

# Les indices Notaires-Insee

## Introduction

Des indices des prix des logements répondent à des besoins divers : de la simple analyse de l'évolution des prix à l'analyse du risque des prêteurs, en passant par la conduite de la politique monétaire et la comptabilité nationale.

L'indice le plus simple, fondé sur la moyenne des prix des ventes observées sur une période donnée, risque d'être biaisé puisque les ventes observées sur cette période ne sont pas nécessairement représentatives de l'ensemble des biens que l'on souhaite étudier. Un indice basé sur le prix médian des transactions est moins sensible aux valeurs extrêmes observées, mais encore sujet à des biais, la qualité des biens vendus pouvant évoluer dans le temps. Dans certains pays comme les États-Unis où la mobilité résidentielle des propriétaires est élevée, la méthode des ventes répétées est utilisée. Un indice des ventes répétées est calculé en comparant le prix observé du même logement lors de deux ventes successives en supposant que sa qualité reste la même. Toutefois, il est difficile et coûteux de vérifier que la qualité des biens est réellement inchangée. De plus, cette méthode n'est pas à l'abri de biais de sélection, car les biens échangés, et en particulier ceux échangés plusieurs fois, peuvent ne pas être représentatifs du marché dans son ensemble.

L'approche hédonique, pratiquée en France, repose sur la caractérisation des logements au travers d'un nombre déterminé de caractéristiques observées, dont chacune, avec sa propre valeur, contribue à la valeur totale. La combinaison des caractéristiques est donc supposée déterminer la valeur du bien. Le modèle est estimé grâce aux prix et aux caractéristiques des logements échangés sur une période dite « d'estimation ». Il est ensuite utilisé pour mesurer au fil du temps la valeur estimée d'un panier choisi de logements. Le modèle permet donc d'estimer la valeur d'un logement, même s'il n'a pas été vendu sur la période d'estimation. Une telle approche résout donc les trois problèmes qui font la difficulté de construire un indice des prix : (1) l'évolution de la qualité des logements : ces derniers ont de nombreuses caractéristiques et certaines d'entre elles peuvent évoluer dans le temps, (2) le fait que le prix d'un logement donné n'est observé que très rarement et (3) le fait que le « panier » de logements à suivre doit être défini avec précision, puisque le type de biens échangés et le stock de logements évoluent dans le temps.

Une telle modélisation a un coût, puisque les prix et les caractéristiques des ventes doivent être observés et enregistrés. Dans les pays où la loi exige que les ventes de logements soient enregistrées devant un notaire, les données sont plus facilement accessibles. Tel est le cas en France, où les données sont collectées et centralisées depuis 1994 ; un indice hédonique trimestriel des prix est ainsi calculé depuis 1998. Les notaires sont responsables de la collecte de données et du calcul des indices, tandis que le Conseil Scientifique définit la démarche scientifique de la méthode hédonique et l'Insee en vérifie la mise en œuvre et valide les résultats.

## Le cadre institutionnel français

Le cadre institutionnel français est caractérisé par un réseau de notaires, qui ont le monopole de l'enregistrement des transactions immobilières, et un organisme statistique national, l'Insee. Le notaire vérifie l'existence des droits de propriété, rédige le contrat de vente, envoie les documents de propriété à la conservation des hypothèques et reçoit les droits de mutation au nom de l'État.

De par sa fonction, le notaire a accès au prix de la transaction et à quelques caractéristiques des logements qui doivent être inscrites dans le contrat de vente. En outre, chaque notaire doit envoyer les informations sur le prix de la vente aux autorités, puisque les droits perçus sur la vente sont fonctions de son prix. En principe, les données couvrent toutes les ventes, fournissent des prix de transaction réels de façon régulière et sur une longue période de temps. La disponibilité rapide des informations est assurée par le délai fixé aux notaires pour faire suivre les taxes, délai réduit à un mois depuis quelques années.

L'Insee est chargé de fournir les statistiques officielles, y compris les comptes nationaux et les indices de prix. Jusqu'à la fin des années 1990, il ne publiait pas d'indices des prix des logements sauf pour la ville de Paris qui disposait depuis 1983 d'une série d'indices des appartements anciens.

En 1997, le Conseil supérieur du notariat (CSN) a souhaité créer un indice de prix pour les logements situés en dehors de la région parisienne. Il s'est alors tourné vers l'Insee et son centre de recherches (Crest) pour des conseils méthodologiques. Des accords formels ont été signés en 1998 et 1999 entre les notaires et l'Insee. Les notaires recueillent les données et calculent les indices chaque trimestre. De nombreux indices sont ainsi mis à disposition du public gratuitement.

