

## 4 – Calculs mis en œuvre

- 4.1 – Les modèles hédoniques
- 4.2 – Les variables caractéristiques
- 4.3 – Les indices élémentaires pour un parc donné
- 4.4 – Les indices agrégés pour un parc donné
- 4.5 – Le chaînage
- 4.6 – Désaisonnalisation des indices

### 4.1 – Les modèles hédoniques

Le calcul des indices des prix des logements anciens repose sur la construction de modèles hédoniques : ceux-ci visent à lier mathématiquement le prix des biens à leurs caractéristiques.

Un modèle distinct est estimé pour chaque strate, le lien entre le prix et les caractéristiques pouvant varier d'un marché à un autre.

Les modèles sont de la forme :

$$\log(p_{s,i}) = \log(p_{0,s}) + \mu_{A-1,s} \cdot Y_{A-1,s,i} + \sum_{m=1}^{11} \theta_{m,s} \cdot M_{m,s,i} + \sum_{k=1}^{K_s} \beta_{k,s} \cdot X_{k,s,i} + \epsilon_{s,i} \quad (1)$$

avec les notations suivantes :

$s$  : la strate (relative à un type de bien donné) sur laquelle on considère le modèle

$p_{s,i}$  : le prix au m<sup>2</sup> (pour les appartements) ou le prix total (pour les maisons) du bien  $i$  de la strate  $s$

$p_{0,s}$  : le prix de référence de la strate  $s$  : il s'agit du prix moyen qu'auraient les biens de la strate pour un jeu de caractéristiques fixé ; on cherchera à suivre l'évolution de ce prix pour calculer un indice sur la strate.

$Y_{A-1,s,i}$  : valeur de l'indicatrice de l'année de mutation pour le bien  $i$  de la strate  $s$  (on estime en pratique les modèles sur deux années de transactions consécutives notées  $A-1$  et  $A$  ; il y a une seule indicatrice pour l'année  $A-1$ , car  $A$  est l'année de référence)

$M_{m,s,i}$  : valeur de l'indicatrice du mois de mutation pour le bien  $i$  de la strate  $s$ , indexée par  $m$  allant de 1 à 11 (le mois de décembre est le mois de référence et ne dispose donc pas d'indicatrice associée)

$X_{k,s,i}$  : valeur de la  $k$ -ième variable caractéristique du modèle pour le bien  $i$  de la strate  $s$ , indexée par  $k$  allant de 1 à  $K_s$  ( $K_s$  étant le nombre de variables caractéristiques utilisées dans le modèle associé à la strate  $s$ )

$\mu_{A-1,s}$  : coefficient associé à l'indicatrice d'année  $Y_{A-1,s}$

$\theta_{m,s}$  : coefficient associé à l'indicatrice de mois  $M_{m,s}$

$\beta_{k,s}$  : coefficient associé à la variable  $X_{k,s}$

$\epsilon_{s,i}$  : terme d'erreur pour le bien  $i$  de la strate  $s$

Ils sont estimés sur une période passée, appelée « période de référence », qui correspond au flux des transactions observées sur deux années. C'est la raison pour laquelle sont introduites les

variables  $Y_{A-1,s}$  et  $M_{m,s}$ , pour prendre en compte les variations de prix au cours de ces deux ans.

La période de référence est constituée des années  $n-3$  et  $n-2$  pour les indices des années  $n$  et  $n+1$ <sup>21</sup> (avec  $n$  pair), le parc étant mis à jour tous les deux ans.

La méthode hédonique implémentée repose sur l'hypothèse que la valorisation des caractéristiques des biens (les coefficients  $\beta_{k,s}$ ) est stable tout au long d'une période comprenant la période de référence sur laquelle les modèles sont estimés (années  $n-3$  et  $n-2$ ) ainsi que la période associée sur laquelle on va calculer les indices (années  $n$  et  $n+1$ ), appelée « période courante »<sup>22</sup>.

On réalise deux phases d'estimation, avec la méthode des moindres carrés ordinaires qui fournit la meilleure estimation non biaisée des coefficients  $\mu_{A-1,s}$ ,  $\theta_{m,s}$  et  $\beta_{k,s}$ .

La première phase a pour but d'exclure les observations atypiques, c'est-à-dire celles pour lesquelles la valeur estimée s'écarte trop de la valeur réelle, qui correspondent souvent à des logements avec des caractéristiques très particulières et auxquels la valorisation décrite par l'équation hédonique s'applique mal. L'objectif poursuivi est de produire une estimation plus robuste des coefficients, donc moins sensible aux caractéristiques particulières de l'échantillon de données exploité.

En notant  $\log(\hat{p}_{0,s,temp})$ ,  $\mu_{A-1,s,temp}$ ,  $\theta_{m,s,temp}$ , et  $\beta_{k,s,temp}$  les estimateurs obtenus à l'issue de cette première phase d'estimation, la relation suivante est vérifiée :

$$\log(p_{s,i}) = \log(\hat{p}_{0,s,temp}) + \mu_{A-1,s,temp} \cdot Y_{A-1,s,i} + \sum_{m=1}^{11} \theta_{m,s,temp} \cdot M_{m,s,i} + \sum_{k=1}^{K_s} \beta_{k,s,temp} \cdot X_{k,s,i} + \epsilon_{s,i,temp}$$

où les  $\epsilon_{s,i,temp}$  sont appelés résidus et correspondent à la différence entre la valeur observée et la valeur estimée de  $\log(p_{s,i})$ .

Il existe plusieurs critères de détection des points atypiques, on les définit ici comme ceux dont la valeur estimée par le modèle s'écarte de la valeur réelle de plus de deux écarts-types.

Plus précisément, on calcule les résidus standardisés  $r_{s,i,temp}$  :

$$r_{s,i,temp} = \frac{\epsilon_{s,i,temp}}{\hat{\sigma}_s \cdot \sqrt{1-h_{s,i}}}$$

<sup>21</sup> On cherche à travailler sur une période de référence proche de celle des indices ; cependant, les données de l'année  $n-1$  ne sont pas encore toutes enregistrées lorsqu'on calcule les modèles utilisés lors de l'année  $n$ , ce pourquoi on s'en tient aux années  $n-3$  et  $n-2$  qui elles sont complètes en matière de capacité de collecte et d'atteinte d'un taux de couverture maximal.

<sup>22</sup> Une autre méthode hédonique, dite par « périodes adjacentes », a été testée dans le cadre de la mise en place de la v4 des indices. Chaque trimestre, un nouveau modèle est estimé sur les données des deux derniers trimestres en incluant une indicatrice temporelle comme variable explicative. L'estimateur associé à cette indicatrice permet alors de rendre compte de l'évolution des prix entre les deux trimestres, toutes choses égales par ailleurs. En se basant uniquement sur les deux derniers trimestres, cette méthode présente l'avantage de s'adapter rapidement si le lien entre le prix et les caractéristiques évolue au cours du temps mais, *a contrario*, la précision des estimateurs est plus faible, car l'échantillon est plus restreint. Au total, les résultats obtenus étaient similaires à ceux produits avec la méthode à qualité fixée, ce qui étaye la robustesse des chiffres produits.

