
**APPREHENDER LA RENTABILITÉ LOCATIVE POUR COMPRENDRE LES
MÉCANIQUES QUI SOUS TENDENT LES LOYERS ET LES PRIX À L'AIDE DE
MÉTHODES DE MACHINE LEARNING**

Martin REGNAUD (*), Marie BREUILLE (*), Julie LE GALLO (**)

(*) CESAER UMR1041, INRAE, AgroSup Dijon, Université de Bourgogne Franche-Comté

(**) CESAER UMR1041, AgroSup Dijon, INRAE, Université de Bourgogne Franche-Comté

mregnaud@meilleursagents.com

Mots-clés : Rentabilité locative, Ratio loyers sur prix, Données d'annonce, Matching

Domaine concerné : Analyse des données et data science, Concepts et mesures

Résumé

Alors qu'il fallait en moyenne, en 2020, 2 ans et 10 mois pour amortir l'achat d'un bien par rapport à la location en France (Meilleurs Agents), cette durée s'élevait à 6 ans et 11 mois à Paris à la même période. Connaître ces chiffres est un élément de décision important pour les ménages en leur permettant d'informer leur choix d'acheter ou de louer leur résidence principale.

Du point de vue opérationnel, le calcul de cette durée se fait via la connaissance des ratios loyers sur prix. L'objectif de cette contribution est d'estimer ces ratios loyers sur prix sur toute la France en croisant des observations de location et de transaction sur les mêmes biens entre 2010 et 2021 puis d'en dégager les déterminants à l'aide de méthodes de machine learning.

A l'aide d'une méthode de *coarsened exact matching*, nous estimons des ratios loyers sur prix sur toute la France puis nous appliquons successivement deux approches pour comprendre leurs déterminants. La première consiste à expliquer les ratios en les régressant avec les caractéristiques du bien et des variables socio-économiques et des mesures des aménités. La deuxième consiste à les expliquer à partir d'un modèle de *gradient boosting d'arbre*. Nous expliquons ensuite l'influence de chaque facteur grâce aux méthodes d'explicabilité de ces modèles : *feature importance* et *shape values*. Pour cette étude, nous nous basons sur des données d'annonces de loyers de la plateforme Meilleurs Agents ayant une géolocalisation précise au niveau de l'adresse. L'utilisation d'annonces de location pour l'étude du marché locatif est légitimée par Chapelle & Eymeoud (2018) [1] qui mettent en évidence l'absence de biais entre les données des plateformes d'annonces et les données issues d'enquêtes utilisées habituellement telles que les données des observatoires locaux (OL) des loyers. Ces enquêtes étant limitées à certaines zones géographiques, nous

cherchons à utiliser une autre source de données pour étendre les études possibles sur tout le territoire.

Nous croisons les données d'annonces de loyers de Meilleurs Agents avec la base de transaction DV3F pour laquelle une géolocalisation à la parcelle est disponible. Le croisement de ces deux bases de données, entre 2010 et 2021, nous permet d'obtenir un jeu de 85'000 couples transaction/location sur toute la France. La base ainsi construite est utilisée pour estimer les ratios loyer sur prix et dégager les différences entre les déterminants selon le territoire, à la fois entre métropoles mais également, grâce à la géolocalisation, au sein même d'un pôle urbain.

L'apport de notre étude est double. Tout d'abord, elle constitue un apport méthodologique, l'utilisation de *gradient boosting* n'ayant encore jamais servi à estimer les déterminants de la variabilité des ratios loyers sur prix. Cette méthode présente un avantage important par rapport aux méthodes classiques car elle permet une meilleure gestion des interactions entre variables et de l'hétérogénéité des effets. Le deuxième apport réside dans l'utilisation de données du même bien pour calculer ces ratios. Par manque de données ayant un niveau de géolocalisation fin, ces ratios sont souvent étudiés au travers de ratios de moyennes de loyers sur des moyennes de prix. Cependant, Hill & Syed (2016) [2] montrent que cette approche peut entraîner une erreur de près de 20% dans l'estimation de la rentabilité. La structure et la qualité des logements servant à calculer la moyenne des loyers peuvent différer de celles des transactions servant à calculer la moyenne des prix. Afin d'éviter l'apparition de ce biais, ils recommandent la prise en compte des différences de caractéristiques de chaque bien dans le calcul. Notre étude constitue ainsi la première étude, en France, permettant un appariement exact en dehors des zones couvertes par les observatoires locaux des loyers sur ce sujet.

Cette méthode nous permet de mettre en évidence une forte hétérogénéité des ratios loyers sur prix à la fois au sein des grandes aires urbaines mais aussi à plus grande échelle sur le territoire national. Cette étude est à notre connaissance, la première à mettre en évidence ce phénomène à grande échelle.

Bibliographie

- [1] Chapelle G., Eymedoud J.-B., « Can Big Data Increase Our Knowledge of Local Rental Markets? Estimating the Cost of Density with Rents », *SciencesPo Mimeo*, 2018
- [2] Hill R.J., Syed I.A., « Hedonic price–rent ratios, user cost, and departures from equilibrium in the housing market », *Regional Science and Urban Economics*, pp 60-72, 2016