Le travail qui mène de la rédaction d'un contrat de vente à la publication d'un indice est important. Chacun des (environ) 5 500 offices notariaux est invité à faire parvenir une copie de chaque acte de vente, soit par courrier, soit par envoi dématérialisé comme cela se développe depuis 2010. La vitesse à laquelle chaque office envoie les données est cruciale pour la qualité de l'indice : un acte qui arrive en retard pour le trimestre considéré ne sera pas utilisé pour le calcul de l'indice. D'où l'intérêt des envois dématérialisés, qui évitent les délais de regroupement d'actes et les délais postaux ; ils représentent désormais la majorité des envois.

Il existe deux bases de données : la base Bien, gérée par la CINP, qui couvre l'Île-de-France (un peu moins de 20 % des transactions) et la base Perval, gérée par le groupe ADSN pour le CSN, qui couvre le reste de la France (un peu plus de 80 % des transactions).

La grande majorité des transactions immobilières portent sur des logements, maisons et appartements, à parts relativement égales. Une partie des transactions restantes portent sur d'autres types de biens immobiliers (garages, locaux d'activité, immeubles entiers, terrains,...), mais ceux-ci ne donnent pas lieu au calcul d'indices de prix.

L'envoi des données par les notaires s'effectuait jusqu'à fin 2016 sur une base volontaire, le taux de couverture de la base de données n'était donc pas de 100 %. La loi oblige depuis le 1<sup>er</sup> janvier 2017 les notaires à alimenter les bases, aussi les taux de couverture ont progressé même si la phase de montée en charge n'est pas terminée. Il est ainsi autour de 60 % (75 % en Île-de-France et 55 % en province) fin 2017. Ce taux de couverture est estimé par comparaison du montant des droits de mutation figurant dans la base de données avec celui perçu par la direction générale des Finances publiques (DGFIP).

Un taux de couverture de 100 % n'est cependant pas nécessaire pour le calcul d'un indice hédonique. En effet, le modèle hédonique est estimé sur une période de référence à partir d'un panier fixe de biens, défini sur un nombre suffisant d'années de transactions. Ce nombre d'années (2 ans dans les dernières versions des indices Notaires-Insee) est déterminé avec un double objectif : d'abord, celui de contenir suffisamment d'observations afin que les estimations soient de bonne qualité ; ensuite, que la période soit récente et relativement courte, afin que les évolutions sur la valorisation des biens immobiliers soient correctement et rapidement prises en compte sur les estimations.

La base de données utilisée dans le cadre des indices est rendue anonyme pour se conformer aux recommandations de la Cnil. Les caractéristiques de localisation sont le code Insee de la commune ainsi que, dans certains cas (grandes villes, littoral, montagne,...) un code de quartier (tels les 80 quartiers administratifs de Paris).

Maisons et appartements sont traités séparément. La localisation est très importante, non seulement en termes d'équipements et d'aménités de quartier, mais également en termes de caractéristiques de construction. Par exemple, la construction haussmannienne du 19<sup>e</sup> siècle à Paris est de meilleure qualité que la construction de la même période dans d'autres zones. C'est pour cela que les régressions hédoniques sont estimées à un niveau local détaillé et que les modèles peuvent inclure des variables de quartier.

## La méthode hédonique

La méthode hédonique s'appuie sur un modèle économétrique liant le prix aux caractéristiques du logement. Le modèle est estimé à partir des prix et caractéristiques des biens vendus sur une période dite d'estimation. Puis il est utilisé pour ajuster la qualité des biens sur les périodes suivantes. Lorsque cela est bien fait, la qualité est contrôlée, la valeur d'un bien peut être estimée en l'absence d'une vente et le panier de biens suivi peut être défini et évoluer dans le temps.

Certains utilisent une approche hédonique alternative. Ils font une régression unique du prix de vente en ajoutant des indicatrices temporelles au modèle, ces indicatrices servant de base au calcul d'un indice. Cette méthode est en général appliquée sur deux périodes adjacentes, glissantes dans le temps. Elle est plus sensible aux biais induits par l'échantillonnage des ventes observées que la méthode consistant à évaluer un panier de biens fixe à chaque date, mais capte simplement l'évolution progressive des prix des caractéristiques.