Avec :

$$\hat{\sigma}_s^2 = \frac{\sum_{i=1}^{n_s} \epsilon_{s,i, \text{temp}}^2}{n_s - (p_s + 1)}$$

$\hat{\sigma}_s$  : la racine carrée de la variance estimée de  $\epsilon_{s,i}$  égale à

$n_s$  : le nombre d'observations dans la strate

$p_s$  : le nombre total de variables (caractéristiques du bien et indicatrices d'année ou de mois) dans le modèle associé à la strate

$h_{s,i} = X_{s,i}' (X_s' X_s)^{-1} X_{s,i}$  : l'effet levier de l'observation  $i$  où  $X_s$  est la matrice de taille  $n_s * (p_s + 1)$  représentant les valeurs des variables du modèle (ainsi que la constante) pour l'ensemble des observations de la strate  $s$  et où  $X_{s,i}$  désigne le vecteur de taille  $1 * (p_s + 1)$  regroupant les valeurs des variables pour l'observation  $i$  de la strate  $s$ .

On supprime alors toutes les observations ayant un résidu standardisé sortant de l'intervalle ]-2 ; 2]. Sur le parc 2015-2016, 4,6 % des appartements anciens et 4,9 % des maisons anciennes ont été supprimés à cette étape aussi bien en Île-de-France qu'en province.

À l'issue de cette étape, l'ensemble des observations encore présentes compose ce que l'on appelle le parc de référence : il s'agit donc de l'ensemble des transactions du champ sur la période de référence et n'ayant pas été détectées atypiques. C'est sur ce parc qu'on va effectuer la seconde phase d'estimation pour produire les modèles finaux.

Ce parc de référence constitue le « panier de biens » ou « portefeuille » dont on va mesurer la variation de prix pour calculer les indices. On s'assure ainsi que les indices retracent l'évolution de prix des mêmes logements et qu'ils ne sont pas sensibles aux variations de la structure du marché (il s'agira alors de valoriser les logements du parc de référence lors de la période courante grâce aux modèles, bien qu'ils n'aient pas été vendus à nouveau).

Finalement, on effectue la seconde phase d'estimation pour obtenir les estimateurs définitifs  $\log(\hat{p}_{0,s})$ ,  $\mu_{A-1,s}$ ,  $\theta_{m,s}$  et  $\beta_{k,s}$  caractérisant le modèle :

$$\log(p_{s,i}) = \log(\hat{p}_{0,s}) + \mu_{A-1,s} \cdot Y_{A-1,s,i} + \sum_{m=1}^{11} \theta_{m,s} \cdot M_{m,s,i} + \sum_{k=1}^{K_s} \beta_{k,s} \cdot X_{k,s,i} + \epsilon_{s,i} \quad (2)$$

## 4.2 – Les variables caractéristiques

Les variables caractéristiques des biens  $X_k$  sont calculées à partir des caractéristiques initiales disponibles dans les bases (voir partie 2.4). Elles peuvent inclure des effets croisés<sup>23</sup> ou des regroupements de valeurs au sein d'une même modalité<sup>24</sup>. Elles sont par ailleurs codifiées dans un format adapté à leur inclusion dans les modèles de régression : elles peuvent être des variables indicatrices (prenant les valeurs 0 ou 1) pour les caractéristiques avec un nombre fini de modalités ou des variables continues.

On définit une valeur de référence pour chaque caractéristique et on appelle « logement de référence » le logement possédant l'ensemble des valeurs de référence. Les indices calculés seront les mêmes quelles que soient les valeurs fixées, mais le choix d'une référence est impératif lorsqu'on souhaite implémenter l'ensemble des modalités d'une même caractéristique au sein d'un modèle afin d'éviter la colinéarité des variables explicatives. Ainsi, pour une caractéristique possédant  $m$  modalités, on définira  $m-1$  variables indicatrices.

Par exemple, la caractéristique *nombre de salles de bains* est scindée en trois modalités : 0, 1 et 2 ou plus. On choisit 1 comme référence et on définit alors deux variables indicatrices pour les autres modalités :

- $I_{0 \text{ salle de bain}}$  qui prendra la valeur 1 si le logement ne possède pas de salle de bains et 0 sinon ;
- $I_{2 \text{ salles de bain ou plus}}$  qui prendra la valeur 1 si le logement possède deux salles de bains ou plus et 0 sinon.

Si ces deux variables sont utilisées dans un modèle, alors on calculera les estimateurs des effets associés :

- $\hat{\beta}_{0 \text{ salle de bain}}$  qui va chiffrer l'impact sur le logarithme du prix de ne pas avoir de salle de bains plutôt que d'en avoir une (une salle de bains étant la référence) ;
- $\hat{\beta}_{2 \text{ salles de bain ou plus}}$  qui va chiffrer l'impact sur le logarithme du prix d'avoir deux salles de bains ou plus plutôt que d'en avoir une.

L'ensemble des caractéristiques introduites dans les modèles sont listées dans les figures 4-1 et 4-2.

---

<sup>23</sup> On croise des variables quand celles-ci sont très corrélées entre elles pour éviter les problèmes de colinéarité.

<sup>24</sup> On regroupe des valeurs pour deux raisons : d'une part pour avoir un nombre suffisant d'observations pour chaque modalité et que l'estimation du coefficient associé soit suffisamment précise, d'autre part pour avoir des modalités ayant un sens économique sur la formation du prix. On ne dénote par exemple pas de différence de prix entre le 5<sup>e</sup> et 6<sup>e</sup> étage, donc il n'est pas nécessaire de les considérer séparément ; en revanche, ce sera le cas entre le RDC et le 1<sup>er</sup> étage.

