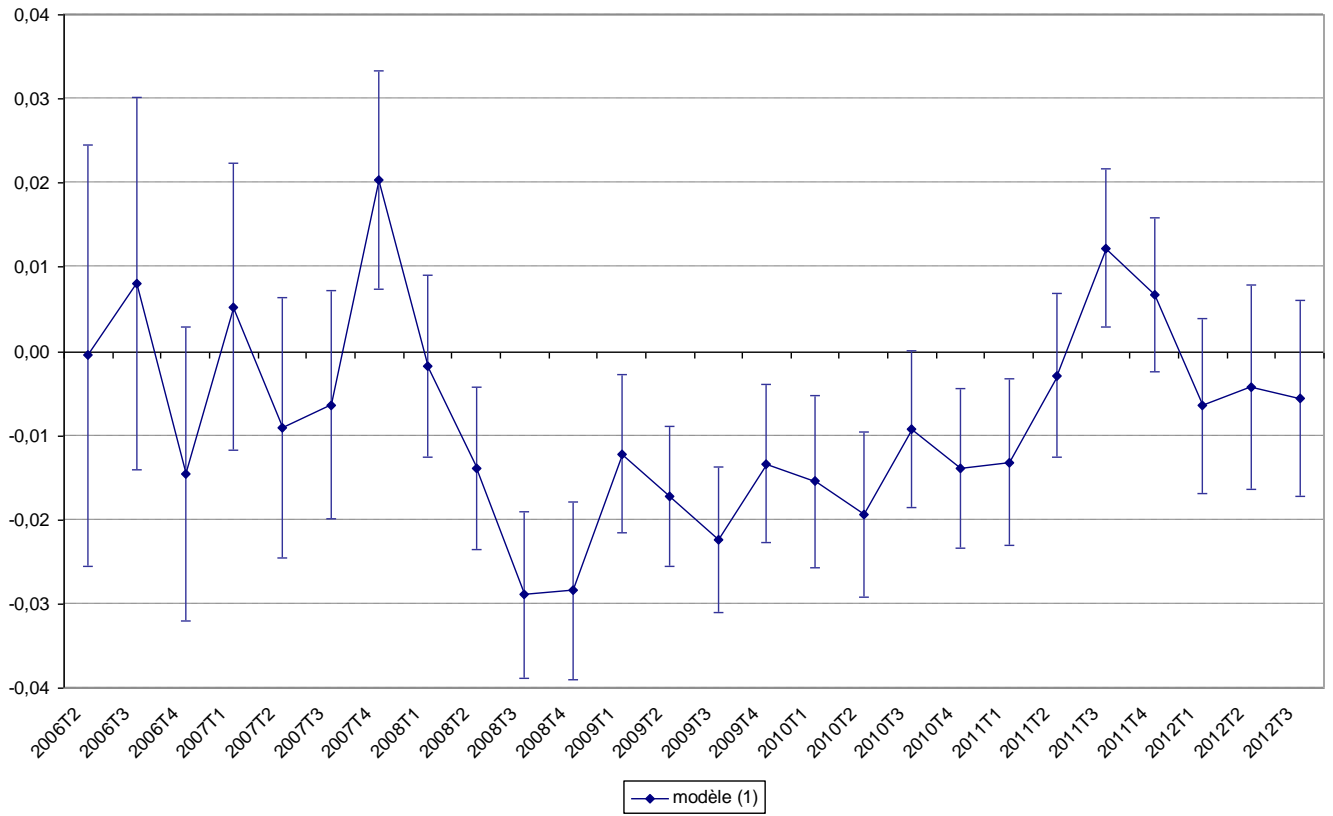
**Graphique 6.37 : La présence d'une piscine dans la résidence**



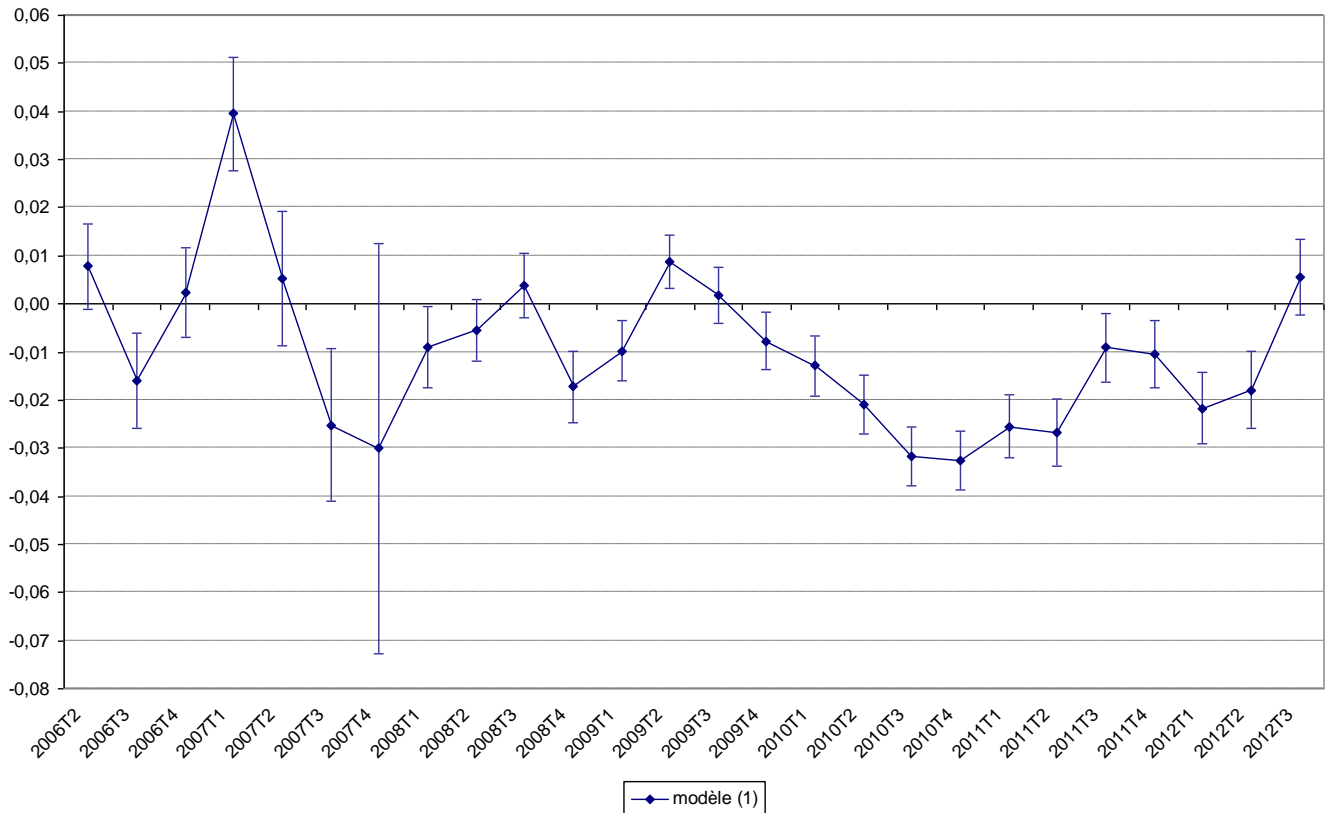
**Graphique 6.38 : La présence d'un dispositif de sécurité**



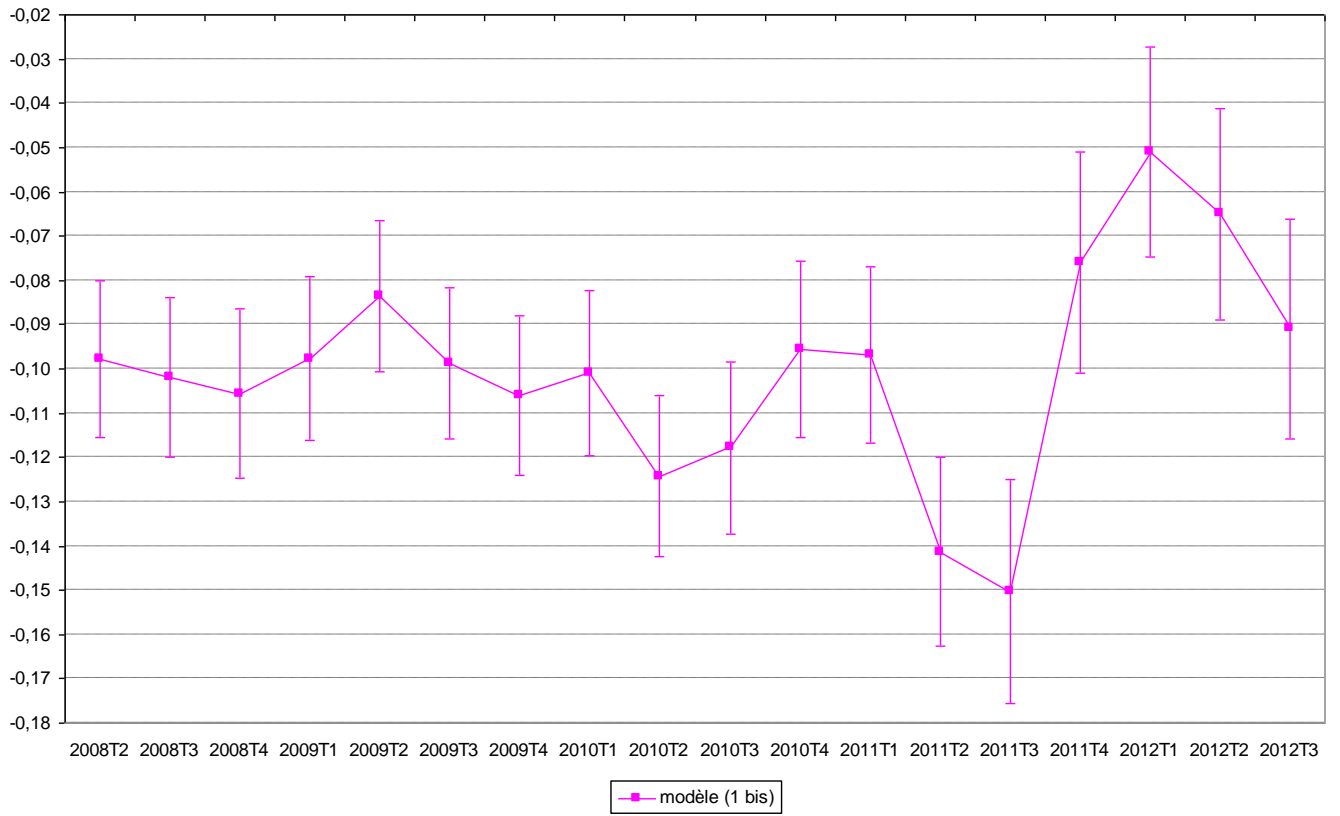
**Graphique 6.39 : La certification QUALITEL**



**Graphique 6.40 : La présence d'un balcon ou d'une terrasse pour les appartements**



**Graphique 6.41 : Les maisons bâties sur un grand terrain**



**Annexe 13 : Les p-value correspondant au test de nullité du paramètre pour les modèles (1) et (1 bis)**

Trimestre	Première classe géographique		Deuxième classe géographique		Troisième classe géographique		Quatrième classe géographique		Cinquième classe géographique		Sixième classe géographique		Septième classe géographique		Arrière-pays littoral	
	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)
2006T2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,56	0,26
2006T3	0,00	0,00	0,00	0,97	0,00	0,00	0,00	0,00	0,03	0,00	0,18	0,00	0,04	0,00	0,13	0,33
2006T4	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,97	0,00	0,03	0,00	0,00	0,01
2007T1	0,00	0,00	0,00	0,01	0,00	0,00	0,08	0,00	0,82	0,00	0,93	0,00	0,53	0,00	0,00	0,00
2007T2	0,08	0,00	0,00	0,04	0,18	0,00	0,80	0,00	0,19	0,00	0,05	0,00	0,04	0,00	0,00	0,00
2007T3	0,08	0,00	0,00	0,00	0,57	0,00	0,78	0,00	0,18	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2007T4	0,54	0,00	0,00	0,00	0,04	0,00	0,32	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2008T1	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,26	0,16
2008T2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,30	0,28
2008T3	0,04	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,04	0,02
2008T4	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,06	0,22
2009T1	0,04	0,00	0,00	0,37	0,03	0,00	0,14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,04
2009T2	0,00	0,00	0,00	0,00	0,00	0,00	0,04	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2009T3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2009T4	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2010T1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2010T2	0,00	0,00	0,00	0,00	0,09	0,00	0,25	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2010T3	0,01	0,00	0,00	0,00	0,17	0,00	0,21	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2010T4	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2011T1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,20
2011T2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,96	0,57
2011T3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,84	0,23
2011T4	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,77	0,14
2012T1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,47	0,69
2012T2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,09	0,16
2012T3	0,00	0,00	0,00	0,58	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,35	0,12

Trimestre	Ville-centre		Banlieue		Ville isolée		Indicatrice temporelle		Catégorie "ordinaire"		Catégorie "très confortable"	
	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)	modèle (1)	modèle (1 bis)	modèle (1)	
2006T2	0,00	0,00	0,06	0,27	0,45	0,00	0,00	0,00	0,00	0,00	0,00	0,09
2006T3	0,00	0,02	0,01	0,36	0,06	0,00	0,01	0,01	0,01	0,01	0,21	0,97
2006T4	0,00	0,15	0,07	0,71	0,00	0,00	0,00	0,00	0,06	0,35	0,40	
2007T1	0,00	0,01	0,09	0,53	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,36
2007T2	0,00	0,01	0,00	0,51	0,01	0,00	0,00	0,00	0,04	0,51	0,28	
2007T3	0,00	0,00	0,00	0,01	0,14	0,00	0,00	0,00	0,00	0,00	0,00	0,14
2007T4	0,00	0,00	0,00	0,00	0,56	0,09	0,86	0,43	0,00	0,00	0,00	0,03
2008T1	0,00	0,00	0,01	0,12	0,02	0,00	0,60	0,52	0,00	0,00	0,00	0,80
2008T2	0,00	0,00	0,00	0,00	0,31	0,01	0,00	0,00	0,00	0,00	0,00	0,01
2008T3	0,00	0,00	0,00	0,02	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2008T4	0,01	0,02	0,35	0,56	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,18
2009T1	0,00	0,00	0,23	0,49	0,00	0,00	0,18	0,01	0,00	0,00	0,00	0,00
2009T2	0,00	0,00	0,00	0,05	0,91	0,08	0,00	0,00	0,00	0,00	0,00	0,00
2009T3	0,00	0,00	0,06	0,00	0,36	0,44	0,82	0,65	0,00	0,00	0,00	0,02
2009T4	0,00	0,00	0,00	0,00	0,32	0,40	0,00	0,00	0,00	0,00	0,00	0,00
2010T1	0,00	0,00	0,00	0,00	0,24	0,08	0,66	0,12	0,00	0,00	0,00	0,00
2010T2	0,00	0,00	0,00	0,00	0,82	0,41	0,00	0,00	0,00	0,00	0,00	0,00
2010T3	0,00	0,00	0,00	0,00	0,00	0,00	0,08	0,00	0,00	0,00	0,00	0,00
2010T4	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2011T1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2011T2	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,01	0,00	0,00	0,00
2011T3	0,00	0,00	0,00	0,00	0,00	0,00	0,09	0,09	0,00	0,00	0,00	0,00
2011T4	0,00	0,00	0,00	0,00	0,00	0,00	0,10	0,16	0,05	0,00	0,00	0,00
2012T1	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00
2012T2	0,00	0,00	0,00	0,00	0,00	0,00	0,15	0,45	0,84	0,00	0,00	0,00
2012T3	0,00	0,00	0,00	0,00	0,03	0,73	0,00	0,00	0,00	0,00	0,00	0,00

Trimestre	Proximité d'un transport collectif par voie ferrée	Présence d'un système de climatisation	Présence d'une piscine	Certification QUALITEL	Présence d'un dispositif de sécurité	Présence d'un balcon ou d'une terrasse pour les appartements
	modèle (1)	modèle (1)	modèle (1)	modèle (1)	modèle (1)	modèle (1)
2006T2	0,00	0,03	<b>0,55</b>	<b>0,97</b>	0,03	<b>0,09</b>
2006T3	0,00	<b>0,05</b>	<b>0,81</b>	<b>0,47</b>	<b>0,51</b>	0,00
2006T4	0,00	0,00	0,00	<b>0,10</b>	<b>0,43</b>	<b>0,61</b>
2007T1	0,00	0,00	0,00	<b>0,54</b>	<b>0,26</b>	0,00
2007T2	0,00	0,01	0,00	<b>0,25</b>	<b>0,67</b>	<b>0,45</b>
2007T3	0,00	0,00	0,00	<b>0,36</b>	<b>0,95</b>	0,00
2007T4	0,01	0,00	<b>0,70</b>	0,00	<b>0,45</b>	<b>0,17</b>
2008T1	0,00	0,00	<b>0,09</b>	<b>0,75</b>	<b>0,30</b>	0,04
2008T2	0,00	<b>0,28</b>	0,00	0,00	0,01	<b>0,09</b>
2008T3	0,00	<b>0,18</b>	0,00	0,00	0,00	<b>0,26</b>
2008T4	<b>0,37</b>	0,05	<b>0,30</b>	0,00	0,00	0,00
2009T1	<b>0,58</b>	<b>0,64</b>	0,00	0,01	0,00	0,00
2009T2	0,00	0,01	0,00	0,00	0,00	0,00
2009T3	0,00	<b>0,98</b>	<b>0,34</b>	0,00	0,00	<b>0,56</b>
2009T4	0,00	0,00	0,02	0,01	0,00	0,01
2010T1	0,00	0,00	<b>0,21</b>	0,00	0,00	0,00
2010T2	0,00	<b>0,06</b>	0,00	0,00	0,00	0,00
2010T3	0,00	0,00	0,00	<b>0,05</b>	0,00	0,00
2010T4	0,00	0,00	0,00	0,00	0,00	0,00
2011T1	0,00	0,01	0,00	0,01	0,00	0,00
2011T2	0,00	<b>0,06</b>	0,00	<b>0,57</b>	0,00	0,00
2011T3	0,00	<b>0,14</b>	0,00	0,01	0,00	0,01
2011T4	0,00	<b>0,43</b>	0,00	<b>0,15</b>	<b>0,56</b>	0,00
2012T1	0,00	0,00	0,00	<b>0,22</b>	<b>0,08</b>	0,00
2012T2	0,00	0,00	0,00	<b>0,50</b>	0,04	0,00
2012T3	0,00	<b>0,28</b>	0,00	<b>0,34</b>	0,00	<b>0,17</b>

Les régresseurs n'apparaissant pas dans ces tableaux sont toujours significatifs pour les modèles (1) et (1 bis) au niveau 5% sur l'ensemble de la période 2006T1-2012T3 (i.e. p-value<0,05).

## Première étude : test de sensibilité de l'indice des prix des logements anciens au choix de méthodes de calcul

### Annexe 14 : Les logements anciens retenus pour le calcul des indices Notaires-Insee

Un logement ancien contenu dans les bases notariales appartient au champ des indices Notaires-Insee s'il répond aux critères suivants:

- il est libre d'occupation au moment de la vente ou le statut d'occupation du logement n'est pas renseigné
- il est destiné à un usage strict d'habitation
- il est acquis en pleine propriété par un particulier, une société civile immobilière ou par un acquéreur dont la qualité n'est pas renseignée

et

- il fait l'objet d'une vente de gré à gré (i.e. hors adjudications)

#### Les maisons anciennes

On ne garde que les maisons anciennes qui correspondent à :

- une ferme,
- une maison de ville ou de village,
- un pavillon,
- une villa ,

ou

- une maison dont le type n'est pas renseigné

En outre, on ne retient que les maisons anciennes ayant un nombre de pièces compris entre 1 et 12, et une surface habitable comprise entre 20 et 300 m<sup>2</sup>. Dans la base « Perval », les maisons anciennes retenues sont également caractérisées par une surface du terrain non nul et un prix net vendeur compris entre 1 500 euros et 15 millions d'euros. Quant à celles considérées dans la base « BIEN », elles ont un prix net vendeur supérieur à 1500 euros.

#### Les appartements anciens

On ne garde que les appartements anciens qui correspondent à :

- un appartement standard,
  - un duplex,
  - un studio,
  - un triplex,
- ou
- un appartement dont le type n'est pas renseigné



En outre, seuls les appartements anciens vérifiant les conditions suivantes sont considérés :

- un nombre de pièces compris entre 1 et 8 ;
- une surface habitable comprise entre 10 et 200 m<sup>2</sup> ;
- un prix net vendeur supérieur à 1 500 euros ;
- un prix net vendeur au m<sup>2</sup> inférieur à 25 000 euros.

## Seconde étude : test de sensibilité de l'indice des prix des logements anciens à la distribution géographique des logements

### Annexe 15 : La limite des distributions géographiques trimestrielles

En observant les distributions géographiques trimestrielles, on constate que l'effectif de logements anciens pour certains triplets (type de construction ; commune ; trimestre) qui comprennent des ventes de logements neufs peut être nul. Ainsi, plus de 13% de l'ensemble des logements neufs vendus entre les années 2006 et 2010 ne peuvent être clonés à l'aide de logements anciens de même type, vendus dans la même commune au même trimestre (cf. tableau 6.9).

**Tableau 6.9 : Répartition des triplets (type de construction ; commune ; trimestre) pour lesquels il y a existence de ventes de logements neufs en fonction de l'existence ou non de ventes de logements anciens (2006-2010)**

Existence de ventes de logements anciens	Part des triplets (type de construction ; commune ; trimestre)	Nombre de logements neufs vendus	Part dans l'ensemble des logements neufs vendus
<b>oui</b>	73,6%	471 549	86,4%
<b>non</b>	26,4%	74 324	<b>13,6%</b>
<b>Total</b>	100,0%	545 873	100,0%

Lecture : 26,4% des triplets (type de construction ; commune ; trimestre) pour lesquels il y a existence de ventes de logements neufs se caractérisent par une absence de ventes de logements anciens.

Afin de réduire ce défaut de couverture, on relâche la granularité de description du temps pour adopter l'année. Ainsi, la part des logements neufs qui ne peuvent pas être clonés se réduit à 6% (cf. tableau 6.10). La contrepartie de cet assouplissement est que les variations infra annuelles de l'indice « clone » ne peuvent pas être comparées à celles des autres indices<sup>56</sup> et doivent donc être ignorées dans l'analyse des différences : seuls les écarts de long terme sont pertinents.

**Tableau 6.10 : Répartition des triplets (type de construction ; commune ; année) pour lesquels il y a existence de ventes de logements neufs en fonction de l'existence ou non de ventes de logements anciens (2006-2010)**

Existence de ventes de logements anciens	Part des triplets (type de construction ; commune ; année)	Nombre de logements neufs vendus	Part dans l'ensemble des logements neufs vendus
<b>oui</b>	85,0%	513 167	94,0%
<b>non</b>	15,0%	32 706	<b>6,0%</b>
<b>Total</b>	100,0%	545 873	100,0%

Lecture : 15,0% des triplets (type de construction ; commune ; année) pour lesquels il y a existence de ventes de logements neufs se caractérisent par une absence de ventes de logements anciens.

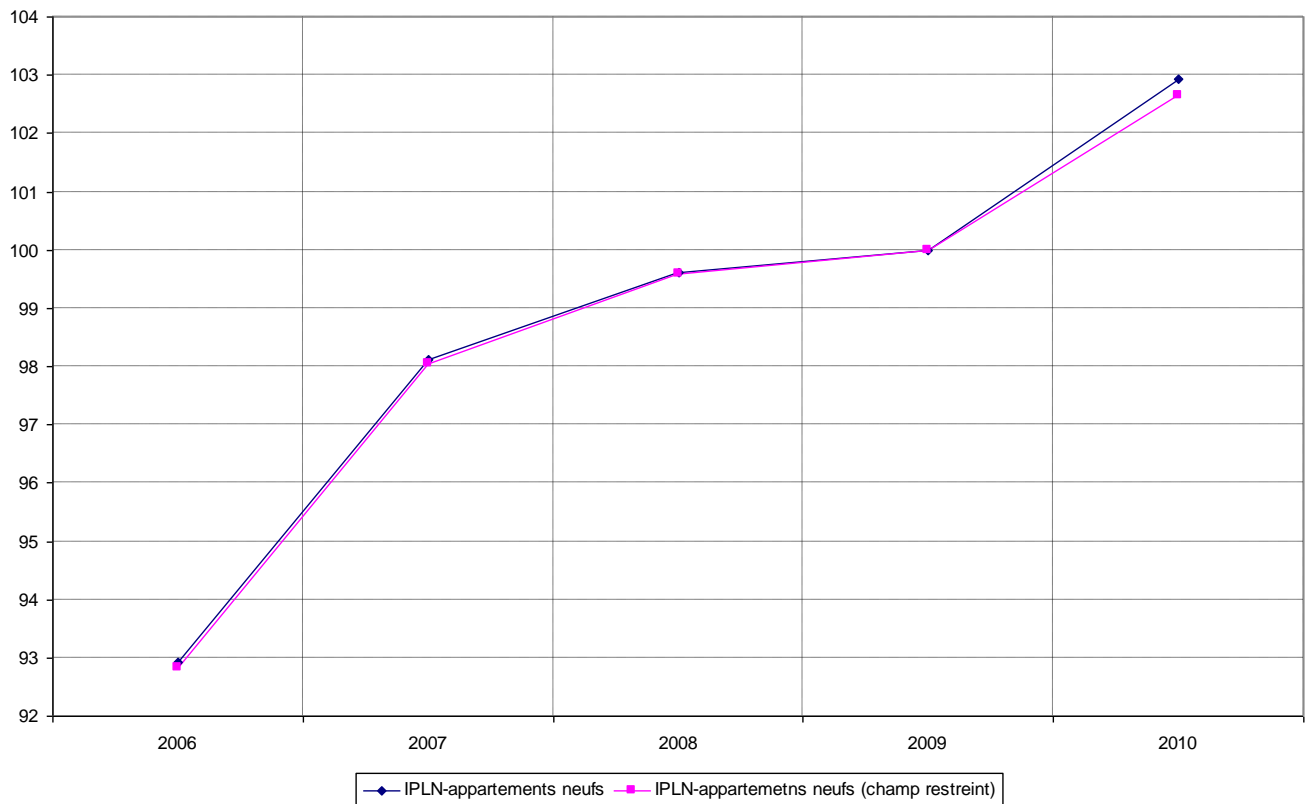
<sup>56</sup> L'IPLN et l'indice « 1<sup>ère</sup> simulation »

La non prise en compte des 6% de logements neufs qui ne peuvent pas être clonés constitue une approximation dans notre raisonnement. Elle est cependant modeste : pour chaque type de construction, l'écart en valeur absolue entre un indice des prix des logements neufs calculé sur l'ensemble de l'échantillon et un indice qui serait calculé sur l'échantillon restreint aux seuls logements neufs pouvant être effectivement clonés s'élève au maximum à 0,6 point d'indice sur la période 2006-2010 (cf. graphiques 6.4 et 6.5). L'indice des prix des maisons neuves, noté IPLN-maisons neuves, et l'indice des prix des appartements neufs, noté IPLN-appartements neufs, sont calculés de la même manière que l'IPLN mais en utilisant deux modèles hédoniques différents. Ces deux modèles sont similaires au modèle **(1)** à la base du calcul de l'IPLN (cf. §1.3.3.2), la seule différence résidant dans la non prise en compte du type de construction dans ces deux modèles (cf. annexe 16).

