

Novembre 2024

CALCUL D'ÉVOLUTIONS DE LOYERS À PARTIR DES DONNÉES DES OBSERVATOIRES LOCAUX DES LOYERS :

méthode et résultats



Co-direction d'études : Clément Pavard, Odile Dubois-Joye
Production statistique et cartographique : Anaïs Aber, Wael Jradi, Coriolan Gout

Introduction

Le réseau des observatoires locaux des loyers réalise des enquêtes annuelles auprès de particuliers et de professionnels dans leur territoire de compétence. Le parc de référence enquêté est constitué de l'ensemble des locaux à usage d'habitation ou à usage mixte, meublés ou non et occupés à titre de résidence principale, à l'exception des logements sociaux ou construits en maîtrise d'ouvrage d'insertion. La méthodologie de recueil et de traitement des données est identique, visant à produire des données harmonisées, fiables et comparables entre les différents territoires couverts par un observatoire. Le recueil et le traitement des données respectent les prescriptions méthodologiques validées par le Comité scientifique du réseau¹.

Le réseau des OLL produit des données en niveau. En effet, les pouvoirs publics ont réorienté en 2012-2013 le principal dispositif public d'observation des loyers vers une observation fiable sur les niveaux des loyers, pour fournir une vision fine et segmentée de chacun des marchés locaux étudiés (notamment selon le type d'habitat, le nombre de pièces, l'époque de construction, l'ancienneté d'emménagement et un zonage local dans les agglomérations grandes et moyennes). Outre la mise en place d'un nouveau dispositif de régulation des loyers (encadrement des loyers promulgué par la loi ALUR, modifié par la loi ELAN), ce choix a été guidé par la nécessité de disposer de données précises pour la conception de politiques de développement de l'offre locative et pour l'information du grand public.

La production de données en niveau offre de nombreuses opportunités d'application pour la mise en place de politiques publiques mobilisant des montants d'investissement élevés (Programmes locaux de l'habitat, ciblage du parc susceptible d'être conventionné Anah, développement de produits locatifs à niveau de loyer intermédiaire voire social, caractérisation de l'offre privée accessible aux ménages modestes...).

L'expression d'une attente d'indicateurs d'évolution demeure cependant fréquente, ne serait-ce que pour mesurer l'impact de certaines politiques publiques sur le marché. Il est possible de lister ici certaines des questions implicitement ou explicitement posées concernant la production d'un tel indicateur :

- Quelle est l'évolution de la charge de loyer des locataires, à qualité constante, en neutralisant effets structurels et de qualité ?
- Quelle est l'évolution des loyers de marché, c'est à dire des biens pris en location, en valeur brute et en neutralisant les effets de structure ?
- Quelle est la proportion de logements dont le loyer dépasse le taux de revalorisation légal (IRL) ? Quelle est l'évolution des loyers des logements ayant fait l'objet de travaux ? Quelle est l'évolution des loyers de logement faisant l'objet d'une procédure de réévaluation du loyer en cas de loyer manifestement sous-évalué ?

Si des enquêtes par panel sur les loyers du parc locatif privé ont montré leur capacité à répondre à la plupart de ces questions², elles supposent d'importants surcoûts.

Aussi, l'ANIL a développé une méthodologie économétrique analogue à celle utilisée par nombre d'instituts statistiques nationaux concernant l'observation des prix de transaction des logements, en l'occurrence par introduction d'indicatrices temporelles dans un modèle hédonique de loyers. Ces

¹ [Consulter les prescriptions méthodologiques](#)

² Cf. travaux de l'OLAP sur les évolutions des loyers pour l'agglomération parisienne et les enquêtes « province » dans 11 grandes agglomérations menées dans les années 90 et 2000

travaux ont été présentés au Comité scientifique des observatoires des loyers qui s'est prononcé favorablement pour un déploiement de la méthode.

Le présent document fait état :

- des éléments de méthodologie, notamment par le rappel des méthodes économétriques sur la mesure des évolutions d'un prix ;
- la méthode retenue et les choix opératoires concernant l'ajustement des modèles ;
- les résultats obtenus.

Sommaire

Introduction	3
I. Revue de littérature des méthodes économétriques sur la mesure des évolutions d'un prix.....	6
1.1. Méthodes basées sur des équations de prix hédonique	6
1.2. Variante écartée : méthode mixte	7
II. Présentation de la méthode retenue	8
2.1. Méthode des indicatrices temporelles	8
2.2. Variables explicatives du loyer intégrées dans le modèle	9
2.2.1. Caractéristiques liées à la localisation	9
2.2.2. Caractéristiques intrinsèques au logement	10
2.2.3. Caractéristiques liées au bail	10
2.2.4. Variables liées à la temporalité de l'enquête	10
2.2.5. Sélection des variables grâce à la régression Lasso	12
2.3. Mise en œuvre d'un modèle de régression quantile.....	13
III. Exemples de mise en œuvre de la méthodologie	15
3.1. Mise en œuvre de la méthodologie sur la ville de Lille	15
3.1.1. Construction d'une variable de localisation par clustering	15
3.1.2. Résultats obtenus pour la période 2021-2023	17
3.2. Mise en œuvre de la méthodologie sur la ville de Toulouse.....	19
3.2.1. Construction d'une variable de localisation par clustering	19
3.2.2. Résultats obtenus pour la période 2021-2023	21
IV. Résultats obtenus sur l'ensemble des observatoires des loyers	24
4.1. Évolution des loyers calculées par quartiles sur 2 années cumulées : 2020-2022	24
4.2. Évolution des loyers calculées par quartiles sur 2 années cumulées : 2021-2023	27
4.3. Modalités de diffusion.....	29
4.3.1. Règles de diffusion	29
4.3.2. Exemple de diffusion des indicateurs en carte.....	30
Conclusion.....	311
Références bibliographiques.....	322

I. Revue de littérature des méthodes économétriques sur la mesure des évolutions d'un prix

Différentes méthodes économétriques peuvent être utilisées pour estimer les évolutions annuelles de prix des logements. Elles permettent de produire des indicateurs à structure ou qualité constantes. Il est ainsi envisageable de les utiliser pour proposer une réponse aux deux premières questions citées en introduction.

1.1. Méthodes basées sur des équations de prix hédonique

Une synthèse des méthodes utilisées par les instituts de statistiques européens a été effectuée par Hill et al. (2018). Cette étude compare les **cinq méthodes statistiques utilisées** : la réévaluation des prix, la méthode des caractéristiques moyennes, la méthode d'imputation hédonique, la méthode des indicatrices temporelles sur période glissante et la méthode des médianes stratifiées.

La méthode de réévaluation des prix repose sur l'utilisation d'un seul modèle hédonique permettant d'estimer un facteur d'ajustement. Dans un premier temps, un modèle semi-log est estimé en ne prenant en compte que l'année 1 permettant d'exprimer les prix virtuels des caractéristiques. L'indice de prix non ajusté de la qualité est le ratio des moyennes géométriques des prix observés des périodes N et N-1. Le facteur d'ajustement de la qualité est le ratio de la moyenne arithmétique des prix des logements estimés pour les périodes N et N-1 utilisant les prix virtuels estimés sur l'année 1. L'indice de prix est le ratio de l'indice de prix non ajusté de la qualité sur le facteur d'ajustement. Le modèle hédonique doit être recalculé régulièrement pour maintenir un niveau de fiabilité suffisant (limiter la dérive).

Pour **la méthode des caractéristiques moyennes**, deux modèles hédoniques semi-log sont estimés : un pour la période N et un pour la période N-1. Les deux modèles hédoniques sont utilisés pour estimer le prix d'un bien présentant les caractéristiques moyennes du panier de biens. L'indice de prix obtenu *in fine* est le ratio du prix du bien moyen estimé grâce au modèle de la période N sur le prix estimé grâce au modèle de la période N-1. Le bien moyen doit être actualisé tous les ans.

La méthode d'imputation hédonique permet d'estimer le prix de vente à la période N d'un logement vendu à la période N-1, afin d'évaluer l'évolution des prix estimés entre la période N-1 et la période N. L'indice d'évolution est une moyenne du ratio d'estimation de prix d'un logement à la période N sur l'estimation du prix sur la période N-1

La méthode des indicatrices temporelles sur période glissante (RTD) estime un modèle hédonique utilisant une fenêtre de temps couvrant plusieurs périodes. L'utilisation de cette méthode suppose que les prix des caractéristiques pour toutes les périodes temporelles de la fenêtre soient égaux. La fenêtre est ensuite avancée d'une période et le modèle réestimé. L'indice de prix est le ratio du prix estimé sur la fenêtre actuelle sur le prix estimé sur la fenêtre précédente. L'enjeu de cette méthode est de trouver un compromis sur la taille de la fenêtre. En effet, une grande fenêtre augmente la taille de l'échantillon et donc la robustesse du modèle mais elle a tendance à lisser les différences entre les périodes.

Enfin, dans **la méthode de médianes stratifiées**, l'échantillon est scindé en strates (géographiques et sur les caractéristiques des logements), puis un modèle hédonique est estimé par strate permettant d'estimer un prix médian par strate. Des moyennes arithmétiques pondérées (par le volume de transaction ou par le volume du parc) sont réalisées sur les résultats médians des strates. Lorsque le nombre de strates augmente, l'indice est de plus en plus précis mais cela accroît le risque d'avoir des strates vides.

Pour l'étude des prix de transactions immobilières, les méthodes décrites ci-dessus donnent des résultats assez similaires pour estimer des évolutions de prix de vente de logement lorsque l'échantillon est assez grand et que les strates de la méthode stratifiée sont assez fines (Hill et al. 2018). Ces conditions amènent ces auteurs à préconiser l'utilisation de l'ancienneté du logement comme caractéristique d'un modèle global plutôt que d'estimer un modèle spécifique pour les logements récents.

1.2. Variante écartée : méthode mixte

Dans les bases de données des OLL, certains logements sont collectés sur deux années consécutives, leur identifiant permet de les isoler. Cela pourrait permettre d'utiliser un modèle avec suivi dans le temps pour ces logements. Mais un modèle mixte comprenant un calcul des évolutions brutes pour les logements suivis lors des deux années étudiées et un calcul des évolutions par une méthode économétrique pour les autres logements a été écartée à cause d'une incompatibilité sur la prise en compte de l'ancienneté d'occupation entre ces deux méthodes. En effet, par exemple, dans le premier cas, un emménagé récent (emménagé depuis moins d'un an) en 2017 ne l'est plus en 2018 ; alors que dans la deuxième méthode les mêmes caractéristiques sont appliquées, il y a donc un « glissement » de l'ancienneté d'occupation, un emménagé récent en 2017 le sera aussi en 2018. De plus, dans de nombreuses études sur le prix de vente de biens, les méthodes basées sur des prix répétés ou des méthodes hybrides ne permettent pas d'obtenir de meilleurs résultats que des méthodes de modélisation hédonique, notamment lorsque les effectifs sont faibles ou lorsque le temps entre deux ventes est faible (Clapp, Giaccotto, et Tirtiroglu 1991; Wallace et Meese 1997; Gatzlaff et Ling 1994; Thouvenin 2010).

Signalons que l'enquête menée par les OLL collecte systématiquement de l'information sur les caractéristiques structurelles (type d'habitat, nombre de pièces, époque de construction, ancienneté d'occupation, localisation), collecte pour un échantillon restreint des caractéristiques détaillées (présence d'annexes au logement, étage, présence d'un ascenseur...), mais ne collecte pas d'éléments sur l'état général du bien ou la réalisation de travaux. Ainsi, en n'utilisant que les informations disponibles systématiquement, il est envisageable de produire des indicateurs d'évolution à structure constante, mais non à qualité constante.

II. Présentation de la méthode retenue

Pour le calcul des évolutions des prix de location utilisant les données des OLL, la méthode retenue a été la plus directe, à savoir la **méthode des indicatrices temporelles**, en estimant un modèle hédonique de loyers agglomération par agglomération. Pour une agglomération donnée, un modèle hédonique unique est ainsi estimé sur des observations collectées pour l'année N et N+1 en introduisant, en plus des variables explicatives usuelles (surface, nombre de pièces, etc.), **une indicatrice temporelle sur l'année d'observation**.

Ceci permet de connaître **l'évolution du loyer à structure constante par l'interprétation directe du coefficient estimé pour l'indicatrice temporelle**.

Cette variante est préférée au calcul de la moyenne des ratios d'évolution des prix unitaires estimés pour chaque bien du panier, ce qui permettrait de connaître la moyenne des évolutions des loyers des logements. Elle est également préférée aux méthodes d'imputation hédonique pour sa plus grande simplicité car elle nécessite l'estimation d'un seul modèle hédonique et, le calcul des intervalles de confiance autour des indicateurs d'évolution estimés est plus direct et ne nécessite pas la mise en œuvre de méthodes de type bootstrap.