**Figure 4-1 : Variables introduites dans les modèles – Appartements**

<b>Variables caractéristiques des appartements</b>	
<b>Caractéristiques</b>	<b>Modalités (les références sont signalées en gras)</b>
Époque de construction	Époque AB : avant 1913 Époque C : de 1914 à 1947 <b>Époque D : de 1948 à 1969</b> Époque E : de 1970 à 1980 Époque F : de 1981 à 1991 Époque G : de 1992 à 2000 Époque HI : depuis 2001 Époque X : non renseigné
Nombre de salles de bains	0 <b>1</b> 2 ou plus
Nombre de garages ou parkings	<b>0</b> 1 2 ou plus
État du bien	<b>B : bon état</b> M : travaux à prévoir V : à rénover XX : non renseigné
Étage croisé avec la présence d'un ascenseur	<b>Rez-de-chaussée</b> 1er étage 2e étage 3e étage Plus que 3e étage avec ascenseur Plus que 3e étage sans ascenseur
Présence d'une cave	<b>Pas de cave</b> Au moins une cave
Présence d'une terrasse ou balcon ou loggia	<b>Sans terrasse ni balcon ni loggia</b> Avec terrasse ou balcon ou loggia Non renseigné
Nombre de pièces croisé avec la surface (sous forme de la surface moyenne par pièce)	Studio - petit : <20 m <sup>2</sup> <b>Studio - moyen : entre 20 et 30 m<sup>2</sup></b> Studio - grand : >30 m <sup>2</sup> 2 pièces - petit : <17 m <sup>2</sup> <b>2 pièces - moyen : entre 17 et 24 m<sup>2</sup></b> 2 pièces - grand : >24 m <sup>2</sup> 3 pièces - petit : <18 m <sup>2</sup> <b>3 pièces - moyen : entre 18 et 23 m<sup>2</sup></b> 3 pièces - grand : >23 m <sup>2</sup> 4 pièces - petit : <17 m <sup>2</sup> <b>4 pièces - moyen : entre 17 et 21 m<sup>2</sup></b> 4 pièces - grand : >21 m <sup>2</sup> 5 pièces ou + - petit : <16 m <sup>2</sup> <b>5 pièces ou + - moyen : entre 16 et 22 m<sup>2</sup></b> 5 pièces ou + - grand : >22 m <sup>2</sup>

Variables caractéristiques des appartements	
Étiquette du diagnostic de performance énergétique	A ou B : ≤90 Kw/h C : 91 à 150 Kw/h D : 151 à 230 Kw/h E : 231 à 330 Kw/h F ou G : >330 Kw/h <b>X : Non renseigné</b>
Distance de la commune d'appartenance à la grande commune <sup>25</sup> la plus proche (en kilomètres, calculée de centre à centre)	variable numérique continue - référence 0 km en province, 15 km en Île-de-France
Zonage détaillé croisé avec le nombre de pièces (en 5 classes : 1, 2, 3, 4 et 5 pièces ou plus)	Il existe 5 modalités par zone (pour chaque modalité du nombre de pièces). <b>Référence : zone 1 pour les 3 pièces</b>

**Figure 4-2 : Variables introduites dans les modèles - Maisons**

Variables caractéristiques des maisons	
Caractéristiques	Modalités (les références sont signalées en gras)
Nombre de niveaux	1 <b>2 ou plus</b>
Nombre de salles de bains	0 <b>1</b> 2 ou plus
Nombre de garages ou parkings	0 <b>1</b> 2 ou plus Non renseigné
Époque de construction	Époque AB : avant 1913 (inclus) Époque C : de 1914 à 1947 Époque D : de 1948 à 1969 Époque E : de 1970 à 1980 Époque F : de 1981 à 1991 Époque GHI : depuis 1992 <b>Époque X : non renseigné</b>
Logarithme de la surface habitable	<i>variable numérique continue - référence : <b>ln(100)</b></i>
Logarithme de la surface du terrain	<i>variable numérique continue - référence : <b>ln(610)</b></i>
État du bien	B : bon état M : travaux à prévoir V : à rénover <b>XX : non renseigné</b>
Présence d'un sous-sol / cave	<b>pas de sous-sol / cave</b> présence d'un sous-sol / cave

<sup>25</sup> Il s'agit de la distance à la commune de 50 000 habitants la plus proche pour les strates de province et de la distance à Paris pour les strates d'Île-de-France.

Variables caractéristiques des maisons	
Étiquette du diagnostic de performance énergétique	A ou B : ≤90 Kw/h C : 91 à 150 Kw/h D : 151 à 230 Kw/h E : 231 à 330 Kw/h F ou G : >330 Kw/h <b>X : Non renseigné</b>
Distance de la commune d'appartenance à la grande commune la plus proche <sup>26</sup> (en kilomètres, calculée de centre à centre)	variable numérique continue - référence 0 km en province, 27 km en Île-de-France
Zonage détaillé croisé avec le nombre de pièces (en 5 classes : moins de 4, 4, 5, 6 et 7 pièces ou plus)	Il existe 5 modalités par zone (pour chaque modalité du nombre de pièces). <b>Référence : zone 1 pour les 4 pièces</b>

Le zonage détaillé croisé avec le nombre de pièces dans l'étape de modélisation<sup>27</sup> n'a pas le même rôle que les strates pour prendre en compte l'effet de la localisation sur les prix : une strate correspond à un périmètre sur lequel on estime un modèle hédonique (cf. partie 3), tandis que la zone (croisée avec le nombre de pièces) correspond à une variable explicative du modèle. En d'autres termes, on suppose qu'au sein d'une strate, un logement peut être valorisé différemment selon la zone détaillée dans laquelle il se trouve et son nombre de pièces, mais que les effets des autres caractéristiques (époque de construction, nombre de salles de bains...) sont les mêmes partout à l'intérieur de la strate (ils peuvent en revanche varier d'une strate à une autre).

Ces zones ont été construites de la manière suivante :

- elles doivent respecter un seuil minimum de 110 transactions en moyenne par an ;
- au sein des strates des appartements parisiens, le zonage détaillé correspond simplement au quartier administratif. Pour les strates des maisons et celles des appartements du reste de l'Île-de-France, le zonage détaillé correspond au département. En effet, comme les strates peuvent regrouper des communes éloignées géographiquement, l'utilisation du département permet de prendre en compte, le cas échéant, des niveaux de prix différents liés à la situation géographique ;
- en province, le zonage détaillé a été construit sur le même principe que celui des strates, à l'exception du seuil minimal de transactions qui a été porté à 110 transactions annuelles. Pour les communes qui forment une strate à elles seules, la segmentation a été établie par dire d'experts : les notaires locaux, interrogés sur la question, ont défini des unités de quartiers d'après leur connaissance du marché immobilier. Pour la strate regroupant les 4 Drom, ainsi que celle comprenant la Corse, le zonage détaillé correspond au département.

<sup>26</sup> Il s'agit de la distance à la commune de 50 000 habitants la plus proche pour les strates de province et de la distance à Paris pour les strates d'Île-de-France.

<sup>27</sup> L'examen des données montre que la variation de prix d'une zone à une autre se différencie selon le nombre de pièces des logements. Par exemple, passer d'un quartier résidentiel à un centre-ville pourra se répercuter différemment sur les prix des studios ou sur les prix des quatre-pièces pour lesquels la demande n'est pas la même.

Toutes les variables fournies dans les figures 4-1 et 4-2 ne feront pas forcément partie du modèle retenu pour une strate donnée. Une caractéristique peut en effet ne pas avoir le même impact d'un marché à un autre et notamment n'est pas toujours explicative du prix dans une strate particulière. On peut par exemple penser que la présence d'un garage ne sera pas déterminante sur le prix des logements dans une zone où il est facile de se garer à l'extérieur tandis que cela sera fortement valorisé dans une zone où il existe très peu de places de stationnement.