Parmi tous les indices des prix des logements hédoniques, l'indice Notaires–Insee a la caractéristique unique de mesurer la variation des prix à partir d'une période de base (« de référence »), représentant un panier de référence. En outre, pour tenir compte de la variabilité du prix des caractéristiques dans le temps et l'espace, le modèle est estimé dans des zones où les prix sont supposés homogènes (strates). Idéalement, ces strates pourraient être de petits sous-marchés. En pratique, toutefois, le nombre de strates a été limité à environ 350 pour assurer un nombre suffisant de ventes au sein de la strate (plus de 440 par an).

Le modèle relie le prix des logements (au m<sup>2</sup> pour les appartements, total pour les maisons) à leurs caractéristiques et leur localisation (un quartier dans une strate). Les caractéristiques sont : le quartier, l'époque de construction, le nombre de pièces, le nombre de salles de bain, le nombre de garages ou parkings, l'état du bien, l'étiquette énergie, la distance à la commune de 50 000 habitants la plus proche (ou à Paris pour l'Île-de-France) ; et de plus, pour les appartements l'étage, la présence d'un balcon (terrasse ou loggia), d'une cave, d'un ascenseur à partir du 4<sup>e</sup> étage, la surface croisée avec le nombre de pièces ; et pour les maisons la présence d'un sous-sol, le nombre de niveaux, la surface du terrain et la surface habitable. La localisation du logement à l'intérieur de la strate, résumée par l'appartenance à un quartier (croisé avec le nombre de pièces), est considérée comme une de ses caractéristiques, au même titre que le nombre de pièces ou sa surface.

On estime un modèle par strate, en différenciant les appartements et les maisons, ce qui implique que toutes les variables sont implicitement en interaction avec la strate. Depuis la version en vigueur en 2011, on utilise une période de deux ans pour la validité des modèles et donc la stabilité des prix relatifs. En conséquence, les modèles sont estimés tous les deux ans, afin de concilier le calcul d'un indice à qualité constante, une bonne prise en compte de l'évolution du marché et les impératifs de la production de données conjoncturelles. Chaque trimestre, on calcule dans chaque strate le prix du logement de référence (*i.e.* celui étant pris en référence dans les régressions) à partir des coefficients du modèle appliqués sur les transactions du trimestre. L'évolution de la valeur de ce bien de référence nous donne donc l'évolution de l'indice élémentaire pour chacune des strates. Ces évolutions élémentaires ne sont pas publiées mais agrégées à des niveaux plus élevés à des fins de publication. Cette agrégation est faite sur la base de moyennes géométriques pondérées à des niveaux géographiques infra-départementaux et de moyennes arithmétiques pondérées à des niveaux départementaux et plus agrégés, les coefficients de pondération utilisés reflétant en valeur la structure du parc de logements.

À un niveau géographique plus élevé, l'indice des prix est un indice de Laspeyres, défini comme le rapport de la valeur estimée d'un panier de logements à sa valeur à la période de l'indice de base. Par exemple, l'indice de la « Province » mesure l'évolution de la valeur de l'ensemble du panier de référence de la province.

Il y a de forts effets saisonniers, en particulier pour les maisons. Ces effets sont liés à la mobilité résidentielle des familles, en particulier au calendrier de l'année scolaire. Des indices désaisonnalisés sont calculés et publiés par l'Insee.

### **Le nombre de transactions**

Le nombre de transactions est estimé par le Conseil général de l'environnement et du développement durable. La méthode de calcul consiste à croiser des informations détenues par la direction générale des Finances publiques (DGFIP) sur l'assiette de l'impôt, qui sont exhaustives (système MEDOC) mais ne distinguent pas les divers types de biens (immeubles à usage d'habitation, immeubles à usage professionnel, terrains), et des informations en provenance des bases de données notariales, qui distinguent les divers types de biens mais ne sont pas exhaustives.

Le nombre de transactions est estimé sur le champ de la France (hors Mayotte).

### **Conclusion : des indices qui s'adaptent**

La demande d'indices fiables des prix des logements est forte et croissante.

Les prix des logements et plus généralement de l'immobilier peuvent évoluer différemment des prix des autres produits de consommation et tous les acteurs économiques ont besoin de connaître ces différences. Les indices Notaires–Insee représentent donc un apport important pour les statistiques officielles.

Pour répondre à la demande, le processus de collecte et la méthodologie de calcul évoluent régulièrement pour fournir des instruments d'analyse toujours plus pertinents. Le travail des notaires est essentiel pour assurer d'une part l'exhaustivité et la représentativité des indicateurs et pour obtenir d'autre part une qualité optimale pour les informations saisies.

La loi du 28 mars 2011 confère d'ailleurs au notariat une mission de service public dans ce domaine. Les données produites conjointement par l'Insee et les notaires sont labellisées par l'Autorité de la statistique publique (ASP).