

6 – Passage de la version 3 à la version 4

La version 4 des indices est entrée en application en 2018. Elle a été définie à l'occasion d'une expertise de la méthodologie existante. Ce travail est effectué par le Conseil scientifique des indices Notaires-Insee périodiquement, afin de faire évoluer la méthode en fonction de l'état des marchés. La version 4 ne provient donc pas d'une refonte complète du processus de production des indices mais d'une actualisation du dispositif utilisé en version 3⁴⁶, auquel ont été apportés un certain nombre d'ajustements. Ceux-ci sont décrits ci-dessous :

- Champ des indices :
 - le champ géographique comprend dorénavant la Corse et les Drom (Mayotte n'étant pas inclus) ;
 - les chalets ont été ajoutés dans le champ des indices des prix des maisons ;
 - les logements dont le nombre de pièces n'est pas renseigné sont maintenant supprimés. Ce traitement était déjà effectué en province en version 3, mais le nombre de pièces était estimé en Île-de-France, lorsque la surface était renseignée (l'observation était supprimée sinon), en fonction de la surface, de la strate et du type de bien. Il a été décidé d'homogénéiser les traitements en supprimant ces observations sur l'ensemble du territoire, étant donné le faible volume concerné (moins de 0,5 % des observations), plutôt que d'en faire une estimation incertaine ;
 - le choix des variables fiscales utilisées pour la définition des logements anciens a été homogénéisé entre l'Île-de-France et la province. L'ancien correspond toujours à la définition fiscale : logements non soumis en partie ou en totalité à la TVA. Les modifications dans la législation française en matière de TVA immobilière peuvent donc avoir un impact sur le champ de l'ancien utilisé pour les indices Notaires-Insee. Par exemple, les biens de moins de 5 ans revendus, auparavant soumis à la TVA et qui ne le sont plus depuis 2011, n'étaient pas comptés dans les logements anciens alors qu'ils le sont maintenant. Cela a pour conséquence d'augmenter les volumes de l'ordre de 2 % (calcul effectué sur le champ des logements anciens de province en 2017).

⁴⁶ Pour plus d'informations, la méthode de calcul des indices en version 3 est décrite dans l'*Insee Méthodes* n° 128 de juillet 2014 (<https://www.insee.fr/fr/information/2569926>).

- Définition du parc de référence : en version 3, le parc de référence correspondait au panier de biens des indices et était composé des transactions du champ sur la période de référence, auxquelles on retirait les 5 % d'observations avec les prix (au m² pour les appartements et total pour les maisons) les plus élevés et les 5 % d'observations avec les prix les plus faibles. L'équation hédonique était cependant estimée sur le « parc d'estimation » correspondant à l'échantillon utilisé pour l'estimation et composé de l'ensemble des transactions du champ auxquelles on retirait les valeurs atypiques détectées avec les résidus standardisés. Ces deux définitions avaient pour but d'exclure les valeurs extrêmes, avec des approches différentes. Dans la pratique, ces deux parcs étaient très proches⁴⁷, on a préféré conserver en version 4 un seul parc endossant ces deux rôles (appelé « parc de référence » mais dont la définition correspond à celle du « parc d'estimation » de la version 3).
- Mise à jour de la stratification en Île-de-France : de nouvelles strates ont été définies en Île-de-France ; la méthode de classification a été enrichie par rapport à la version 3 en ajoutant notamment des variables relatives aux communes, externes aux bases notariales (voir partie 3) et en modifiant les critères de volume afin d'obtenir des strates plus fines. Elle a de plus été éditée sur des données récentes.
- Le choix des variables candidates à l'inclusion dans les modèles : les variables incluses dans le modèle ont été homogénéisées entre Île-de-France et province. Des variables de distances communales (distance à la commune de plus de 50 000 habitants la plus proche pour la province, distance à Paris pour l'Île-de-France) et l'étiquette énergie ont été ajoutées aux variables candidates pour les appartements comme pour les maisons. Par ailleurs, le nombre de bâtiments, utilisé en version 3 uniquement pour les indices des maisons d'Île-de-France, a été retiré. Enfin, l'état du bien, pour les appartements et pour les maisons, ainsi que la présence d'une terrasse ou d'un balcon (ou d'une loggia depuis la version 4) pour les appartements, étaient utilisés seulement en province en version 3 et sont maintenant aussi utilisés en Île-de-France.
- Le traitement de la non-réponse des variables a également été homogénéisé : l'imputation des variables « époque de construction » et « présence d'un ascenseur » pour les appartements à partir de l'historique des ventes (voir partie 2.4) a été mis en place pour la province. En version 3, ce traitement n'était effectué que pour les appartements d'Île-de-France. Par ailleurs, le choix des modalités et le traitement de la non-réponse de certaines variables ont été actualisés. Ces changements sont présentés dans les figures 6-1 et 6-2.