Est d'abord présentée la méthode des indicatrices temporelles, puis les choix réalisés pour l'ajustement des modèles concernant la construction des variables explicatives et les règles de sélection des variables par la méthode du Lasso.

2.1. Méthode des indicatrices temporelles

La méthode retenue consiste à regrouper plusieurs périodes et à estimer un unique modèle hédonique auquel on ajoute des indicatrices temporelles. Sous sa forme semi-logarithmique, le modèle s'écrit de la manière suivante :

$$\text{Log}(\text{Loyer}) = \alpha + \beta \cdot X + \delta \cdot D + \varepsilon$$

Où $\text{Log}(\text{Loyer})$ est le logarithme du loyer, α est la constante, X les caractéristiques associées aux logements, D les indicatrices temporelles (qui prennent la valeur 0 pour la période de référence et 1 pour la période courante) et ε un terme d'erreurs supposées i.i.d.. β sont les paramètres à estimer associés aux caractéristiques des logements et δ celui à estimer pour l'indicatrice temporelle.

L'indicateur mesuré en % s'obtient de la manière suivante³ :

$$Tx \text{ évol} = 100 \times (\text{Exp}(\widehat{\delta}_t) - 1)$$

Cette méthode présente l'avantage d'un indicateur qui ne dépend pas directement du panier de biens de référence. Par ailleurs, la variance et les intervalles de confiance associés s'obtiennent directement à partir de la delta méthode. Pour un indicateur mesuré en %, la formule de la variance est la suivante :

$$\text{Var} = \left(100 \times \text{Exp}(\widehat{\delta}_t)\right)^2 \times \mathbb{V}(\widehat{\delta}_t)$$

La méthode pose en revanche l'hypothèse de valeurs identiques des $\hat{\beta}$ associés aux caractéristiques, quelle que soit la période⁴.

³ Cette formule est valable lorsque l'année de référence est fixée en t . Si l'année de référence retenue est fixée en $N+1$, la formule devient : $100 \times (\text{Exp}(-\widehat{\delta}_t) - 1)$

⁴ Pour une période de deux ans, cette hypothèse semble acceptable car il n'y a pas de raisons particulières pour que la valorisation des caractéristiques des logements varie de manière sensible d'une année à l'autre. Les estimations réalisées sur des modèles séparés par années le corroborent.

L'intervalle de confiance à 95 % est calculé à partir de la formule suivante :

$$IC = [Tx \text{ évol} - 1,96 \times \sqrt{Var} ; Tx \text{ évol} + 1,96 \times \sqrt{Var}]$$

2.2. Variables explicatives du loyer intégrées dans le modèle

Les variables intégrées dans le modèle ont été sélectionnées à partir d'une revue de littérature, mais aussi à partir de raisons plus techniques : leur disponibilité systématique dans les bases de données des OLL et les qualités d'ajustements des modèles obtenues.

2.2.1. Caractéristiques liées à la localisation

Pour capter un effet de localisation du logement sur le loyer, une sectorisation au moyen de clusters a été établie afin d'obtenir un compromis entre taille de la maille géographique utilisée dans le modèle et significativité de cet effet géographique. Au sein de chaque périmètre concerné, la méthode retenue *in fine* pour constituer des clusters est une classification hiérarchique ascendante à partir d'une analyse en composantes principales. Les individus sont les IRIS des villes-centres.

Les 22 variables suivantes sont retenues pour discriminer les zones où les loyers sont homogènes (tableau 1).

Tableau 1 : description des variables retenues pour la construction du maillage

Description de la variable	Source (année 2021)
Part du parc locatif privé construits avant 1946 (%)	Insee, RP
Part du parc locatif privé construits entre 1946 et 1970 (%)	Insee, RP
Part du parc locatif privé construits entre 1971 et 1990 (%)	Insee, RP
Part du parc locatif privé construits entre 1991 et 2005 (%)	Insee, RP
Part du parc locatif privé construits après 2005 (%)	Insee, RP
Part des propriétaires occupants dans les résidences principales (%)	Insee, RP
Part des locataires du locatif privé dans les résidences principales (%)	Insee, RP
Part des locataires du locatif privé meublé dans les résidences principales (%)	Insee, RP
Part des studios et des 2 pièces dans le parc locatif privé (%)	Insee, RP
Part des résidences secondaires dans les logements ordinaires (%)	Insee, RP
Part des logements vacants dans les logements ordinaires (%)	Insee, RP
Nombre moyen de personnes par ménage dans le locatif privé	Insee, RP
Revenu médian	Insee, Filosofi
Taux de pauvreté au seuil de 60 % du revenu (%)	Insee, Filosofi
Taux d'évolution du nombre de ménages entre 2013 et 2019 (%)	Insee, RP
Taux de chômage (%)	Insee, RP
Part des 65 ans ou plus dans la population totale (%)	Insee, RP
Part des 18-24 ans dans la population totale (%)	Insee, RP
Part des 25-39 ans dans la population totale (%)	Insee, RP
Part des étudiants 15-64 ans dans la population totale (%)	Insee, RP
Densité de population (nb d'habitants/ha)	Insee, populations légales
Taux de mobilité des ménages	Insee, RP

Pour s'assurer que ces variables décrivent bien le parc locatif privé et ses locataires, une Analyse en composantes principales (ACP) est menée pour résumer de la manière la plus pertinente possible les données initiales en projetant les observations dans un espace plus restreint. En fonction des

caractéristiques du périmètre géographique sur lequel la modélisation est appliquée, un nombre de composantes principales est retenu selon le critère de la part d'inertie cumulée. Le plus souvent, les deux premiers axes cumulent plus de 50% de l'inertie et sont retenus.

On utilise ensuite une classification hiérarchique ascendante (sur les axes factoriels retenus à l'issue de l'ACP) qui délimite des groupes d'individus, soit des typologies, à partir de leurs caractéristiques.

En dernier lieu, ces classes sont expertisées au moyen d'une visualisation cartographique et de quelques statistiques descriptives par cluster. La carte obtenue est mise en regard de la connaissance locale du territoire et aussi de référentiels urbanistiques (représentation des quartiers prioritaires de la ville par exemple). À l'issue de cette expertise, des regroupements de classes « à la main » peuvent être opérés, en particulier lorsqu'une classe ne comporte pas assez d'IRIS et de biens en location privée.

À partir de la classification finale, on procède à une repondération des données collectées par l'OLL dans le nouveau zonage, puis ces nouvelles classes géographiques sont intégrées dans le modèle. Cette pondération est réalisée en application de la méthodologie de calage sur marges définie dans les prescriptions méthodologiques du Comité scientifique des OLL.

2.2.2. Caractéristiques intrinsèques au logement

Un ensemble de caractéristiques intrinsèques au logement est intégrée au modèle. Il s'agit de la surface totale et de la surface moyenne par pièce passées en logarithme (la variable expliquée du loyer étant elle-même passée en logarithme). Ces deux variables sont réputées avoir une incidence très forte sur le loyer, pesant pour plus de la moitié de la formation du prix du loyer.

Sont également intégrées deux indicatrices : l'une distinguant les studios (appartements d'une pièce) et appartements de deux pièces et plus et l'autre distinguant maisons et appartements. Enfin, une variable sur l'époque de construction du logement est également intégrée. Cette dernière permet notamment d'approcher la qualité énergétique du logement. En effet, en l'absence de travaux de rénovation importants, la performance énergétique d'un logement est étroitement liée à sa période de construction du fait de normes de constructions qui ont évolué au fil de temps.

2.2.3. Caractéristiques liées au bail

Les OLL collectent également des informations relatives au bail. Certaines d'entre elles peuvent avoir un effet significatif sur le prix du loyer. Ainsi, on ajoute au modèle l'ancienneté d'occupation du logement par le locataire (plus l'ancienneté d'occupation sera importante, moins le loyer sera élevé en moyenne). Deux autres variables de contrôle aux effets cependant moindres sur le loyer sont également intégrées : le type de gestion (directe ou déléguée) et le trimestre d'emménagement.

2.2.4. Variables liées à la temporalité de l'enquête

Année d'enquête :

Une indicatrice sur l'année d'enquête est intégrée au modèle. **Elle constitue la variable d'intérêt du modèle. Elle s'interprète directement comme l'évolution du loyer à structure constante.**

Mois de référence du loyer :

Le loyer collecté par les OLL correspond le plus souvent à un loyer au 1^{er} janvier de l'année n. Toutefois, les collectes locales peuvent s'étaler sur plusieurs mois, et le loyer au 1^{er} janvier n'est pas

systématiquement renseigné ; il peut être celui de la date d'enquête. L'irrégularité du pas des enquêtes peut induire un biais sensible sur le coefficient d'évolution.

Afin de corriger ce biais, on introduit une variable captant l'impact du mois de référence⁵ de l'enquête par rapport au mois de janvier par l'intermédiaire de fonctions *splines*. Ces fonctions permettent de régulariser les variations saisonnières du loyer et de les corriger. **Dans l'interprétation des modèles, le loyer modélisé correspondra à un loyer au 1^{er} janvier.**

Ces fonctions peuvent être interprétées comme des ensembles de polynômes cubiques qui sont ajustés aux données. Dans notre cas d'application, ce sont des fonctions polynômiales par morceaux à 3 degrés captant un impact non-linéaire des mois sur le loyer. La relation entre le mois de référence de l'enquête et le niveau de loyer est ainsi représentée selon l'équation suivante :

$$\text{Log}(\text{Loyer}_{m_2}) = \beta \cdot P(x_{\text{mois}}) + c$$

Où P est un polynôme de degré 3 et c une constante.

Lorsque l'on applique par exemple cette fonction aux échantillons de collecte de la ville de Marseille, P et c sont définis par :

	$P(x)$	c
En 2019/2020 :	$P(x) = 2.6 \cdot 10^{-16} + 0.00501(x - 1) - 0.00682(x - 1)^2 + 0.000771(x - 1)^3$	$c = 2.510$
En 2020/2021 :	$P(x) = 1.25 \cdot 10^{-16} - 0.0181(x - 1) - 8.45 \cdot 10^{-05}(x - 1)^2 + 0.000284(x - 1)^3$	$c = 2.525$

En appliquant cette formule entre 2019 et 2020, on obtient :

- pour le 1^{er} mois pris en référence (janvier) : $\text{Log}(\text{Loyer}_{m_2}) = 0 + 2,510 = 2,510$ (soit $\text{Loyer}_{m_2} = 12,305$, ce qui correspond au loyer d'un « bien type »)
- pour le 2^e mois (février) : $\text{Log}(\text{Loyer}_{m_2}) = -0,016 + 2,510 = 2,494$ (soit $\text{Loyer}_{m_2} = 12,108$)
- pour le 10^e mois (octobre) : $\text{Log}(\text{Loyer}_{m_2}) = 0,055 + 2,510 = 2,565$ (soit $\text{Loyer}_{m_2} = 12,997$)

Ces valeurs sont cohérentes avec les loyers observés sur cette ville.

Le loyer médian constaté par l'observatoire des loyers pour l'ensemble du parc locatif privé dans la ville de Marseille est calculé à l'échelle de plusieurs zones de prix situées dans différents quartiers. La médiane des prix à la location varie de 11,7 € dans la zone la moins chère (Marseille 4) à 13,2 € dans la zone la plus chère (Marseille 1)⁶.

Les courbes des fonctions *splines* sur le mois de référence sont tracées en vert dans les graphiques ci-contre.

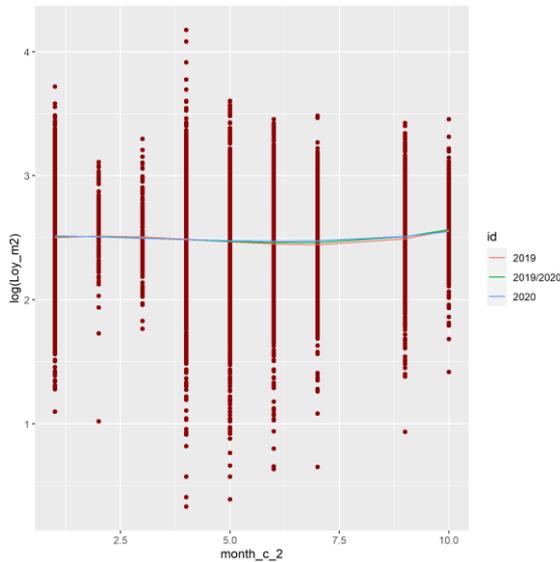
On note que la relation entre le mois de référence de l'enquête et le niveau de loyer sur deux années de collecte combinées est proche de celle illustrée indépendamment pour chacune des années de collecte.

