

Revue de littérature sur les indices hédoniques de prix immobiliers : l'expérience française des notaires et de l'INSEE

RÉSUMÉ CRITIQUE BASÉ SUR LAFERRÈRE (2003)

25 novembre 2025

Résumé

Cet article propose une revue de littérature centrée sur l'expérience française de construction d'indices hédoniques de prix immobiliers, telle que décrite par Anne Laferrère dans « Hedonic housing price indexes : the French experience ». Cette contribution s'inscrit dans le champ plus large de la mesure des prix immobiliers et des méthodes hédoniques appliquées aux biens hétérogènes. Nous présentons d'abord le contexte économique et statistique qui motive le recours à des indices de prix de l'immobilier fondés sur des modèles hédoniques, puis la spécificité de l'approche française : zonage fin, définition de paniers de référence et hypothèse de stabilité temporelle de la structure des prix. Nous discutons ensuite l'organisation institutionnelle des données (« notaires-INSEE »), les choix méthodologiques (stratification géographique, traitement des non-réponses, hypothèses d'échantillonnage) et leurs implications pour l'analyse économique et la politique publique. Enfin, nous mettons en évidence les forces et limites de cette approche, ainsi que les pistes de recherche qu'elle ouvre pour l'étude des marchés immobiliers et de la dynamique des prix d'actifs.

1 Introduction

La mesure de l'évolution des prix immobiliers soulève des problèmes spécifiques par rapport à celle des biens de consommation courante. Le logement constitue à la fois un bien de consommation durable et un actif patrimonial, représentant une part majeure de la richesse des ménages et un poste important dans leur budget. Les pouvoirs publics interviennent massivement sur ce marché (régulation des loyers, aides au logement, politique fiscale sur les transactions et les plus-values, zonage urbain, subventions à la construction, etc.), ce qui rend indispensable la disponibilité d'indicateurs fiables de l'évolution

des prix.¹

Or, le marché du logement est caractérisé par une forte hétérogénéité des biens (localisation, qualité, âge, équipements) et par la rareté des observations de prix : la plupart des logements ne sont vendus que très occasionnellement, et la plupart des transactions portent sur des biens anciens plutôt que neufs. La comparaison de simples prix moyens d'une période à l'autre mêle ainsi effets de variation de qualité et véritable inflation immobilière. Dans ce contexte, les méthodes hédoniques, qui décomposent le prix en contributions de caractéristiques observables, sont devenues un outil central pour la construction d'indices de prix de l'immobilier.

L'expérience française décrite par Laferrère constitue une mise en oeuvre à grande échelle de ces méthodes, fondée sur une collaboration originale entre les notaires, qui collectent et centralisent les données de transaction, et l'INSEE, qui développe et garantit la méthodologie économétrique. L'objet de cette revue de littérature est de synthétiser et de discuter cette contribution dans le cadre plus général des travaux sur les indices de prix hédoniques.

2 Cadre théorique et positionnement dans la littérature

2.1 Hétérogénéité des logements et approche hédonique

La littérature économique sur les indices de prix des biens hétérogènes s'appuie sur l'idée qu'un bien est défini par un vecteur de caractéristiques ou de services (surface, nombre de pièces, localisation, qualité de construction, équipements, etc.). Dans ce cadre, le prix d'un logement est interprété comme le résultat de la valorisation de ces attributs par le marché. Un modèle de prix hédonique relie ainsi le logarithme du prix (ou du prix au mètre carré) à un ensemble de variables décrivant ces caractéristiques.

Laferrère souligne deux hypothèses structurantes communes aux approches hédoniques d'indices de prix immobiliers :

- chaque logement est supposé décrit par une combinaison finie de caractéristiques, la « qualité » non observée ou émergente étant assimilée à une variation de prix plutôt qu'à un changement de nature du bien ;
- la relation entre prix et caractéristiques est supposée stable dans le temps, au moins sur une période de référence limitée, ce qui permet d'utiliser des coefficients estimés sur une période pour valoriser des paniers de logements à d'autres dates.