On souhaite donc, pour chaque strate, faire entrer dans l'équation associée uniquement les variables les plus corrélées avec le prix. Limiter le nombre de variables explicatives permet d'accroître la validité et la robustesse du modèle<sup>28</sup> et réduit aussi par la même occasion le risque de colinéarité entre les variables. Pour cela, on a recours à une procédure de sélection de variables<sup>29</sup>. Il existe diverses méthodes de sélection ; on utilise celle dite ascendante (*méthode Forward*)<sup>30</sup>.

La méthode Forward est une méthode de pas à pas, où l'on injecte une variable après l'autre. On initialise un modèle sans variables, puis à chaque itération on retient, parmi l'ensemble des variables proposées, celle qui améliore le plus le modèle selon un critère donné. Le critère utilisé ici, le plus classique, est un test de Fischer. Lors de l'ajout de la j-ème variable, la statistique de test pour une variable v est :

$$F_v = \frac{SSE_{M(j-1)} - SSE_{M(j-1,v)}}{\frac{SSE_{M(j-1,v)}}{n-j-1}}$$

avec  $SSE_{M(j-1)}$  : la somme des carrés des résidus du modèle avec les j-1 variables précédemment retenues ;

$SSE_{M(j-1,v)}$  : la somme des carrés des résidus du modèle avec les j-1 variables précédemment retenues et la variable v ;

n : le nombre d'observations.

On retient la variable pour laquelle on obtient la plus faible p-value associée au test. La procédure se termine lorsqu'on atteint une itération où aucune variable n'est jugée suffisamment pertinente pour intégrer le modèle selon un critère d'entrée ; dans notre cas, la procédure s'arrête lorsqu'aucune variable ne produit de p-value inférieure à 0,05.

---

<sup>28</sup> En incluant des variables peu explicatives du prix, on risque de sur-ajuster le modèle, c'est-à-dire d'obtenir un modèle avec un grand nombre d'estimateurs qui s'adaptera trop spécifiquement aux données de l'échantillon utilisé pour la régression (ici le parc de référence) et qui s'appliquera alors beaucoup moins bien pour tous les logements n'ayant pas fait partie de cet échantillon. Pour jauger de la qualité générale d'un modèle dans ce cadre, on cherche à obtenir à la fois un R<sup>2</sup> élevé, représentant la qualité de l'estimation sur l'échantillon, et un PRESS faible, représentant la qualité de prévision hors échantillon par une forme de validation croisée.

<sup>29</sup> Il est parfois préféré de tester manuellement l'inclusion des variables. Le recours à une procédure de sélection est ici très avantageux par son caractère automatique, étant donné le grand nombre de modèles à estimer.

<sup>30</sup> Différentes méthodes de sélection et différents critères ont été testés dans le cadre de la mise en place de la version 4 des indices, aboutissant à des résultats très similaires, ce qui est positif quant à la robustesse des variables sélectionnées.



### 4.3 – Les indices élémentaires pour un parc donné

Les indices élémentaires sont les indices calculés au niveau géographique le plus fin, c'est-à-dire au niveau des strates : ils représentent le rapport entre la valeur courante d'un parc de logements de référence et sa valeur sur une période de base.

On note  $I_{t/b}(s)$  l'indice de la strate  $s$  pour la période  $t$  par rapport à la base  $b$ . Cette période de base, que l'on va noter  $0$ , est fixée au dernier trimestre de l'année  $n-1$  pour les indices des années  $n$  et  $n+1$  (avec  $n$  pair). La valeur des indices en  $t$  représente l'évolution de la valorisation du parc de référence entre  $0$  et  $t$  (ils seront ensuite chaînés pour établir les séries longues et représenteront l'évolution de prix par rapport à une base 100 fixée par ailleurs).

Les logements du parc de référence peuvent être valorisés en période courante à travers la relation hédonique. On décline pour cela le modèle (1) (voir partie 4-1), appliqué uniquement sur la période de référence, afin de l'adapter à l'ensemble de la période d'application du parc ; le prix d'un bien  $j$  de la strate  $s$  au cours d'une période  $t$  peut alors s'exprimer par :

$$\log(p_{s,j,t}) = \log(p_{0,s,t}) + \sum_{k=1}^{K_t} \beta_{k,s} \cdot X_{k,s,j} + \epsilon_{s,j,t} \quad (3)$$

On remarquera que le modèle (1) utilisé uniquement dans la phase d'estimation sur la période de référence est compatible avec ce dernier modèle. L'introduction des indicatrices d'années et de mois vient du fait que le parc de référence couvre une période longue sur laquelle peuvent avoir lieu des variations de prix. De façon équivalente, le prix de référence de la strate  $s$  pour le mois  $(a,m)$  de la période de référence serait :

$$\log(p_{0,s,a,m}) = \log(p_{0,s}) + \mu_{a,s} + \theta_{m,s}$$

Avec  $\mu_{a,s}=0$  si  $a$  est l'année de référence et  $\theta_{m,s}=0$  si  $m$  est le mois de référence.

On cherche maintenant à estimer  $\log(p_{0,s,t})$  en période courante. On ne va pas estimer le modèle (3) par les moindres carrés ordinaires, comme on l'a fait précédemment pour le modèle (1), car on va travailler en période courante sur des périodes plus courtes (un trimestre), qui de plus sont récentes par rapport à la date du calcul et donc pour lesquelles toutes les transactions n'auront pas encore été enregistrées. On a alors peu d'observations, ce qui engendrerait des estimateurs peu précis.

On va en revanche pouvoir réutiliser les estimateurs  $\hat{\beta}_{k,s}$ , plus fiables, obtenus lors de l'estimation (2) sur le parc de référence tronqué des valeurs atypiques, vu qu'on suppose que les effets des caractéristiques  $\beta_{k,s}$  sont stables tout au long de la période.

Tout d'abord, on introduit  $\tilde{p}_{s,j,t}$  le « prix équivalent bien de référence » du bien  $j$  de la strate  $s$  au temps  $t$ , tel que :

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

D'après la relation (3), cette quantité vaut également :

$$\log(\tilde{p}_{s,j,t}) = \log(p_{s,j,t}) - \sum_{k=1}^{K_t} \beta_{k,s} \cdot X_{k,s,j} = \log(p_{0,s,t}) + \epsilon_{s,j,t}$$

En passant à la moyenne sur l'ensemble des  $J_{s,t}$  biens échangés dans la strate  $s$  au cours de la période  $t$ , on obtient<sup>31</sup> :

$$\frac{\sum_{j=1}^{J_{s,t}} \log(p_{s,j,t})}{J_{s,t}} = \frac{\sum_{j=1}^{J_{s,t}} \left[ \log(p_{s,j,t}) - \sum_{k=1}^{K_t} \beta_{k,s} \cdot X_{k,s,j} \right]}{J_{s,t}} = \frac{\sum_{j=1}^{J_{s,t}} [\log(p_{0,s,t}) + \epsilon_{s,j,t}]}{J_{s,t}} = \log(p_{0,s,t})$$

