
Estimation de la valeur du patrimoine immobilier des ménages à partir de données exhaustives

Mathias André (*), Odran Bonnet (*), et Olivier Meslin (*)

(*) Insee, Département des Études Économiques

mathias.andre@insee.fr

odran.bonnet@insee.fr

olivier.meslin@insee.fr

Mots-clés. : Modèles hédoniques, *Machine Learning*, données administratives, données massives, prix immobilier, patrimoine, imputation.

Domaines. Données massives, appariements et données administratives

Résumé

L'essor de l'exploitation et de la diffusion des données administratives rend aujourd'hui possible la constitution de bases de données exhaustives sur le patrimoine immobilier des ménages. À l'instar de plusieurs études récentes menées notamment en Norvège et aux États-Unis ([3], [4]), ce travail vise à estimer et étudier le patrimoine immobilier des ménages en France. La constitution d'une base exhaustive sur le patrimoine immobilier des ménages permet d'étudier finement la distribution de ce dernier et notamment de mesurer l'importance des plus hauts patrimoines. Les sources traditionnelles, comme l'*enquête Patrimoine*, sont limitées par la taille de leur échantillon et ne permettent pas d'étudier les extrémités de la distribution comme les centiles ou les milliles de patrimoine immobilier. Une telle source permet également d'étudier la répartition géographique des propriétés immobilières.

La première étape du travail consiste à modéliser les valeurs de marché des logements ayant fait l'objet d'une transaction à partir des données relatives aux ventes immobilières (données *Demande de Valeurs Foncières*). La seconde étape consiste à utiliser cette modélisation pour imputer une valeur de marché à l'ensemble des logements possédés par les ménages. Ce travail s'appuie sur une base de données nouvellement constituée qui décrit l'ensemble des biens immobiliers possédés par les ménages résidents¹ [1]. Les étapes de modélisation et d'imputation s'appuient sur les caractéristiques observables du logement disponibles dans le cadastre : localisation (coordonnées géographiques), surface, étages, présence de dépendances, etc.

La contribution de ce travail est double :

- La première contribution est d'ordre méthodologique. Il s'agit de confronter plusieurs méthodes d'estimation de la valeur des biens immobiliers en comparant notamment les

1. La constitution de cette base de données a mobilisé trois sources : les données cadastrales (fichiers Majic), le répertoire Fidéli et le registre du commerce et des sociétés (RCS).

régressions hédoniques à des algorithmes de *machine learning*. Un des enjeux est de définir les indicateurs pertinents de la qualité des algorithmes, dans la mesure où l'objectif n'est pas uniquement de minimiser l'erreur moyenne de prédiction, mais aussi, et surtout, d'aboutir à une distribution des patrimoines stable et peu sensible aux changements de spécification, notamment en haut de la distribution. Un autre défi méthodologique est de mesurer la valeur de la localisation. Cette dernière est un facteur de premier ordre dans la détermination des prix immobiliers, mais est estimée plus ou moins précisément selon le nombre de ventes réalisées dans la localité. Cette étude permet d'illustrer et de mesurer l'apport de nouveaux algorithmes pour la statistique publique vis-à-vis des méthodes économétriques usuelles. [6].

- La seconde contribution est d'éclairer le débat public sur les inégalités relatives au patrimoine immobilier. Les prix immobiliers ont fortement progressé au cours des deux dernières décennies, ce qui a tiré à la hausse la valeur globale des patrimoines ([7], [5]). Cependant, l'augmentation des prix n'a pas été homogène sur le territoire, ce qui rend difficile l'estimation de la concentration des plus hauts patrimoines. Le développement d'une base exhaustive sur les patrimoines immobiliers est donc un prérequis crucial pour étudier les disparités géographiques du patrimoine immobilier et estimer la concentration des plus hauts patrimoines (sans avoir recours à des hypothèses paramétriques sur leur distribution [2]).

Bibliographie

Références

- [1] ANDRÉ, M., AND MESLIN, O. Et pour quelques appartements de plus : Étude de la propriété immobilières et des caractéristiques redistributives de la taxe foncière sur une base exhaustive des patrimoines immobiliers. *Document de travail* (à paraître 2021).
- [2] BLANCHET, T., FOURNIER, J., AND PIKETTY, T. Generalized pareto curves : theory and applications. *Review of Income and Wealth* (2017).
- [3] FAGERENG, A., HOLM, M. B., AND TORSTENSEN, K. N. Housing wealth in norway, 1993–2015. *Journal of Economic and Social Measurement*, Preprint (2018), 1–17.
- [4] GALLIN, J., MOLLOY, R., NIELSEN, E., SMITH, P., AND SOMMER, K. Measuring aggregate housing wealth : New insights from machine learning. *Journal of Housing Economics* 51 (2021), 101734.
- [5] KNOLL, K., SCHULARICK, M., AND STEGER, T. No price like home : Global house prices, 1870-2012. *American Economic Review* 107, 2 (2017), 331–53.
- [6] MULLAINATHAN, S., AND SPIESS, J. Machine learning : an applied econometric approach. *Journal of Economic Perspectives* 31, 2 (2017), 87–106.
- [7] PIKETTY, T., AND ZUCMAN, G. Capital is back : Wealth-income ratios in rich countries 1700–2010. *The Quarterly Journal of Economics* 129, 3 (2014), 1255–1310.