⁵ Si l'enquête porte sur les caractéristiques du logement et du bail au 1^{er} janvier de l'année, le mois de référence sera alors janvier, sinon il s'agira du mois de la date (la plus proche possible du 1^{er} janvier) pour laquelle les caractéristiques du logement et du bail sont connues.

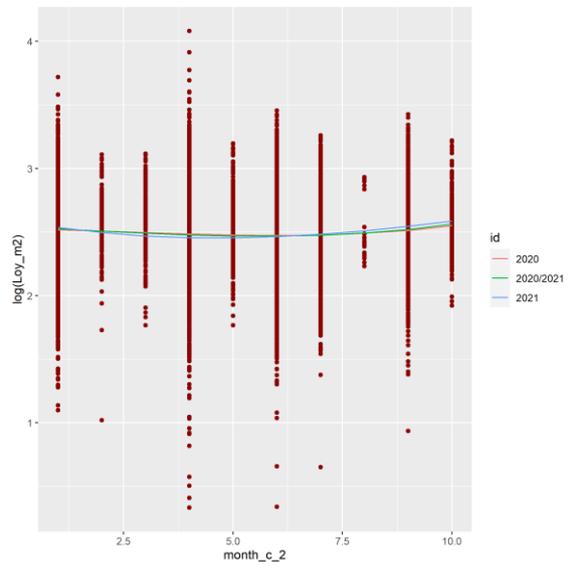
⁶ Chiffres disponibles sur le site observatoires-des-loyers.org : <https://www.observatoires-des-loyers.org/connaitre-les-loyers/carte-des-niveaux-de-loyers/agglomeration-daix-marseille/marseille-13055?annee=2022>

Graphique 1 : courbes des fonctions *splines* illustrant la relation entre le mois de référence de l'enquête (ici de janvier à octobre) et le niveau de loyer

Collectes 2019-2020 (ville de Marseille)



Collectes 2020-2021 (ville de Marseille)



2.2.5. Sélection des variables grâce à la régression Lasso

En dernière étape, une sélection de variables à intégrer dans chaque modèle hédonique est mise en œuvre. **Cette sélection est réalisée à partir d'une régression Lasso** (*Least Absolute Shrinkage and Selection Operator*). Cette méthode a été introduite par Tibshirani (1996) pour la sélection et la pénalisation des paramètres dans une régression linéaire. Elle appartient avec les régressions Ridge et Elasticnet à la famille des modèles de régressions pénalisées ou régularisées.

Bien que l'estimateur par régression linéaire soit l'estimateur sans biais en termes de compromis biais/variance, la régression pénalisée admet un certain biais pour réduire la variance. Cela signifie que la fonction de minimisation de cette dernière a deux composantes : l'erreur quadratique moyenne et la pénalité pour les paramètres.

Pour un ensemble d'observations N indicées par $i = 1, 2, \dots, N$ et un ensemble de p variables indicées par $j = 1, 2, \dots, p$, les prédicteurs standardisés x_{ij} et la variable dépendante y_i , la régression Lasso trouve les β_j qui minimise la fonction d'objectif suivante :

$$\sum_{i=1}^N \left(y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

La première partie de cette fonction qui minimise la somme des carrés des résidus est identique à celle utilisée en régression linéaire classique et la seconde correspond au terme de pénalité introduit par la régression Lasso. Ce terme de pénalité est ajusté par le paramètre λ qui est déterminé de manière exogène par l'utilisateur le plus souvent par validation croisée. Ainsi, lorsque que certaines variables sont introduites dans le Lasso en diminuant très légèrement la somme des résidus au carré, l'impact de la pénalité augmente et impose des coefficients nuls. Cette propriété est utilisée pour sélectionner les variables pertinentes à introduire dans le modèle.

La robustesse des résultats a été testée en utilisant une méthode alternative de sélection séquentielle à l'aide du critère AIC. Cette méthode donne des résultats très proches.

Tableau 2 : Récapitulatif et description des variables explicatives intégrées au modèle

Variable à expliquer	Code	Variabes explicatives	Unité/modalités
Logarithme du loyer	Effet géographique	Clusters créés par ACP	Code du cluster
	Surface	Log (Surface)	m ²
	Surf_moy_piece	Log (Surface par pièce)	m ²
	Indic_studio	Studio	1 : oui 0 : non
	Indic_maison	Maison	0 : oui 1 : non
	Epoque_cons	Epoque de construction	1 : avant 1946 2 : 1946-1970 3 : 1971-1990 4 : 1991-2005 5 : après 2005
	Duree_occup	Ancienneté d'occupation	1 : moins d'un an 2 : de 1 à 3 ans 3 : plus de 3 ans
	Type_gestion	Type de gestion	1 : déléguée ; 2 : directe
	Trim_emm	Trimestre d'emménagement	1 : T1 2 : T2 3 : T3 4 : T4
	Mois_ref	Mois de réf. du loyer	Fonction spline
	Année	Année d'enquête	1 : N-1 0 : N

Les croisements de variables suivants sont également proposés et peuvent être sélectionnés lors de la procédure Lasso :

- « Cluster » et « Indicatrice studio »
- « Durée d'occupation » et « Trimestre d'emménagement »
- « Durée d'occupation » et « Cluster »
- « Durée d'occupation » et « Indicatrice studio »
- « Années » et « Type de gestion » (pour les OLL sans clusters)

2.3. Mise en œuvre d'un modèle de régression quantile

Le modèle a d'abord été testé en appliquant la méthode des moindres carrés ordinaires. Cependant, les intervalles de confiance calculés autour des indicateurs d'évolution moyens s'avèrent élevés. Ce niveau d'incertitude limite l'interprétation des résultats et est pour partie lié à la nature des données, notamment en raison des poids élevés attribués aux biens en gestion directe.

Mais cette incertitude résulte aussi de l'hypothèse d'une hétérogénéité des situations d'évolutions de loyers, car de pratiques diverses des bailleurs en matière de gestion locative et non maîtrisées statistiquement parlant.

Aussi, la méthode a été orientée vers une mesure robuste de l'évolution de la médiane des loyers, ainsi que du premier quartile et du troisième quartile permettant de mieux rendre compte de cette hétérogénéité et obtenir une information plus complète sur les dynamiques locales d'évolutions des loyers. En outre, comme les indicateurs de loyers produits par les OLL sont prioritairement diffusés en médiane, obtenir un niveau d'évolution entre deux médianes paraît plus opportun.

Les régressions quantiles déterminent comment varie chaque décile de la variable d'intérêt en fonction des déterminants auxquels on s'intéresse.

La variable d'intérêt est ici le loyer (exprimé en log), et les variables explicatives sont les mêmes que celles utilisées pour la régression linéaire (cf. tableau 1).

Il est fait l'hypothèse que la même forme fonctionnelle avec les mêmes régresseurs est adaptée pour décrire les différents déciles de loyer.

Avant mise en œuvre du modèle, les observations sont ordonnées selon un ordre croissant du prix du loyer. Le modèle s'écrit de la manière suivante :

$$\text{Décile} \left(\text{Log}(\text{Loyer}_j) \right) = \alpha + \beta_S f(\text{Mois}_{\text{enquête}_j}) + \beta_C \cdot X_j + \delta_T \cdot D_j + \varepsilon$$

Où $\text{Log}(\text{Loyer})$ est le logarithme du loyer, α est la constante, X les caractéristiques associées aux logements, D l'indicatrice temporelle (qui prend la valeur 0 pour la période de référence et 1 pour la période courante) et ε un terme d'erreurs supposées i.i.d.

$f(\text{Mois}_{\text{enquête}_j})$ est une fonction *spline* du mois d'enquête. Il s'agit d'une fonction polynomiale par morceaux permettant de capter un impact non linéaire du mois d'enquête sur le loyer. Le paramètre β_S est celui associé à cette fonction.

Les paramètres β_C correspondent au changement du $j^{\text{ième}}$ décile de la distribution conditionnelle de loyer suite à une augmentation d'une unité de X , par exemple une augmentation d'une unité de surface, toutes choses égales par ailleurs (les autres variables restant constantes).

Dans le cas de l'indicatrice temporelle sur l'année de collecte, le paramètre δ_T mesure simplement l'écart entre le $j^{\text{ième}}$ décile de la distribution des loyers pour l'année $N - 1$ et le $j^{\text{ième}}$ décile de la distribution des loyers de l'année N (conditionnelle à l'ensemble des autres variables explicatives).

Ainsi, pour le cas de la médiane, l'indicatrice temporelle peut se lire comme une évolution des loyers médians entre deux années (toutes choses égales par ailleurs). Plus précisément, on interprète une évolution entre janvier $N - 1$ et janvier N (premier mois de collecte pris en référence dans la fonction *spline*).

La régression quantile est robuste aux valeurs aberrantes ou à des erreurs très dispersées.

Cette propriété est due au fait que les quantiles sont moins sensibles que la moyenne à la présence de valeurs extrêmes. Dans le cas d'une modélisation du loyer, l'utilisation de la régression quantile se justifie d'autant plus que la variable d'intérêt ne suit pas une loi Normale (queue de distribution étalée sur les valeurs les plus élevées).

Le critère de qualité d'ajustement du modèle établi avec la méthode en régression quantile est analogue à celui avec la méthode MCO, basée sur la mesure R2. Le pseudo R2, noté R1, mesure la qualité du modèle avec la méthode RQ pour le quantile spécifié, en termes de somme pondérée des valeurs absolues des résidus. Contrairement au R2 qui mesure une qualité globale sur l'ensemble de la distribution conditionnelle, le R1 mesure une qualité locale pour chaque quantile. Le R1 est donc en mesure d'explorer plus d'informations à différents niveaux de la distribution conditionnelle de la variable d'intérêt.

III. Exemples de mise en œuvre de la méthodologie

Cette partie illustre la mise en œuvre de la méthode sur deux périmètres. Sont ici retenues les villes de Lille et de Toulouse.

Les différentes étapes présentées s'enchainent de la manière suivante :

- Construction d'une variable de localisation par clustering ;
- Sélection des variables à intégrer dans le modèle par régression Lasso ;
- Estimation d'un modèle en régression quantile ;
- Présentation et interprétation des résultats des régressions

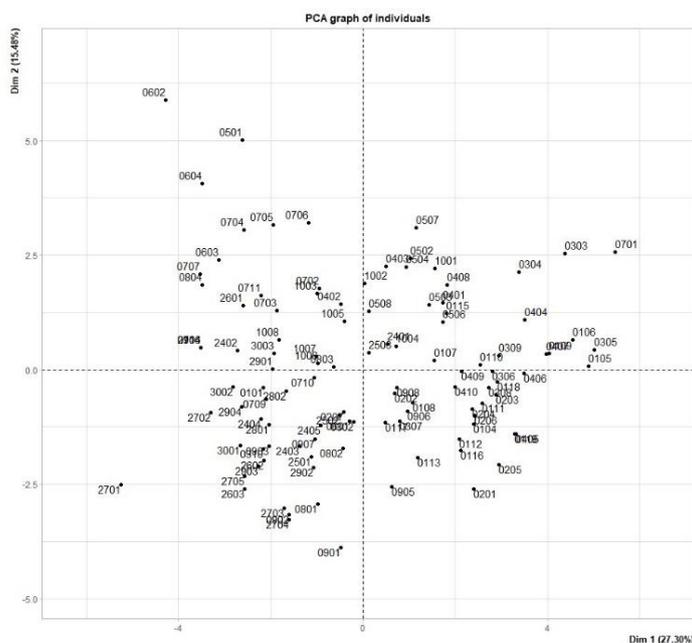
3.1. Mise en œuvre de la méthodologie sur la ville de Lille

3.1.1. Construction d'une variable de localisation par clustering

La première étape consiste à construire un maillage basé sur les caractéristiques des logements et des habitants au moyen d'une Analyse en composantes principales (ACP), dont les individus sont les IRIS qui constituent la maille élémentaire définie par l'INSEE. Les communes d'au moins 10 000 habitants et une forte proportion des communes de 5 000 à 10 000 habitants sont découpées en IRIS. Ce découpage constitue une partition de leur territoire. La France compte environ 16 100 IRIS dont 650 dans les DOM.

On obtient ainsi une représentation des IRIS des villes-centres des agglomérations couvertes par un OLL en fonction de leurs caractéristiques dans un espace bidimensionnel comme le montre le graphique ci-dessous.

