

apur

CESAER

OLAP
OBSERVATOIRE
DES LOYERS
agglomération parisienne

SeLoger

Evaluation d'impact de l'encadrement des loyers à PARIS¹

Marie Breuillé², Camille Grivault³, Julie Le Gallo², Yoann Morin², Martin Regnaud^{2,4}

En partenariat avec
l'Observatoire des Loyers de l'Agglomération Parisienne (OLAP)
et
le groupe Aviv (Meilleurs Agents et SeLoger)

Novembre 2023

¹ Cette étude a grandement bénéficié de la mise à disposition de données par l'OLAP et le groupe AVIV (Meilleurs Agents et SeLoger) ainsi que de l'expertise de Stéphanie Jankel et Jean-François Arènes de l'APUR, Geneviève Prandi (OLAP), Agnès Guerin-Battesti de la DLH, Jean-Claude Driant (EUP), Pierre Madec (OFCE) et plus largement des membres du comité scientifique.

² Centre d'Économie et de Sociologie appliquées à l'Agriculture et aux Espaces Ruraux (UMR CESAER)

³ Géographe, indépendant

⁴ Meilleurs Agents

Table des matières

1. INTRODUCTION	3
2. REVUE DE LA LITTÉRATURE	4
3. MESURE D'IMPACT DE L'ENCADREMENT DES LOYERS A PARIS SUR DONNEES DU GROUPE SELOGER..	6
3.1. DONNEES DU GROUPE SELOGER	6
3.1.1. <i>Collecte et contenu des bases de données</i>	6
3.1.2. <i>Traitements réalisés</i>	8
3.1.3. <i>Analyse statistique des dépassements de loyers</i>	9
3.1.4. <i>Evolution de la probabilité de dépassement avec l'encadrement des loyers</i>	19
3.2. METHODOLOGIE	23
3.2.1. <i>Avantages et limites des principales méthodes d'inférence causale pour évaluer l'impact de l'encadrement des loyers</i>	23
3.2.2. <i>La méthode de la différence-de-différences</i>	24
3.3. RESULTATS	28
3.3.1. <i>Evaluation de l'impact de l'encadrement des loyers</i>	28
3.3.2. <i>Analyse d'hétérogénéité par rapport aux caractéristiques</i>	29
3.3.3. <i>Analyse d'hétérogénéité par niveau de contrainte des loyers</i>	34
3.3.4. <i>Effet de l'encadrement si tous les biens avaient un loyer inférieur ou égal au seuil maximum autorisé</i>	36
3.3.5. <i>Analyse de robustesse</i>	37
4. MESURE D'IMPACT DE L'ENCADREMENT DES LOYERS A PARIS SUR DONNEES DES OBSERVATOIRES LOCAUX DES LOYERS ACCESSIBLES SOUS CASD	39
4.1. DONNEES DES OLL	39
4.2. METHODOLOGIE	41
4.3. RESULTATS D'EVALUATION DE L'IMPACT DE L'ENCADREMENT DES LOYERS	41
5. MESURE D'IMPACT DE L'ENCADREMENT DES LOYERS A PARIS SUR DONNEES « REFERENCES DE VOISINAGE » DE L'OLAP	45
5.1. DONNEES REFERENCES DE VOISINAGE	45
5.2. METHODOLOGIE	45
5.2.1. <i>Modèle de différence-de-différences sans prise en compte des externalités</i>	46
5.2.2. <i>Modèle de différence-de-différences avec prise en compte des externalités</i>	47
5.3. RESULTATS	48
5.3.1. <i>Evaluation de l'impact de l'encadrement des loyers sans prise en compte des externalités</i> ..	48
5.3.2. <i>Evaluation de l'impact de l'encadrement des loyers avec prise en compte des externalités</i> .	50
6. CONCLUSION	53
RÉFÉRENCES	54
ANNEXES	55