Cette stabilité de la fonction hédonique est une hypothèse forte : elle permet de réduire les coûts de recalibrage des modèles (pas de ré-estimation trimestrielle), mais elle suppose

1. Laferrère souligne notamment l'importance de ces indices pour la mesure de l'inflation, l'analyse de la distribution spatiale des prix et l'évaluation des politiques publiques.

que les préférences et la structure du marché ne se modifient pas trop rapidement. Dans la littérature internationale, de nombreux travaux discutent la robustesse de cette hypothèse en présence de chocs de demande, de changements réglementaires ou d'innovations dans l'offre de logements.

2.2 Indices de prix et problème de la qualité

La question centrale pour tout indice de prix est la séparation entre variation de prix « pure » et variation de qualité. Pour les biens ordinaires, des indices de type Laspeyres ou Paasche peuvent être construits à partir de paniers de consommation relativement stables. Pour le logement, la combinaison de l'hétérogénéité des biens, du caractère durable du capital immobilier et de la rareté des transactions rend cette approche très fragile.

Les modèles hédoniques résolvent ce problème en évaluant la valeur d'un « panier de référence » de logements, supposé constant en termes de caractéristiques, dont on suit le prix implicite dans le temps. L'idée, déjà présente dans la littérature sur les indices hédoniques, est de construire un indice de prix pour un ensemble fixe de logements, même si ces logements ne sont pas effectivement revendus chaque période.

Laferrère s'inscrit dans cette tradition, mais propose une mise en œuvre spécifique : la construction, pour chaque zone géographique, d'un stock de référence composé de toutes les ventes observées sur une période de plusieurs années, puis la valorisation trimestrielle de ce stock à partir des coefficients hédoniques et des transactions courantes.

3 Spécificités méthodologiques de l'expérience française

3.1 Stratification géographique et définition des zones

Une caractéristique majeure de l'approche française est le découpage géographique fin du territoire en zones de prix supposées homogènes et supportant chacune un modèle hédonique propre. En pratique, environ 300 zones élémentaires sont définies, couvrant l'ensemble du territoire métropolitain, avec des distinctions entre :

- Paris intra-muros, la *Petite Couronne* et la *Grande Couronne* en région parisienne ;
- grandes villes, villes moyennes, périphéries urbaines ;
- zones rurales et littorales, avec un traitement spécifique pour les stations de ski ou les marchés très atypiques.

L'objectif de cette stratification est de travailler dans des sous-marchés où les prix sont relativement homogènes et suivent des trajectoires parallèles, tout en conservant un nombre suffisant de transactions pour estimer des modèles économétriques robustes. Les zones sont définies à partir d'expertise locale (entretiens avec des professionnels de l'immo-

bilier) et de techniques statistiques de type analyse hiérarchique (arbres de classification) pour regrouper des quartiers similaires.

Par rapport à la littérature, cette approche illustre une tension classique entre granularité géographique (nécessaire pour capter les différences locales de prix et de structure des caractéristiques) et taille de l'échantillon (nécessaire pour obtenir des estimations stables des coefficients hédoniques).

3.2 Stocks de référence et modèle hédonique

Pour chaque zone, l'INSEE définit un *stock de référence* constitué de l'ensemble des ventes observées sur une période de référence de trois à cinq ans (par exemple, 1994–1996 pour les appartements en province). Les ventes dont le prix au mètre carré se situe dans les quantiles extrêmes (par exemple les 5 ou 10% supérieur et inférieur) sont exclues afin de réduire l'influence des valeurs aberrantes.

La taille de ces stocks de référence est importante : en moyenne environ 2 000 logements par zone, et plus de 600 000 logements pour la France entière. Ces logements sont décrits par un vecteur de caractéristiques relativement riche : surface, nombre de pièces, surface moyenne par pièce, période de construction, nombre de salles de bain, présence de parking ou de garage, étage et existence d'un ascenseur pour les appartements, nombre de niveaux, sous-sol et surface de terrain pour les maisons, etc.

Le modèle hédonique estimé dans chaque zone est de la forme :

$$\log p_i = \log p_0 + \sum_a \alpha_a Y_{a,i} + \sum_t \theta_t T_{t,i} + \sum_{k=1}^K \beta_k X_{k,i} + \varepsilon_i, \quad (1)$$

où p_i est le prix au mètre carré du logement i , $Y_{a,i}$ et $T_{t,i}$ sont des variables indicatrices pour l'année et le trimestre de vente, et $X_{k,i}$ représente les caractéristiques structurelles du logement. Le terme $\log p_0$ correspond au prix (au mètre carré) d'un logement de référence, défini par une combinaison de caractéristiques standard (par exemple, appartement de trois pièces, rez-de-chaussée, sans parking ni balcon, construit entre 1948 et 1980).