Graphique 2 : Représentation des individus (IRIS) sur les deux premiers axes de l'ACP - ville de Lille

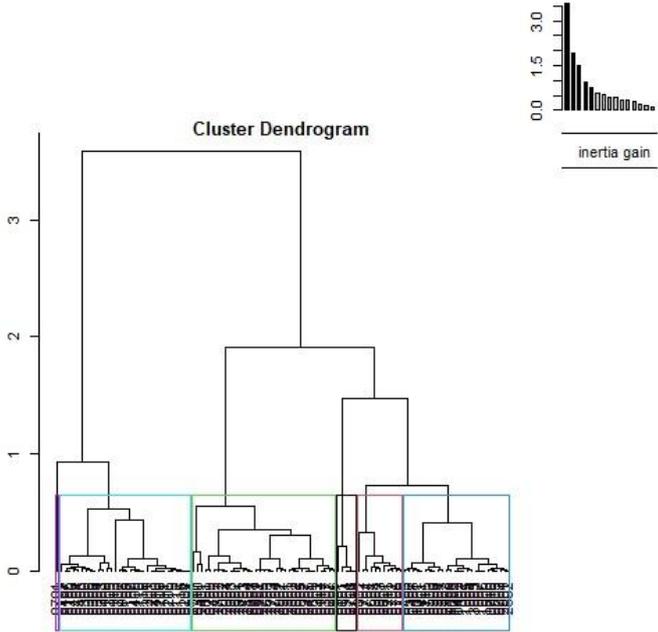


Dans cet exemple, les deux premiers axes cumulent environ 43 % de l'inertie expliquée. On se restreindra à ces deux premiers axes pour réaliser une classification hiérarchique ascendante (CAH).

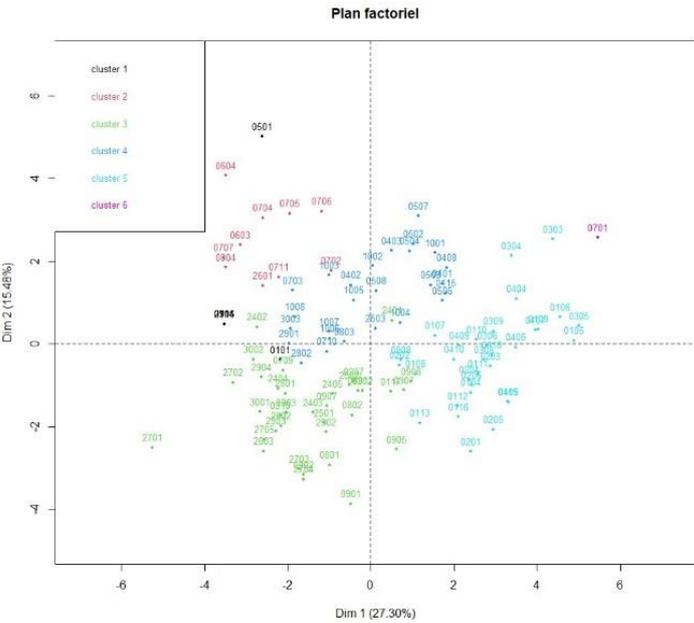
La CAH délimite des groupes d'individus, soit des typologies, à partir des caractéristiques de ces individus. Cette méthode va permettre de distinguer des ensembles au sein desquels les IRIS se ressemblent. Elle produit un arbre hiérarchique à partir duquel est déterminé le nombre de classes le plus pertinent. Pour cela, la méthode du saut est appliquée ; elle consiste à identifier le nombre de classes à partir duquel le pourcentage d'inertie diminue beaucoup plus lentement lorsque l'on parcourt le diagramme de l'inertie cumulée.

Pour l'exemple ci-dessous, la lecture de ce diagramme et du diagramme d'inertie cumulée conduit à choisir 6 classes (ou clusters). Ces 6 classes sont visualisées sur l'arbre hiérarchique. Les clusters qui en découlent sont représentés ci-dessous sur les deux premiers axes de l'ACP.

Graphique 3 : Arbre hiérarchique et diagramme de l'inertie cumulée - ville de Lille



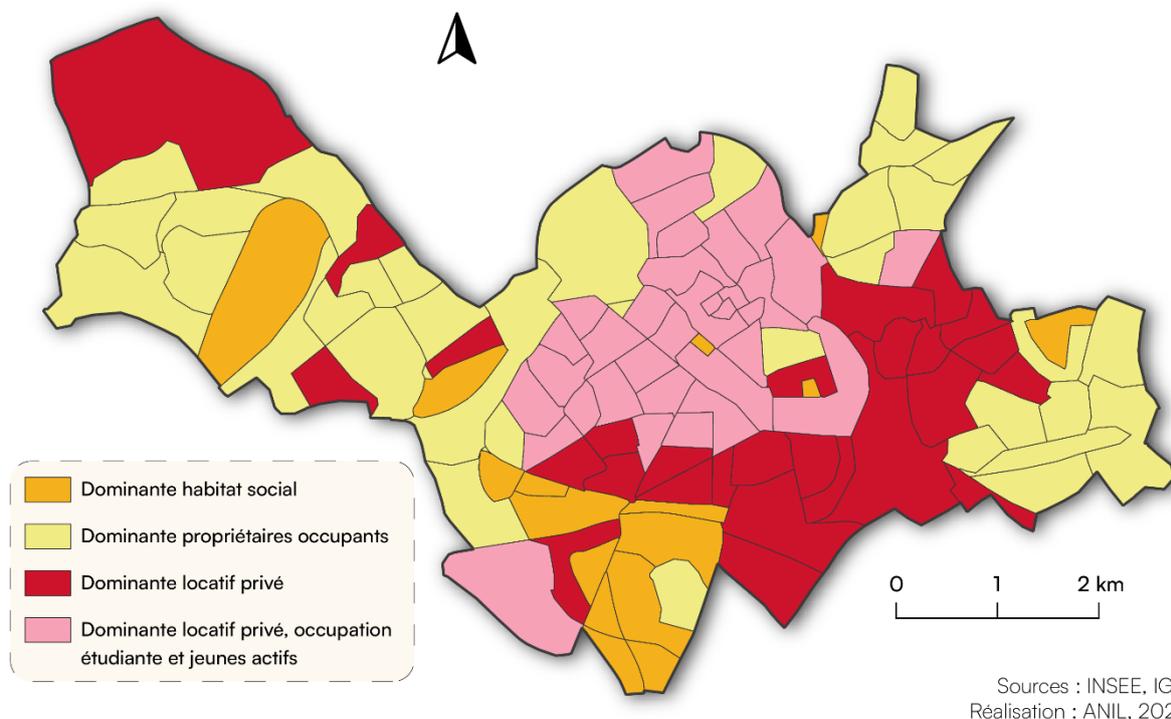
Graphique 4 : Représentation des clusters sur les deux premiers axes de l'ACP - ville de Lille



Etant donné la faiblesse du nombre individus (IRIS) classés dans les clusters 1 et 6, ces derniers sont regroupés respectivement avec les clusters 2 et 5.

In fine, on conserve 4 clusters. Ces derniers sont représentés sur une carte (cf. ci-dessous).

Carte 1 : Représentation des classes en carte - ville de Lille



En résumé, sur la ville de Lille, la variable localisation a été définie au moyen d'une ACP, qui établit quatre clusters, distinguant :

- Les quartiers à dominante d'habitat social, concentrés au Sud de la ville ;
- Les quartiers dont les logements sont très majoritairement occupés par leurs propriétaires ;
- Les quartiers d'habitat privé à dominante locative ;
- L'hypercentre, dont le parc locatif est largement occupé par des jeunes et des étudiants ;

La base de données est ensuite pondérée dans ce nouveau zonage en application de la méthodologie de calage sur marges définie dans les prescriptions méthodologiques du Comité scientifique des OLL.

3.1.2. Résultats obtenus pour la période 2021-2023

À partir des variables prêtes à intégrer au modèle, une régression Lasso est appliquée afin de sélectionner les variables retenues.

Une régression quantile est ensuite modélisée sur trois quantiles : le premier quartile (Q1), la médiane (Q2) et le troisième quartile (Q3).

Les modèles obtenus ainsi que leurs interprétations sont présentés ci-dessous :

Tableau 3 : Régression quantile appliquée aux Q1, Q2 et Q3 - données 2021 / 2023 - ville de Lille

Variables explicatives	Q1	Q2	Q3
(Intercept)	4.203***	4.299***	4.279***
`log(surfpiece)`	-0.180***	-0.216***	-0.249***
`log(Surface)`	0.621***	0.657***	0.714***
`bs(month_c_2, df = 3)1`	0.034***	0.051***	0.036***
`bs(month_c_2, df = 3)3`	0.040***	0.070***	0.080***
indic_maison (réf = appartement)	0.011	0.009	0.022
duree_occup moins de 1 an (réf = 3 ans et plus)	0.130***	0.119***	0.079***
duree_occup 1 à 2 ans (réf = 3 ans et plus)	0.034	0.070***	0.080***
indic_studio (réf = appartements de 2 pièces et plus)	0.140***	0.157***	0.163***
Epoque_cons avant 1946 (réf = après 2005)	-0.057***	-0.038***	-0.022***
Epoque_cons 1946-1970 (réf = après 2005)	-0.072***	-0.053***	-0.052***
Epoque_cons 1971-1990 (réf = après 2005)	-0.060***	-0.042***	-0.033***
Epoque_cons 1991-2005 (réf = après 2005)	0.013***	0.014***	0.014***
type_gestion directe (réf = gestion déléguée)	-0.002	0.013***	0.013***
trim_emm2 (référence = 1 ^{er} trimestre)	0.025	0.053***	0.020**
trim_emm3 (référence = 1 ^{er} trimestre)	0.034***	0.031***	0.030***
trim_emm4 (référence = 1 ^{er} trimestre)	0.016***	0.026***	0.020***
annee2021 (référence = année 2023)	-0.025***	-0.024***	-0.028***
annee2022 (référence = année 2023)	-0.016***	-0.019***	-0.021***
clust2 (référence= cluster 1)	0.187***	0.128***	0.136***
clust3 (référence= cluster 1)	0.074	0.110***	0.087***
clust4 (référence= cluster 1)	0.244***	0.230***	0.236***
`indic_studio1:clust3`	0.003	0.015**	0.041***
`duree_occup2:clust2`	0.071*	0.019	-0.022
`duree_occup1:clust3`	0.092*	0.0001	0.035**
`duree_occup2:clust3`	0.179***	0.036	0.013
`duree_occup1:clust4`	0.013	-0.020	-0.009
`duree_occup2:clust4`	0.096**	0.009	-0.018
`duree_occup1:trim_emm2`	-0.006	-0.038***	-0.016*
`duree_occup2:trim_emm2`	-0.033	-0.057***	-0.034***
`duree_occup2:trim_emm3`	-0.027***	-0.017***	-0.020***
`duree_occup1:indic_studio1`	-0.082***	-0.076***	-0.044***
`duree_occup2:indic_studio1`	-0.065***	-0.059***	-0.033***

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

La majorité des variables est significative au seuil de 1%.

Voici les différentes interprétations issues du modèle **sur la médiane du loyer (Q2)** :

- Une augmentation de 1 % de la surface entraîne une augmentation du loyer médian de 6,6 % ;
- Les studios présentent en moyenne des loyers médians 15,7 % plus chers que les loyers des autres typologies d'appartements, toutes choses égales par ailleurs ;
- Les logements occupés depuis moins de 1 an se louent 11,9 % plus chers que ceux occupés depuis plus de 3 ans. Cet écart traduit le saut des loyers à la relocation ;
- Les logements construits sur la période 1991-2005 se louent 1,4 % plus chers que ceux de la période d'après 2005 toutes choses égales par ailleurs, ce qui traduit un effet régulateur des dispositifs d'investissement locatif ;
- La saisonnalité de la période d'emménagement a quelques effets sur le loyer, les logements se louent 5,3 % plus chers au deuxième trimestre qu'au premier trimestre, 3,1 % plus cher au troisième trimestre qu'au premier trimestre et 2,6 % plus chers au quatrième trimestre du fait d'une rotation étudiante plus importante sur ces périodes ;
- Les loyers de 2023 se louent 1,9 % plus chers que ceux de 2022 et 2,4 % plus chers que ceux de 2021 ;
- Enfin, la modélisation montre que comparés aux logements des quartiers à dominante d'habitat social, les logements se louent plus chers dans les quartiers à dominante de

propriétaires, dans les quartiers d'habitat privé à dominante locative, ainsi que dans l'hypercentre, dont le parc locatif est largement occupé par des jeunes et des étudiants.

Lorsque l'on traduit les coefficients associés à l'indicateur temporelle en pourcentage (cf. partie 2.1. Présentation détaillée de la méthode des indicatrices temporelles), on obtient les résultats présentés dans le tableau 4 ci-dessous.