En utilisant les estimateurs  $\hat{\beta}_{k,s}$ , on va pouvoir estimer  $\log(p_{s,j,t})$  par :

$$\log(\hat{p}_{s,j,t}) = \log(p_{s,j,t}) - \sum_{k=1}^{K_t} \hat{\beta}_{k,s} \cdot X_{k,s,j}$$

puis  $\log(p_{0,s,t})$  par la moyenne des  $\log(\hat{p}_{s,j,t})$ . Par souci de robustesse, on n'utilise pas pour cela l'ensemble des  $J_{s,t}$  biens : on retire, au sein de chaque zone détaillée de la strate, les 2 % des observations avec  $\log(\hat{p}_{s,j,t})$  le plus élevé et les 2 % avec  $\log(\hat{p}_{s,j,t})$  le plus faible. En notant  $J'_{s,t}$  le nombre d'observations restantes, on obtient :

$$\log(\hat{p}_{0,s,t}) = \frac{\sum_{j=1}^{J'_{s,t}} \log(\hat{p}_{s,j,t})}{J'_{s,t}} = \frac{\sum_{j=1}^{J'_{s,t}} \left[ \log(p_{s,j,t}) - \sum_{k=1}^{K_t} \hat{\beta}_{k,s} \cdot X_{k,s,j} \right]}{J'_{s,t}}$$

On note  $p_{0,s,t} = \exp(\log(\hat{p}_{0,s,t}))$  l'estimation du prix de référence de la strate  $s$  à la période  $t$ .

On peut maintenant estimer la valeur du parc de référence pendant une période  $t$ .

On considère un logement  $i$  du parc de référence sur la strate  $s$ . On va noter  $\log(\hat{p}_{s,i,t})$  l'estimation du logarithme de son prix à la période  $t$  calculé à partir des précédents estimateurs :

$$\log(\hat{p}_{s,i,t}) = \log(\hat{p}_{0,s,t}) + \sum_{k=1}^{K_t} \hat{\beta}_{k,s} \cdot X_{k,s,i}$$

et  $p_{s,i,t} = \exp(\log(\hat{p}_{s,i,t}))$  l'estimation de son prix (au mètre carré pour les appartements, total pour les maisons) à la période  $t$  :

$$p_{s,i,t} = \exp(\log(\hat{p}_{0,s,t}) + \sum_{k=1}^{K_t} \hat{\beta}_{k,s} \cdot X_{k,s,i}) = p_{0,s,t} \cdot \exp\left(\sum_{k=1}^{K_t} \hat{\beta}_{k,s} \cdot X_{k,s,i}\right)$$

Pour les appartements, on multiplie de plus le prix au mètre carré estimé par la surface pour obtenir l'estimation du prix total.

Afin d'homogénéiser les notations, on introduit  $P_{s,i,t}$  l'estimation du prix total :

$$P_{s,i,t} = p_{s,i,t} \cdot A_{s,i} = p_{0,s,t} \cdot A_{s,i} \cdot \exp\left(\sum_{k=1}^{K_t} \hat{\beta}_{k,s} \cdot X_{k,s,i}\right)$$

<sup>31</sup> Les termes d'erreur  $\epsilon_{s,j,t}$  sont de moyenne nulle par définition.

Avec  $A_{s,i}$  valant la surface du bien  $i$  en mètres carrés pour les strates d'appartements et  $A_{s,i}=1$  pour les maisons.

L'estimation de l'évolution du prix d'un logement, entre deux périodes  $t_1$  et  $t_2$ , est la même pour tous les logements de la strate :

$$\frac{\hat{P}_{s,i,t_2}}{\hat{P}_{s,i,t_1}} = \frac{p_{0,s,t_2} \cdot A_{s,i} \cdot \exp\left(\sum_{k=1}^{K_s} \hat{\beta}_{k,s} \cdot X_{k,s,i}\right)}{p_{0,s,t_1} \cdot A_{s,i} \cdot \exp\left(\sum_{k=1}^{K_s} \hat{\beta}_{k,s} \cdot X_{k,s,i}\right)} = \frac{p_{0,s,t_2}}{p_{0,s,t_1}}$$

La valorisation du parc de référence en  $t$  pour une strate  $s$  donnée, notée  $\hat{W}_{s,t}$ , correspond à la somme des prix estimés en  $t$  de tous les logements composant le parc de référence pour cette strate :

$$\hat{W}_{s,t} = \sum_{i=1}^{N_s} \hat{P}_{s,i,t} = \sum_{i=1}^{N_s} \left( p_{0,s,t} \cdot A_{s,i} \cdot \exp\left(\sum_{k=1}^{K_s} \hat{\beta}_{k,s} \cdot X_{k,s,i}\right) \right) = p_{0,s,t} \cdot \sum_{i=1}^{N_s} \left( A_{s,i} \cdot \exp\left(\sum_{k=1}^{K_s} \hat{\beta}_{k,s} \cdot X_{k,s,i}\right) \right) \quad (4)$$

avec  $N_s$  le nombre de logements composant le parc de référence pour la strate  $s$ .

Ainsi, en calculant l'évolution de la valorisation du parc de référence entre deux périodes  $t_1$  et  $t_2$ , ou, en d'autres termes, l'évolution de l'indice élémentaire de la strate  $s$  entre  $t_1$  et  $t_2$ , on retrouve :

$$\frac{\hat{W}_{s,t_2}}{\hat{W}_{s,t_1}} = \frac{p_{0,s,t_2} \cdot \sum_{i=1}^{N_s} \left( A_{s,i} \cdot \exp\left(\sum_{k=1}^{K_s} \hat{\beta}_{k,s} \cdot X_{k,s,i}\right) \right)}{p_{0,s,t_1} \cdot \sum_{i=1}^{N_s} \left( A_{s,i} \cdot \exp\left(\sum_{k=1}^{K_s} \hat{\beta}_{k,s} \cdot X_{k,s,i}\right) \right)} = \frac{p_{0,s,t_2}}{p_{0,s,t_1}}$$

De là, on peut déduire la valeur des indices élémentaires : il reste à choisir une période de base par rapport à laquelle on mesure la variation de la valeur du parc de référence<sup>32</sup>. Pour les indices trimestriels des années  $n$  et  $n+1$ , calculés sur le parc de référence contenant les transactions des années  $n-3$  et  $n-2$ , on place la période de base au 4<sup>e</sup> trimestre de l'année  $n-1$  que l'on appelle  $t_0$ <sup>33</sup>

<sup>32</sup> Le choix de la période de base n'a pas d'impact sur les évolutions de prix calculées, il s'agit simplement d'une référence. Il est très simple de changer la base une fois un indice calculé, on divise pour cela toutes les valeurs de l'indice par la valeur qu'il prend à la nouvelle période de base souhaitée.