Les coefficients des variables de caractéristiques β_k sont interprétés comme des prix implicites relatifs à ce logement de référence. Les coefficients temporels α_a et θ_t permettent de reconstituer des indices temporels à partir de la valorisation du stock de référence. La qualité des régressions, mesurée par le R^2 , est jugée satisfaisante pour des données microéconomiques (individuelles), avec des valeurs typiquement comprises entre 0,25 et 0,40 dans la plupart des zones.

3.3 Valorisation du stock et construction de l'indice

Une fois le modèle hédonique estimé sur la période de référence, l'indice de prix est obtenu en comparant la valeur estimée du stock de référence à différentes dates. Pour

une période courante τ , le prix du logement de référence est approché en combinant les prix des logements effectivement vendus durant τ et les coefficients estimés sur la période de référence. On en déduit ensuite la valeur estimée de chaque logement du stock de référence, puis la valeur totale du stock dans la zone, en multipliant par la surface et en agrégeant sur l'ensemble des logements.

L'indice de prix pour la zone est alors défini comme le ratio entre la valeur estimée du stock à la date courante et sa valeur à la période de base. La construction de l'indice national s'obtient par agrégation (pondérée) des indices de zones.

Un point important dans la littérature est la fréquence de mise à jour des indices et des modèles. Dans l'expérience française, les premiers indices publiés étaient semestriels à Paris et annuels en province, avant de devenir purement trimestriels pour les grandes zones, avec une fréquence de révision des modèles et des stocks de référence de l'ordre de cinq ans. Cette architecture vise à concilier réactivité (construction d'indices fréquents) et stabilité (éviter de recalibrer en permanence les modèles hédoniques).

4 Organisation des données, couverture et aspects institutionnels

4.1 La base de données notaires-INSEE

L'originalité de l'expérience française réside aussi dans la structure institutionnelle de la collecte de données. Toutes les ventes immobilières doivent être enregistrées devant notaire, qui rédige l'acte et collecte les droits de mutation. À partir du début des années 1990, les notaires ont progressivement centralisé ces informations dans une base de données informatique couvrant l'ensemble des transactions immobilières (habitations, parkings, terrains, bâtiments neufs, etc.).

Pour le segment du logement, la base de données comprend plusieurs millions de transactions, dont environ 510 000 ventes de logements en une seule année au début des années 2000. La couverture n'est pas parfaite : elle est estimée à 80–90% pour Paris, autour de 80% pour la proche banlieue et environ 60–70% pour le reste de la France. Ces différences de couverture justifient un repondérage du stock de référence dans certaines zones.

Du point de vue de la littérature, l'expérience française illustre une stratégie réaliste : viser une couverture élevée mais pas nécessairement totale, en considérant que les transactions enregistrées constituent un échantillon suffisamment large et proche du hasard pour que l'approximation d'un indice de prix soit pertinente, surtout lorsque l'on travaille sur un stock de référence agrégé plutôt que sur les transactions de la seule période courante.

4.2 Traitement des non-réponses et qualité des données

La qualité de l'indice dépend crucialement de la qualité des variables de caractéristiques. Laferrère décrit des procédures de traitement des non-réponses : certaines observations sont exclues (par exemple, absence d'information sur le nombre de pièces), d'autres variables sont imputées à partir d'informations disponibles ou de distributions observées. La base de données est anonymisée, les adresses exactes n'étant pas publiées ni utilisées pour la construction des indices, même si la localisation est disponible via des unités statistiques (zones de recensement).

La littérature sur les indices hédoniques souligne le rôle critique des variables de localisation fine (quartier, accessibilité aux transports et équipements) et des caractéristiques de qualité du bâti. L'approche française atteint un compromis entre richesse de l'information et contraintes de collecte, mais ouvre aussi des pistes pour des enrichissements futurs (par exemple, ajout de variables de nuisances environnementales ou de qualité des écoles).