⁴⁷ Des tests ont été effectués sur le parc des appartements anciens de province 2011-2012. Sur les indices 2014, le maximum des différences départementales est de 0,25 point, ce qui a très peu d'impact sur les évolutions d'indices.

Figure 6-1 : Synthèse des changements opérés sur le traitement de la non-réponse – Appartements

Appartements		
Modification	Version 3	Version 4
Le modèle de régression utilisé pour l'imputation de la surface habitable a été harmonisé entre l'Île-de-France et la province.	Province : les observations avec surface manquante étaient supprimées lors de la construction des parcs d'estimation ou de référence. En période courante, estimation en fonction du nombre de salles de bains, du nombre de pièces et de l'époque de construction. Île-de-France : estimation en fonction du nombre de salles de bains, du nombre de pièces, de l'époque de construction, de l'étage et du nombre de garages.	Province et Île-de-France : estimation de la surface (pour la construction du parc de référence ainsi qu'en période courante), en fonction du nombre de salles de bains, du nombre de pièces, de l'époque de construction, de l'étage et du nombre de garages.
Actualisation des modalités de l' époque de construction .	Modalités : Époque A : avant 1850 Époque B : de 1850 à 1913 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 Époque X : non renseigné	Modalités : 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 G : de 1992 à 2000 Époque HI : depuis 2001 Époque X : non renseigné
Modification de l'imputation du nombre de salles de bains .	Modalités : (0, 1, 2 ou plus, non renseigné) Imputation déterministe à non renseigné	Modalités : (0, 1, 2 ou plus) Imputation déterministe à 1
L'imputation de la présence d'une cave est modifiée.	Modalités : (Pas de cave, Au moins une cave, Non renseigné) Imputation déterministe à "Non renseigné"	Modalités : (Pas de cave, Au moins une cave) Imputation déterministe à "Pas de cave"
La présence d'une terrasse ou d'un balcon est élargie à la présence d'une terrasse ou d'un balcon ou d'une loggia . L'imputation est de plus modifiée.	Modalités : (Sans terrasse ni balcon, Avec terrasse ou balcon) Imputation déterministe à "Sans terrasse ni balcon"	Modalités : (Sans terrasse ni balcon ni loggia, Avec terrasse ou balcon ou loggia, Non renseigné) Imputation déterministe à "Non renseigné"

Appartements		
Modification	Version 3	Version 4
Les modalités des classes définies par le croisement entre le nombre de pièces et la surface moyenne par pièce ont été modifiées.	Modalités : Studio - petit : <20 m ² Studio - moyen : entre 20 et 30 m ² Studio - grand : >30 m ² 2 pièces - petit : <17 m ² 2 pièces - moyen : entre 17 et 24 m ² 2 pièces - grand : >24 m ² 3 pièces - petit : <18 m ² 3 pièces - moyen : entre 18 et 22 m ² 3 pièces - grand : >22 m ² 4 pièces ou + - petit : <17 m ² 4 pièces ou + - moyen : entre 17 et 21 m ² 4 pièces ou + - grand : >21 m ²	Modalités : Studio - petit : <20 m ² Studio - moyen : entre 20 et 30 m ² Studio - grand : >30 m ² 2 pièces - petit : <17 m ² 2 pièces - moyen : entre 17 et 24 m ² 2 pièces - grand : >24 m ² 3 pièces - petit : <18 m ² 3 pièces - moyen : entre 18 et 23 m ² 3 pièces - grand : >23 m ² 4 pièces - petit : <17 m ² 4 pièces - moyen : entre 17 et 21 m ² 4 pièces - grand : >21 m ² 5 pièces ou + - petit : <16 m ² 5 pièces ou + - moyen : entre 16 et 22 m ² 5 pièces ou + - grand : >22 m ²