**Graphique 6.42 : L'IPLN-maisons neuves et l'IPLN-maisons neuves calculé sur l'échantillon restreint aux seuls logements neufs pouvant être effectivement clonés (2006-2010)**



**Graphique 6.43 : L'IPLN-appartements neufs et l'IPLN-appartements neufs calculé sur l'échantillon restreint aux seuls logements neufs pouvant être effectivement clonés (2006-2010)**



En conséquence, la population des « clones » est tirée dans la population des logements anciens mutés contenus dans les bases notariales à partir des distributions géographiques annuelles.

## Annexe 16 : L'IPLN-maisons neuves et l'IPLN-appartements neufs

Pour les maisons neuves, le modèle hédonique considéré s'écrit sous la forme suivante :

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

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

où  $\text{nb\_obs\_mais}(t)$  est le nombre d'observations issues de l'ECLN correspondant à des maisons neuves vendues au trimestre  $t$

Pour les appartements neufs, on utilise le modèle suivant :

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

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

où  $\text{nb\_obs\_appart}(t)$  est le nombre d'observations issues de l'ECLN correspondant à des appartements neufs vendus au trimestre  $t$

Les indicatrices correspondant aux caractéristiques géographiques des logements figurant dans ces deux modèles sont exactement les mêmes que celles apparaissant dans le modèle **(1)**. Celles correspondant aux caractéristiques physiques sont uniquement définies à partir du nombre de pièces des logements (2 indicatrices) :

- logement d'une pièce ;
- logement de 2 à 4 pièces ;
- logement de plus de 5 pièces (modalité de référence).

En outre, une indicatrice signalant la présence d'un balcon ou d'une terrasse figure dans le modèle hédonique pour les appartements neufs.

En appliquant la méthode sur périodes adjacentes aux maisons neuves, il vient que l'indice des prix des maisons neuves d'un trimestre  $T_1$ , exprimé en base 100 au trimestre  $T_{\text{base}}$  (avec  $T_{\text{base}} < T_1$ ), est donné par la relation suivante :

$$\text{IPLN}_{T_1}^{\text{mais}, T_{\text{base}}} = \exp \left( \sum_{t=T_{\text{base}}}^{T_1-1} \hat{\delta}_{t+1}^{\text{mais}} \right) \times 100$$

où  $\forall t = T_{\text{base}}, \dots, T_1 - 1$ ,  $\hat{\delta}_{t+1}^{\text{mais}}$  est l'estimateur des MCO du paramètre  $\delta_{t+1}^{\text{mais}}$  obtenu en pondérant chaque observation  $i$  d'un trimestre  $t$  par le nombre de logements réservés correspondant  $n_{i,t}$ .

De même, il vient que l'indice des prix des appartements neufs d'un trimestre  $T_1$ , exprimé en base 100 au trimestre  $T_{base}$  (avec  $T_{base} < T_1$ ), est donné par la relation suivante :

$$\text{IPLN}_{T_1}^{\text{appart}, T_{base}} = \exp \left( \sum_{t=T_{base}}^{T_1-1} \hat{\delta}_{t+1}^{\text{appart}} \right) \times 100$$

où  $\forall t = T_{base}, \dots, T_1 - 1$ ,  $\hat{\delta}_{t+1}^{\text{appart}}$  est l'estimateur des MCO du paramètre  $\delta_{t+1}^{\text{appart}}$  obtenu en pondérant chaque observation  $i$  d'un trimestre  $t$  par le nombre de logements réservés correspondant  $n_{i,t}$ .

## Diagnostics

### Annexe 17 : Les variables explicatives du modèle (4) présentant un problème de colinéarité au sens du VIF sur la période 2006T1-2012T3

Variable j	Nombre de régressions pour lesquelles $VIF_j \geq 10$	Pourcentage correspondant	VIF <sub>j</sub> moyen	VIF <sub>j</sub> minimum	VIF <sub>j</sub> maximum
12 <sup>ème</sup> classe géographique	25	96,15%	52,88	9,19	152,30
11 <sup>ème</sup> classe géographique	25	96,15%	46,56	9,30	126,04
8 <sup>ème</sup> classe géographique	23	88,46%	45,56	8,57	133,23
7 <sup>ème</sup> classe géographique	20	76,92%	35,43	6,19	107,19
6 <sup>ème</sup> classe géographique	20	76,92%	27,98	6,30	84,36
3 <sup>ème</sup> classe géographique	20	76,92%	20,34	6,70	51,62
4 <sup>ème</sup> classe géographique	17	65,38%	14,19	6,73	30,21
9 <sup>ème</sup> classe géographique	16	61,54%	19,57	3,72	58,41
5 <sup>ème</sup> classe géographique	16	61,54%	19,06	5,21	48,07
2 <sup>ème</sup> classe géographique	16	61,54%	13,68	4,58	32,06
10 <sup>ème</sup> classe géographique	15	57,69%	13,23	3,20	36,42
villes de banlieue	9	34,62%	8,94	6,26	11,68
villes-centres	5	19,23%	8,33	5,56	11,38
13 <sup>ème</sup> classe géographique	5	19,23%	6,81	2,22	22,39
1 <sup>ère</sup> classe géographique	1	3,85%	5,68	2,52	14,61

Lecture : pour 20 des 26 régressions effectuées sur la période 2006T1-2012T3 (soit 76,92% de l'ensemble des régressions), le VIF associé au régresseur « 7<sup>ème</sup> classe géographique » est supérieur ou égal à 10.

**Annexe 18 : Les indices de conditionnement et les proportions de variance associées correspondant au modèle (4) et révélant un problème de colinéarité (2006T1-2012T3)**

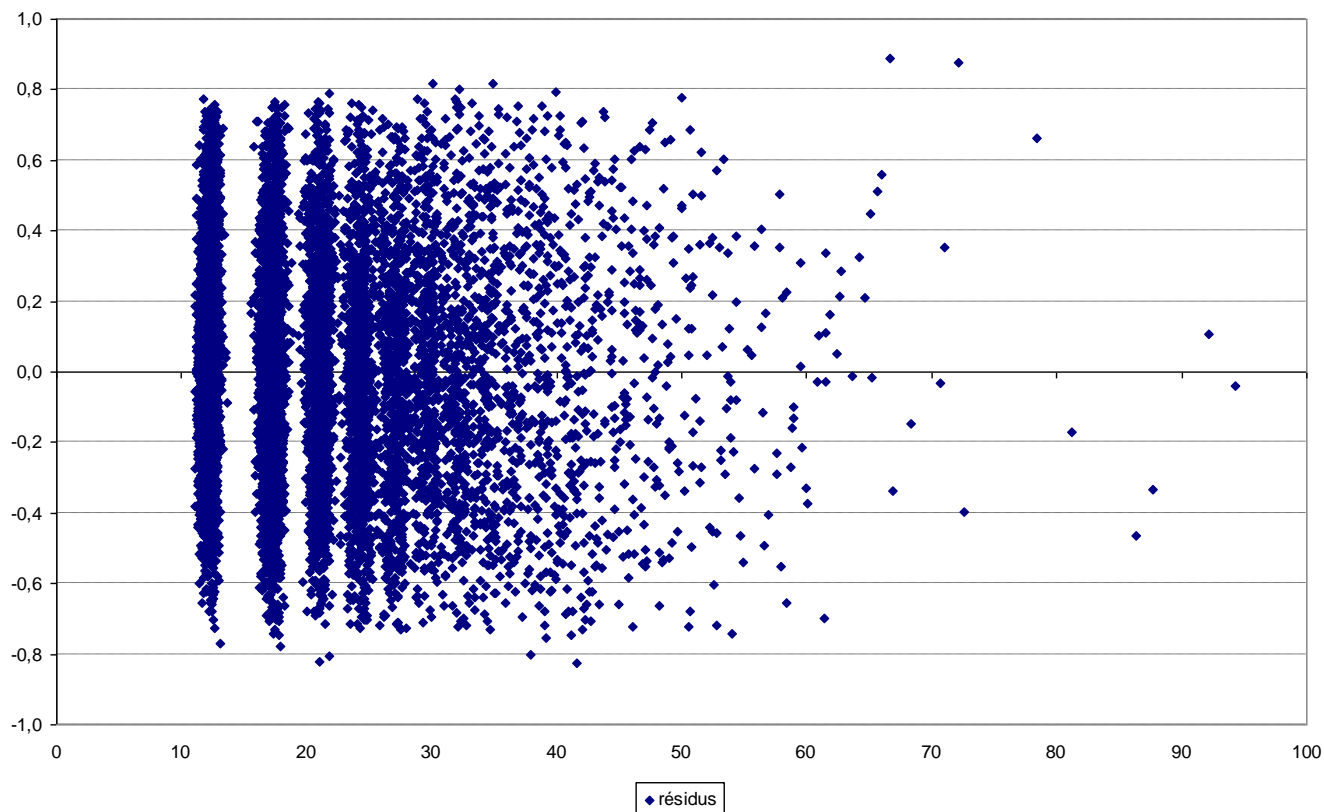
Trimestre	k	$\lambda_k$	$Cl_k$	1 <sup>ère</sup> classe géographique	2 <sup>ème</sup> classe géographique	3 <sup>ème</sup> classe géographique	4 <sup>ème</sup> classe géographique	5 <sup>ème</sup> classe géographique	6 <sup>ème</sup> classe géographique	7 <sup>ème</sup> classe géographique
2009T1	35	0,002	30,618	0,862	0,942	0,967	0,953	0,964	0,975	0,979
2009T2	35	0,002	31,871	0,849	0,941	0,971	0,951	0,965	0,976	0,983
2010T1	35	0,002	31,111	0,863	0,939	0,965	0,942	0,963	0,977	0,981
2010T4	35	0,003	30,679	0,844	0,942	0,965	0,923	0,962	0,975	0,984
2011T1	35	0,002	34,950	0,865	0,954	0,970	0,937	0,970	0,980	0,987
2011T2	35	0,001	41,987	0,900	0,967	0,979	0,954	0,977	0,986	0,990
2011T3	35	0,001	46,949	0,932	0,968	0,982	0,965	0,981	0,990	0,992
2012T2	35	0,001	42,150	0,896	0,963	0,972	0,956	0,978	0,987	0,988
2012T3	35	0,002	36,778	0,853	0,947	0,957	0,946	0,969	0,983	0,984

Trimestre	k	$\lambda_k$	$Cl_k$	8 <sup>ème</sup> classe géographique	9 <sup>ème</sup> classe géographique	10 <sup>ème</sup> classe géographique	11 <sup>ème</sup> classe géographique	12 <sup>ème</sup> classe géographique	13 <sup>ème</sup> classe géographique	14 <sup>ème</sup> classe géographique
20091	35	0,002	30,618	0,982	0,956	0,950	0,982	0,988	0,831	0,758
20092	35	0,002	31,871	0,983	0,963	0,951	0,983	0,990	0,824	0,770
20101	35	0,002	31,111	0,983	0,965	0,937	0,983	0,988	0,848	0,841
20104	35	0,003	30,679	0,984	0,963	0,943	0,981	0,989	0,842	0,723
20111	35	0,002	34,950	0,987	0,969	0,953	0,986	0,992	0,885	0,764
20112	35	0,001	41,987	0,991	0,980	0,967	0,991	0,994	0,921	0,839
20113	35	0,001	46,949	0,993	0,984	0,973	0,991	0,995	0,932	0,875
20122	35	0,001	42,150	0,991	0,978	0,966	0,992	0,993	0,879	0,823
20123	35	0,002	36,778	0,986	0,970	0,954	0,989	0,992	0,848	0,755



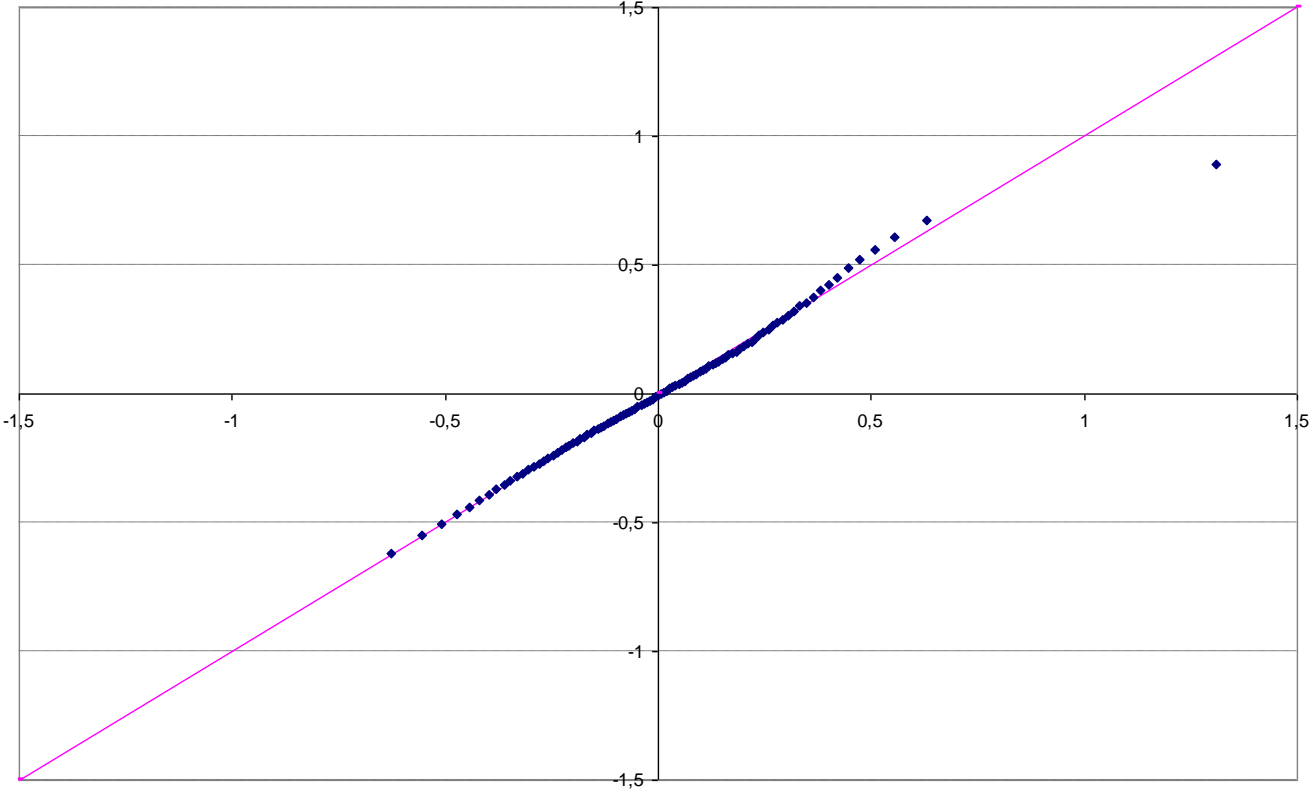
**Annexe 19 : Graphiques des résidus  $\tilde{u}_{i,t}$  en fonction des valeurs estimées  $\hat{y}_{i,t}$  pour le modèle (4) (2006T1-2012T3)**

**Graphique 6.44 : Les résidus en fonction des valeurs estimées (2009T1-2009T2)**



**Annexe 20 : Les QQ-plots des résidus pour le modèle (4) (2006T1-2012T3)**

**Graphique 6.45 : QQ-plot des résidus (2009T1-2009T2)**



**Annexe 21 : Le nombre d'observations aberrantes et influentes pour chaque régression pour le modèle (1) (2006T1-2012T3)**

Trimestre	Observations aberrantes				Observations influentes			
	Nombre d'observations	Pourcentage	Nombre de logements	Pourcentage	Nombre d'observations	Pourcentage	Nombre de logements	Pourcentage
2006T2	911	5,42%	9 447	15,40%	0	0,00%	0	0,00%
2006T3	823	4,96%	9 640	15,97%	0	0,00%	0	0,00%
2006T4	902	5,18%	10 057	15,89%	0	0,00%	0	0,00%
2007T1	608	3,21%	8 296	12,09%	1	0,01%	31	0,05%
2007T2	617	3,15%	8 327	12,08%	1	0,01%	31	0,04%
2007T3	967	5,04%	9 771	16,00%	0	0,00%	0	0,00%
2007T4	939	5,04%	8 677	15,33%	0	0,00%	0	0,00%
2008T1	629	3,45%	5 978	11,10%	2	0,01%	72	0,13%
2008T2	522	3,07%	4 777	10,11%	2	0,01%	72	0,15%
2008T3	659	4,57%	4 771	12,84%	0	0,00%	0	0,00%
2008T4	602	4,89%	4 211	13,79%	0	0,00%	0	0,00%
2009T1	757	5,36%	5 481	13,85%	0	0,00%	0	0,00%
2009T2	951	5,48%	7 185	13,71%	0	0,00%	0	0,00%
2009T3	987	5,59%	7 916	14,75%	0	0,00%	0	0,00%
2009T4	914	5,35%	7 909	15,14%	0	0,00%	0	0,00%
2010T1	848	5,08%	7 659	14,66%	0	0,00%	0	0,00%
2010T2	831	4,96%	8 855	15,85%	0	0,00%	0	0,00%
2010T3	852	5,08%	9 452	16,01%	0	0,00%	0	0,00%
2010T4	805	4,96%	8 475	14,76%	0	0,00%	0	0,00%
2011T1	706	4,74%	6 830	13,75%	0	0,00%	0	0,00%
2011T2	687	4,90%	6 355	14,31%	0	0,00%	0	0,00%
2011T3	769	5,22%	7 394	15,24%	0	0,00%	0	0,00%
2011T4	833	5,21%	7 911	14,41%	0	0,00%	0	0,00%
2012T1	790	5,14%	6 946	14,40%	0	0,00%	0	0,00%
2012T2	721	5,01%	5 432	13,78%	0	0,00%	0	0,00%
2012T3	788	5,24%	5 804	14,36%	0	0,00%	0	0,00%
<b>Moyenne</b>	<b>785</b>	<b>4,82%</b>	<b>7 444</b>	<b>14,21%</b>	<b>0</b>	<b>0,00%</b>	<b>8</b>	<b>0,01%</b>

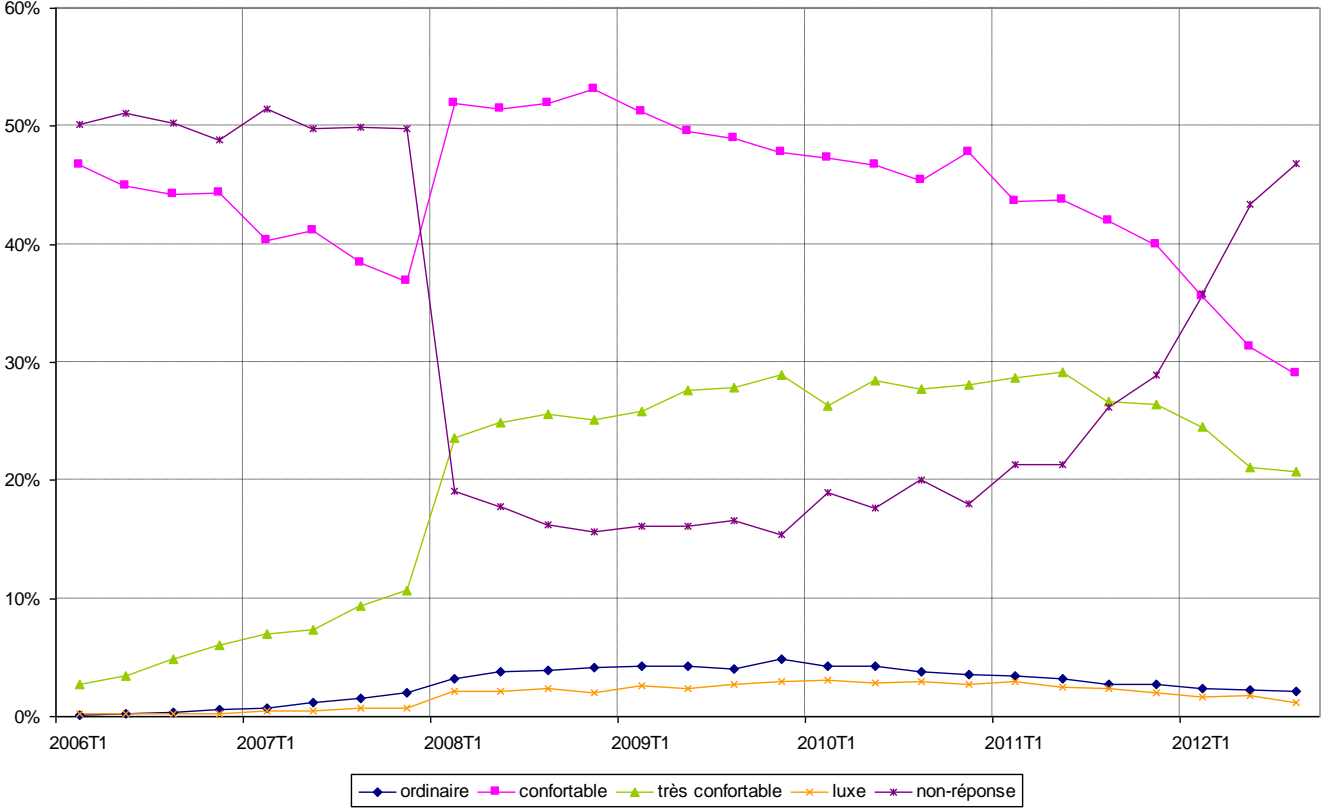
## Annexe 22 : La composition des 12 nouvelles classes géographiques

Classe géographique	ZEAT		TUU	
1	2	Bassin parisien	1	2 000 à 4 999 habitants
	4	Est	3	10 000 à 19 999 habitants
	4	Est	4	20 000 à 49 999 habitants
2	1	Région parisienne	4	20 000 à 49 999 habitants
	3	Nord	2	5 000 à 9 999 habitants
	3	Nord	6	100 000 à 199 999 habitants
	8	Centre-est	3	10 000 à 19 999 habitants
	9	Méditerranée	0	commune rurale
	9	Méditerranée	1	2 000 à 4 999 habitants
3	9	Méditerranée	2	5 000 à 9 999 habitants
	2	Bassin parisien	2	5 000 à 9 999 habitants
	2	Bassin parisien	5	50 000 à 99 999 habitants
	3	Nord	5	50 000 à 99 999 habitants
	4	Est	5	50 000 à 99 999 habitants
	5	Ouest	0	commune rurale
	5	Ouest	1	2 000 à 4 999 habitants
	7	Sud-ouest	3	10 000 à 19 999 habitants
	7	Sud-ouest	6	100 000 à 199 999 habitants
4	9	Méditerranée	6	100 000 à 199 999 habitants
	5	Ouest	2	5 000 à 9 999 habitants
	5	Ouest	3	10 000 à 19 999 habitants
	5	Ouest	5	50 000 à 99 999 habitants
	7	Sud-ouest	0	commune rurale
	7	Sud-ouest	1	2 000 à 4 999 habitants
5	7	Sud-ouest	2	5 000 à 9 999 habitants
	2	Bassin parisien	3	10 000 à 19 999 habitants
	4	Est	0	commune rurale
	4	Est	1	2 000 à 4 999 habitants
	4	Est	2	5 000 à 9 999 habitants
	5	Ouest	6	100 000 à 199 999 habitants
6	8	Centre-est	5	50 000 à 99 999 habitants
	2	Bassin parisien	0	commune rurale
	2	Bassin parisien	7	200 000 à 1 999 999 habitants
	3	Nord	1	2 000 à 4 999 habitants
	4	Est	7	200 000 à 1 999 999 habitants
	<b>7</b>	<b>Sud-ouest</b>	<b>4</b>	<b>20 000 à 49 999 habitants</b>
	7	Sud-ouest	5	50 000 à 99 999 habitants
9	Méditerranée	4	20 000 à 49 999 habitants	
7	2	Bassin parisien	4	20 000 à 49 999 habitants
	2	Bassin parisien	6	100 000 à 199 999 habitants
	4	Est	6	100 000 à 199 999 habitants
	5	Ouest	4	20 000 à 49 999 habitants
	5	Ouest	7	200 000 à 1 999 999 habitants