<sup>33</sup> L'Insee publie uniquement les indices trimestriels (les seuls ayant l'appellation « Notaires-Insee »). ADNOV et PNS calculent et publient également des indices mensuels sur trimestre glissants calculés selon la même méthodologie et dénommés par le dernier mois de la période. Par exemple, l'indice de février d'une année  $A$  sera calculé sur une période comprenant décembre  $A-1$ , janvier  $A$  et février  $A$ . Les indices de mars, juin, septembre et décembre correspondent alors aux indices trimestriels. Dans ce cadre et par souci de continuité, les indices de janvier, avril, juillet et octobre des années  $n$  et  $n+1$  sont alors en base  $t_0'$  (août, septembre et octobre de l'année  $n-1$ ). De même les indices de février, mai, août et novembre des années  $n$  et  $n+1$  sont en base  $t_0''$  (septembre, octobre et novembre de l'année  $n-1$ ).

On obtient alors les indices élémentaires :

$$I_{t/t_0}(s) = \frac{\hat{W}_{s,t}}{\hat{W}_{s,t_0}} = \frac{\hat{p}_{0,s,t}}{\hat{p}_{0,s,t_0}}$$

#### 4.4 – Les indices agrégés pour un parc donné

Une fois les indices élémentaires connus, ceux-ci sont agrégés à différents échelons pour obtenir les indices sur les niveaux géographiques souhaités. Une agrégation des indices des prix des appartements et de ceux des maisons a aussi lieu selon le même procédé pour obtenir les indices sur l'ensemble des logements.

##### 4.4.1 Correction de la non-exhaustivité des bases

Les bases notariales ne couvrent pas la totalité des biens échangés (cf. partie 2.2). Pour corriger cette non-exhaustivité, on introduit des coefficients de redressement. Ces coefficients sont estimés à partir de données fiscales (montants des droits de mutation établis par la Direction générale des Finances publiques, DGFIP). Comme celles-ci donnent peu de précisions sur les biens échangés, on estime les coefficients sur des agrégats assez larges : par département et par année de mutation, appartements et maisons confondus. On obtient le coefficient  $\hat{\delta}_{a,d}$  en divisant le montant des assiettes des droits de mutation des transactions dans le département d pour l'année a, d'après les données fiscales, par ce même montant enregistré dans la base notariale. On estime un décalage de 2 mois entre la date de l'acte de vente et la comptabilisation des droits par la DGFIP.

Chaque coefficient est appliqué à tous les biens du croisement (département x année de mutation) dans le calcul des termes qui pondèrent les indices.

##### 4.4.2 Calcul des poids

Le poids des différents indices d'un zonage géographique au sein d'un agrégat plus large représente leur part de la valeur totale du parc de référence sur l'agrégat. Cette valeur du parc est calculée sur le dernier trimestre du parc de référence, que l'on appelle période de pondération q, en prenant en compte les coefficients de correction de la non-exhaustivité<sup>34</sup>. Par exemple, pour une strate s :

$$\hat{W}_{s,q}^{\delta} = \sum_{i=1}^{N_s} \delta_{a_i,d_i} \cdot \hat{P}_{s,i,q} = \sum_{i=1}^{N_s} \left( \delta_{a_i,d_i} \cdot \hat{p}_{0,s,q} \cdot A_{s,i} \cdot \exp\left(\sum_{k=1}^{K_s} \hat{\beta}_{k,s} \cdot X_{k,s,i}\right) \right)$$

avec  $a_i$  et  $d_i$  l'année et le département du bien i.

<sup>34</sup> La prise en compte ou non de ces coefficients dans la phase de calcul des indices élémentaires n'a pas d'impact sur les résultats obtenus. Elle est cependant nécessaire dans la phase de pondération pour que celle-ci ne soit pas faussée par des taux de couverture inégaux sur différents territoires.

Toujours en prenant l'exemple d'une strate  $s$ , son poids au sein d'un agrégat  $A$  composé de  $n_A$  strates vaut alors :

$$\frac{W_{s,q}^{\hat{\delta}}}{\sum_{u=1}^{n_A} W_{u,q}^{\hat{\delta}}}$$

#### 4.4.3 Indices agrégés aux niveaux infra-départementaux et départementaux

Pour obtenir des indices agrégés à des niveaux infra-départementaux ou départementaux, on calcule une moyenne géométrique<sup>35</sup> sur les indices élémentaires<sup>36</sup>. Pour un agrégat  $A$ , infra-départemental ou départemental, composé de  $n_A$  strates, l'indice agrégé vaut :

$$I_{t/t_0}(A) = \prod_{s=1}^{n_A} I_{t/t_0}(s)^{\left(\frac{W_{s,q}^{\hat{\delta}}}{\sum_{u=1}^{n_A} W_{u,q}^{\hat{\delta}}}\right)}$$

#### 4.4.4 Indices agrégés aux niveaux supra-départementaux

Pour obtenir des indices agrégés à des niveaux supra-départementaux, on calcule une moyenne arithmétique. Pour un agrégat  $B$  supra-départemental, composé de  $n_B$  départements, l'indice agrégé vaut :

$$I_{t/t_0}(B) = \sum_{d=1}^{n_B} \left( \frac{W_{d,q}^{\hat{\delta}}}{\sum_{e=1}^{n_B} W_{e,q}^{\hat{\delta}}} \right) \cdot I_{t/t_0}(d)$$

<sup>35</sup> Ce mode de calcul est utilisé dans la construction des indices statistiques lorsque l'on pense qu'il peut exister un arbitrage entre différents biens de nature voisine. Dans notre cas, il peut y avoir un choix d'acquisition entre des biens appartenant à des strates géographiques voisines. On utilise des moyennes arithmétiques pour les niveaux plus agrégés que les niveaux départementaux, car l'acheteur a moins de chances d'arbitrer entre des strates géographiques plus éloignées pour s'installer (région, province, France entière).

<sup>36</sup> Certaines strates peuvent être composées de logements appartenant à différents départements. Dans ce cas, le poids de l'indice de la strate au sein d'un département est calculé en utilisant uniquement les logements de la strate appartenant à ce département.

## 4.5 – Le chaînage

Les indices calculés jusqu'ici sont associés à un parc de référence et sont calculés par rapport à une date  $t_0$ , date qui varie avec le changement de parc de référence tous les deux ans. Les séries d'indices sur l'ensemble de la période sont reconstituées par chaînage.

On introduit une base 100<sup>37</sup>, qui sera la base de la série longue : cette base 100 est fixée en V4 à la moyenne annuelle des indices de 2015<sup>38</sup>. La valeur de l'indice en t en cette base 100 représentera l'évolution des prix entre t et la période choisie pour la base 100.