Figure 6-2 : Synthèse des changements opérés sur le traitement de la non-réponse – Maisons

Maisons		
Modification	Version 3	Version 4
Le modèle de régression utilisé pour l'imputation de la surface habitable a été harmonisé entre l'Île-de-France et la province.	Province : les observations avec surface manquante étaient supprimées lors de la construction des parcs d'estimation ou de référence. En période courante, estimation en fonction du nombre de salles de bains, du nombre de pièces et de l'époque de construction. Île-de-France : estimation en fonction du nombre de salles de bains, du nombre de pièces, de l'époque de construction, du nombre de niveaux et du nombre de garages.	Province et Île-de-France : estimation de la surface (pour la construction du parc de référence ainsi qu'en période courante), en fonction du nombre de salles de bains, du nombre de pièces, de l'époque de construction et du nombre de niveaux.
Les modalités du nombre de niveaux ont été harmonisées entre la province et l'Île-de-France	Modalités province : (1, 2, 3 ou plus) Modalités IDF (1, 2 ou plus)	Modalités province et IDF : (1, 2 ou plus)
Les modalités et l'imputation du nombre de garages ont été modifiées.	Modalités : (0, 1, 2 ou plus) Imputation déterministe à 1 pour la province, à 0 pour l'IDF	Modalités : (0, 1, 2 ou plus, Non renseigné) Imputation déterministe à "Non renseigné" pour la province et l'IDF

Maisons		
Modification	Version 3	Version 4
Actualisation des modalités de l' époque de construction	Modalités : Époque AB : avant 1913 (inclus) Époque C : de 1914 à 1947 Époque D : de 1948 à 1969 Époque E : de 1970 à 1980 Époque FGHI : depuis 1981 Époque X : non renseigné	Modalités : 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 Époque X : non renseigné
La présence d'une cave et celle d'un sous-sol sont regroupées en présence d'une cave ou d'un sous-sol et harmonisées entre la province et l'Île-de-France. L'imputation est modifiée.	Modalités province : (pas de sous-sol, présence d'un sous-sol) Imputation déterministe à "pas de sous-sol" Modalités IDF : (pas de cave, présence d'une cave) Imputation déterministe à "pas de cave"	Modalités province et IDF : (pas de sous-sol ni de cave, présence d'un sous-sol ou d'une cave) Imputation déterministe à "pas de sous-sol ni de cave"
Modification de l'imputation du nombre de salles de bains.	Modalités : (0, 1, 2 ou plus) Imputation déterministe à 0	Modalités : (0, 1, 2 ou plus) Imputation déterministe à 1

- Correction des variations saisonnières : deux années de prévisions sont calculées avec les modèles ARIMA, contre une en version 3.
- Base 100 : les indices sont maintenant diffusés en base *moyenne annuelle 2015=100*, contre *2010 T1=100* en version 3.

Afin d'évaluer l'écart global entre les versions 3 et 4, un historique en version 4 a été recalculé. Les figures 6-3, 6-4 et 6-5 présentent la comparaison des indices entre les deux versions du premier trimestre 2010 au premier trimestre 2018.

Figure 6-3 : Indice des prix des logements anciens en France métropolitaine V3 / V4