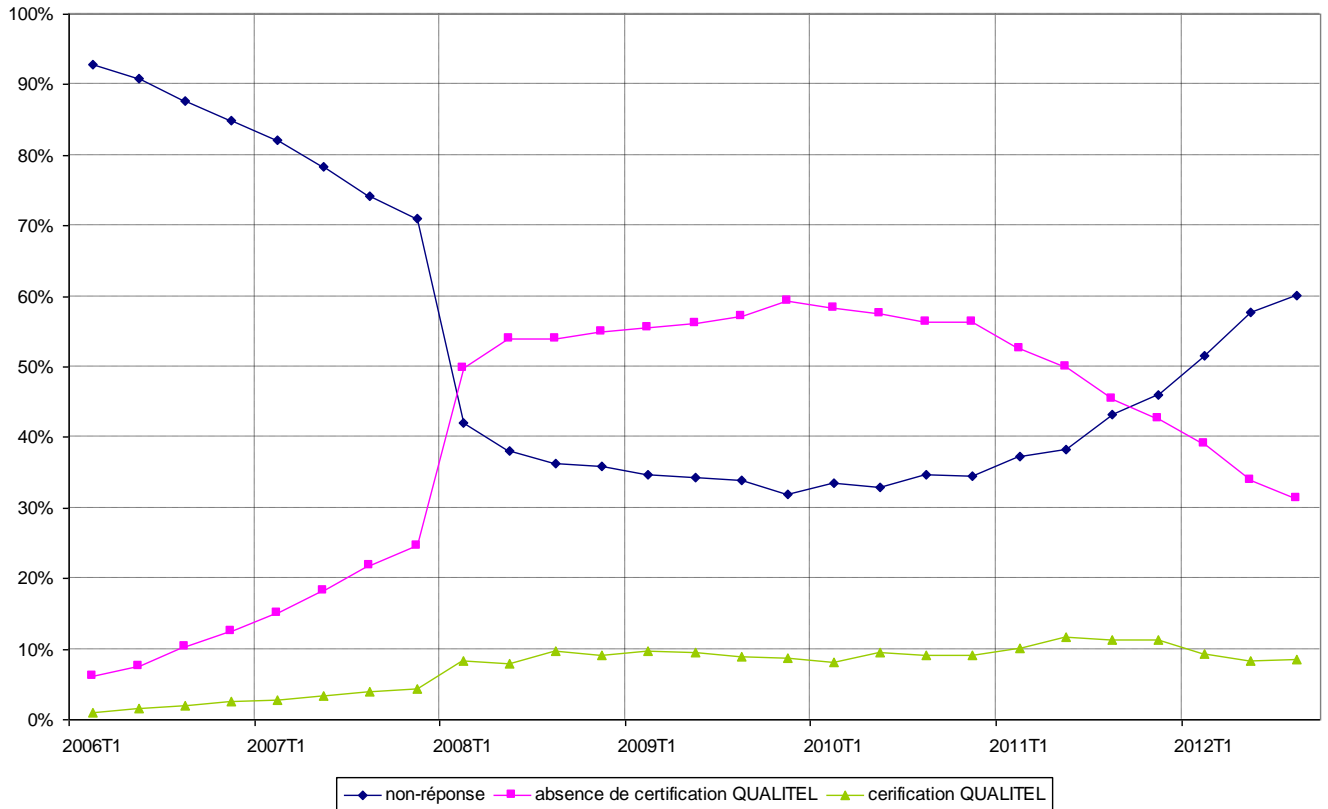
Classe géographique	ZEAT		TUU	
8	3	Nord	0	commune rurale
	3	Nord	3	10 000 à 19 999 habitants
	8	Centre-est	2	5 000 à 9 999 habitants
	<b>8</b>	<b>Centre-est</b>	<b>0</b>	<b>commune rurale</b>
	9	Méditerranée	5	50 000 à 99 999 habitants
	9	Méditerranée	7	200 000 à 1 999 999 habitants
9	1	Région parisienne	1	2 000 à 4 999 habitants
	8	Centre-est	1	2 000 à 4 999 habitants
	8	Centre-est	6	100 000 à 199 999 habitants
	9	Méditerranée	3	10 000 à 19 999 habitants
10	1	Région parisienne	5	50 000 à 99 999 habitants
	3	Nord	7	200 000 à 1 999 999 habitants
	8	Centre-est	4	20 000 à 49 999 habitants
11	1	Région parisienne	0	commune rurale
	<b>1</b>	<b>Région parisienne</b>	<b>2</b>	<b>5 000 à 9 999 habitants</b>
	<b>1</b>	<b>Région parisienne</b>	<b>3</b>	<b>10 000 à 19 999 habitants</b>
	1	Région parisienne	8	unité urbaine de Paris
12	3	Nord	4	20 000 à 49 999 habitants
	7	Sud-ouest	7	200 000 à 1 999 999 habitants
	8	Centre-est	7	200 000 à 1 999 999 habitants

**Annexe 23 : Les variables de confort du modèle (1)**

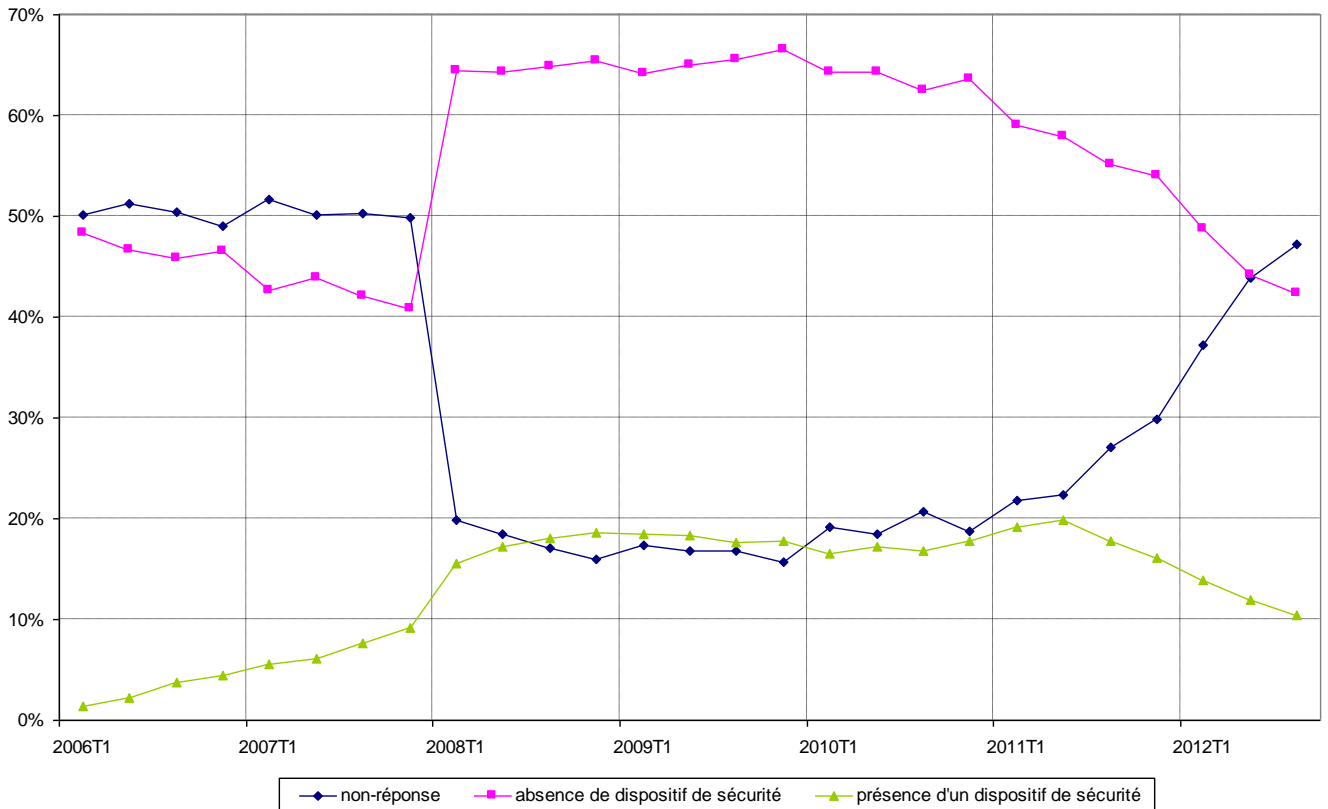
**Graphique 6.46 : La catégorie du programme de construction**



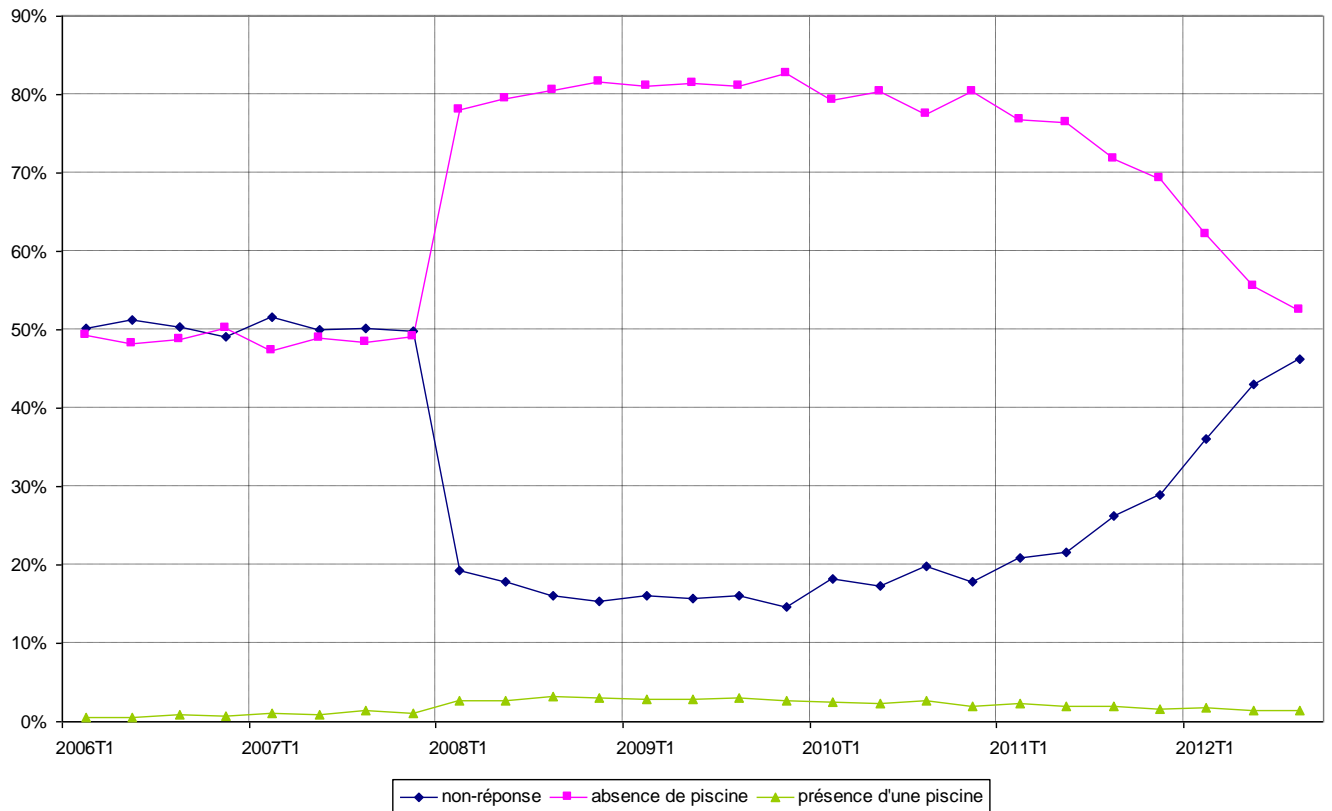
**Graphique 6.47 : Certification QUALITEL**



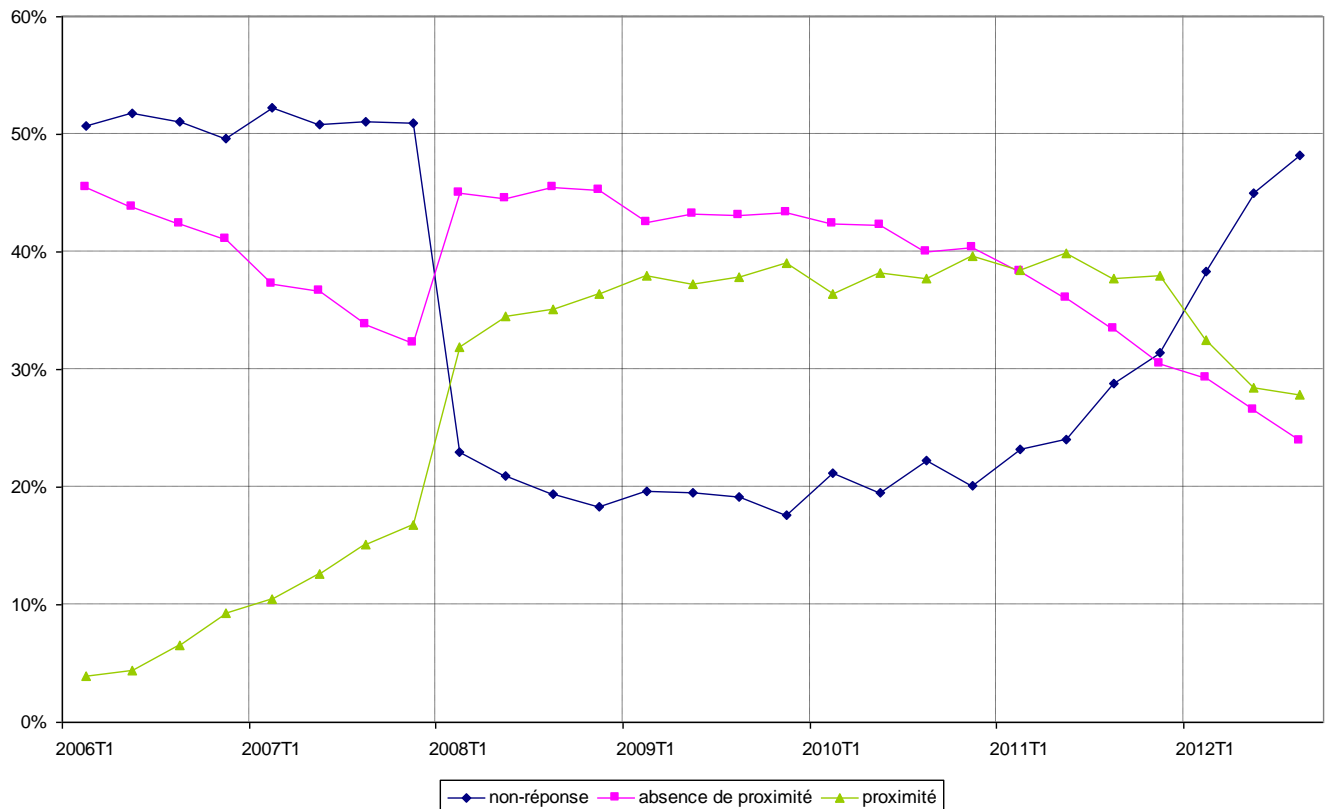
**Graphique 6.48 : La présence d'un dispositif de sécurité**