Il faut tout d'abord initialiser les séries. Si on calculait un indice pour la première fois, on pourrait par exemple utiliser la base  $t_0$  du premier parc de référence comme base 100 et les séries calculées sur ce parc comme séries initiales<sup>39</sup>. Dans notre cas, les évolutions de prix, calculées en version 4 des indices à partir du T1 2018, sont chaînées aux séries existantes calculées à partir des versions précédentes des indices. On utilise donc lors du premier chaînage ces séries, référencées en moyenne annuelle 2015, jusqu'au T4 2017.

On note  $I_{t/t_0}(z)$  l'indice en t de la zone z référencé sur la base 100.

Pour distinguer les différents parcs sur lesquels on travaille, on va noter p(i) le i-ème parc de référence utilisé en version 4 des indices et  $t_0[p(i)]$  la période de base associée au parc p(i). On a par exemple p(1) le parc composé des transactions des années 2015 et 2016, utilisé pour calculer les indices de 2018 et 2019 avec  $t_0[p(1)] = T 42017$ , puis p(2) le parc composé des transactions des années 2017 et 2018, utilisé pour calculer les indices de 2020 et 2021 avec  $t_0[p(2)] = T 42019$  ...

Avec ces nouvelles notations, un indice non chaîné calculé sur le parc p(i) en base  $t_0[p(i)]$ , pour une zone z et une période t, est représenté par  $I_{t/t_0[p(i)]}^{p(i)}(z)$ .

L'indice chaîné en base 100 pour une zone z et une période t vaut alors :

$$I_{t/t_0}(z) = I_{t/t_0[p(i)]}^{p(i)}(z) \cdot \prod_{j=1}^{i-1} I_{t_d[p(j+1)]/t_0[p(j)]}^{p(j)}(z) \cdot I_{2017T4/t_0}(z)$$

<sup>37</sup> Les évolutions de prix décrites par les indices sont les mêmes quelle que soit la base 100 choisie et il est facilement possible de changer de base 100 postérieurement aux calculs. Son nom vient du fait que les indices précédemment décrits sont multipliés par 100, ce uniquement pour en faciliter la lecture.

<sup>38</sup> En d'autres termes, la base est choisie telle que la moyenne des quatre indices trimestriels bruts de 2015 vaille 100. C'est notamment en moyenne annuelle 2015 qu'Eurostat diffuse différents indices européens.

<sup>39</sup> Les notaires calculent également des indices mensuels sur trimestre glissant (chaque indice mensuel est calculé avec trois mois de données). Pour une valeur du parc = 100 en  $t_0$ , l'indice en  $t_0'$  prendrait la valeur

$$100 \cdot \frac{\hat{W}_{s,t_0'}}{\hat{W}_{s,t_0}} \quad \text{et l'indice en } t_0'' \text{ prendrait la valeur} \quad 100 \cdot \frac{\hat{W}_{s,t_0''}}{\hat{W}_{s,t_0}}$$

Les indices de janvier, avril, juillet et octobre sont chaînés sur les bases  $t_0'$  des différents parcs. De même, les indices de février, mai, août et novembre sont chaînés sur les bases  $t_0''$  des parcs.

### Exemple numérique pour une zone z :

L'indice au 2017T4 de la série préexistante en base 100 (moyenne annuelle 2015) est connu.

Supposons qu'il vaille 101 :  $I_{2017T4/0}(z) = 101$

Lorsqu'on se place en 2018T1 : l'indice non chaîné au 2018T1, en base 2017T4, est calculé d'après le parc  $p(1)=[2015-2016]$ .

Supposons qu'il vaille 1,03 :  $I_{2018T1/2017T4}^{[2015-2016]}(z) = 1,03$

L'indice chaîné en base 100 au 2018T1 vaut alors :

$$I_{2018T1/0}(z) = I_{2018T1/2017T4}^{[2015-2016]}(z) \cdot I_{2017T4/0}(z) = 1,03 * 101 = 104,03$$

De même, lorsqu'on se place en 2019T4 : l'indice non chaîné au 2019T4, en base 2017T4, est calculé d'après le parc  $p(1)=[2015-2016]$ .

Supposons qu'il vaille 1,1 :  $I_{2019T4/2017T4}^{[2015-2016]}(z) = 1,1$

L'indice chaîné en base 100 au 2019T4 vaut alors :

$$I_{2019T4/0}(z) = I_{2019T4/2017T4}^{[2015-2016]}(z) \cdot I_{2017T4/0}(z) = 1,1 * 101 = 111,1$$

Plaçons-nous cette fois au 2020T1 : l'indice non chaîné au 2020T1, en base 2019T4, est calculé d'après le parc  $p(2)=[2017-2018]$ .

Supposons qu'il vaille 0,95 :  $I_{2020T1/2019T4}^{[2017-2018]}(z) = 0,95$

L'indice chaîné en base 100 au 2020T1 vaut alors :

$$I_{2020T1/0}(z) = I_{2020T1/2019T4}^{[2017-2018]}(z) \cdot I_{2019T4/0}(z) = 0,95 * 111,1 = 105,545$$

équivalent à :

$$I_{2020T1/0}(z) = I_{2020T1/2019T4}^{[2017-2018]}(z) \cdot I_{2019T4/0}(z) = 0,95 * 111,1 = 105,545$$

## **4.6 – Désaisonnalisation des indices**

Les séries d'indices sont publiées en deux versions : une version brute, obtenue après l'étape de chaînage, et une version corrigée des variations saisonnières visant à présenter l'évolution des prix après élimination de l'effet des fluctuations saisonnières au cours de l'année. On observe en effet une saisonnalité sur les indices des prix des logements anciens, même si elle n'est pas très marquée : de manière générale, on observe une légère augmentation des prix au troisième trimestre de l'année, soit pendant la période des vacances plus propice aux déménagements en vue de la rentrée scolaire.

La correction des variations saisonnières (CVS) des séries s'effectue avec la méthode X12-ARIMA<sup>40</sup>, *via* le logiciel Demetra, sur les séries d'indices trimestriels glissants diffusées

---

<sup>40</sup> Cette méthode repose sur un principe itératif d'estimation des différentes composantes, cette estimation étant faite à chaque étape grâce à des moyennes mobiles adéquates.

mensuellement par les notaires<sup>41</sup>. Cette méthode permet notamment de prolonger les séries à l'aide de prévisions par des modèles ARIMA, ce qui permet également de bonnes estimations des coefficients CVS sur les périodes postérieures à l'estimation. Il est également possible d'appliquer une correction pour jours ouvrables, mais les indices n'en font pas l'objet, la composition en jours ouvrables des trimestres n'ayant pas d'impact notable sur la mesure des prix.