Tableau 4 : Taux d'évolution en % par quartile de loyer entre 2021 et 2023 - ville de Lille

	Taux d'évolution (en %)	Intervalle de confiance	R1
Q1	2,51	[1.799 , 3.221]	0,53
Q2	2,41	[1.750 , 3.070]	0,55
Q3	2,83	[2.118 , 3.542]	0,57

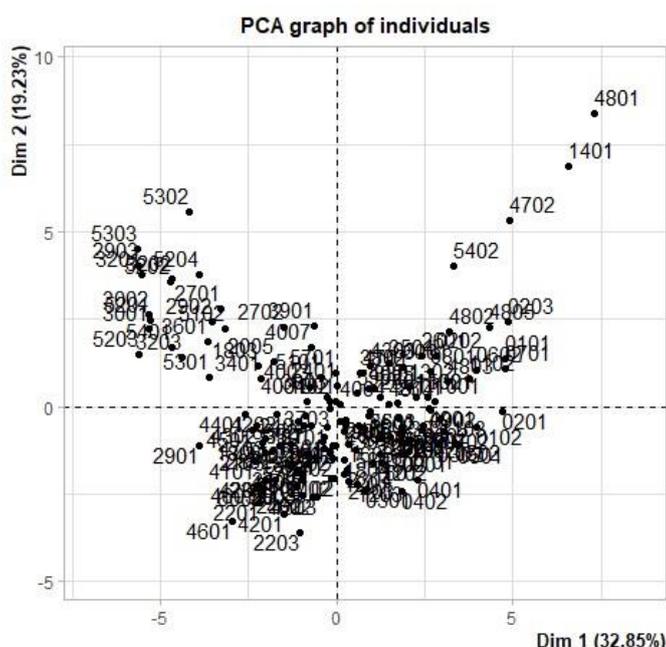
Le taux d'évolution calculé sur les loyers médians entre 2021 et 2023 s'établit à **2,41 %**. Les indicateurs liés aux premier et troisième quartiles de loyers montrent des taux d'évolution équivalents pour les 25 % des loyers les plus bas (+ 2,51 %) et des taux d'évolution plus élevés que pour les loyers médians pour les 25 % des loyers les plus hauts (+ 2,83 %). La qualité d'ajustement du modèle, établie par le R1, est satisfaisante.

3.2. Mise en œuvre de la méthodologie sur la ville de Toulouse

3.2.1. Construction d'une variable de localisation par clustering

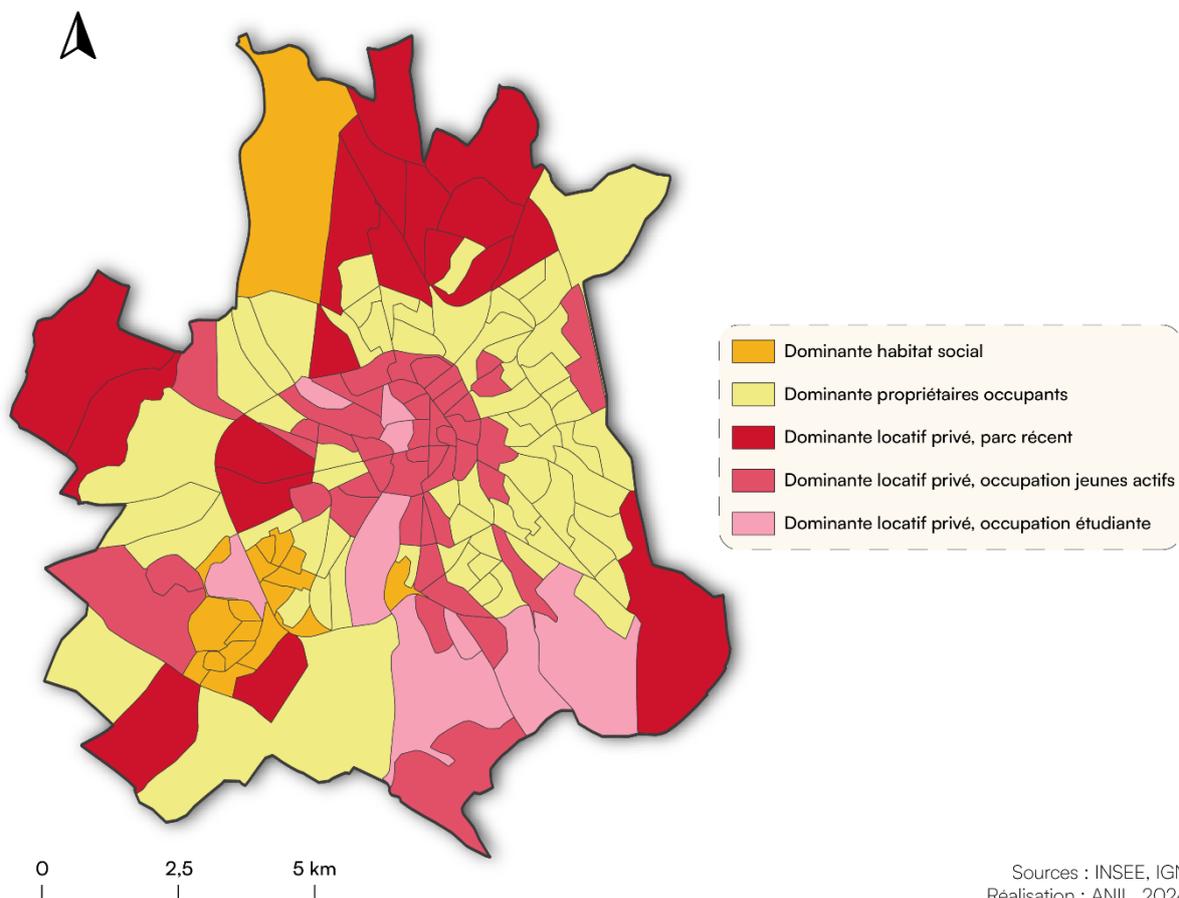
Comme pour la ville de Lille, la première étape consiste à construire un maillage basé sur les caractéristiques des logements et des habitants par ACP. On obtient ainsi une représentation des IRIS en fonction de leurs caractéristiques dans un espace bidimensionnel comme le montre le graphique ci-dessous.

Graphique 5 : Représentation des individus (IRIS) sur les deux premiers axes de l'ACP - ville de Toulouse



Enfin, on représente ces clusters sur une carte (cf. ci-après).

Carte 2 : Représentation des classes en carte - ville de Toulouse



En résumé, à Toulouse, la variable localisation a été définie au moyen d'une ACP, qui a établi cinq clusters, distinguant :

- Les quartiers à dominante d'habitat social ;
- Les quartiers dont les logements sont très majoritairement occupés par leurs propriétaires ;
- Les quartiers de faubourg au sein desquels on retrouve majoritairement un parc locatif privé récent ;
- Les quartiers plus centraux au sein desquels est majoritairement localisé le parc locatif privé ancien ;
- Les quartiers universitaires, dont le parc locatif est largement occupé par des étudiants ;

La base de données est ensuite pondérée dans ce nouveau zonage en application de la méthodologie de calage sur marges définie dans les prescriptions méthodologiques du Comité scientifique des OLL.

3.2.2. Résultats obtenus pour la période 2021-2023

À partir des variables prêtes à intégrer au modèle, une régression Lasso est appliquée afin de sélectionner les variables retenues.

Une régression quantile est ensuite modélisée sur trois quantiles : le premier quartile (Q1), la médiane (Q2) et le troisième quartile (Q3).

Les modèles obtenus ainsi que leurs interprétations sont présentés ci-dessous :

Tableau 5 : Régression quantile appliquée aux Q1, Q2 et Q3 - données 2021 / 2023 - ville de Toulouse

Variables explicatives	Q1	Q2	Q3
(Intercept)	4.228***	4.277***	4.222***
`log(surfpiece)`	-0.113***	-0.175***	-0.196***
`log(Surface)`	0.581***	0.636***	0.682***
`bs(month_c_2, df = 3)1`	0.010	0.009	0.018**
indic_maison (réf = appartement)	0.118***	0.114***	0.123***
duree_occup moins de 1 an (réf = 3 ans et plus)	-0.013	-0.013	-0.027***
indic_studio (réf = appartements de 2 pièces et plus)	0.012	0.088***	0.150***
Epoque_cons avant 1946 (réf = après 2005)	0.029***	0.066***	0.112***
Epoque_cons 1946-1970 (réf = après 2005)	-0.044***	-0.023***	-0.005
Epoque_cons 1971-1990 (réf = après 2005)	-0.011**	0.009*	0.020***
Epoque_cons 1991-2005 (réf = après 2005)	0.018***	0.026***	0.032***
type_gestion directe (réf = gestion déléguée)	-0.000	0.009***	0.019***
trim_emm3 (référence = 1 ^{er} trimestre)	0.023***	0.017***	0.019***
trim_emm4 (référence = 1 ^{er} trimestre)	0.023***	0.013***	0.013**
annee2021 (référence = année 2023)	-0.032***	-0.030***	-0.031***
annee2022 (référence = année 2023)	-0.023***	-0.021***	-0.021***
clust3 (référence= cluster 1)	0.051***	0.041***	0.030**
clust4 (référence= cluster 1)	0.036**	0.077***	0.103***
clust5 (référence= cluster 1)	0.136***	0.139***	0.131***
`indic_studio1:clust2`	0.043**	0.023	0.003
`indic_studio1:clust3`	-0.008	-0.017	-0.034***
`indic_studio1:clust4`	0.055***	0.028*	-0.033***
`duree_occup1:clust2`	0.157***	0.128***	0.137***
`duree_occup2:clust2`	0.132***	0.109***	0.096***
`duree_occup1:clust3`	0.062**	0.053***	0.061***
`duree_occup2:clust3`	0.043***	0.030***	0.019**
`duree_occup1:clust4`	0.158***	0.102***	0.136***
`duree_occup2:clust4`	0.111***	0.063***	0.057***
`duree_occup2:clust5`	0.004	-0.006	-0.026**
`duree_occup1.trim_emm2`	0.023***	0.016**	0.016

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

La majorité des variables est significative au seuil de 1%.

Voici les différentes interprétations issues du modèle sur la **médiane du loyer** :

- Une augmentation de 1 % de la surface entraîne une augmentation du loyer médian de 6,4 % ;
- Les maisons se louent 11,4 % plus chères que les appartements toutes choses égales par ailleurs ;
- Les studios présentent en moyenne des loyers médians 8,8 % plus chers que les autres typologies d'appartements toutes choses égales par ailleurs ;
- Hormis ceux construits sur la période 1946-1970, les logements se louent plus chers que ceux de la période d'après 2005 chers toutes choses égales par ailleurs ;
- Le mode de gestion et la saisonnalité de la période d'emménagement ont des effets modérés sur le loyer ;
- Les loyers de 2023 sont 2,1 % plus chers que ceux de 2022 et 3,0 % plus chers que ceux de 2021 ;
- Enfin, la modélisation montre que comparés aux logements des quartiers à dominante d'habitat social, les logements se louent plus chers dans les quartiers de faubourg au sein desquels se situe majoritairement un parc locatif privé récent, dans les quartiers plus centraux au sein desquels est majoritairement localisé un parc locatif privé ancien ainsi que dans les quartiers universitaires, dont le parc locatif est largement occupé par des étudiants.

Lorsque l'on traduit les coefficients associés à l'indicatrice temporelle en pourcentage (cf. partie 2.1. *Méthode des indicatrices temporelles*), on obtient les résultats présentés dans le tableau suivant :

Tableau 6 : Taux d'évolution en % par quartile de loyer entre 2021 et 2023 - ville de Toulouse

	Taux d'évolution (en %)	Intervalle de confiance	R1
Q1	3,26	[2.493 , 4.027]	0,50
Q2	3,05	[2.349 , 3.751]	0,53
Q3	3,13	[2.482 , 3.778]	0,56

Le taux d'évolution calculé sur le loyer médian entre 2021 et 2023 s'établit à **3,05 %**. Les indicateurs liés aux premier et troisième quartiles de loyers montrent des taux d'évolution légèrement plus élevés que pour les loyers médians pour les 25 % des loyers les plus bas (+ 3,26 %) ainsi que pour les 25 % des loyers les plus hauts (+ 3,13 %). La qualité d'ajustement du modèle, établie par le R1, est satisfaisante.

IV. Résultats obtenus sur l'ensemble des observatoires des loyers

Les indicateurs d'évolution ont été calculés pour l'ensemble des OLL sur deux périodes de deux années glissantes : 2020 – 2022 et 2021 – 2023.

Ce choix de calcul sur deux années glissantes a été conduit par l'interprétation parfois limitée de taux d'évolution sur un an en présence d'évolutions faibles. En effet, les évolutions de loyers antérieures à 2023 ont été lestées par un IRL bas pendant plusieurs années. Cela conduisait parfois à obtenir des taux d'évolution qui n'étaient pas significativement différents de 0. De manière plus générale, cette présentation d'un taux sur deux ans glissants est de nature à robustifier les indicateurs présentés en réduisant les intervalles de confiance.