**Graphique 6.49 : La présence d'une piscine dans la résidence**

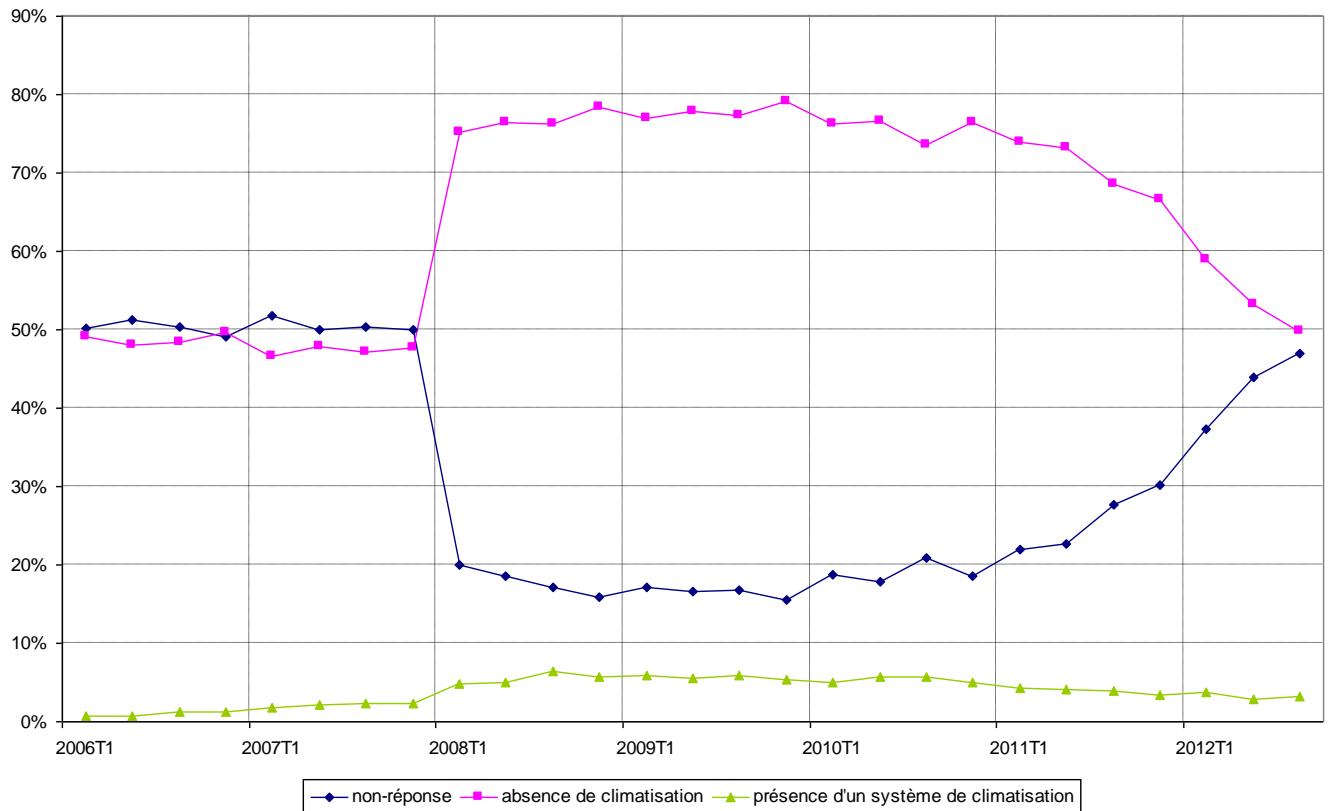


**Graphique 6.50 : La proximité d'un transport collectif par voie ferrée**

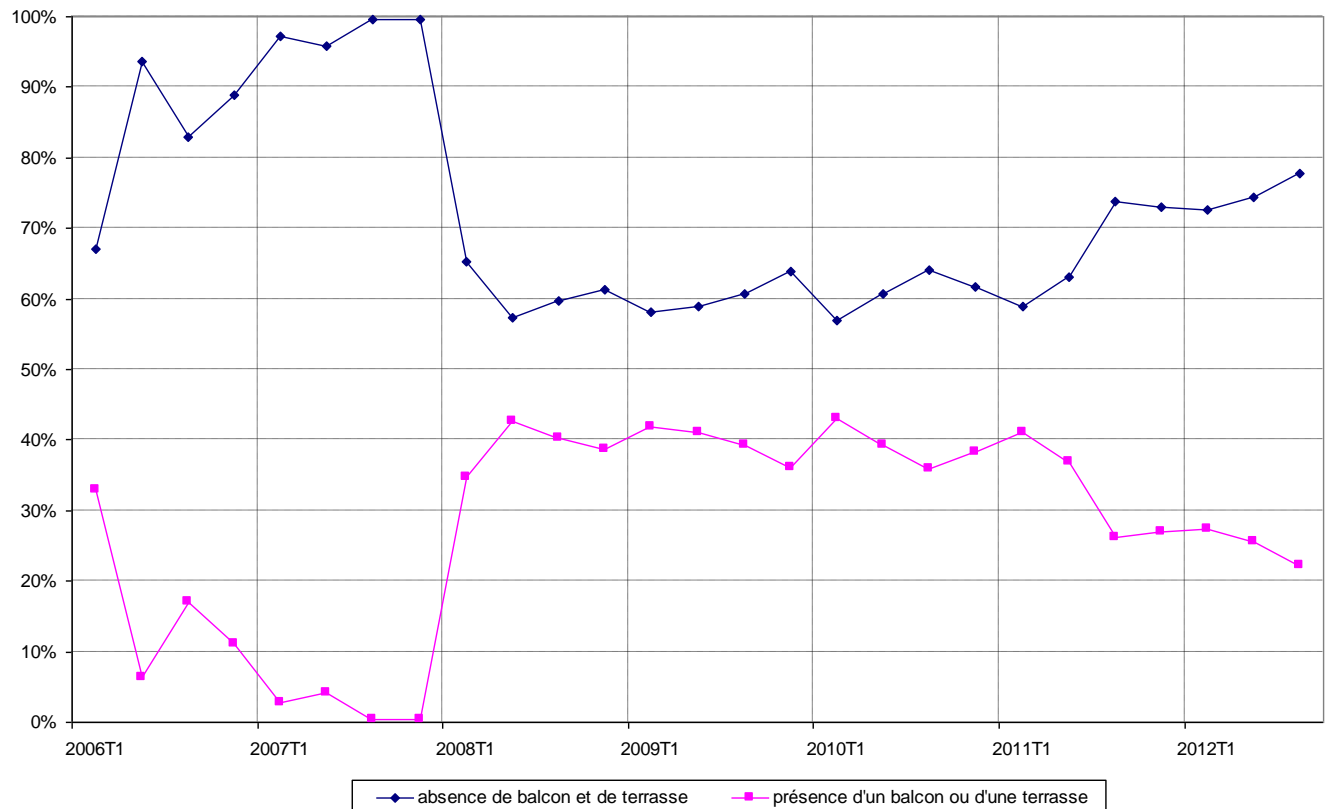




**Graphique 6.51 : La présence d'un système de climatisation**

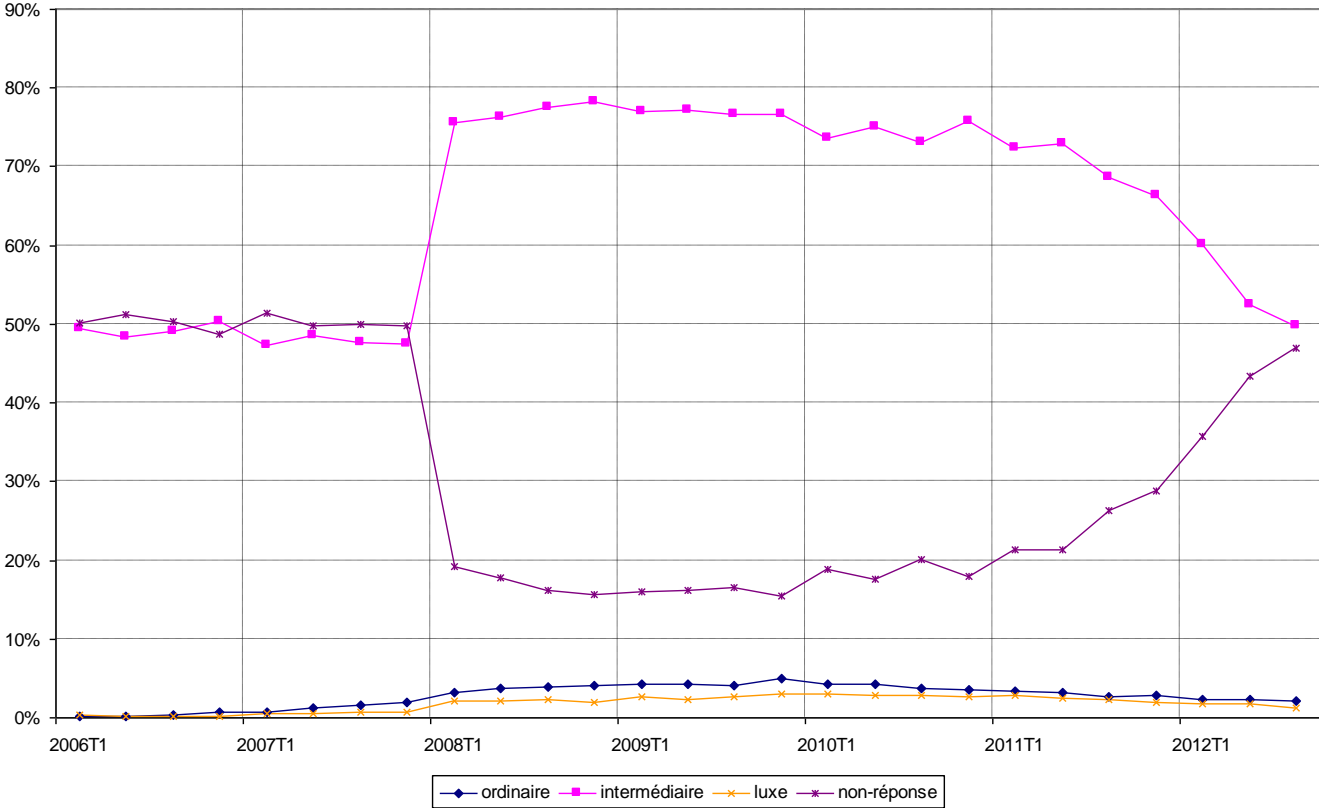


**Graphique 6.52 : La présence d'un balcon ou d'une terrasse pour les appartements**

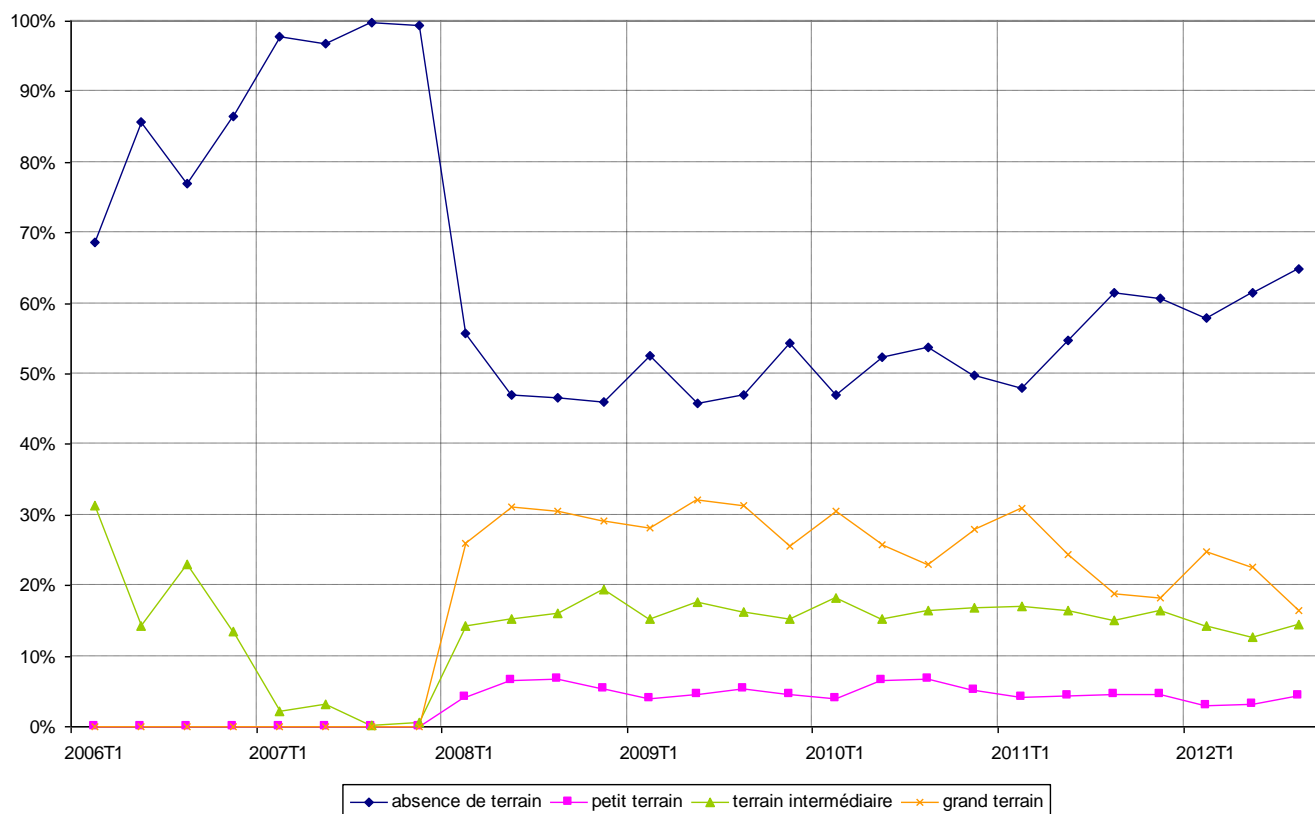


Annexe 24 : Les nouvelles variables apparaissant dans le modèle (1 bis)

Graphique 6.53 : La nouvelle variable « catégorie du programme »



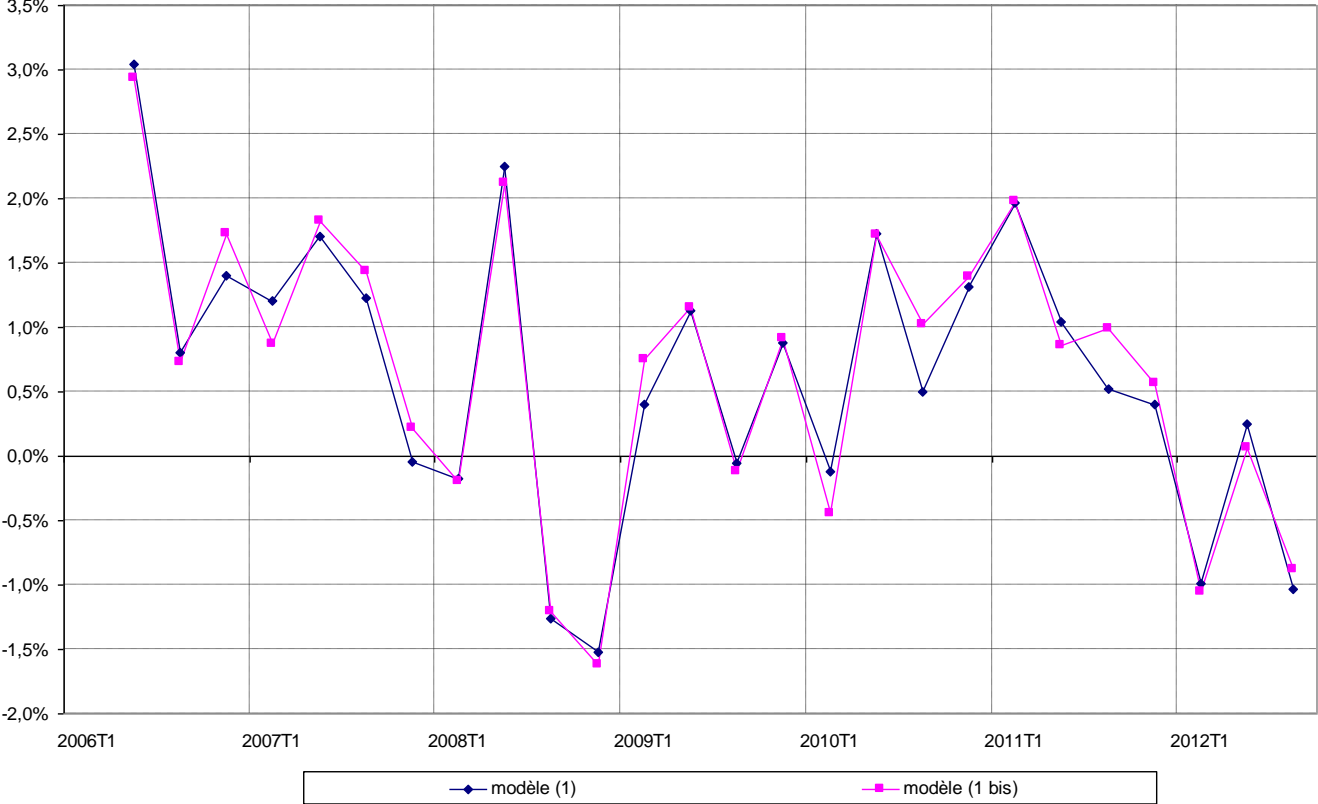
**Graphique 6.54 : La taille des terrains pour les maisons**



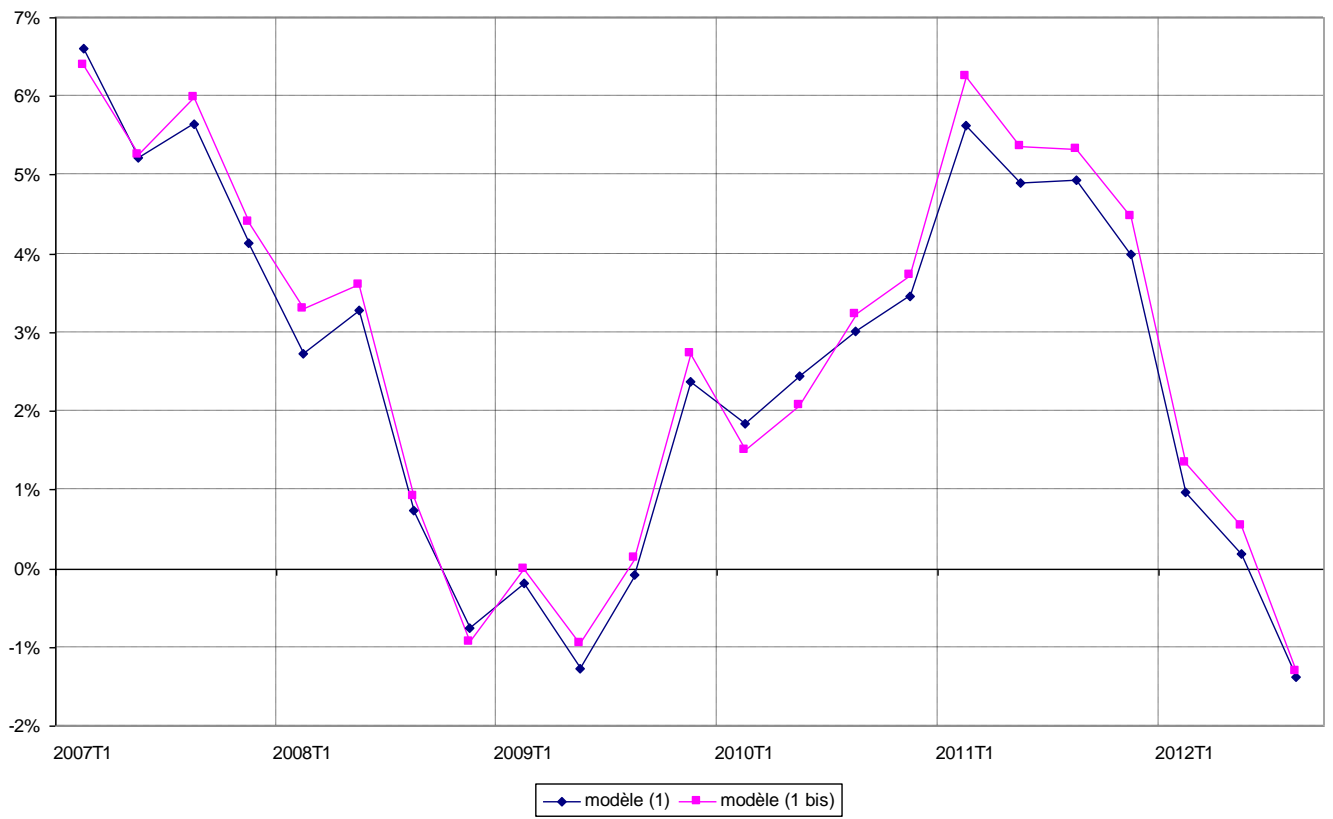
**Note :** un petit terrain correspond à une surface inférieure à 10 m<sup>2</sup> et un terrain intermédiaire à une surface comprise entre 10 et 200 m<sup>2</sup>

**Annexe 25 : Les taux de croissance en glissement trimestriel et en glissement annuel des indices obtenus à partir des modèles (1) et (1 bis)**

**Graphique 6.55 : Les taux de croissance en glissement trimestriel**



**Graphique 6.56 : Les taux de croissance en glissement annuel**



## Annexe 26 : La fiche méthodologique de l'indice du coût de la construction (ICC)

Organisme producteur : Insee

### Caractéristiques

- Correction : Non corrigé (brut)
- Période de référence de l'indice : 2010
- Périodicité : Trimestrielle
- Zone géographique : France métropolitaine

### Présentation synthétique

Le Ministère de l'écologie, du développement durable et de l'énergie mène une enquête « Prix de revient des logements neufs » auprès d'environ 1 600 permis de construire tirés dans la base Sit@del2. Les éléments tirés de cette enquête permettent à l'Insee d'en déduire, actuellement grâce à un modèle hédonique, un « indice du coût de la construction » à valeur réglementaire depuis 1953, et un indice de prix de production des logements neufs à usage d'habitation, pour un usage proprement statistique. Dans les deux cas, il s'agit en fait d'un indice de prix de production dans l'activité de construction de logements (partie de la classe d'activités « 41.20 - construction de bâtiments résidentiels et non résidentiels ») : les différences tiennent principalement dans le changement de la période de référence (1er trimestre de 1953 vs. année 2010) et dans la politique de révision des séries publiées.

### Définition

L'indice du coût de la construction ([ICC](#)) est un indice trimestriel, référence 100 au quatrième trimestre 1953, date de sa création. L'ICC mesure l'évolution du prix de construction des bâtiments neufs à usage principal d'habitation non communautaire en France métropolitaine. Il est publié au Journal officiel vers le milieu du quatrième mois suivant le trimestre sous revue. Il est calculé par l'Insee en collaboration avec le ministère de l'Écologie, du Développement durable et de l'Énergie. Bien que consacrée par l'usage, la dénomination « indice du coût de la construction » est impropre puisque c'est un indice de prix, fondé sur l'observation des marchés de construction conclus entre les maîtres d'ouvrage et les entreprises assurant les travaux de bâtiment, à l'exclusion des autres composantes qui entrent dans le prix de revient des logements (charge foncière, frais annexes de promotion, frais financiers, etc.). Le coût de la construction proprement dit est appréhendé par d'autres indicateurs, notamment les « [index BT](#) ».

### Méthode de calcul

Un indice de prix est un instrument de mesure de l'évolution des prix. Un tel objectif suppose d'isoler les variations pures de prix, en écartant tout effet lié à des modifications de contenu des produits en question, par exemple, dans ce cas précis, le renchérissement des logements consécutif à l'augmentation de la surface des pièces, à la mise en pratique de normes d'isolation plus contraignantes, à l'installation d'équipements plus perfectionnés. Le plus souvent, les indices de prix sont élaborés comme suit : on relève la valeur courante d'un échantillon intangible d'articles, puis on la rapporte à ce qu'elle était à une date de référence. Et lorsqu'un article est remplacé par un modèle

différent, on s'efforce d'évaluer un « effet qualité », c'est-à-dire l'incidence de ce changement sur le nouveau prix. On peut ainsi déterminer une variation pure de prix, en défalquant l'effet qualité. La problématique en matière de construction est particulière car chaque ouvrage est unique. Il n'est donc pas possible de comparer à différentes périodes le prix d'une même construction. Cela interdit de fonder la mesure de la variation des prix sur la simple observation de ceux-ci, et brouille la perception qu'on peut avoir de l'inflation dans le secteur de la construction. En effet, les prix dont il est ici question dépendent d'un grand nombre de paramètres : configuration de la construction, matériaux employés, équipements installés, sujétions de chantier, localisation et ampleur de l'opération, nature des rapports entre les intervenants (promoteurs, pavillonneurs, entreprises générales, sous-traitants, maîtres d'ouvrage), calendrier de réalisation, circonstances de la conclusion du marché. Le dispositif de calcul doit s'adapter à ces caractéristiques.

### **1. Jusqu'au 4<sup>e</sup> trimestre 2009 : la décomposition a posteriori (métrage)**

Le principe du calcul consiste, pour chaque opération de construction, à rapporter le prix du marché à un prix fictif obtenu en valorisant chaque élément de la construction par son prix à une date de référence. Ces prix de référence sont regroupés dans le bordereau général d'évaluation élaboré en 1987. La décomposition a posteriori est réalisée par des métreurs du ministère en charge de l'équipement.

Par sa méthode de calcul, l'indice s'apparente à un indice chaîne de type Paasche. Les pondérations associées aux différents types de logements et aux régions de construction reflètent la réalité du trimestre en cours. De ce fait, un indice de Paasche ne nécessite pas de « rebasage ». La méthode assure la prise en compte de l'évolution de la qualité des logements et la fiabilité à moyen terme de la mesure. L'interprétation des variations trimestrielles est rendue complexe du fait notamment de la longueur du processus de construction. Le calcul est effectué chaque trimestre à partir d'un échantillon représentatif retraçant l'évolution des prix de quelque 6 000 logements neufs au travers du suivi de 320 dossiers.

Les indices du troisième trimestre sont calculés économétriquement à partir des évolutions d'un indice de coût des facteurs de la construction (index BT01). Le pourcentage de variation de l'ICC entre le deuxième et troisième trimestre 2009 est donné par la formule : «  $ICC = 0,69953 \times x(t)$  », où  $x(t)$  désigne le pourcentage de variation de l'index BT01 sur la même période.

L'indice du coût de la construction du quatrième trimestre 2009 est calculé de façon semblable. Le pourcentage de variation de l'ICC entre le troisième et quatrième trimestre 2009 est donné par la formule : «  $ICC = 0,71978 \times x(t)$  », où  $x(t)$  désigne le pourcentage de variation de l'index BT01, entre juillet et octobre 2009.

### **2. À partir du 1<sup>er</sup> trimestre 2010 : la méthodologie hédonique**

La méthode hédonique consiste à établir une relation, à l'aide d'un modèle économétrique, entre le prix du marché de construction et les caractéristiques de l'ouvrage, ce qui permet d'évaluer la valeur implicite de ces différentes caractéristiques (surface, nombre d'étages, niveau de confort, localisation géographique et niveau de vie du lieu de construction par exemple). L'évolution temporelle du prix est captée dans le modèle à l'aide de variables indicatrices de la date. Cette méthode est déjà employée à l'Insee, notamment pour l'indice des prix des logements anciens. Elle assure la prise en compte de l'évolution de la qualité des logements.

Le champ couvert par l'ICC est très large. Il comprend les trois grands types de construction à usage d'habitation : l'individuel pur, l'individuel groupé et les logements collectifs. Pour rendre compte de l'hétérogénéité des constructions, un modèle hédonique est défini pour chaque type de construction. Les coefficients des modèles sont ré-estimés chaque trimestre ; les modèles eux-mêmes sont revus périodiquement. L'ICC hédonique pour l'ensemble de la construction de logements neufs résulte de l'agrégation des trois sous-indices proportionnellement à l'importance de chaque type de construction.

Les données nécessaires au calcul de l'ICC proviennent de l'enquête statistique sur « l'indice du coût de la construction et le prix de revient des logements neufs » (ICC-PRLN), réalisée par le service statistique du ministère de l'Écologie, du Développement durable et de l'Énergie. Cette enquête permet de retracer l'évolution des prix de la construction à travers le suivi de 500 dossiers représentant, selon les trimestres, 7 000 à 8 000 logements neufs.

## Utilisations

L'Indice du coût de la construction est notamment utilisé pour réviser les loyers des baux commerciaux (code de commerce loi n° 2001-1168 du 11 décembre 2001 art. 33 V, Journal officiel du 12 décembre 2001, articles L145-33 et L145-34). Selon la Loi n° 2011-525 du 17 mai 2011 de simplification et d'amélioration de la qualité du droit (article 63), trois indices peuvent être utilisés pour la révision des baux professionnels :

- [l'indice du coût de la construction](#) ou, s'ils sont applicables...
- [l'indice des loyers commerciaux](#) (activités commerciales et artisanales);
- [l'indice des loyers des activités tertiaires](#) (autres que commerciales).

Le décret n° 2008-1139 du 4 novembre 2008 relatif à l'indice des loyers commerciaux et le décret n°2011-2028 du 29 décembre 2011 relatif à l'indice des loyers des activités tertiaires définissent les activités concernées ainsi que les modalités de calcul et de publication de ces indices.

L'indice du coût de la construction, l'indice des loyers commerciaux et l'indice des loyers des activités tertiaires font chacun l'objet d'une publication au Journal Officiel. C'est la date de publication de l'indice au Journal Officiel qui sert de référence.

Les évolutions de l'ICC peuvent être comparées ou rapprochées de celles d'indicateurs de coûts des facteurs de production, comme le [BT01](#).

L'ICC donne également lieu pour des fins statistiques à l'[indice des prix de production de la construction neuve à usage d'habitation - référence 100 en 2010](#), notamment utilisé comme déflateur des comptes nationaux de la construction, à côté des indices de prix de l'entretien et de l'amélioration du logement ([IPEA](#)), élaborés désormais chaque trimestre par l'Insee et publiés dans la collection « Informations Rapides ».

## Bibliographie

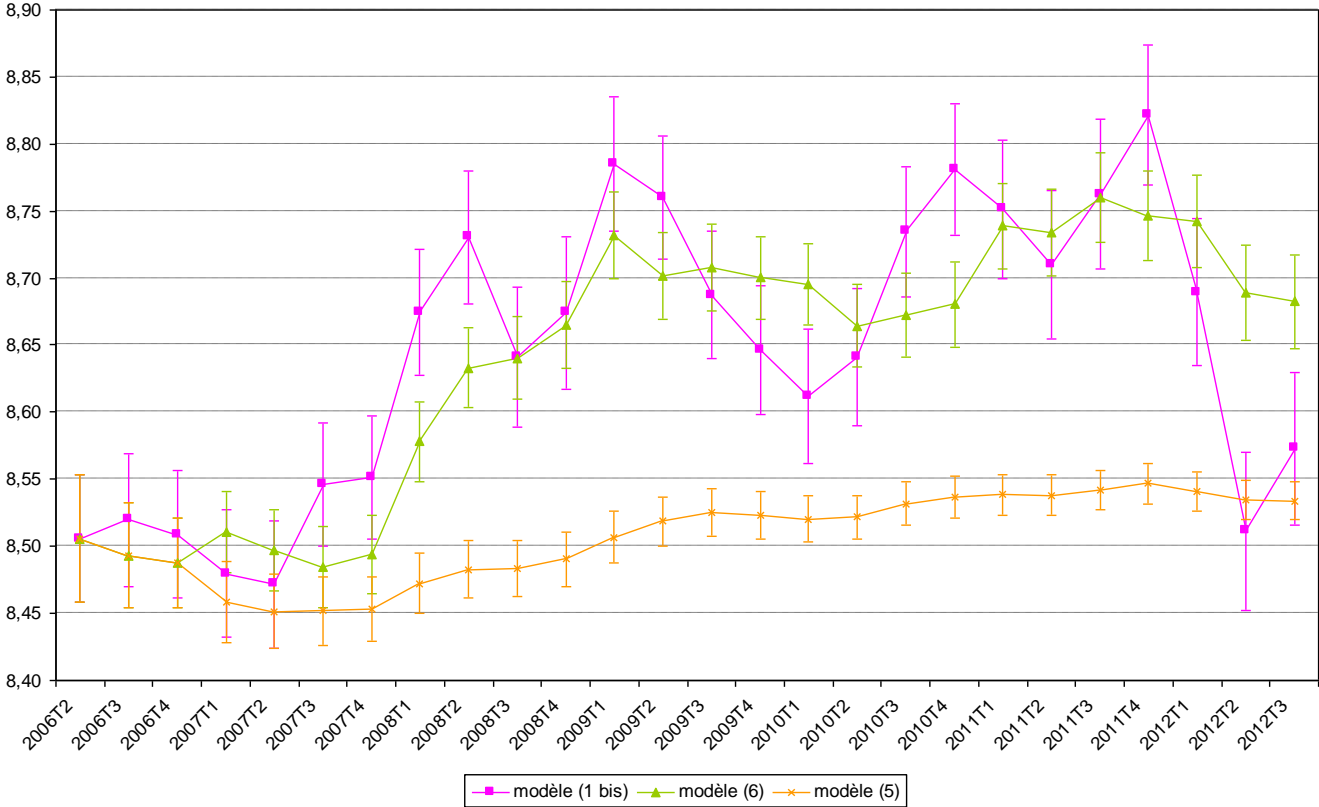
- Bulletin Mensuel de la Statistique - supplément Janvier/Mars 1954 - Note méthodologique : Un nouvel indice du coût de la construction des immeubles à usage d'habitation.
- Protocole d'accord entre le Ministère de l'Économie et le Ministère de l'Environnement et du Cadre de Vie concernant l'indice du coût de la construction - Fascicule n° 42 (5 novembre 1980) texte n° 974 du Bulletin Officiel du Ministère de l'Environnement ;