5 Apports, limites et perspectives de recherche

5.1 Apports majeurs

Du point de vue de la littérature sur les indices de prix immobiliers, l'expérience française apporte plusieurs contributions importantes :

- **Mise en oeuvre à grande échelle** d'indices hédoniques couvrant l'ensemble d'un pays, avec une stratification géographique fine (plus de 300 zones) et une fréquence de publication trimestrielle.
- **Utilisation systématique de stocks de référence** pour abstraire des changements de composition des transactions : l'indice suit le prix d'un panier fixe de logements, et non le prix moyen des transactions observées.
- **Architecture institutionnelle robuste** combinant expertise statistique publique (INSEE) et information de marché détenue par un acteur privé/professionnel (notaires), permettant une production régulière d'indices à faible coût pour le contribuable.
- **Stabilité empirique de la méthodologie** : la première révision des modèles et des stocks de référence ne modifie pas substantiellement le profil des indices, ce qui renforce la crédibilité de l'approche.

Ces apports rendent possible une série de travaux microéconomiques et macroéconomiques sur le marché du logement : étude de la mobilité résidentielle, des arbitrages localisation/surface, de la capitalisation des équipements publics dans les prix, ou encore comparaison de l'évolution des prix immobiliers et des prix à la consommation.

5.2 Limites et questions ouvertes

Plusieurs limites relevées dans la littérature, implicites ou explicites dans le texte de Laferrère, ouvrent des pistes de débat :

- **Hypothèse de stabilité temporelle des coefficients** : si les préférences des ménages ou la structure de l'offre évoluent rapidement (par exemple, apparition de nouvelles normes énergétiques), les prix implicites des caractéristiques peuvent changer, remettant en cause la validité des modèles sur une longue période.
- **Sélection des transactions** : même si la couverture est élevée, l'échantillon des ventes peut ne pas être aléatoire (propension à vendre plus élevée pour certains types de biens ou de ménages), ce qui pourrait biaiser l'évaluation de la valeur du stock de référence.
- **Variables de qualité non observées** : certaines dimensions de la qualité (bruit, qualité du voisinage, environnement local) ne sont pas directement prises en compte dans les modèles ; leur variation dans le temps peut alors être interprétée à tort comme une variation de prix.
- **Granularité géographique** : l'utilisation de zones relativement larges peut masquer des hétérogénéités intra-zones ; à l'inverse, un zonage trop fin souffre du manque de transactions. Le compromis retenu est raisonnable mais discutable, et susceptible d'être repensé à la lumière de nouvelles données ou méthodes (géostatistiques, modèles hiérarchiques, etc.).

5.3 Perspectives de recherche

Laferrère insiste sur le fait que la disponibilité d'indices hédoniques fiables, comparables dans l'espace et dans le temps, ouvre un large champ de recherches. Parmi les perspectives possibles :

- intégration des indices hédoniques dans des modèles structurels de choix résidentiels, de mobilité ou d'investissement immobilier ;
- analyse fine des cycles immobiliers et de leur interaction avec la politique monétaire et budgétaire ;
- enrichissement des modèles hédoniques par des variables issues de systèmes d'information géographique (proximité des transports, des écoles, des espaces verts) ou de bases de données environnementales ;
- comparaison internationale des méthodologies d'indices hédoniques, notamment en ce qui concerne la fréquence de ré-estimation des modèles et la gestion des changements structurels.

6 Conclusion

L'expérience française des indices hédoniques de prix du logement, telle que décrite par Laferrère, constitue un exemple abouti de mise en oeuvre d'une méthodologie économétrique sophistiquée au service de la statistique publique. Elle illustre la manière dont la combinaison d'une base de données riche en transactions, d'un appareil méthodologique cohérent (zonage, stocks de référence, modèles hédoniques) et d'une coopération institutionnelle stable (notaires-INSEE) peut produire des indicateurs de prix immobiliers fiables, utiles tant pour la recherche que pour la décision publique.

Du point de vue de la littérature, cette contribution confirme la pertinence des indices de prix hédoniques pour les biens hétérogènes, tout en rappelant l'importance de tester empiriquement les hypothèses sous-jacentes (stabilité des coefficients, représentativité des données, qualité de l'information sur les caractéristiques). Les développements récents en économétrie, en analyse spatiale et en sciences des données offrent de nombreuses possibilités pour prolonger et enrichir cette approche, en particulier dans un contexte où les marchés immobiliers sont au cœur des préoccupations macroéconomiques et sociales.