En outre, il a été choisi de calculer les indicateurs à l'échelle des villes centres des périmètres des OLL. Le choix de cette maille se justifie déjà par le fait que le parc locatif y est proportionnellement plus important. Ensuite, le calcul d'un taux d'évolution à l'échelle d'une agglomération impliquerait une vision unique d'une situation largement hétérogène. En effet, les dynamiques de marché entre les villes centres des agglomérations et les communes périphériques peuvent être incomparables du fait de contextes de marchés différents, voire de contraintes réglementaires différenciées (encadrement des loyers en évolution et encadrement des loyers en niveau). Un indicateur mesuré à l'échelle de l'agglomération serait plus difficilement interprétable à l'aune de cette hétérogénéité de situations qui se répercute sur les résultats statistiques (plus forte dispersion des données et donc des marges d'incertitude plus élevées).

4.1. Évolution des loyers calculées par quartiles sur 2 années cumulées : 2020-2022

Cf. tableau page 26.

Pour la période 2020-2022, les niveaux d'évolution des loyers médians sont cohérents et représentatifs des dynamiques des marchés immobiliers. Les intervalles de confiance recourent néanmoins quelques fois zéro du fait d'évolutions contenues par un IRL bas sur cette période. Les R1 indiquent une qualité d'ajustement du modèle satisfaisante.

Les évolutions des 1^{er} et 3^e quartiles de la distribution des loyers rendent compte que les niveaux et les évolutions de loyers ne sont pas homogènes dans le haut (Q3) et le bas (Q1) de la distribution. L'évolution de loyer n'est pas toujours croissante à mesure que l'on prend en compte les déciles supérieurs. Une piste de travail complémentaire pourrait s'attacher à comparer les niveaux d'évolution par quartile pour tester leur égalité statistique afin d'identifier dans quelles agglomérations les niveaux d'évolution de loyers diffèrent effectivement en fonction de la position dans la distribution des loyers.

Leur analyse spatiale confirme la cohérence des résultats avec les dynamiques des marchés locatifs constatées localement. Plus précisément, on observe des évolutions de prix de loyers modérées dans les agglomérations les plus tendues, et donc concernées par la taxe sur les logements vacants (TLV) au sein desquelles les bailleurs sont soumis à un encadrement de loyer en cas de relocation ou de renouvellement du bail⁷. Les Indices de référence des loyers (IRL) moyens sur les années 2020 et 2021 s'établissent respectivement à 0,84 % et 0,40 % (soit 1,24 % au cumul). Ces niveaux s'avèrent être en correspondance étroite avec les variations annuelles de loyers dans le secteur libre effectivement calculées par l'Insee à l'échelle nationale via les indices des loyers d'habitation qui s'établissent respectivement à +0,55 % en moyenne en 2021 et + 0,90 % en moyenne en 2022 (soit +1,45 % au cumul).

⁷ Cf. Analyse juridique de l'ANIL : <https://www.anil.org/aj-encadrement-evolution-loyers-zones-tendues-2020>

Des taux d'évolutions sont en-deçà de cet IRL moyen sur deux ans au sein de deux villes appartenant à des agglomérations classées en zone tendue (Lyon et Montpellier). Des niveaux d'évolution inférieurs à 1,24 % sur deux ans sont par ailleurs observés dans quelques villes localisées dans le quart nord-est de l'Hexagone et inscrites dans un marché peu tendu (Arras, Nancy, Montbéliard, Mâcon).

Une part importante des évolutions de loyers mesurées pour les villes couvertes par un OLL s'établit légèrement au-dessus du niveau de l'IRL sans toutefois dépasser +2 % sur deux ans : ce groupe inclut de grandes agglomérations concernées par la réglementation sur l'encadrement de l'évolution des loyers⁸ (Toulouse, Bayonne, Bordeaux, La Rochelle). La qualité des indicateurs pour ces villes s'avère de surcroît très bonne. Les évolutions des loyers parisiens, calculées directement par l'OLAP selon une méthodologie différente, relèvent également de cette tranche.

Quelques villes concernées par ce classement TLV enregistrent cependant des niveaux d'évolutions plus importants au-delà de + 2 % sur deux ans (Nice, Marseille, Nantes, Sète, Strasbourg, Ajaccio, Lille). Mais l'ampleur reste à mettre en perspective avec des intervalles de confiance plus élevés, excepté pour Marseille et Lille.

Pour ce qui est des agglomérations qui ne sont pas concernées par ce classement TLV, ou ne l'étaient pas encore au moment de la collecte des données, un groupe d'agglomérations se distingue par des évolutions modérées sur deux ans eu égard à l'IRL (Arras, Brest, Alençon Nancy) d'un autre groupe d'agglomérations attractives, où les évolutions sont sensiblement plus marquées. Il s'agit de Lorient (+ 5,5 %), Vannes (+ 4,7 %), Rennes (+ 3 %) ou encore Nîmes (+3,3 %). Les agglomérations du Nord-Ouest de La Réunion sont également concernées par de telles hausses, puisque les loyers ont respectivement augmenté de 4,7 % et 4,5 % entre 2020 et 2022 à Saint-Denis et à Saint-Paul. Ces observations sont cohérentes avec d'autres mesures plus larges sur les tensions immobilières. Les deux agglomérations incluant Rennes et Nîmes viennent effectivement d'être classées comme tendues par décret ministériel, et le marché immobilier breton est soumis à une certaine pression ces dernières années, tous segments confondus, qui se traduit également par une forte hausse des prix dans le marché de l'achat-vente.

⁸ À ne pas confondre avec l'expérimentation de l'encadrement des loyers applicable dans quelques agglomérations de zones tendues (cf. Analyse juridique de l'ANIL : <https://www.anil.org/aj-experimentation-encadrement-niveau-loyers-zones-tendues/>)

Tableau 7 : Évolutions des loyers calculées par quartiles sur 2 années cumulées : 2020-2022

Ville	2020-2022								
	Q1	IC	R1	Q2	IC	R1	Q3	IC	R1
L2A00 - Ajaccio	5,05	[2.892 , 7.208]	0,45	4,86	[2.937 , 6.783]	0,47	4,3	[2.121 , 6.479]	0,49
L0600 – Nice	2,78	[1.348 , 4.212]	0,44	2,28	[1.056 , 3.504]	0,48	1,64	[0.455 , 2.825]	0,50
L1300 - Marseille	2,79	[1.562 , 4.018]	0,47	2,15	[1.429 , 2.871]	0,50	1,64	[0.664 , 2.616]	0,51
L1700 - La Rochelle	1,23	[-0.429 , 2.889]	0,54	1,78	[0.193 , 3.367]	0,58	2,42	[0.452 , 4.388]	0,59
L2500 - Besançon	4,97	[3.511 , 6.429]	0,49	2,86	[1.561 , 4.159]	0,51	3,17	[1.973 , 4.367]	0,53
L2501 - Montbéliard	0	[-5.46 , 5.46]	0,34	0,45	[-2.011 , 2.911]	0,41	3,58	[-0.807 , 7.967]	0,42
L2900 - Brest	1,16	[-0.819 , 3.139]	0,51	2,65	[0.166 , 5.134]	0,50	4,13	[1.429 , 6.831]	0,53
L3000 - Nîmes	3,05	[1.848 , 4.252]	0,51	3,25	[2.188 , 4.312]	0,54	3,6	[2.191 , 5.009]	0,57
L3100 - Toulouse	1,66	[1.048 , 2.272]	0,50	1,94	[1.275 , 2.605]	0,54	2,19	[1.531 , 2.849]	0,56
L3300 - Bordeaux	1,4	[0.553 , 2.247]	0,52	1,47	[0.657 , 2.283]	0,55	3,2	[2.351 , 4.049]	0,58
L3400 - Montpellier	0,25	[-0.508 , 1.008]	0,56	0,78	[0.203 , 1.357]	0,59	-0,68	[-1.594 , 0.234]	0,62
L3401 – Sète	1,56	[-0.625 , 3.745]	0,52	3,87	[1.508 , 6.232]	0,54	3,34	[1.024 , 5.656]	0,55
L3500 - Rennes	2,25	[1.555 , 2.945]	0,58	3,01	[2.287 , 3.733]	0,60	3,68	[2.826 , 4.534]	0,61
L3501 – Vitré	1,36	[-1.948 , 4.668]	0,54	0,47	[-1.403 , 2.343]	0,59	1,62	[-0.59 , 3.83]	0,62
L3502 - Saint-Malo	3,19	[1.158 , 5.222]	0,51	2,03	[0.312 , 3.748]	0,58	1,17	[-0.618 , 2.958]	0,59
L3700 - Tours	2,6	[1.623 , 3.577]	0,61	2,08	[0.835 , 3.325]	0,61	3,77	[2.417 , 5.123]	0,63
L3800 - Grenoble	2,37	[1.623 , 3.117]	0,56	1,61	[0.987 , 2.233]	0,57	1,48	[0.738 , 2.222]	0,57
L4400 - Nantes	2,95	[2.012 , 3.888]	0,59	3,54	[2.5 , 4.58]	0,60	4,3	[3.693 , 4.907]	0,61
L5400 - Nancy	1,59	[0.773 , 2.407]	0,54	1,08	[0.103 , 2.057]	0,58	0,57	[-0.216 , 1.356]	0,63
L5600 - Vannes	3,64	[2.578 , 4.702]	0,65	4,7	[3.443 , 5.957]	0,66	7,45	[5.939 , 8.961]	0,67
L5601 - Lorient	4,59	[2.959 , 6.221]	0,53	5,53	[4.346 , 6.714]	0,55	6,09	[3.959 , 8.221]	0,55
L5900 – Lille	2,08	[1.341 , 2.819]	0,53	2,69	[1.946 , 3.434]	0,55	2,54	[1.833 , 3.247]	0,58
L6100 - Alençon	2,39	[1.054 , 3.726]	0,51	1,94	[0.612 , 3.268]	0,54	2,47	[1.044 , 3.896]	0,57
L6200 - Arras	-1,54	[-2.992 , -0.088]	0,64	-1,27	[-2.495 , -0.045]	0,66	-1,69	[-2.874 , -0.506]	0,67
L6300 - Clermont-Ferrand	2,42	[1.36 , 3.48]	0,53	2,97	[1.827 , 4.113]	0,55	3,03	[2.101 , 3.959]	0,58
L6400 - Bayonne	0,54	[-1.483 , 2.563]	0,46	1,24	[-0.008 , 2.488]	0,51	2,65	[1.396 , 3.904]	0,54
L6700 - Strasbourg	3,35	[2.035 , 4.665]	0,56	2,68	[1.49 , 3.87]	0,57	2,24	[1.083 , 3.397]	0,58
L6900 - Lyon	0,08	[-0.998 , 1.158]	0,49	0,47	[-0.337 , 1.277]	0,52	0,43	[-0.662 , 1.522]	0,54
L7100 - Chalon sur Saône	0,92	[-1.413 , 3.253]	0,50	1,58	[0.067 , 3.093]	0,51	1,04	[-0.46 , 2.54]	0,54
L7101 - Mâcon	-0,23	[-1.95 , 1.49]	0,51	0,68	[-1.044 , 2.404]	0,53	0,08	[-1.43 , 1.59]	0,55
L8500 - La Roche sur Yon	6,03	[4.185 , 7.875]	0,57	5,66	[3.876 , 7.444]	0,59	8,18	[5.745 , 10.615]	0,63
L8501 – Les Sables d'Olonne	3,26	[1.495 , 5.025]	0,57	3,42	[1.814 , 5.026]	0,63	5,17	[3.426 , 6.914]	0,62
L8502 - Challans-Gois	3,05	[1.71 , 4.39]	0,67	2,71	[1.349 , 4.071]	0,68	2,76	[1.404 , 4.116]	0,65
L9740 - Saint-Denis (CINOR)	5,94	[4.287 , 7.593]	0,60	4,48	[3.256 , 5.704]	0,64	4,2	[2.906 , 5.494]	0,65
L9741 - Saint Paul (TCO)	5,66	[1.689 , 9.631]	0,47	4,71	[2.469 , 6.951]	0,53	4,13	[1.844 , 6.416]	0,56
L9742 - Saint Pierre (CIVIS)	0,63	[-0.927 , 2.187]	0,75	0,94	[-0.743 , 2.623]	0,71	2,68	[0.296 , 5.064]	0,67

4.2. Évolution des loyers calculées par quartiles sur 2 années cumulées : 2021-2023

Cf. tableau page 28.