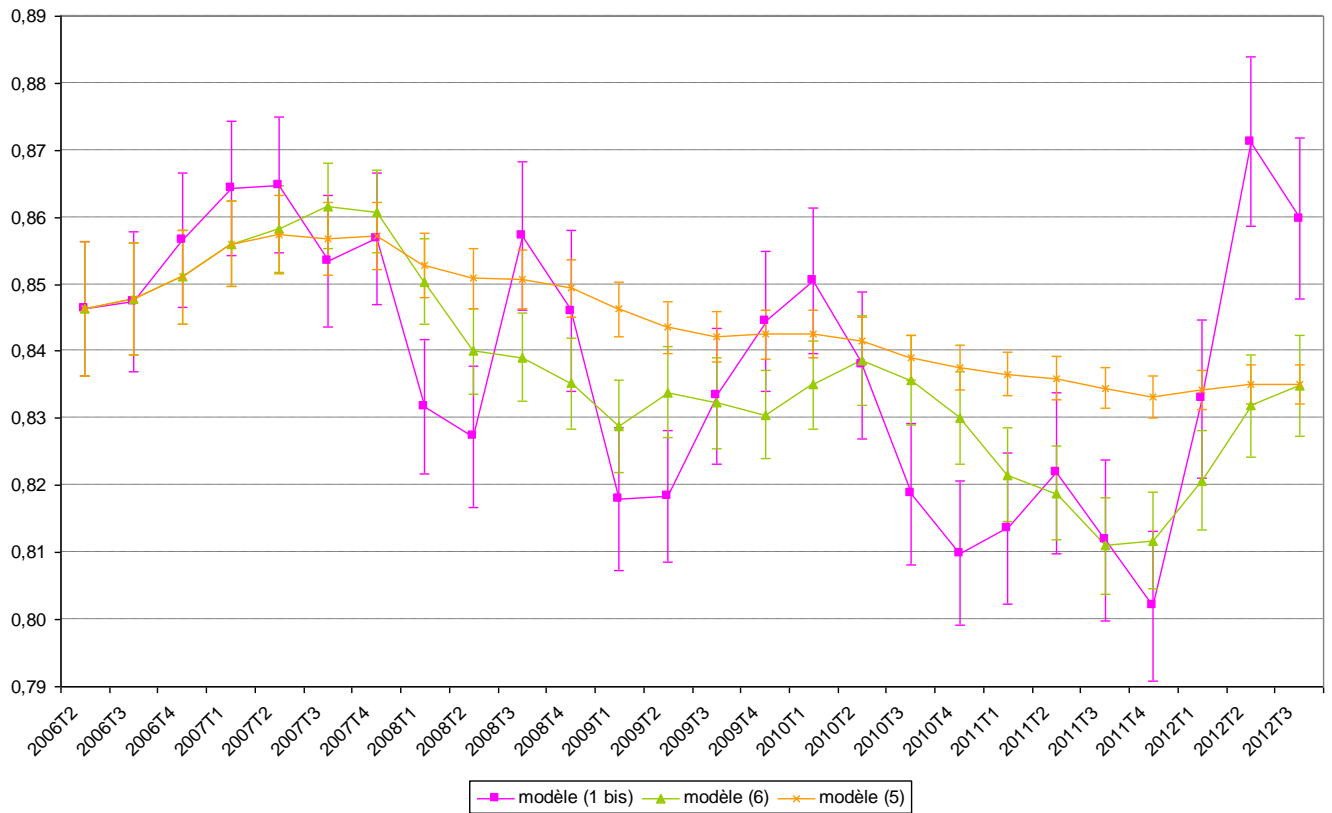
- Les comptes du logement en base 1971, par R. Brocard et C. Hesse - Collections de l'Insee n° C43 - 1975 ;
- [Les indices du coût de la construction](#), par J. Guet - Économie et Statistique n° 2 - juin 1969 ;
- [L'indice du coût de la construction](#), par J. Lienhardt et J. Rouchet - Courrier des Statistiques n° 17 - janvier 1981 ;
- [L'ICC, indice du coût de la construction](#), par L. Cases - Courrier des Statistiques n° 108 - décembre 2003 ;

**Annexe 27 : Les valeurs des estimateurs des paramètres et les intervalles de confiance associés au niveau 5% pour les modèles (1 bis), (5) et (6)**

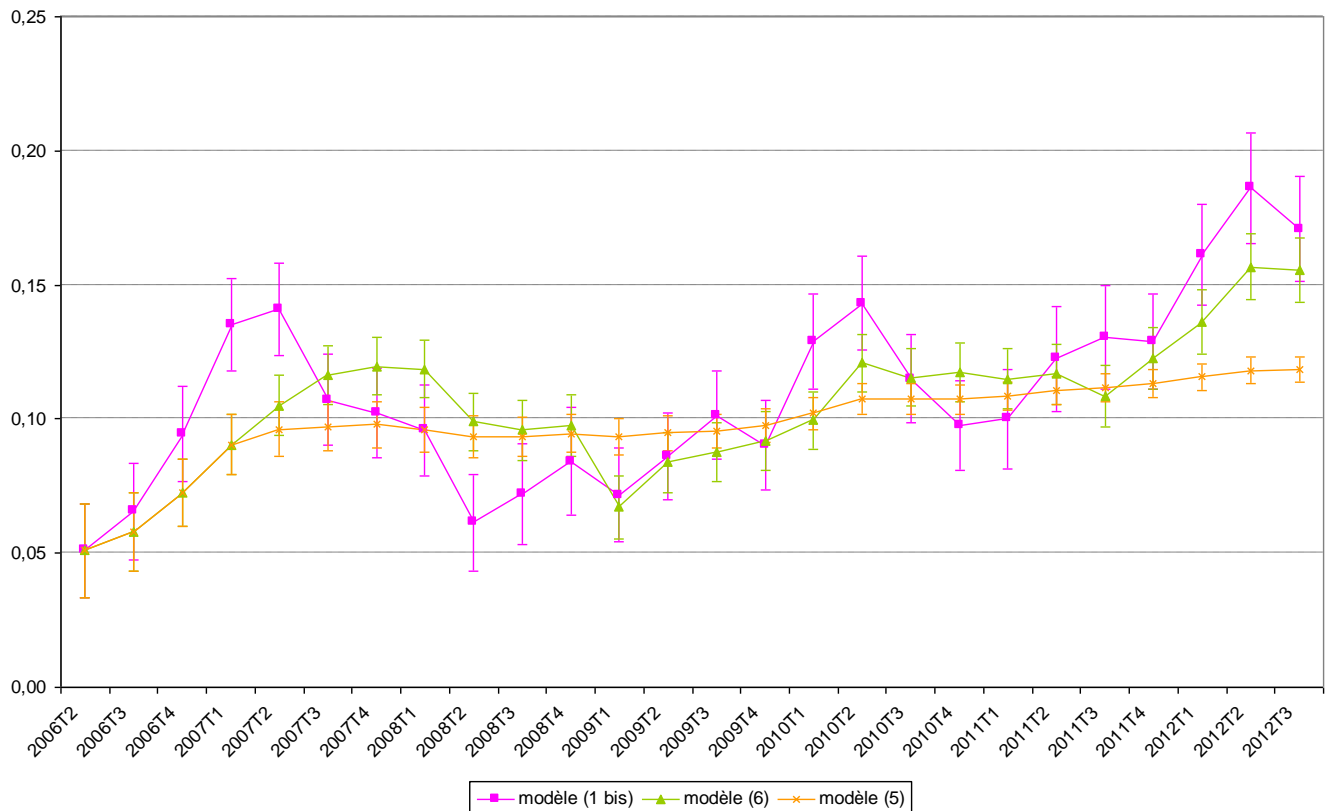
**Graphique 6.57 : La constante**



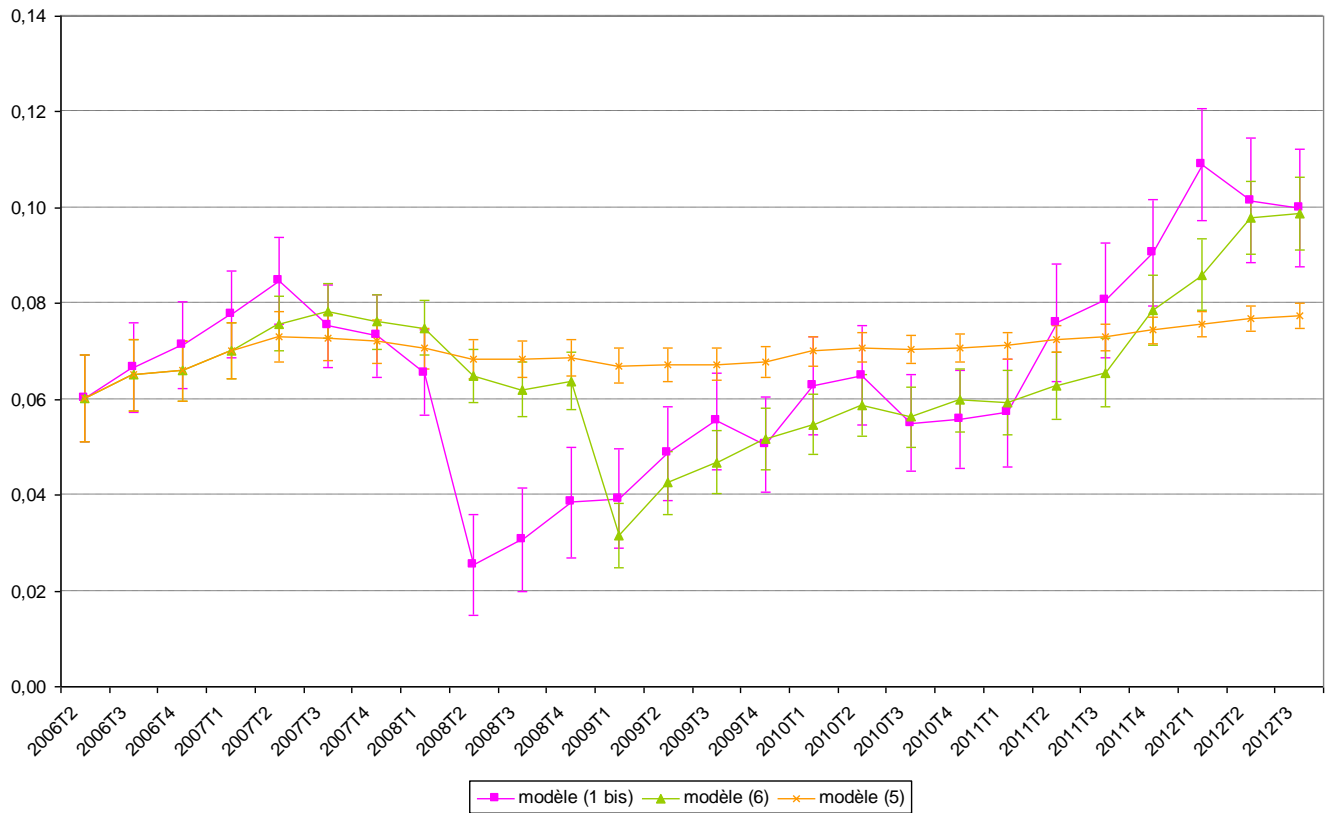
**Graphique 6.58 : La surface habitable moyenne en m<sup>2</sup> du logement**



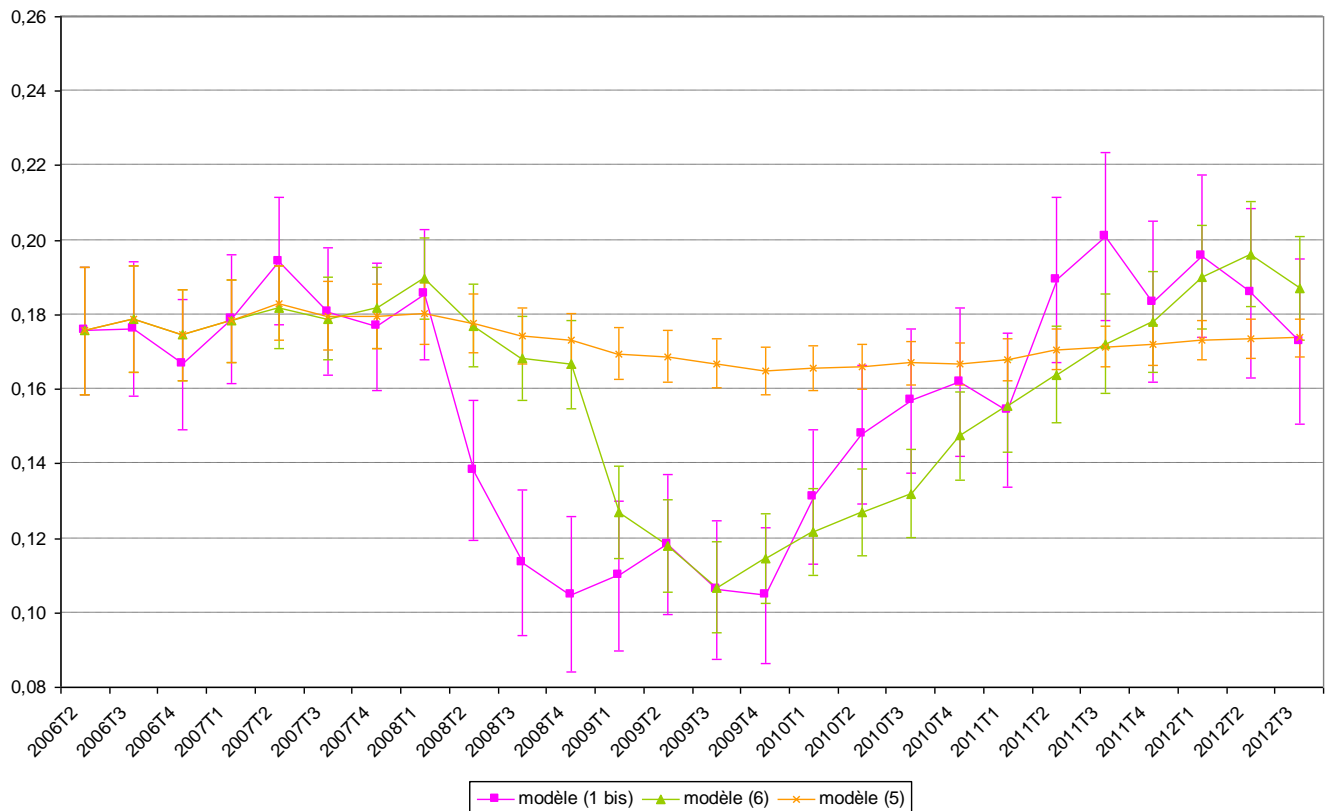
**Graphique 6.59 : Les appartements d'une pièce**



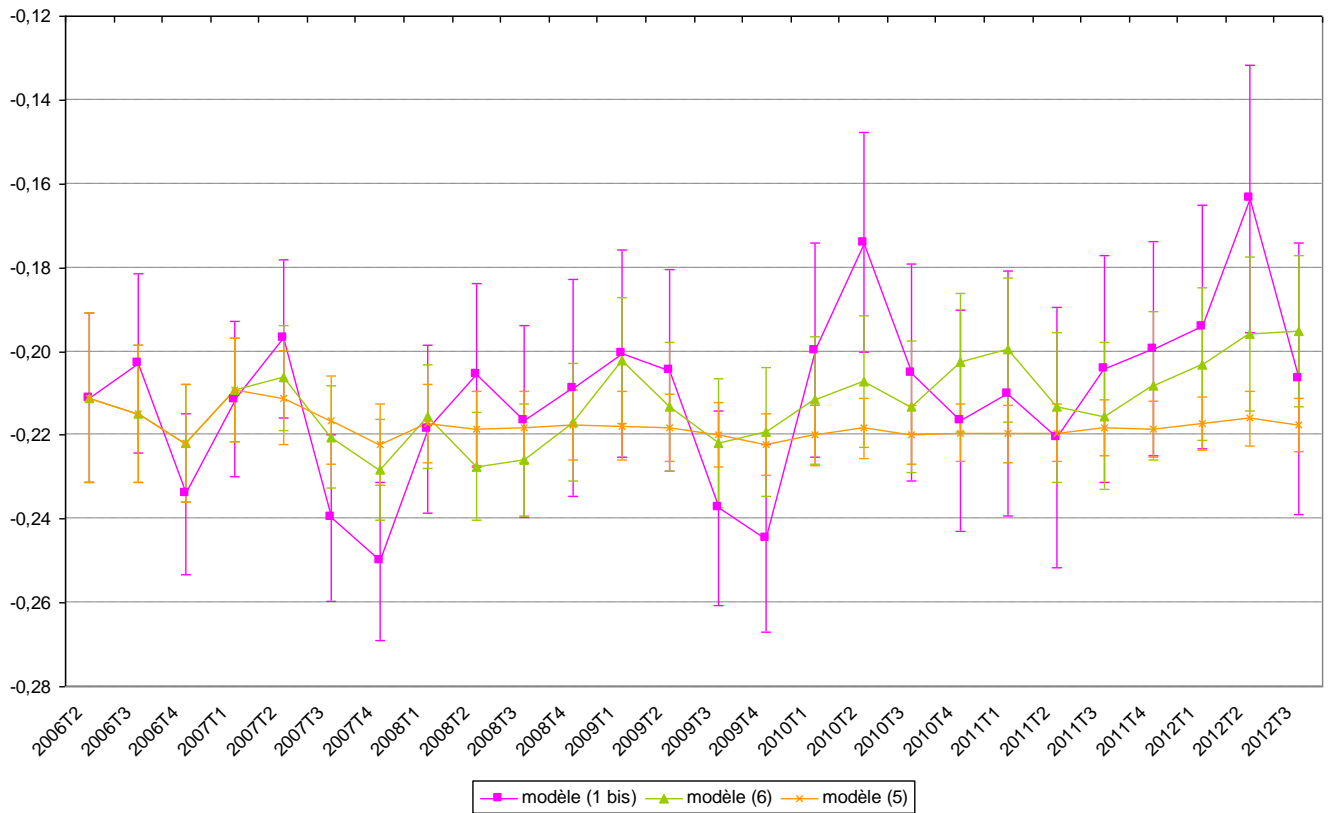
**Graphique 6.60 : Les appartements de 2 à 4 pièces**



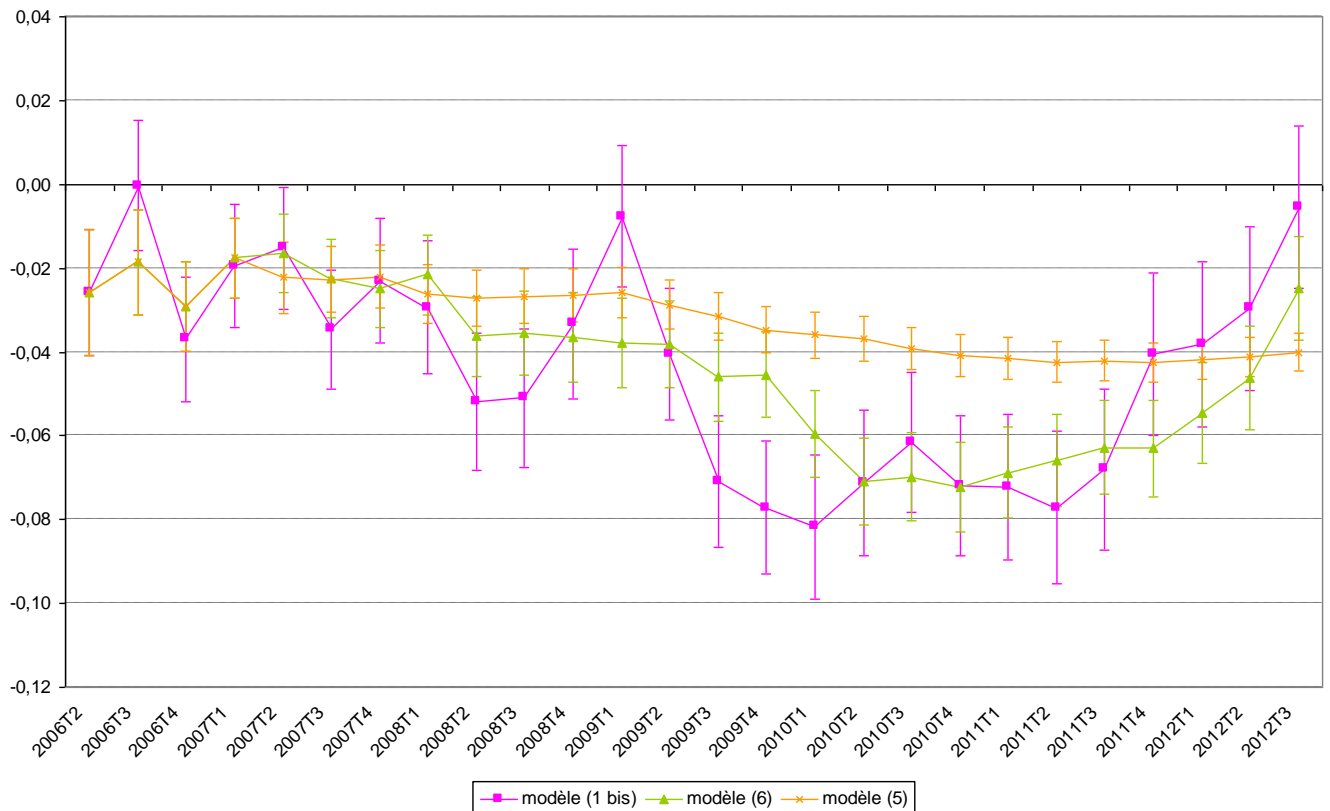
**Graphique 6.61 : Les appartements de plus de 5 pièces**



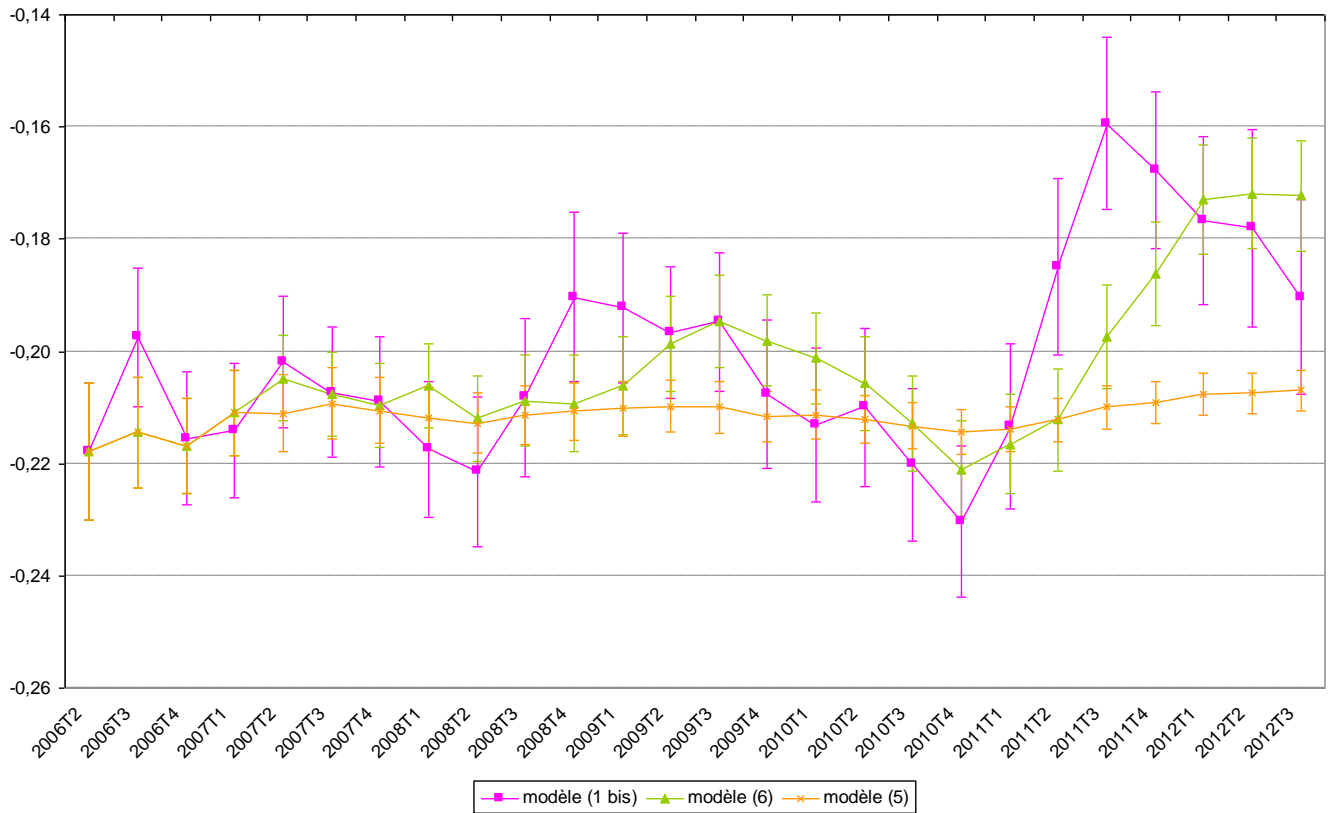
**Graphique 6.62 : Première classe géographique**



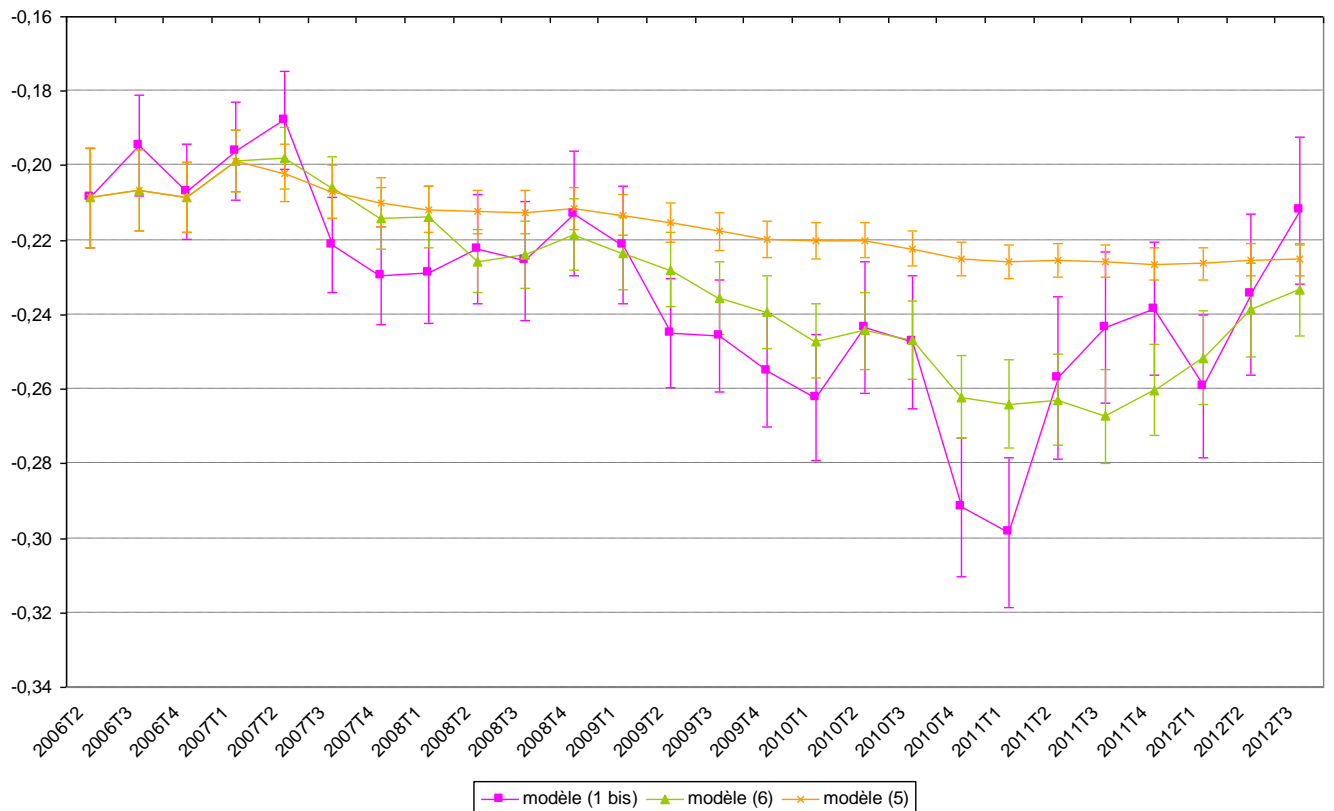
**Graphique 6.63 : Deuxième classe géographique**