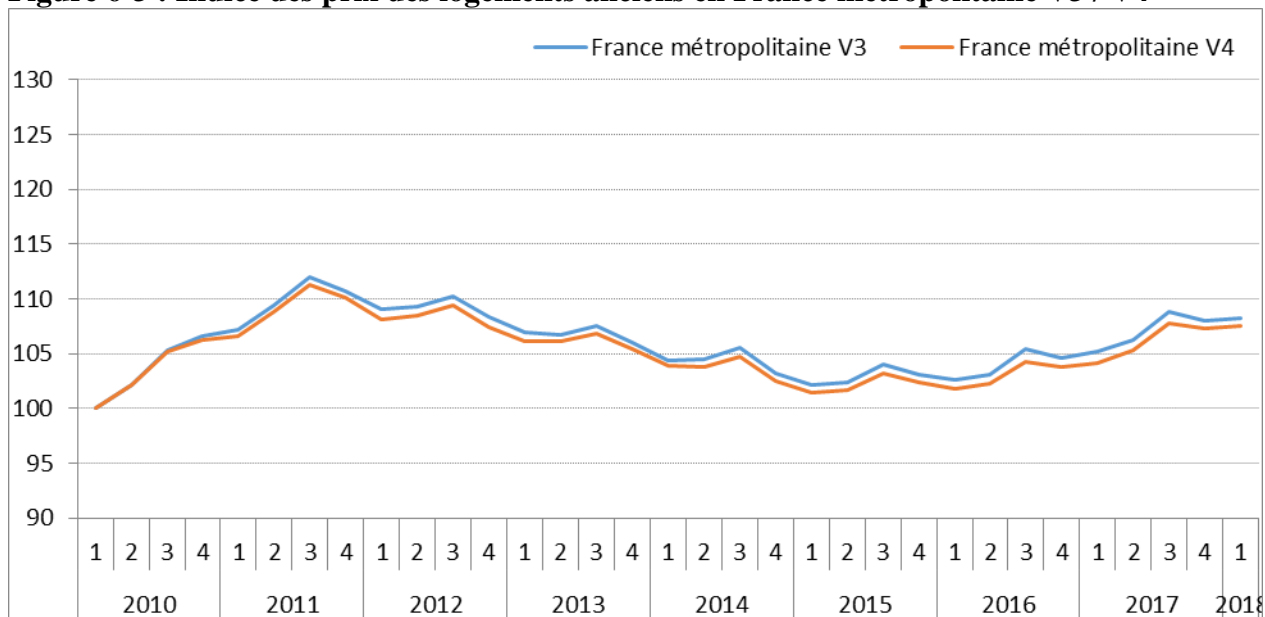


Figure 6-4 : Indice des prix des logements anciens en Île-de-France V3 / V4

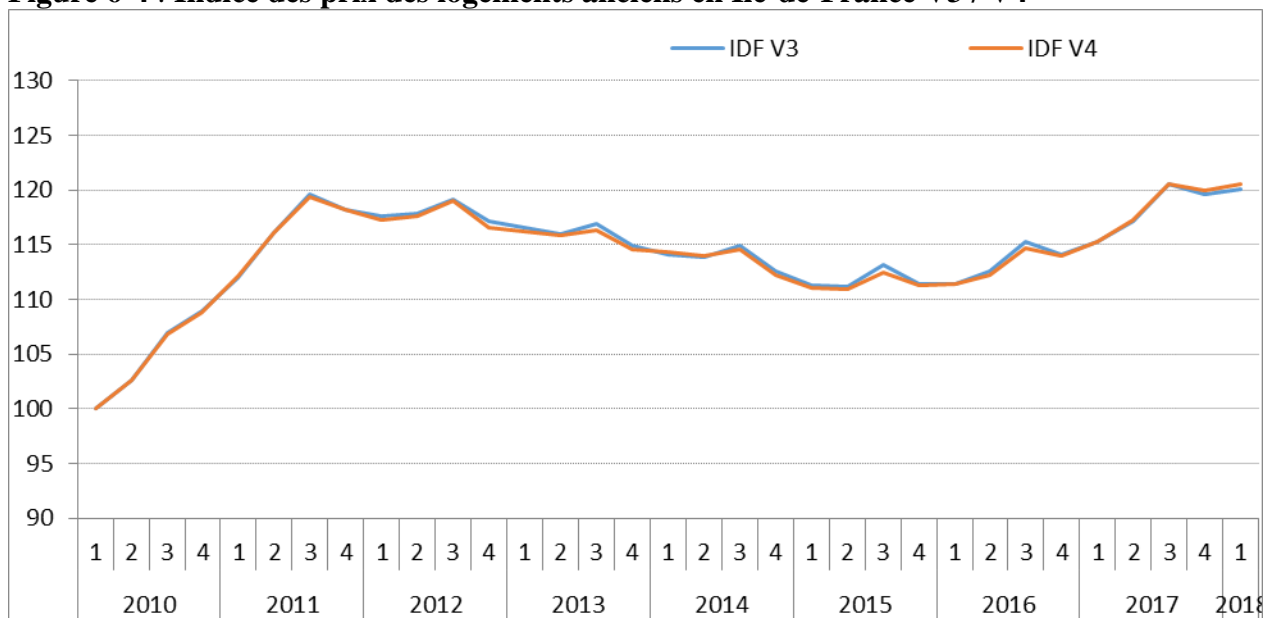


Figure 6-5 : Indice des prix des logements anciens en province V3 / V4