Entre le 1^{er} janvier 2021 et le 1^{er} janvier 2023, les niveaux d'évolution des loyers médians connaissent une hausse assez significative comparée à la période précédente (2020-2022). Pour cause, le contexte inflationniste a été répercuté sur l'IRL dès le début de l'année 2022 suite au conflit ukrainien. Quand l'IRL moyen s'établissait à +0,40 % en 2021, il a bondi à +2,40 % en moyenne sur l'année 2022, avec un relèvement rapide des taux au deuxième trimestre (jusqu'à +3,60 %). À partir du troisième trimestre, les IRL ont d'ailleurs été régulés par la puissance publique. On se référera donc à la somme de ces deux IRL moyens (soit +2,80 %) avec prudence du fait d'une hétérogénéité importante de niveaux d'IRL qui ont pu s'appliquer sur la période en fonction des dates de reconductions ou de renouvellements des baux. Ces niveaux s'avèrent être en correspondance étroite avec les variations annuelles de loyers dans le secteur libre effectivement calculées par l'Insee à l'échelle nationale via les indices des loyers d'habitation qui s'établissent respectivement à +0,90 % en moyenne en 2022 et + 1,63 % en moyenne en 2023 (soit +2,53 % au cumul).

Du fait de taux d'évolution des loyers plus marqués sur cette période, les intervalles de confiance recourent peu fréquemment zéro ce qui permet d'interpréter les indicateurs avec plus de certitudes quant au sens de l'évolution. Les R1 témoignent d'une qualité d'ajustement du modèle satisfaisante.

Une analyse fine des résultats permet des constats comparables à ceux dressés sur les évolutions de loyers sur la période 2021-2023.

Ainsi, les villes avec des évolutions en dessous de l'IRL moyen sont des villes localisées sur un axe Nord/Sud-Est, composé essentiellement par de grandes agglomérations concernées par la TLV : Lille, Grenoble, Aix-en-Provence/Marseille, Montpellier. La qualité de l'indicateur de l'indicateur d'évolution est par ailleurs de bonne qualité. Des villes situées dans des plus petites agglomérations connaissent également des évolutions inférieures à l'IRL moyen : Arras, Montbéliard, Besançon, Chalon/Saône, Mâcon

Un autre groupe se détache du fait d'évolutions légèrement supérieures à l'IRL principalement constitué de grandes agglomérations concernées par la taxe sur les logements vacants : Lyon, Toulouse, Nantes, Bastia, Saint-Denis de La Réunion (nouvellement classée TLV). C'est également le cas de Paris, dont l'indicateur est issu de calculs propres à l'OLAP selon une autre méthodologie et de villes appartenant à des agglomérations avec tensions de marché modérées (Nancy, Auray, Alençon).

Enfin, certaines villes se distinguent par des évolutions plus nettes. Elles sont essentiellement localisées à l'Ouest du territoire et marquées par une attractivité résidentielles génératrice de pression immobilière. Les taux d'évolutions restent néanmoins en deçà de +5 % pour les villes situées au sein des agglomérations concernées par la TLV (Toulon, Strasbourg, Bordeaux, Rennes, ainsi que deux agglomérations de La Réunion : Saint-Paul et Saint-Pierre). Les villes connaissant les évolutions de loyers les plus rapides sont situées hors du zonage TLV : Lorient, Vannes, Saint-Malo, Vitré, excepté Ajaccio, La Rochelle, et trois communes du cœur du Pays Basque : Bayonne, Anglet et Biarritz.

Tableau 8 : Évolutions des loyers calculées par quartiles sur 2 années cumulées : 2021-2023

Ville	2021-2023								
	Q1	IC	R1	Q2	IC	R1	Q3	IC	R1
L2A00 - Ajaccio	4,25	[0.824 , 7.676]	0,45	6,63	[4.166 , 9.094]	0,46	5,82	[3.614 , 8.026]	0,49
L2B00 - Bastia	1,38	[-0.856 , 3.616]	0,42	3,56	[0.911 , 6.209]	0,43	4,38	[2.211 , 6.549]	0,43
L1300 - Marseille	1,79	[0.754 , 2.826]	0,47	2,15	[1.207 , 3.093]	0,49	2,69	[1.613 , 3.767]	0,51
L1302 - Aix en Provence	1,03	[-0.659 , 2.719]	0,57	1,91	[-0.16 , 3.98]	0,60	0,98	[-0.817 , 2.777]	0,63
L1700 - La Rochelle	10,05	[8.123 , 11.977]	0,52	8,27	[6.572 , 9.968]	0,56	11,16	[8.828 , 13.492]	0,56
L2500 - Besançon	3,51	[2.284 , 4.736]	0,49	2,61	[1.387 , 3.833]	0,51	3,44	[2.46 , 4.42]	0,53
L2501 - Montbéliard	0,42	[-3.079 , 3.919]	0,33	1,87	[-0.703 , 4.443]	0,39	2,23	[-1.718 , 6.178]	0,41
L2900 - Brest	3,56	[0.305 , 6.815]	0,51	4,94	[2.332 , 7.548]	0,52	1,87	[-0.187 , 3.927]	0,52
L3000 - Nîmes	0,65	[-1.031 , 2.331]	0,51	1,96	[0.68 , 3.24]	0,54	1,43	[0.265 , 2.595]	0,57
L3100 - Toulouse	3,26	[2.493 , 4.027]	0,50	3,05	[2.349 , 3.751]	0,53	3,13	[2.482 , 3.778]	0,56
L3300 - Bordeaux	4,1	[3.108 , 5.092]	0,50	4,59	[3.619 , 5.561]	0,54	6,64	[5.475 , 7.805]	0,56
L3400 - Montpellier	2,89	[2.171 , 3.609]	0,56	2,05	[1.433 , 2.667]	0,60	2,23	[1.555 , 2.905]	0,62
L3500 - Rennes	4,78	[3.92 , 5.64]	0,56	4,09	[3.076 , 5.104]	0,59	3,78	[3.093 , 4.467]	0,61
L3501 - Vitré	4,01	[-0.055 , 8.075]	0,52	5,39	[3.107 , 7.673]	0,58	5,22	[2.594 , 7.846]	0,60
L3502 - Saint-Malo	3,29	[1.427 , 5.153]	0,50	5,38	[3.697 , 7.063]	0,57	6,95	[5.299 , 8.601]	0,57
L3700 - Tours	3,14	[2.177 , 4.103]	0,63	2,04	[0.837 , 3.243]	0,61	4,83	[3.719 , 5.941]	0,61
L3800 - Grenoble	2,35	[1.146 , 3.554]	0,56	1,93	[0.91 , 2.95]	0,56	2,55	[1.453 , 3.647]	0,57
L4200 - Saint Etienne	5,73	[4.312 , 7.148]	0,41	4,27	[2.235 , 6.305]	0,44	4,43	[2.61 , 6.25]	0,47
L4400 - Nantes	2,53	[1.888 , 3.172]	0,58	3,18	[2.238 , 4.122]	0,60	4,34	[3.203 , 5.477]	0,60
L5400 - Nancy	2,91	[1.591 , 4.229]	0,54	3,43	[2.501 , 4.359]	0,59	4,2	[3.334 , 5.066]	0,63
L5600 - Vannes	4,37	[2.886 , 5.854]	0,63	5,09	[3.82 , 6.36]	0,64	5,2	[3.66 , 6.74]	0,65
L5601 - Lorient	5,97	[3.973 , 7.967]	0,51	6,68	[4.648 , 8.712]	0,53	8,64	[6.659 , 10.621]	0,54
L5602 - Auray	1,96	[-0.003 , 3.923]	0,56	3,62	[2.493 , 4.747]	0,59	5,75	[4.257 , 7.243]	0,61
L5900 - Lille	2,51	[1.799 , 3.221]	0,53	2,41	[1.75 , 3.07]	0,55	2,83	[2.118 , 3.542]	0,57
L6100 - Alençon	2,93	[1.361 , 4.499]	0,51	3,77	[2.393 , 5.147]	0,54	5,2	[3.458 , 6.942]	0,56
L6200 - Arras	2,83	[0.891 , 4.769]	0,62	2,18	[0.596 , 3.764]	0,65	3,62	[2.223 , 5.017]	0,66
L6300 - Clermont-Ferrand	4,01	[3.183 , 4.837]	0,54	4,4	[3.245 , 5.555]	0,56	4,56	[3.638 , 5.482]	0,59
L6400 - Bayonne / Anglet / Biarritz	5,13	[4.076 , 6.184]	0,45	5,94	[5.095 , 6.785]	0,50	6,08	[5.184 , 6.976]	0,53
L6700 - Strasbourg	4,6	[3.874 , 5.326]	0,52	4,54	[3.937 , 5.143]	0,55	5,31	[4.643 , 5.977]	0,56
L6900 - Lyon	3,37	[2.361 , 4.379]	0,50	3,3	[1.986 , 4.614]	0,53	3,02	[1.809 , 4.231]	0,55
L7100 - Chalon sur Saône	4,03	[2.128 , 5.932]	0,50	2,84	[1.507 , 4.173]	0,50	2,73	[1.444 , 4.016]	0,53
L7101 - Mâcon	0,32	[-2.255 , 2.895]	0,50	1,33	[-1.027 , 3.687]	0,53	0,99	[-1.199 , 3.179]	0,54
L8300 - Toulon	3,76	[1.589 , 5.931]	0,43	4,21	[2.38 , 6.04]	0,44	6,8	[5.329 , 8.271]	0,46
L8500 - La Roche-sur-Yon	3,08	[0.627 , 5.533]	0,57	5,33	[3.287 , 7.373]	0,61	4,78	[2.242 , 7.318]	0,62
L8501 - Les Sables d'Olonne	4,56	[2.023 , 7.097]	0,57	6,45	[4.691 , 8.209]	0,63	7,07	[5.449 , 8.691]	0,63
L8502 - Challans	2,04	[0.132 , 3.948]	0,66	2,78	[1.264 , 4.296]	0,67	4,05	[1.898 , 6.202]	0,64
L8503 - St-Gilles-Croix-de-Vie / St-Hilaire-de-Riez	0,75	[-3.375 , 4.875]	0,59	2	[-0.369 , 4.369]	0,56	2,43	[0.623 , 4.237]	0,51
L9740 - Saint-Denis (CINOR)	4,51	[3.06 , 5.96]	0,61	3,17	[2.068 , 4.272]	0,64	4,17	[2.932 , 5.408]	0,63
L9741 - Saint Paul (TCO)	3,77	[1.279 , 6.261]	0,46	4,86	[1.835 , 7.885]	0,51	3,38	[-0.414 , 7.174]	0,54
L9742 - Saint Pierre (CIVIS)	2,18	[-0.163 , 4.523]	0,72	4,38	[3.179 , 5.581]	0,67	7,72	[5.23 , 10.21]	0,64

4.3. Modalités de diffusion

4.3.1. Règles de diffusion

Tableau 9 : Classes de niveaux de confiance sur l'indicateur pour les évolutions 2021-2023

Ville	Taux d'évolution du loyer médian	Intervalle de confiance	Largeur de l'IC	Niveau de confiance associé à l'indicateur
L1300 - Marseille	2,2	[1.2 , 3.1]	1,9	Elevé
L1302 - Aix-en-Provence	1,9	[-0.1 , 4.0]	4,1	Intermédiaire
L1700 - La Rochelle	8,3	[6.6 , 10.0]	3,4	Intermédiaire
L2500 - Besançon	2,6	[1.4 , 3.8]	2,4	Intermédiaire
L2501 - Montbéliard	1,9	[-0.7 , 4.4]	5,1	Faible
L2900 - Brest	4,9	[2.3 , 7.5]	5,2	Faible
L2A00 - Ajaccio	6,6	[4.2 , 9.1]	4,9	Faible
L2B00 - Bastia	3,6	[0.9 , 6.2]	5,3	Faible
L3000 - Nîmes	2,0	[0.7 , 3.3]	2,6	Intermédiaire
L3100 - Toulouse	3,1	[2.3 , 3.7]	1,4	Elevé
L3300 - Bordeaux	4,6	[3.6 , 5.5]	1,9	Elevé
L3400 - Montpellier	2,1	[1.4 , 2.6]	1,2	Elevé
L3500 - Rennes	4,1	[3.1 , 5.1]	2,0	Elevé
L3501 - Vitré	5,4	[3.1 , 7.7]	4,6	Faible
L3502 - Saint-Malo	5,4	[3.7 , 7.1]	3,4	Intermédiaire
L3700 - Tours	2,0	[0.8 , 3.2]	2,4	Intermédiaire
L3800 - Grenoble	1,9	[0.9 , 2.9]	2,0	Elevé
L4200 - Saint Etienne	4,3	[2.3 , 6.3]	4,0	Intermédiaire
L4400 - Nantes	3,2	[2.2 , 4.1]	1,9	Elevé
L5400 - Nancy	3,4	[2.5 , 4.4]	1,9	Elevé
L5600 - Vannes	5,1	[3.8 , 6.3]	2,5	Intermédiaire
L5601 - Lorient	6,7	[4.7 , 8.7]	4,0	Intermédiaire
L5602 - Auray	3,6	[2.5 , 4.8]	2,3	Intermédiaire
L5900 - Lille	2,4	[1.8 , 3.1]	1,3	Elevé
L6100 - Alençon	3,8	[2.4 , 5.2]	2,8	Intermédiaire
L6200 - Arras	2,2	[0.6 , 3.8]	3,2	Intermédiaire
L6300 - Clermont-Ferrand	4,4	[3.2 , 5.5]	2,3	Intermédiaire
L6400 - Bayonne / Anglet / Biarritz	5,9	[5.1 , 6.8]	1,7	Elevé
L6700 - Strasbourg	4,5	[3.9 , 5.1]	1,2	Elevé
L6900 - Lyon	3,3	[2.0 , 4.6]	2,6	Intermédiaire
L7100 - Chalon sur Saône	2,8	[1.5 , 4.2]	2,7	Intermédiaire
L7101 - Mâcon	1,3	[-1.0 , 3.7]	4,7	Faible
L8300 - Toulon	4,2	[2.4 , 6.1]	3,7	Intermédiaire
L8500 - La Roche-sur-Yon	5,3	[3.3 , 7.4]	4,1	Intermédiaire
L8501 - Les Sables d'Olonne	6,5	[4.7 , 8.2]	3,5	Intermédiaire
L8502 - Challans	2,8	[1.3 , 4.3]	3,0	Intermédiaire
L8503 - Saint-Gilles-Croix-de-Vie et Saint-Hilaire-de-Riez	2,0	[-0.4 , 4.3]	4,7	Faible
L9740 - Saint-Denis (CINOR)	3,2	[2.1 , 4.3]	2,2	Intermédiaire
L9741 - Saint Paul (TCO)	4,9	[1.8 , 7.9]	6,1	Faible
L9742 - Saint Pierre (CIVIS)	4,4	[3.2 , 5.6]	2,4	Intermédiaire