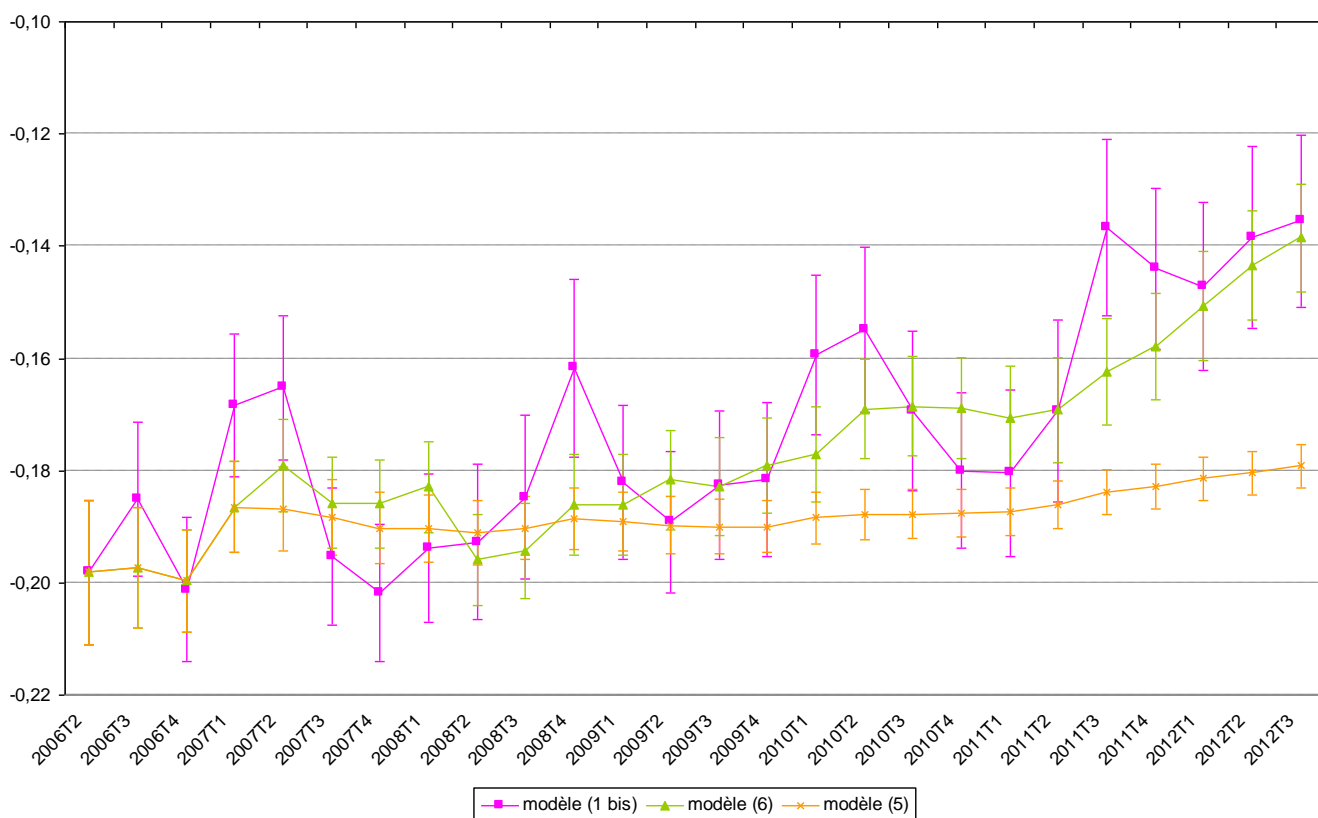
**Graphique 6.64 : Troisième classe géographique**



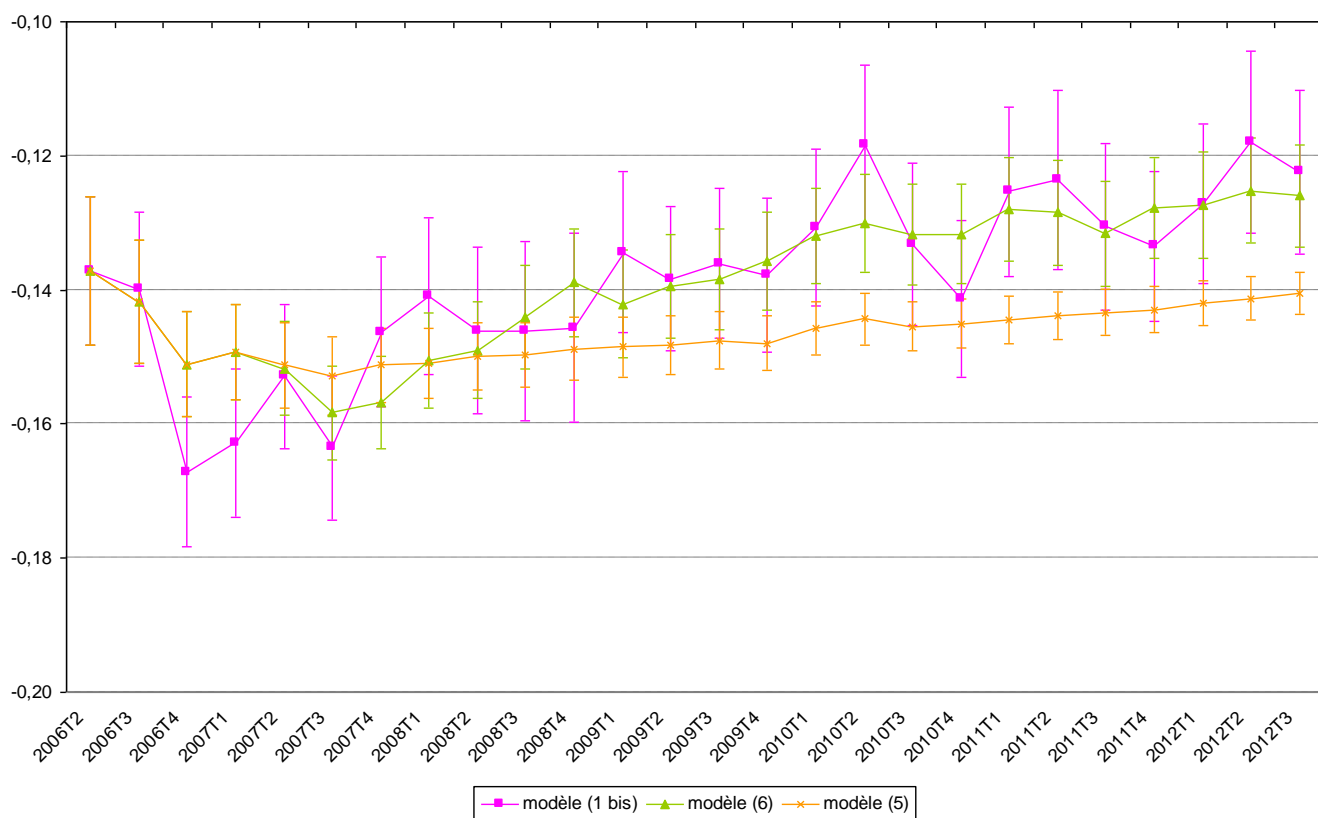
**Graphique 6.65 : Quatrième classe géographique**



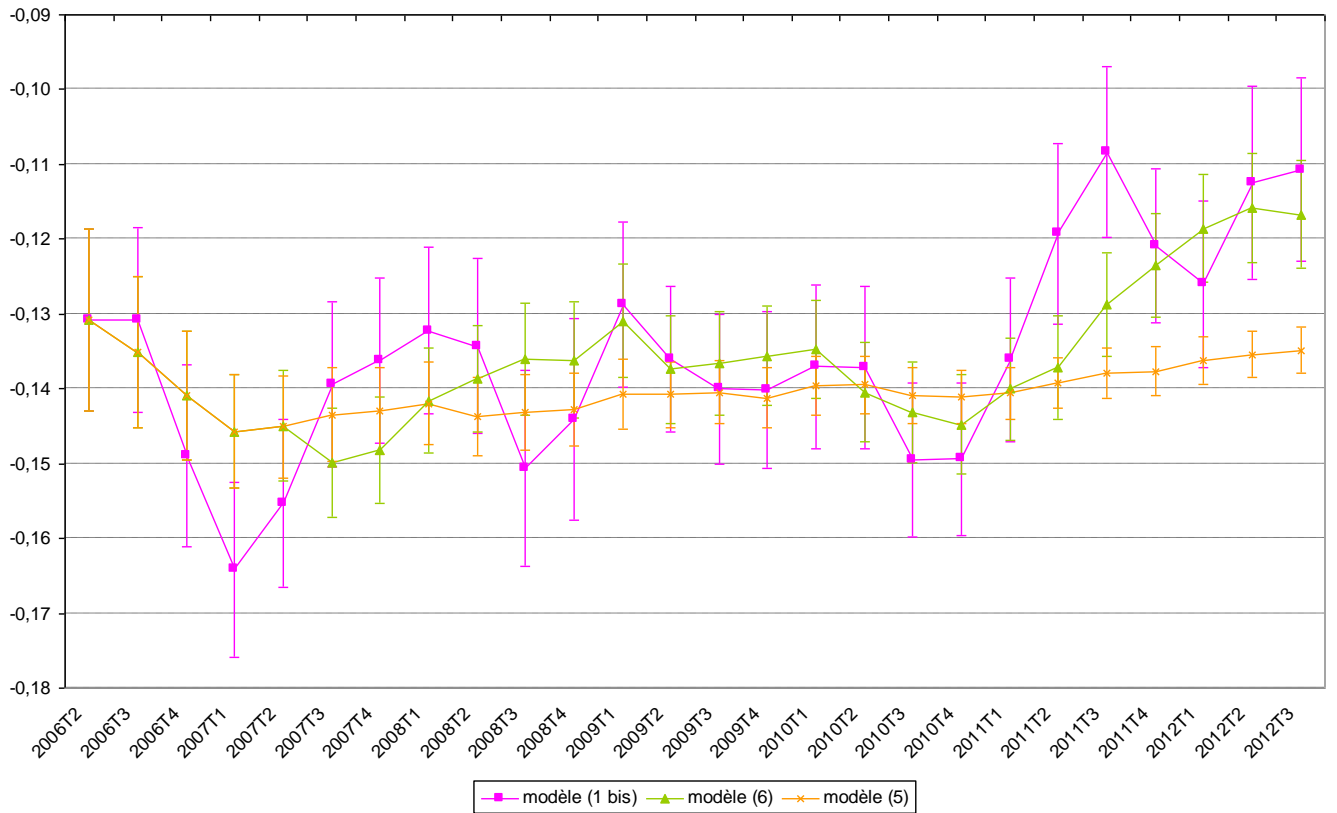
**Graphique 6.66 : Cinquième classe géographique**



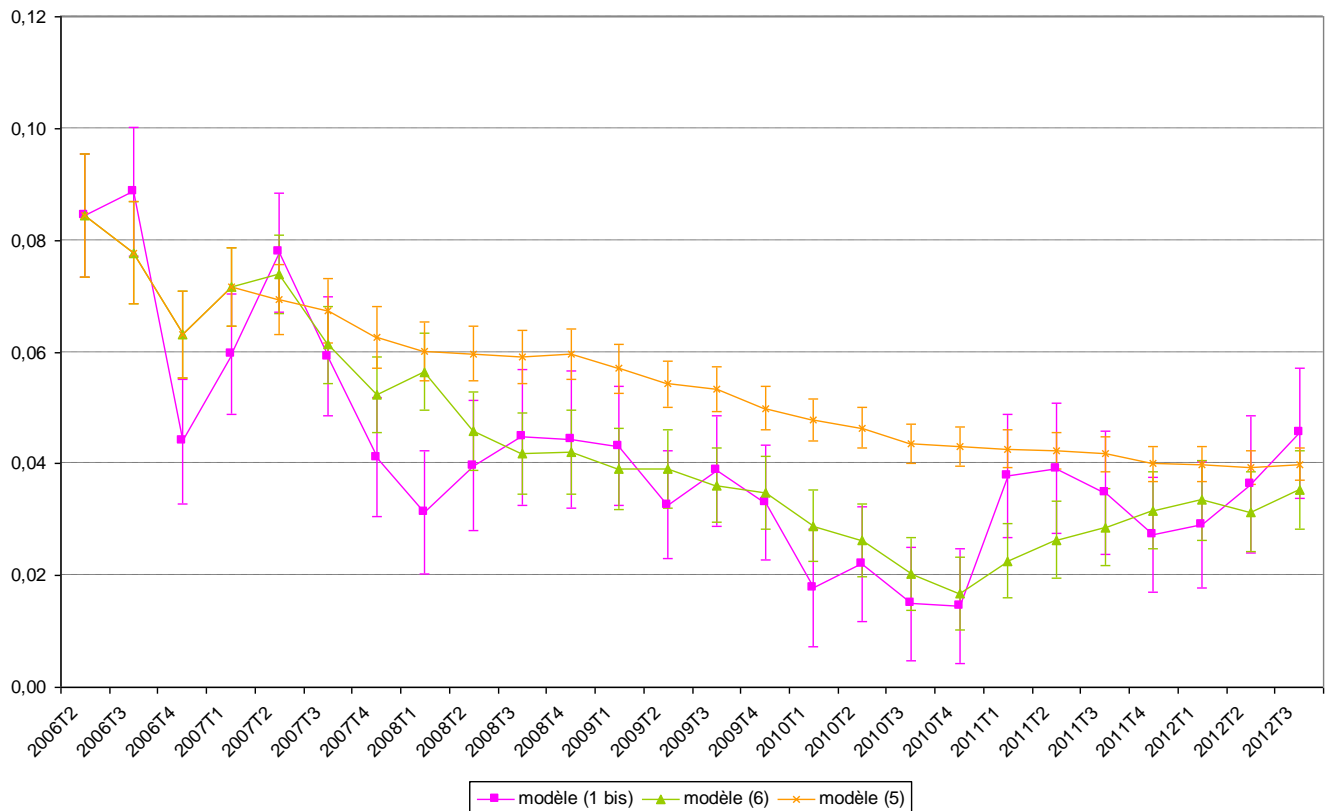
**Graphique 6.67 : Sixième classe géographique**



**Graphique 6.68 : Septième classe géographique**



**Graphique 6.69 : Huitième classe géographique**

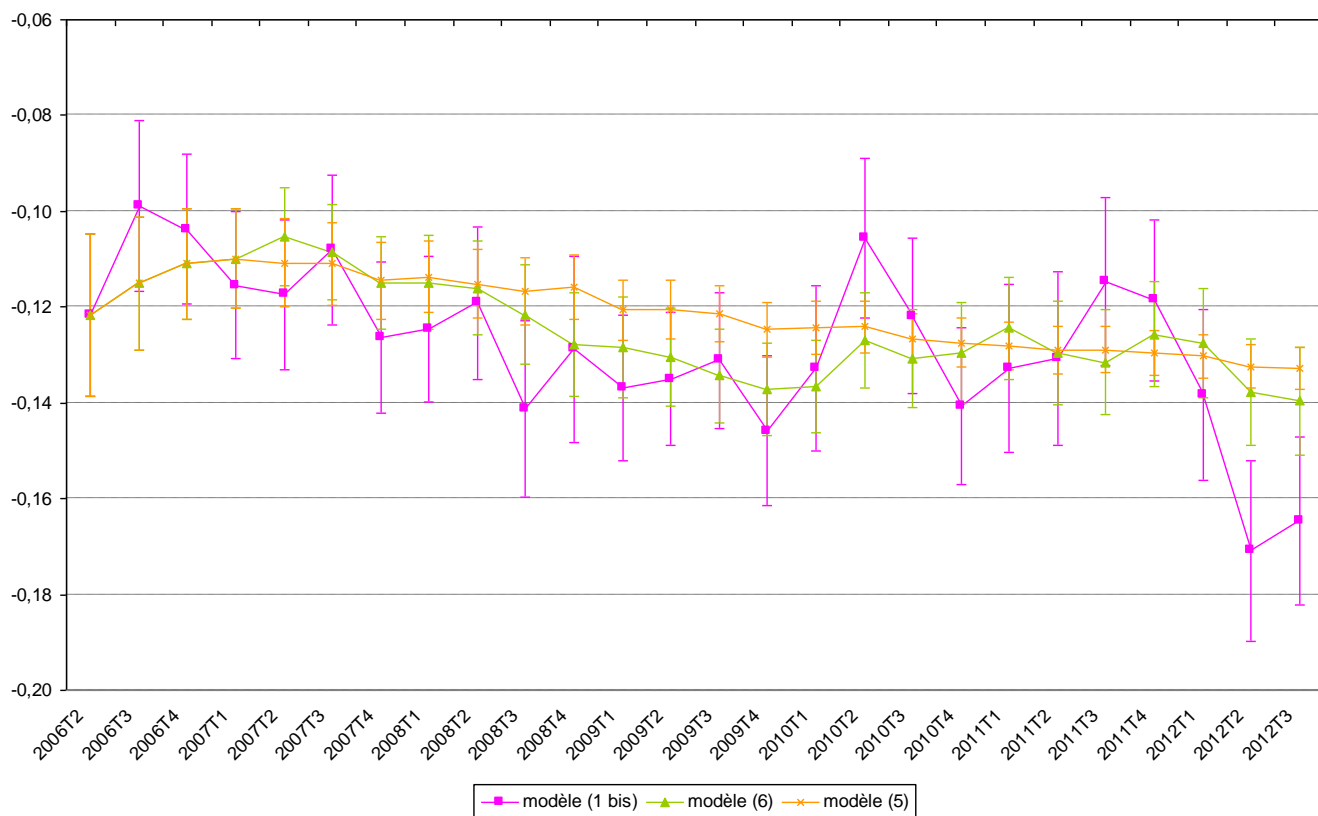




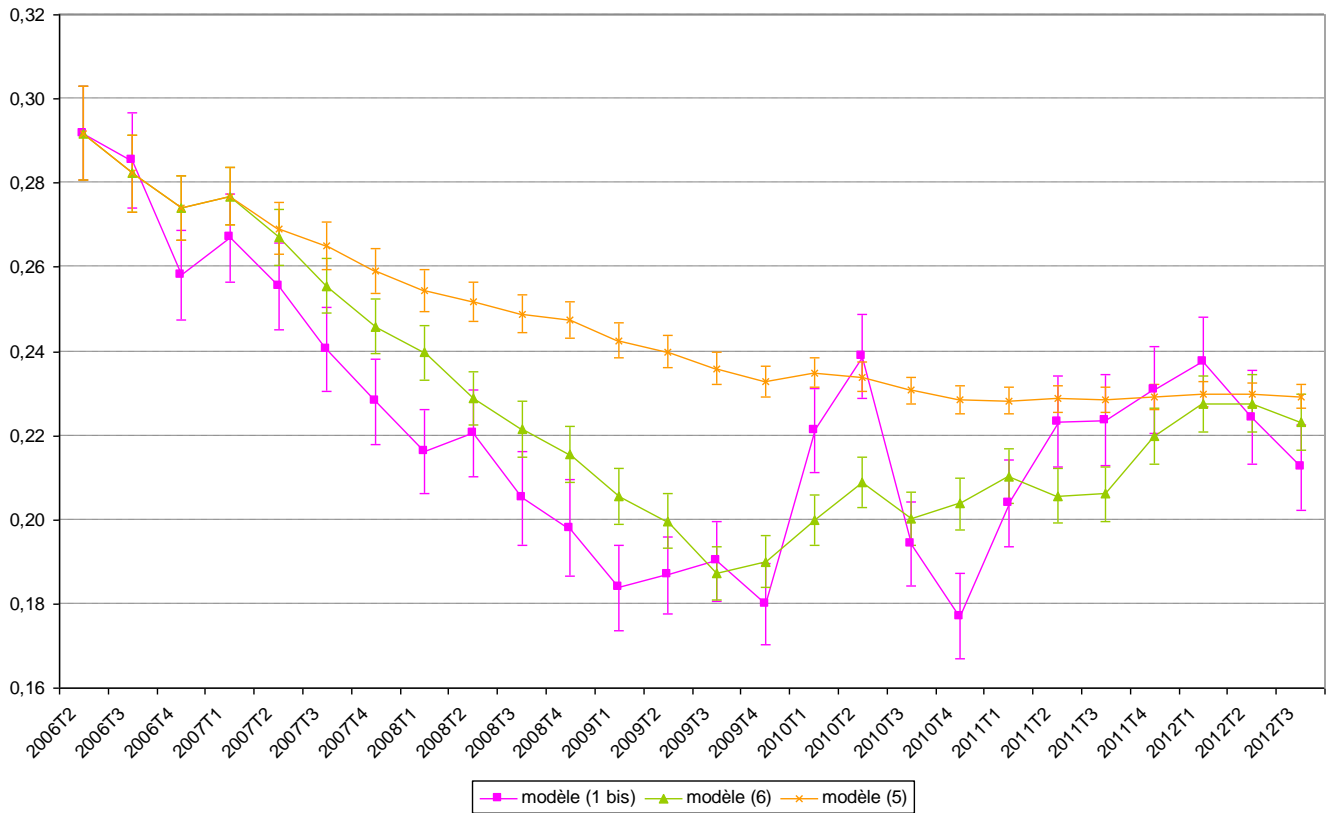
**Graphique 6.70 : Neuvième classe géographique**



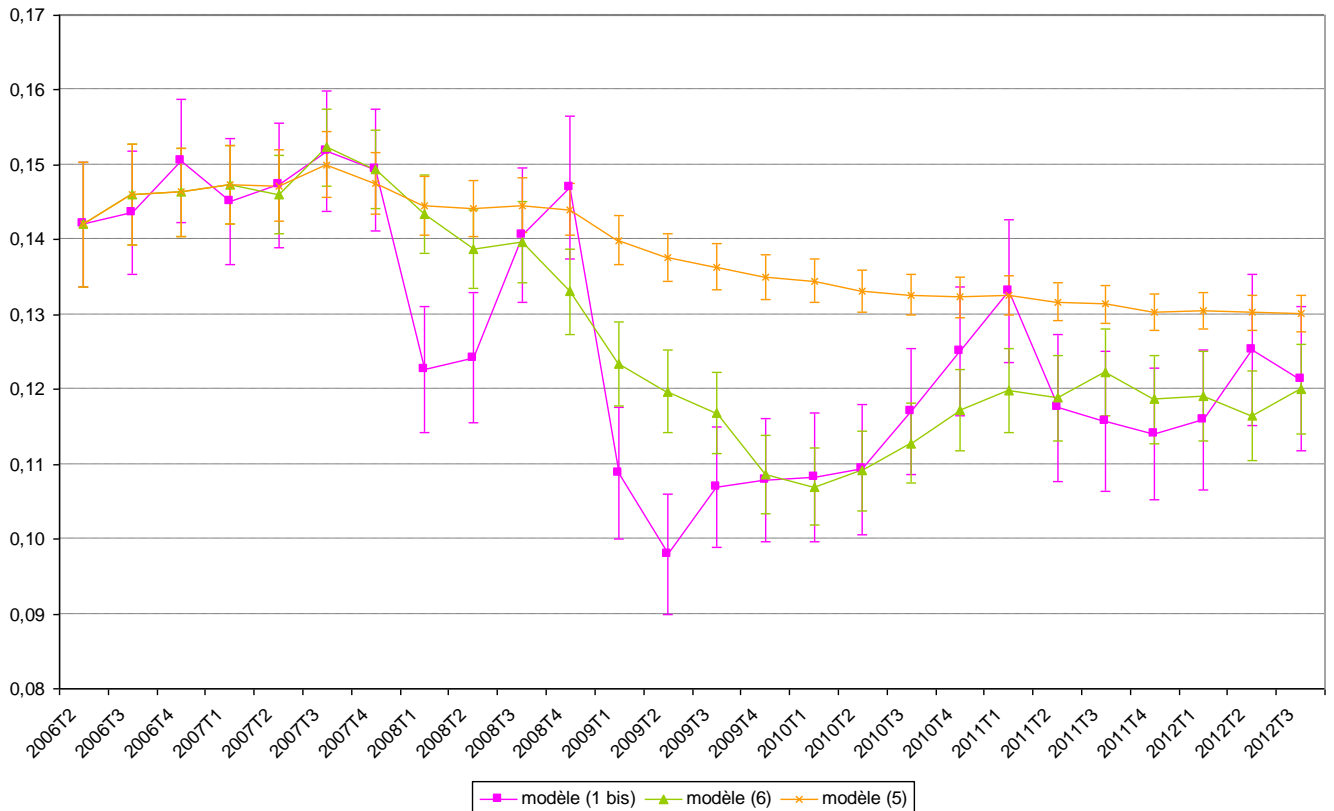
**Graphique 6.71 : Dixième classe géographique**