Les techniques de désaisonnalisation reposent sur le principe de décomposition des séries en différentes parties : une composante de tendance traduisant l'évolution de fond de la série, une composante saisonnière représentant les fluctuations au sein de l'année et une composante irrégulière comprenant les fluctuations résiduelles. Pour le calcul des coefficients de désaisonnalisation, on peut choisir un modèle additif ou multiplicatif. Le modèle additif suppose que les composantes de la série sont indépendantes les unes des autres. Ainsi, le niveau des variations saisonnières est indépendant du niveau de la série. Le modèle multiplicatif suppose, quant à lui, que les composantes de la série sont dépendantes les unes des autres. Pour les prix des logements, on préfère utiliser le schéma multiplicatif sur l'ensemble des séries<sup>42</sup>.

Pour les indices agrégés, il existe deux méthodes de désaisonnalisation. Dans la première, dite directe, on désaisonnalise chaque série indépendamment, quel que soit son niveau d'agrégation. La méthode indirecte, au contraire, consiste à désaisonnaliser d'abord les séries élémentaires, puis à les agréger entre elles pour obtenir les séries désaisonnalisées au niveau supérieur. D'un point de vue théorique, aucune méthode n'est meilleure que l'autre ; cependant, la méthode indirecte présente l'avantage d'assurer la cohérence entre les évolutions sur les différents niveaux d'agrégation. Les désaisonnalisations sont donc ici réalisées de manière indirecte. Pour l'agrégation, les indices élémentaires CVS sont pondérés de la même manière que les indices bruts (cf. partie 4.4). Les séries élémentaires utilisées pour le calcul des séries agrégées sont fournies dans la figure 4-3.

**Figure 4-3 : Séries élémentaires utilisées pour le calcul des séries agrégées CVS**

Séries agrégées	Séries élémentaires correspondantes
<b>Province</b>	
Appartements des agglomérations de plus de 10 000 habitants	- Appartements villes-centres - Appartements banlieue
Appartements	- Appartements villes-centres - Appartements banlieue - Appartements rural
Appartements et maisons	- Appartements villes-centres - Appartements banlieue

<sup>41</sup> Les modèles CVS sont déterminés sur les séries d'indices « trimestriels glissants » et les coefficients saisonniers sont donc mensuels. Pour les CVS trimestrielles, les coefficients saisonniers appliqués sont ceux du dernier mois de chaque trimestre.

<sup>42</sup> Le modèle optimal n'est pas forcément le même sur les différentes zones de calcul. Pour certaines séries, le schéma additif peut s'avérer meilleur, mais les différences entre les séries CVS produites avec un schéma additif et celles produites avec un schéma multiplicatif sont tellement faibles qu'on préfère appliquer ce dernier sur l'ensemble des séries par souci d'homogénéité de traitement.



	<ul style="list-style-type: none"> <li>- Appartements rural</li> <li>- Maisons</li> </ul>
Appartements et maisons : Rhône-Alpes	<ul style="list-style-type: none"> <li>- Appartements Rhône-Alpes</li> <li>- Maisons Rhône-Alpes</li> </ul>
Appartements et maisons : Provence-Alpes-Côte d'Azur	<ul style="list-style-type: none"> <li>- Appartements Provence-Alpes-Côte d'Azur</li> <li>- Maisons Provence-Alpes-Côte d'Azur</li> </ul>
Appartements et maisons : Nord-Pas-de-Calais	<ul style="list-style-type: none"> <li>- Appartements Nord-Pas-de-Calais</li> <li>- Maisons Nord-Pas-de-Calais</li> </ul>
Appartements et maisons : Auvergne-Rhône-Alpes	- Appartements Provence-Alpes-Côte d'Azur
	- Maisons Provence-Alpes-Côte d'Azur
	- Appartements Auvergne
	- Maisons Auvergne
Appartements et maisons : Hauts-de-France	- Appartements Nord-Pas-de-Calais
	- Maisons Nord-Pas-de-Calais
	- Appartements Picardie
	- Maisons Picardie
<b>Île-de-France</b>	
Appartements	- Appartements par département
Appartements Île-de-France hors Paris	- Appartements par département
Appartements grande couronne	- Appartements par département (départ. 77, 78, 91 et 95)
Appartements petite couronne	- Appartements par département (départ. 92, 93 et 94)
Maisons	- Maisons par département
Maisons grande couronne	- Maisons par département (départ. 77, 78, 91 et 95)
Maisons petite couronne	- Maisons par département (départ. 92, 93 et 94)
Appartements et maisons	<ul style="list-style-type: none"> <li>- Appartements par département</li> <li>- Maisons par département</li> </ul>
<b>France métropolitaine</b>	
Appartements	<ul style="list-style-type: none"> <li>- Appartements villes-centres</li> <li>- Appartements banlieue</li> <li>- Appartements rural</li> <li>- Appartements Île-de-France par département</li> </ul>
Maisons	<ul style="list-style-type: none"> <li>- Maisons province</li> <li>- Maisons Île-de-France par département</li> </ul>
Appartements et maisons	<ul style="list-style-type: none"> <li>- Appartements villes-centres</li> <li>- Appartements banlieue</li> <li>- Appartements rural</li> <li>- Maisons province</li> <li>- Appartements Île-de-France par département</li> <li>- Maisons Île-de-France par département</li> </ul>

France hors Mayotte	
Appartements	<ul style="list-style-type: none"> <li>- Appartements villes-centres</li> <li>- Appartements banlieue</li> <li>- Appartements rural</li> <li>- Appartements Drom hors Mayotte</li> <li>- Appartements Île-de-France par département</li> </ul>
Maisons	<ul style="list-style-type: none"> <li>- Maisons Province</li> <li>- Maisons Drom hors Mayotte</li> <li>- Maisons Île-de-France par département</li> </ul>
Appartements et maisons	<ul style="list-style-type: none"> <li>- Appartements villes-centres</li> <li>- Appartements banlieue</li> <li>- Appartements rural</li> <li>- Appartements Drom hors Mayotte</li> <li>- Maisons province</li> <li>- Maisons Drom hors Mayotte</li> <li>- Appartements Île-de-France par département</li> <li>- Maisons Île-de-France par département</li> </ul>

Les coefficients saisonniers sont actualisés une fois par an pour prendre en compte les derniers indices dans les séries brutes utilisées pour le calcul. Leur mise à jour a lieu courant juillet une fois disponibles les données définitives du 4<sup>e</sup> trimestre de la dernière année écoulée. Une mise à jour annuelle des CVS apparaît préférable à une mise jour trimestrielle ou mensuelle, dans la mesure où cela permet de limiter les révisions de coefficients, d'autant plus que ces derniers peuvent se révéler fragiles s'ils sont estimés à partir de données non définitives.

À l'occasion de la révision annuelle des coefficients CVS, on calcule également, *via* des projections, les coefficients CVS utilisés pour le calcul d'indices postérieurs à la période de désaisonnalisation (ces coefficients seront ensuite actualisés sur données réelles lors d'une prochaine mise à jour). Ces projections sont faites sur deux ans.

Un exemple de série brute / CVS est fourni en figure 4-4.

**Figure 4-4 : Indices bruts et CVS des maisons de province**