Les taux d'évolution des loyers médians seront diffusés en valeur arrondie au dixième, et assortis d'un indicateur qui qualifie le niveau de confiance à considérer. Lorsque la marge d'erreur est trop élevée, l'indicateur n'est pas diffusé. Cet indicateur de qualité du taux d'évolution est défini à partir de la largeur de l'intervalle de confiance. Ainsi les classes de niveaux de confiance sur l'indicateur ont été établies pour la diffusion des taux d'évolutions sur la période 2021/2023 :

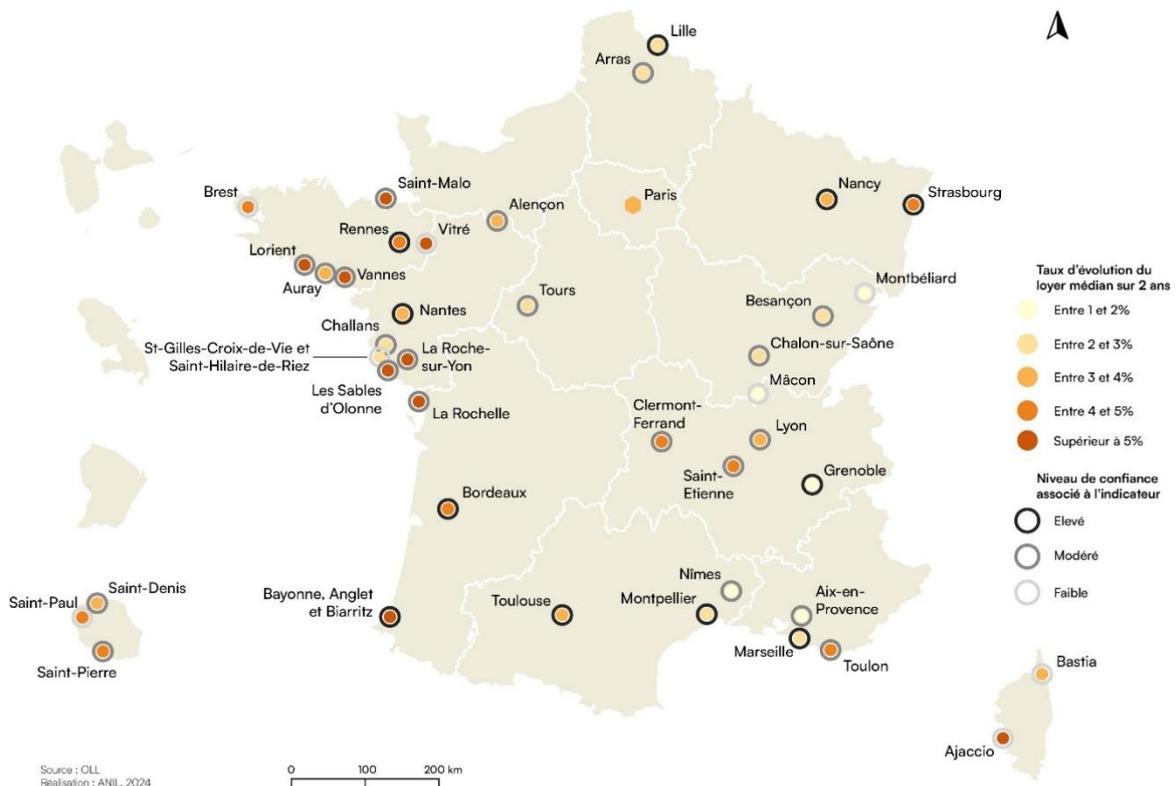
- lorsque la largeur de l'intervalle de confiance (IC) est inférieure ou égale à 2 pts de pourcentage, alors le niveau de confiance associé à l'indicateur est qualifié d'élevé ;
- lorsque la largeur de l'IC est située entre 2,1 et 4,1 pts de pourcentage alors le niveau de confiance associé à l'indicateur est qualifié d'intermédiaire ;
- lorsque la largeur de l'IC est supérieure à 4,1 pts de pourcentage, le niveau de confiance associé à l'indicateur est qualifié de faible.

4.3.2. Exemple de diffusion des indicateurs en carte

La carte ci-dessous illustre la diffusion des indicateurs selon les règles précisées ci-avant.

Cette représentation cartographique combine une information liée au taux d'évolution selon cinq classes à celle sur la fiabilité de l'indicateur. Elle permet également d'identifier plus visuellement des dynamiques différentes qui s'expriment au sein de différents territoires.

Carte 3 : Evolution des loyers médians par ville entre 2021 et 2023



NB : Pour Paris, les calculs ont été réalisés selon une méthodologie propre à l'OLAP grâce à un suivi des logements en panel. L'OLAP a ainsi calculé des taux d'évolution annuels de +0,6 % entre le 1^{er} janvier 2021 et le 1^{er} janvier 2022⁹ et de +2,5 % entre le 1^{er} janvier 2022 et le 1^{er} janvier 2023¹⁰.

⁹ [Evolution en 2021 des loyers d'habitation du secteur locatif privé dans l'agglomération parisienne](#)

¹⁰ [Evolution en 2022 des loyers d'habitation du secteur locatif privé dans l'agglomération parisienne](#)

Conclusion

Depuis 2013, le réseau des Observatoires locaux des loyers (OLL) améliore la connaissance du parc locatif privé en produisant des données fiables de niveaux de loyers au sein des différents territoires étudiés (62 agglomérations début 2024). Ces indicateurs de niveaux de loyers sont produits en vue de produire une vision segmentée des prix selon le type d'habitat, le nombre de pièces, l'époque de construction, l'ancienneté d'emménagement ou encore un zonage local dans les agglomérations grandes et moyennes. Cette méthode d'observation orientée par les pouvoirs publics sert la mise en œuvre de politiques de l'habitat, que ce soit pour alimenter des dispositifs réglementaires (calculs des loyers de référence pour application d'un encadrement des loyers promulgué par la loi ALUR et modifié par la loi ELAN, calcul de loyers plafonds du dispositif Loc'Avantages, ...), ou pour la programmation de l'offre locative.

Pour répondre à l'expression d'une attente d'indicateurs d'évolution des loyers, ne serait-ce que pour mesurer l'impact de certaines politiques publiques sur le marché, l'ANIL a développé une méthodologie économétrique analogue à celle utilisée par nombre d'instituts statistiques nationaux concernant l'observation des prix de transaction des logements, en l'occurrence par introduction d'indicateurs temporelles dans un modèle hédonique de loyers.

Pour une agglomération donnée, un modèle hédonique unique est ainsi estimé sur des observations collectées pour l'année N et N+1 en introduisant, en plus des variables explicatives usuelles (surface, nombre de pièces, etc.), une indicatrice temporelle sur l'année d'observation. Ceci permet de connaître l'évolution du loyer à structure constante par l'interprétation directe du coefficient estimé pour l'indicatrice temporelle.

Il a été choisi de modéliser l'effet des variables retenues sur les différents déciles de loyer plutôt que sur le loyer moyen afin de mieux rendre compte de l'hétérogénéité du niveau de tension sur les différents segments du marché et des pratiques des bailleurs en matière de revalorisations de loyers et aussi pour obtenir une information plus complète sur les dynamiques locales d'évolutions des loyers.

Cette méthode permet d'interpréter - à partir de l'indicatrice temporelle - des taux d'évolutions de loyers pour différents déciles de loyers et notamment un taux d'évolution sur le loyer médian.

L'interprétation des résultats obtenus montre que les niveaux d'évolution des loyers médians sont cohérents et représentatifs des dynamiques de marchés immobiliers constatées localement. En outre, les qualités d'ajustement des modèles ainsi que les niveaux de significativité des coefficients, en particulier ceux des indicatrices temporelles, confirment la robustesse de la méthode.

Les travaux de l'ANIL sur le développement d'un indicateur d'évolution des loyers ont été présentés au Comité scientifique des OLL au fur et à mesure de leur développement. À l'aune des résultats obtenus à partir de la méthodologie consolidée présentée dans cette note, le Comité scientifique s'est prononcé favorablement pour l'inscription de cette méthodologie au sein des prescriptions méthodologiques du réseau. Le calcul de taux d'évolutions des loyers sera ainsi systématisé et les OLL pourront diffuser ce nouvel indicateur complémentaire aux niveaux de loyers en rendant compte de visions plus dynamiques des marchés locatifs étudiés.

Les auteurs remercient les membres du Comité scientifique des OLL pour leurs précieux avis et conseils qui ont permis d'aboutir les travaux de recherche.

Références bibliographiques

- Balcone, T. 2013. « La Construction d'un Indice des Prix des Logements Neufs », Mimeo.
- Clapp, John M., Carmelo Giaccotto, et Dogan Tirtiroglu. 1991. « Housing Price Indices Based on All Transactions Compared to Repeat Subsamples ». *Real Estate Economics* 19 (3): 270-85. <https://doi.org/10.1111/1540-6229.00553>.
- Gatzlaff, Dean H., et David C. Ling. 1994. « Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies ». *Journal of Urban Economics* 35 (2): 221-44. <https://doi.org/10.1006/juec.1994.1014>.
- Hill, Robert J., Michael Scholtz, Chihiro Shimizu, et Miriam Steurer. 2018. « An evaluation of the methods used by European countries to compute their official house price indices ». *Economie et Statistique / Economics and Statistics*, n° 500t (octobre): 221-38. <https://doi.org/10.24187/ecostat.2018.500t.1953>.
- Lefebvre, L., et Rouquette, C. 2012. « Farmland Prices under Pressure from Urbanization ». *Economie & Statistique*.
- Mackinnon, J. G., & White, H. 1985. « Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties ». *Journal of econometrics*, 29(3), 305-325.
- Pusparum, Murih, Anang Kurnia, et Aam Alamudi. 2017. « Winsor approach in regression analysis with outlier ». *Applied Mathematical Sciences* 11 (41): 2031–2046.
- Rosen, Sherwin. 1974. « Hedonic prices and implicit markets: product differentiation in pure competition ». *Journal of political economy* 82 (1): 34–55.
- Straszheim, Mahlon. 1974. « Hedonic estimation of housing market prices: A further comment ». *The review of economics and statistics*, 404–406.
- Thouvenin, Nicolas. 2010. « La formation des prix des logements anciens ». 2010. <https://livre.fnac.com/a4756706/N-Thouvenin-La-formation-des-prix-des-logements-anciens>.
- Tibshirani, R. 1996. « Regression shrinkage and selection via the lasso ». *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
- Wallace, Nancy E., et Richard A. Meese. 1997. « The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches ». *The Journal of Real Estate Finance and Economics* 14 (1): 51-73. <https://doi.org/10.1023/A:1007715917198>.
- White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, 817-838.
- Wu, C. F. J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14(4), 1261-1295.