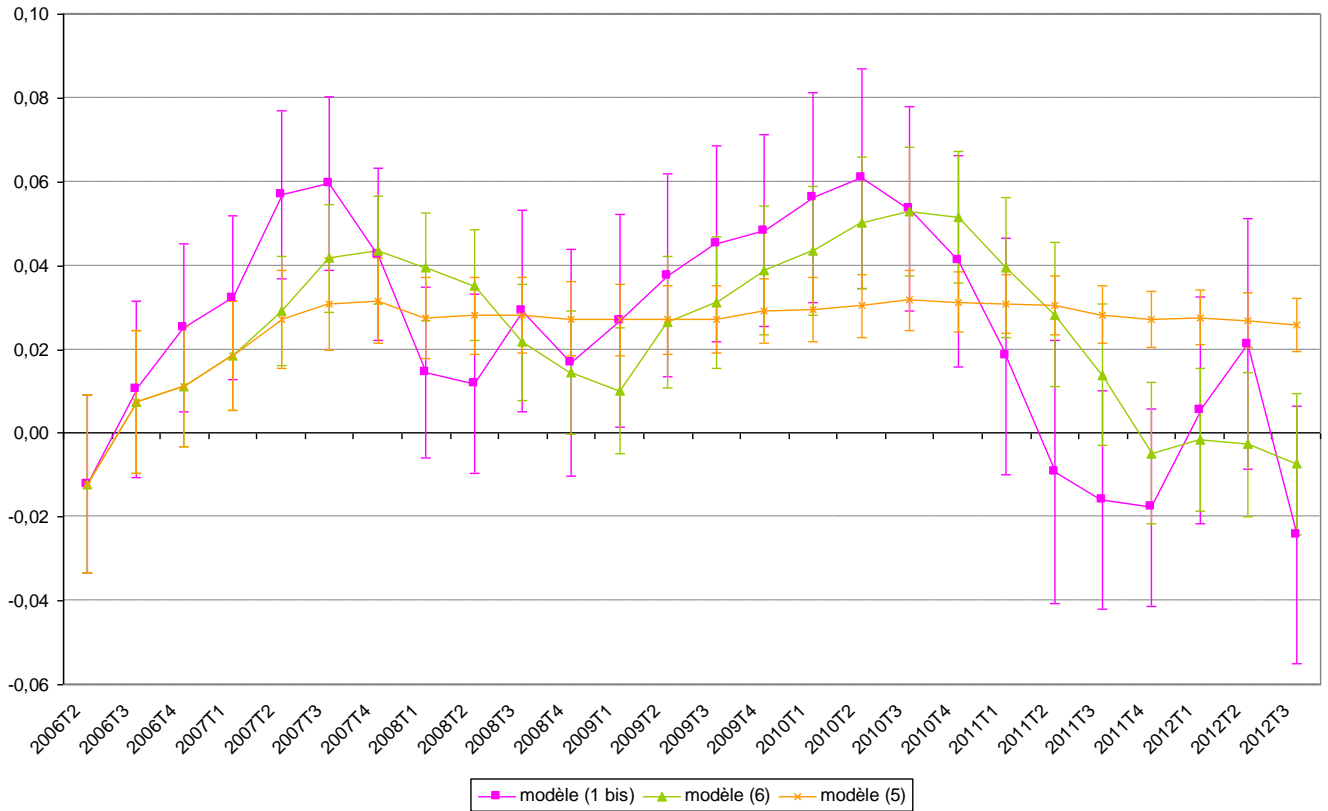
**Graphique 6.72 : Onzième classe géographique**



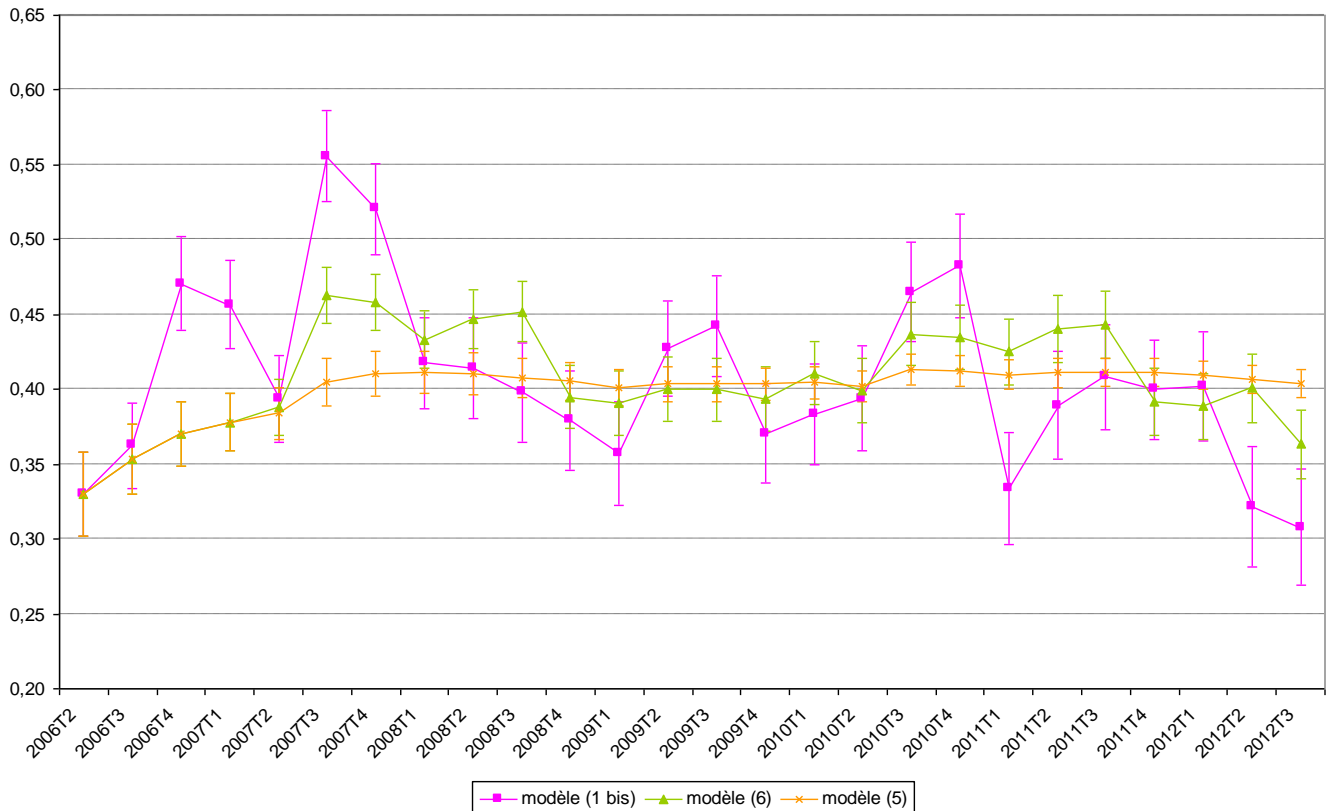
**Graphique 6.73 : Les communes maritimes**



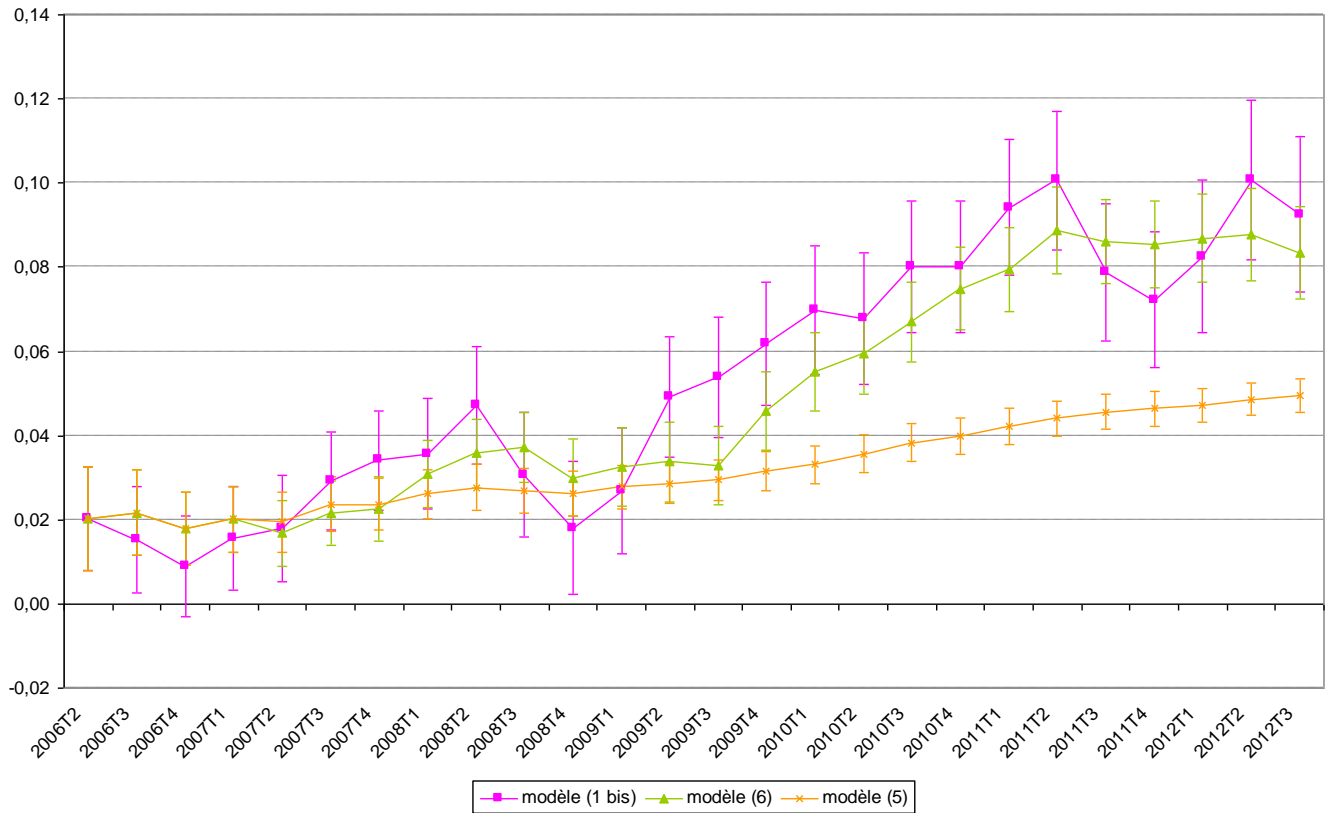
**Graphique 6.74 : L'arrière-pays littoral**



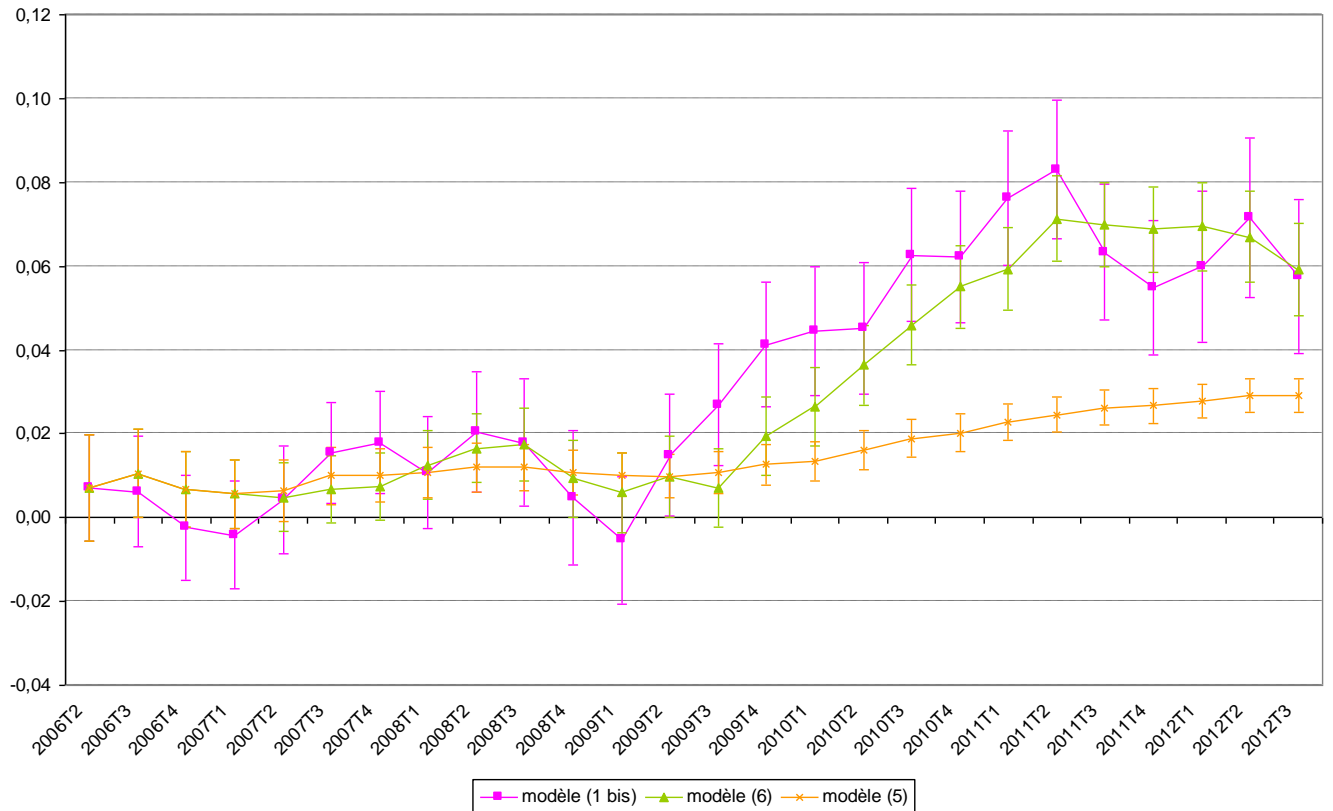
**Graphique 6.75 : Les stations de sport d'hiver et d'alpinisme**



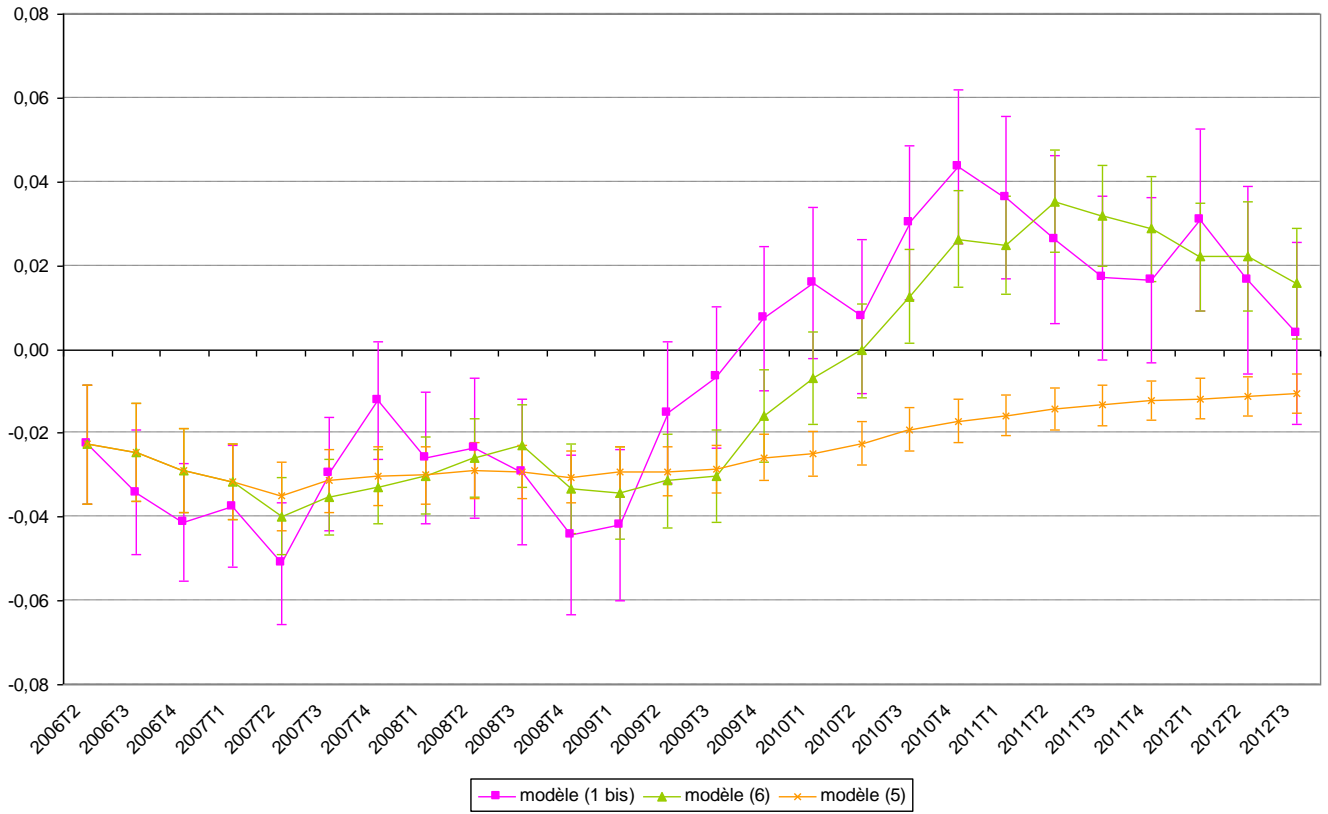
**Graphique 6.76 : Les villes-centres**



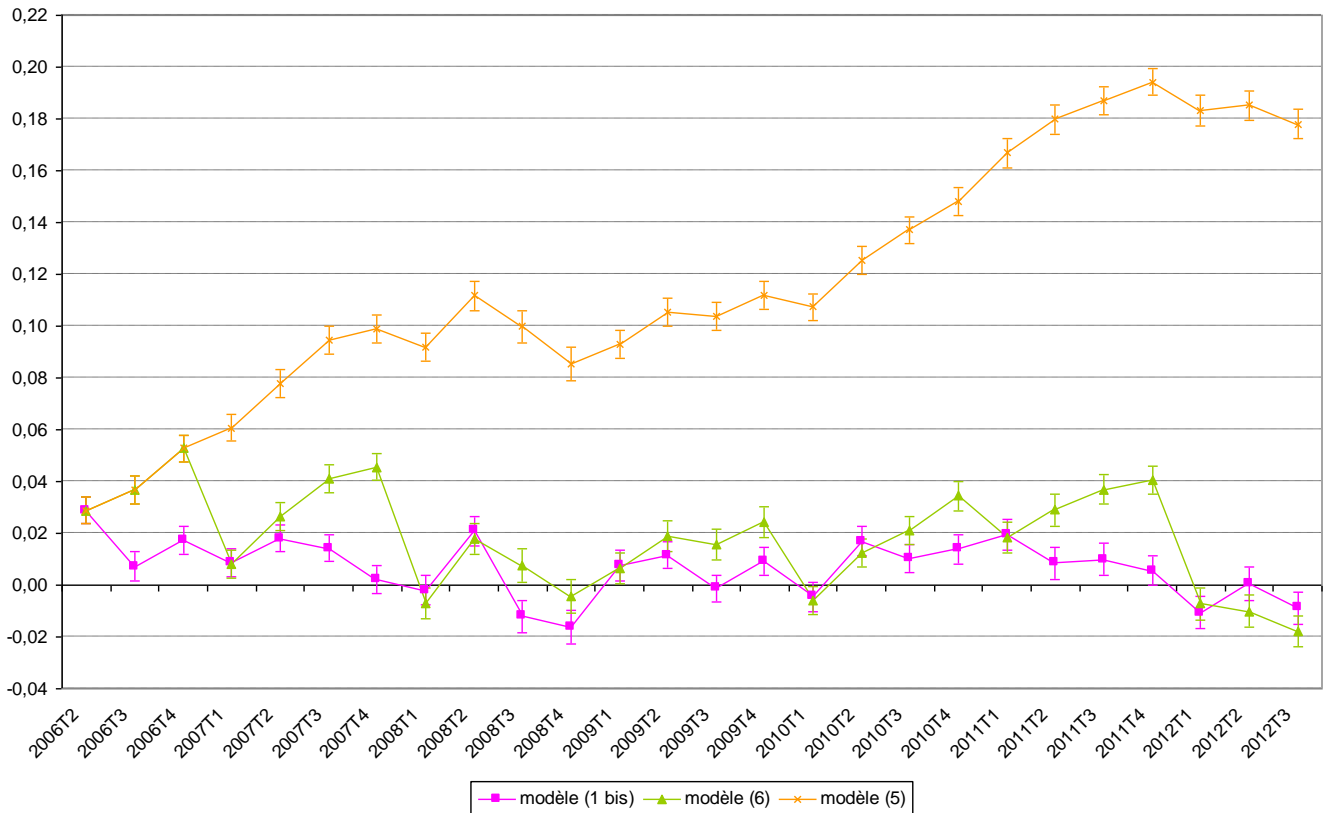
**Graphique 6.77 : Les villes de banlieue**



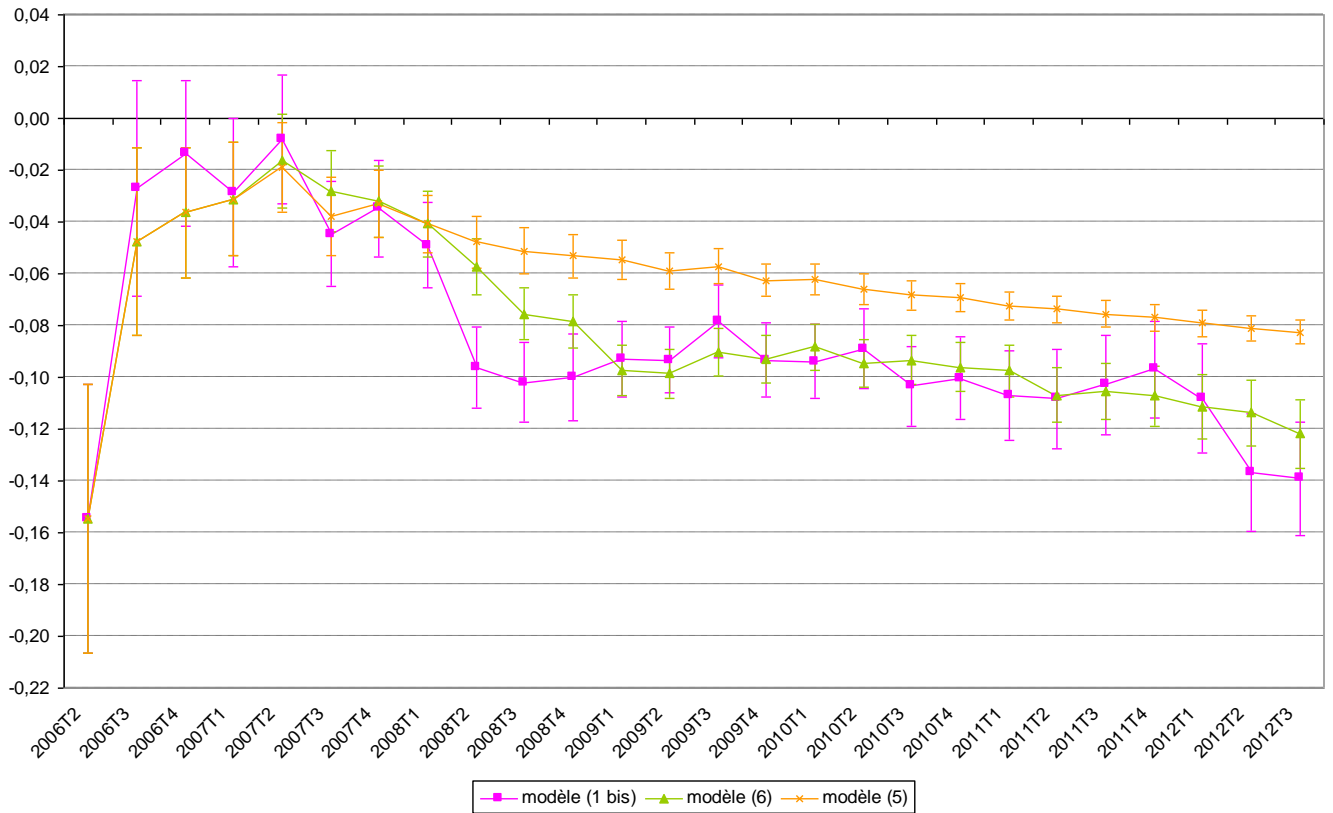
**Graphique 6.78 : Les villes isolées**



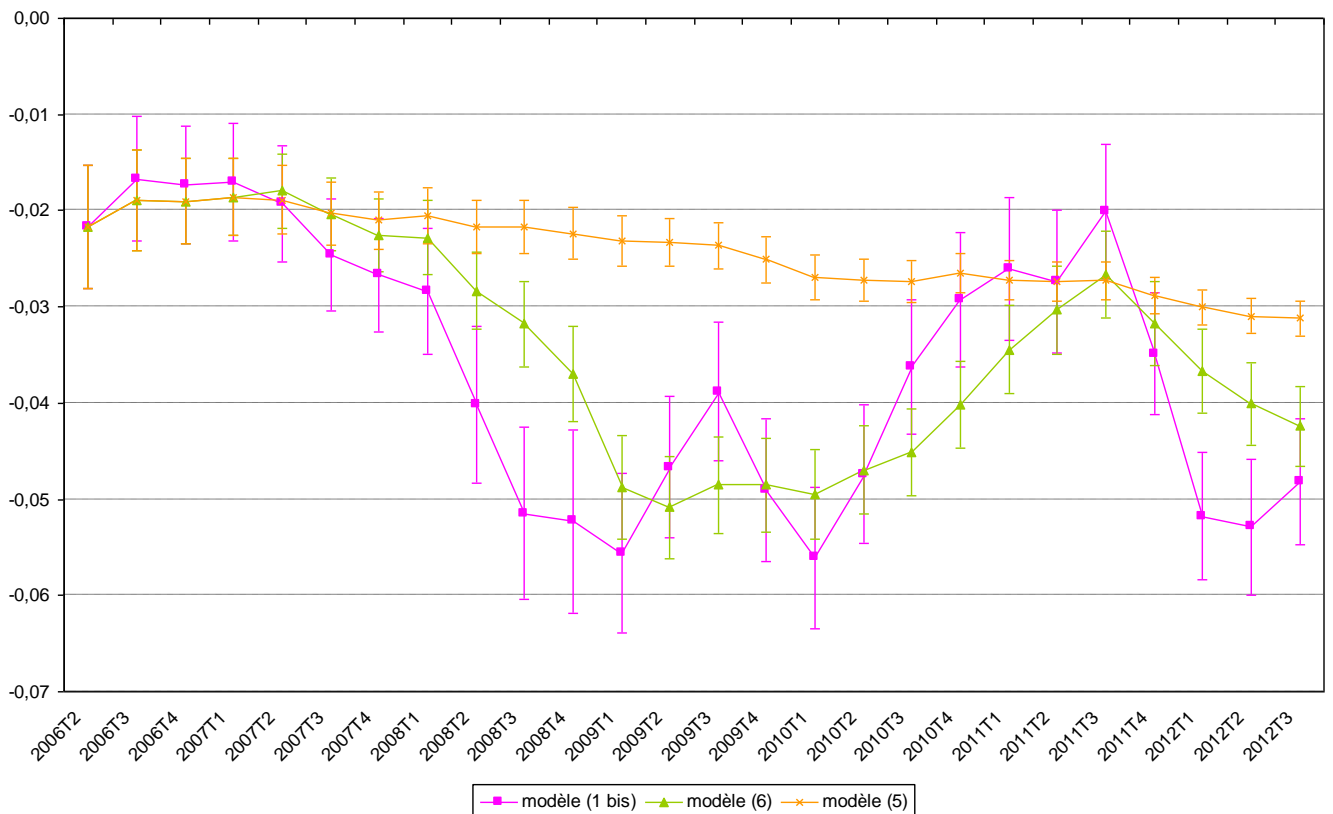
**Graphique 6.79 : L'indicateur temporelle**



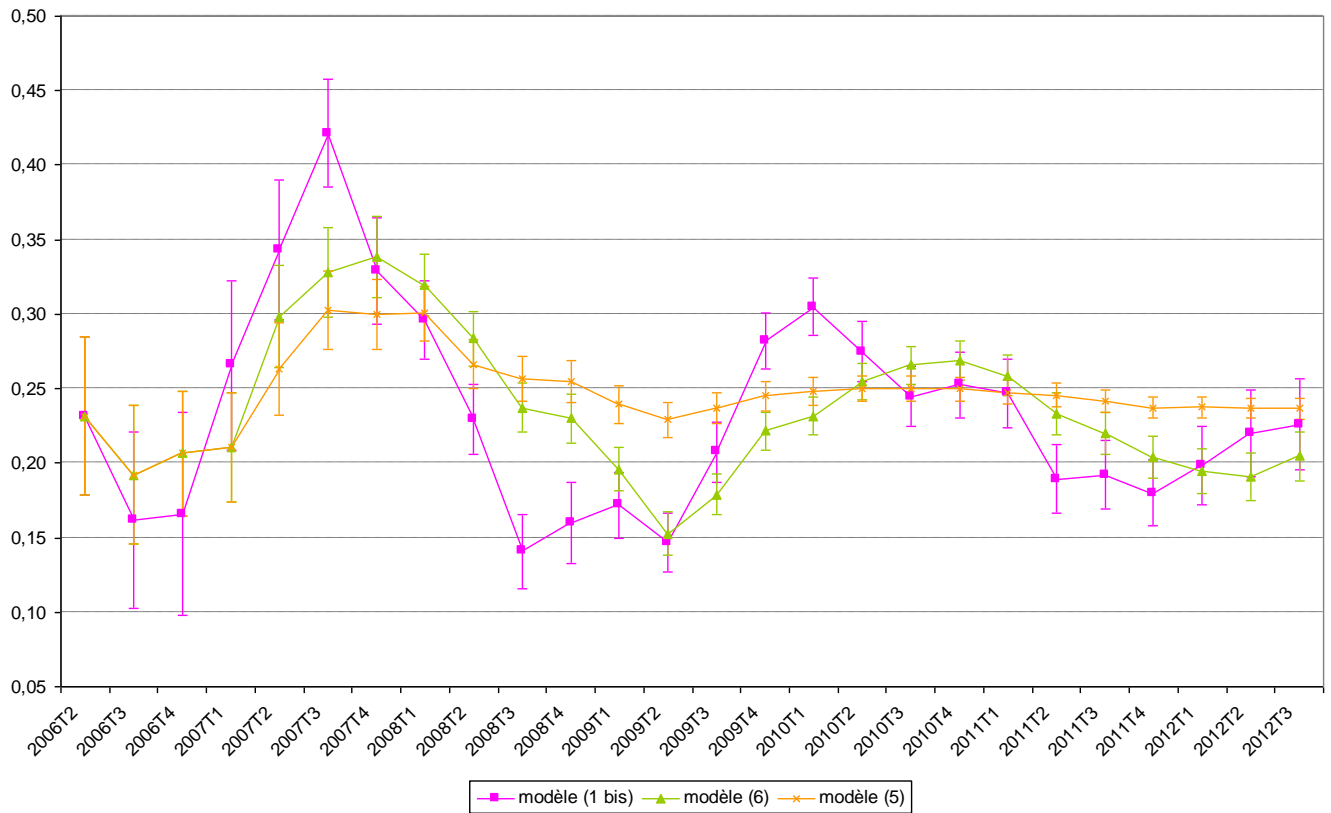
**Graphique 6.80 : La catégorie « ordinaire »**



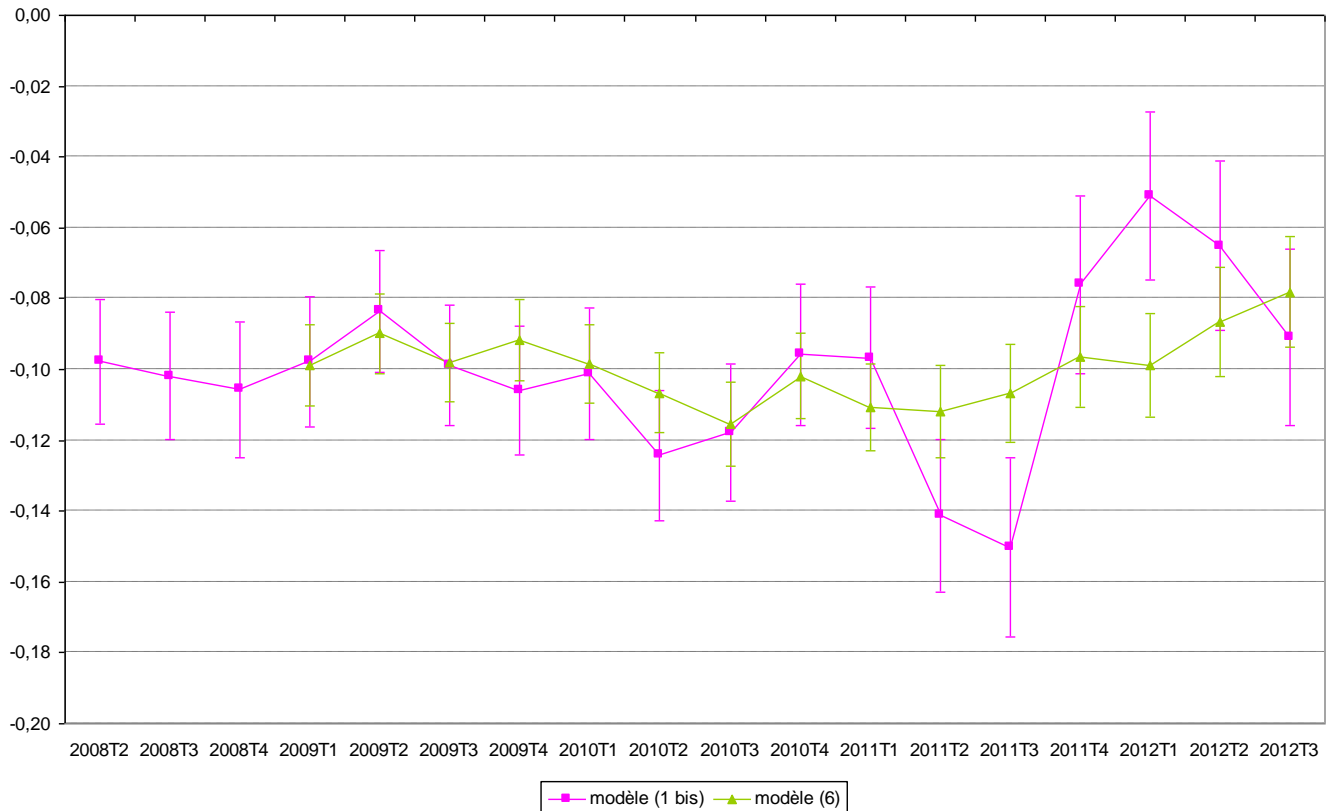
**Graphique 6.81 : La catégorie « intermédiaire »**



**Graphique 6.82 : La catégorie « luxe »**

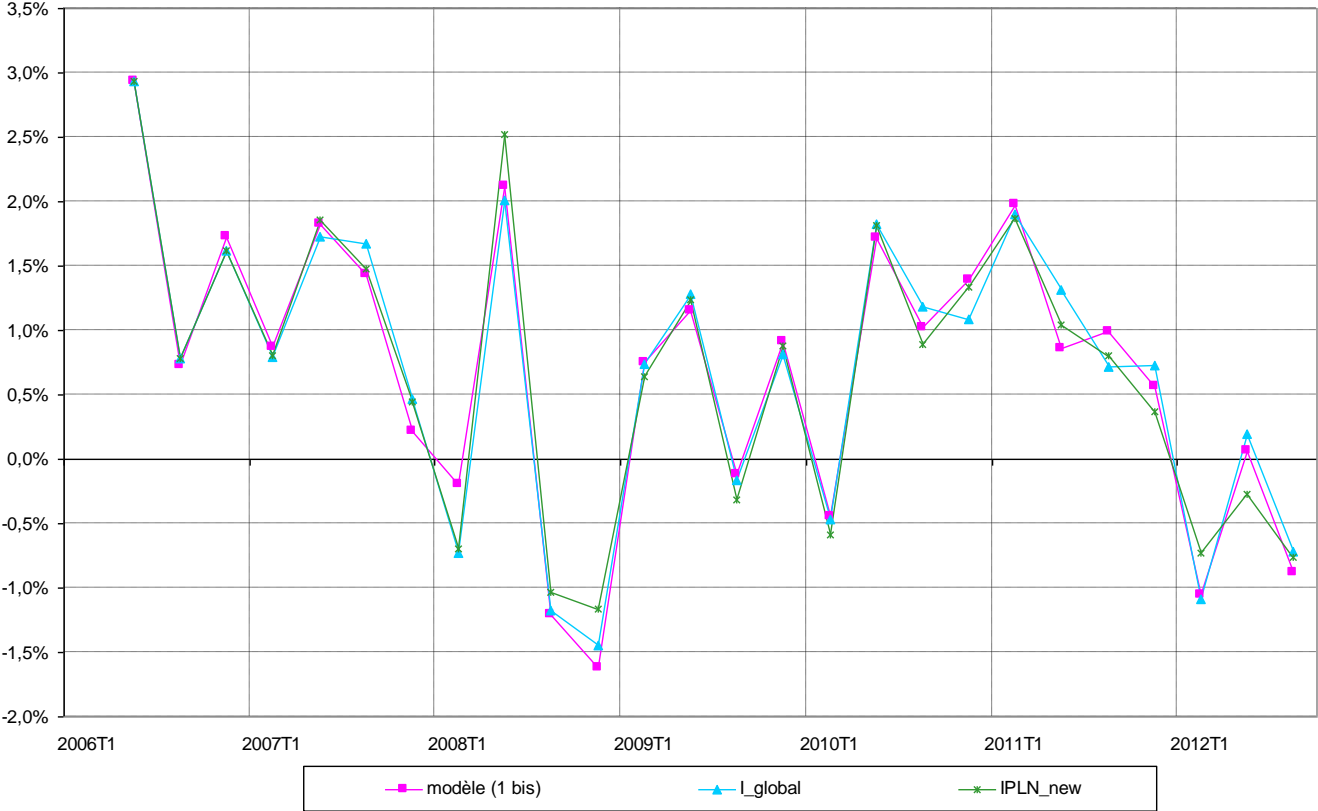


**Graphique 6.83 : Les maisons bâties sur un grand terrain**



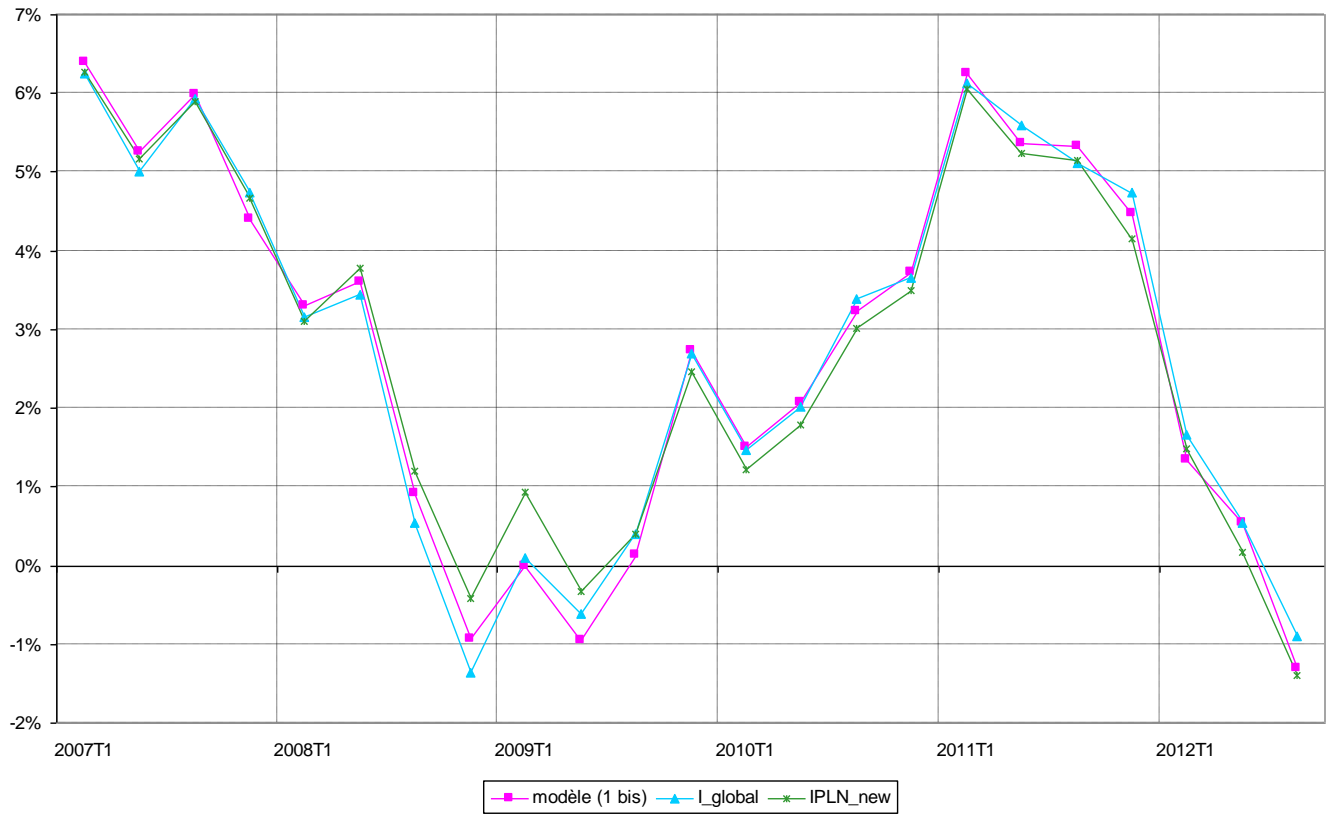
**Annexe 28 : Les taux de croissance en glissement trimestriel et en glissement annuel des indices obtenus à partir des modèles (1 bis), (5) et (6)**

**Graphique 6.84 : Les taux de croissance en glissement trimestriel**





**Graphique 6.85 : Les taux de croissance en glissement annuel**



## Bibliographie

- [1] R. LECAT, « La prise en compte des services de logement dans l'indice des prix à la consommation : une comparaison internationale », *Bulletin de la Banque de France*, n°115, juillet 2003, pages 43-50
- [2] *Technical manual on OWNER-OCCUPIED HOUSING for Harmonised Index of Consumer Prices, v 2.0, March 2012*, Eurostat
- [3] *Handbook on Residential Property Prices Indices (RPPIs)*, Eurostat Methodologies and Working papers
- [4] *Les indices Notaires-Insee de prix des logements anciens, version 3 des modèles hédoniques*, Insee méthodes n°128 de juillet 2014
- [5] *Probabilités, analyse des données et statistique*, G. SAPORTA, Editions TECHNIP
- [6] *Analyse de régression appliquée*, Y. DODGE – V. ROUSSON, Dunod
- [7] *Econométrie linéaire*, B. CREPON, Cepe, 2005
- [8] *Premiers pas en régression linéaire avec SAS*, J. CONFAIS – M. LE GUEN, documents de travail du Centre d'économie de la Sorbonne
- [9] *Les techniques de sondage*, P. ARDILLY, Editions TECHNIP
- [10] T. BALCONE, A.LAFERRERE, « New or Old, Why would Housing Price Indices Differ ? », 2015 : [http://www.jms-insee.fr/2015/S06\\_2\\_ACTE\\_BALCONE\\_JMS2015.PDF](http://www.jms-insee.fr/2015/S06_2_ACTE_BALCONE_JMS2015.PDF)
- [11] T. BALCONE, « La construction de l'indice des propriétaires-occupants (OOH) et de l'indice des prix des logements (HPI) : la nécessité de l'élaboration d'un indice des prix des logements neufs », 2012 : [http://www.jms-insee.fr/2012/S12\\_3\\_ACTE\\_BALCONE\\_JMS2012.PDF](http://www.jms-insee.fr/2012/S12_3_ACTE_BALCONE_JMS2012.PDF)
- [12] T. BALCONE, Mémoire de master de statistique publique, 2013, Miméo, ENSAI

## Série des documents de travail de la DSDS

**F1807** : « Estimer les effectifs de couples de personnes de même sexe au recensement : expérimentation d'une solution de validation du sexe par le prénom » - Élisabeth Algava et Sébastien Hallépée

**F1806** : « L'effet d'une variation du montant de certains transferts du système socio-fiscal sur le niveau de vie : résultats sur 2016 à partir du modèle de microsimulation Ines (Cahier de variantes) » - Maëlle Fontaine et Michaël Sicsic

**F1805** : « Family, Firms and the Gender Wage Gap in France » - Elise Coudin, Sophie Maillard et Maxime Tô

**F1804** : « Mode de collecte et questionnaire, quels impacts sur les indicateurs européens de l'enquête Emploi ? » - Klara Vinceneux  
How do the collection mode and questionnaire used affect the European indicators in the French Labour Force Survey ? - Klara Vinceneux

**F1803** : « Trois versions du taux d'effort en matière de logement » - Pascal Godefroy

**F1802** : « Heterogeneous exposure to labor earnings risk » - Pierre Pora et Lionel Wilner

**F1801** : « L'espérance de vie par niveau de vie Méthode et principaux résultats » - Nathalie Blanpain

**F1708** : « Les données fiscales de l'EDP : une nouvelle source d'informations sur les couples et les familles ? » - Vianney Costemalle

**F1707** : « La situation matrimoniale dans le recensement : impact de la refonte du questionnaire de 2015. » - Guillemette Buisson

**F1706** : « Indices de prix à la consommation » - Patrick Sillard

**F1705** : Effet d'un choc d'inflation sur le revenu disponible et ses composantes deux ans après : une approche par microsimulation - Anne-Lise Biotteau et Maëlle Fontaine

**F1704** : Scanner data and quality adjustment - Isabelle Léonard, Patrick Sillard, Gaëtan Varlet et Jean-Paul Zoyem

**F1703** : Les structures familiales en France : comparaison entre le recensement, l'enquête famille et logements et l'enquête emploi - Guillemette BUISSON et Aude LAPINTE

**F1702** : Projections de la population active à l'horizon 2070 - Malik KOUBI et Anis MARRAKCHI

**F1701** : Les taux marginaux effectifs de prélèvement pour les personnes en emploi en France en 2014 - Juliette FOURCOT et Michaël SICSIC

**F1606** : Projections de population 2013-2070 pour la France : méthode et principaux résultats - Nathalie BLANPAIN et Guillemette BUISSON

**F1605** : Les durées passées en famille monoparentale - Méthode d'estimation des durées et résultats - Vianney COSTEMALLE

**F1604** : ESeG = European Socio economic Groups - Nomenclature socio-économique européenne - Monique MERON, Michel AMAR, Charline BABET, Milan BOUCHET-VALAT, Fanny BUGEJA-BLOCH, François GLEIZES, Frédéric LEBARON, Cédric HUGRÉE, Étienne PENISSAT et Alexis SPIRE

**F1603** : Catégorie sociale d'après les déclarations annuelles de données sociales et catégorie sociale d'après le recensement : quels effets sur les espérances de vie par catégorie sociale ? Comparaison entre les déclarations annuelles de données sociales et les recensements de la population. Comparaison de méthodes d'estimation des espérances de vie - Vianney COSTEMALLE

**F1602** : L'espérance de vie par catégorie sociale et par diplôme - Méthode et principaux résultats - Nathalie BLANPAIN

**F1601** : Échantillonnage des agglomérations de l'IPC pour la base 2015 - Laurence JALUZOT et Patrick SILLARD

**F1508** : Worker-firm matching and the family pay gap: Evidence from linked employer-employee data - Lionel WILNER

**F1507** : Effet des nouvelles mesures sociales et fiscales sur le niveau de vie des ménages : méthodologie de chiffrage avec le modèle de microsimulation Ines - Mathias ANDRÉ, Marie-Cécile CAZENAVE, Maëlle FONTAINE, Juliette FOURCOT et Antoine SIREYJOL

**F1506** : Nowcasting du taux de pauvreté par la micro-simulation - Maëlle FONTAINE et Juliette FOURCOT

**F1505/376-501** : Bilan du projet EDP++ - division Camap et division Enquêtes et études démographiques

**F1504** : Contrôles des rémunérations dans les déclarations annuelles de données sociales (DADS) - Une analyse exploratoire pour améliorer la détection des points atypiques - Claire JACOD

**F1503** : Précision de l'enquête Patrimoine 2010 - Pierre LAMARCHE et Laurianne SALEMBIER

**F1502** : Pourquoi l'indicateur de pauvreté en conditions de vie baisse malgré la crise économique ouverte en 2008 ? Jean-Louis PAN KÉ SHON

**F1501** : Évolution de la population de la France entre 1981 et 2011 : contributions de la fécondité, de la mortalité, du solde migratoire et de la structure de la pyramide des âges - Catherine BEAUMEL et Pascale BREUIL-GENIER

**F1410** : "Personal network" and retirement: Is retirement bad for friendship and good for family relationships ? Anne LAFERRÈRE

**F1409** : Retraités mais pas en retrait : La retraite pousse-t-elle à de nouvelles activités ? Anne LAFERRÈRE

**F1407** : Production "aval" de l'enquête emploi en continu EEC2 2013 - 20XX - Fabien GUGGEMOS

**F1406 bis** : La constitution de l'échantillon démographique permanent de 1968 à 2012 - Stéphane JUGNOT

**F1405 (tome 1)** : Hommes et femmes vivant en couple en 2009, 1999 et aux recensements précédents - Fabienne DAGUET

**F1405 (tome 2)** : Hommes et femmes vivant en couple en 2009, 1999 et aux recensements précédents - Fabienne DAGUET

**F1404** : L'addition est-elle moins salée ? La réponse des prix à la baisse de TVA dans la restauration en France - Quentin LAFFÉTER et Patrick SILLARD

**F1403** : Estimer les flux d'entrées sur le territoire à partir des enquêtes annuelles de recensement - Chantal BRUTEL

**F1402** : Une rotation de la main d'œuvre presque quintuplée en 30 ans : plus qu'un essor des formes particulières d'emploi, un profond changement de leur usage - Claude PICART

**F1401** : Calculs statistiques de stock et de flux sur la révision électorale 2012-2013 - Christelle RIEG