1. Introduction

Paris est la première ville française à avoir instauré l'encadrement des loyers, et ce à deux reprises. Dans le cadre de la loi ALUR du 24 mars 2014, l'encadrement des loyers parisiens est tout d'abord entré en vigueur du 1^{er} août 2015 jusqu'au 28 novembre 2017, date de son annulation par décision du tribunal administratif de Paris (confirmée en appel en juin 2018). Depuis le 1^{er} juillet 2019, un nouveau dispositif d'encadrement a pris la forme d'une expérimentation pour une durée initialement prévue pour 5 années dans le cadre de la loi ELAN du 23 novembre 2018, qui a été prolongée par la loi « 3DS » du 21 février 2022 de 3 années supplémentaires, soit jusqu'au 23 novembre 2026. Dans la continuité de Paris, d'autres villes et des établissements publics de coopération intercommunale (EPCI) appartenant à des « zones tendues » ont instauré le dispositif expérimental d'encadrement des loyers : Lille depuis le 1^{er} mars 2020, l'EPCI de Plaine Commune depuis le 1^{er} juin 2021, Lyon et Villeurbanne depuis le 1^{er} novembre 2021, l'EPCI d'Est Ensemble depuis le 1^{er} décembre 2021, Montpellier depuis le 1^{er} juillet 2022 et enfin Bordeaux depuis le 15 juillet 2022, soit 24 villes au total.

Une évaluation de ce dispositif expérimental est explicitement prévue par la loi, ce qui a conduit l'APUR à l'inscrire dans son programme de travail 2022/2023 établi avec ses différents partenaires, notamment les services de l'Etat et la Ville de Paris, et à missionner notre équipe de recherche pour la réaliser. L'évaluation de cette politique publique requiert en effet la mise en œuvre de méthodes capables d'identifier l'effet causal de l'encadrement, en particulier sur les montants des loyers et des transactions, parmi les autres facteurs susceptibles d'exercer une influence (e.g., crise de la Covid-19, loi Climat et résilience, crise énergétique ou encore guerre en Ukraine). Plus précisément, ces méthodes consistent à reconstituer un contrefactuel dans le cadre d'une expérience quasi-naturelle pour mesurer l'effet causal en comparant les loyers (ou transactions) observés avec ceux qui l'auraient été en l'absence du dispositif. Si ces méthodes ont déjà été mobilisées pour évaluer des dispositifs d'encadrement des loyers à l'étranger, elles le sont pour la première fois sur le dispositif français actuellement en vigueur.

Les conséquences du premier dispositif d'encadrement avaient été analysées par Malard et Poulhès (2020) pour le Commissariat général au développement durable, sur la base de données collectées par l'OLAP. A l'aide d'un modèle logistique, ils prédisent un taux de dépassement théorique sur la période d'encadrement et le comparent avec le taux observé. Cette méthode fige les effets fixes temporels aux périodes antérieures, et risque ainsi de conduire à un biais dans la mesure de l'effet de l'encadrement, qui peut être surestimé ou sous-estimé par l'effet simultané d'autres chocs (Covid-19, loi Climat et résilience, etc.). Sans utilisation de contrefactuel qui aurait subi tous les autres chocs à l'exclusion de l'encadrement, il n'est pas possible d'isoler l'effet causal de l'encadrement. De plus, ils rétroplacent à la période antérieure (entre le 1^{er} janvier 2014 et le 31 juillet 2015) les seuils de l'encadrement au 1^{er} août 2015 afin de calculer le taux de dépassement pré-encadrement, avec pour conséquence de potentiellement sous-estimer ce taux en raison du trend de hausse des loyers.

Un suivi régulier de l'évolution des loyers depuis leur encadrement est réalisé par l'Observatoire des Loyers de l'Agglomération Parisienne (OLAP) qui publie chaque année une note sur l'encadrement des loyers à Paris. Dans la dernière note de décembre 2022, l'OLAP observe que 30% des loyers excèdent le plafond, avec un dépassement moyen de 12% du loyer total. Par ailleurs, un baromètre de l'encadrement des loyers, élaboré conjointement par la Ville de Paris et la fondation Abbé Pierre en 2021, est diffusé chaque année, fournissant les résultats de l'exploitation de l'extension informatique « encadrement » mise en place pour permettre aux locataires de vérifier la conformité des annonces avec la loi⁵.

⁵ <https://www.fondation-abbe-pierre.fr/nos-publications/communiqués-de-presse/2eme-barometre-de-lobservatoire-de-lencadrement-des-loyers>

Notre étude est menée en partenariat avec l'OLAP et le groupe AVIV (Meilleurs Agents – SeLoger). Trois bases de données complémentaires sont mobilisées dans le cadre de cette étude : 1) les annonces diffusées sur les sites du groupe SeLoger depuis le 1^{er} janvier 2018, 2) les données collectées par les observatoires locaux des loyers et disponibles sous le CASD et 3) les données individuelles issues de la base « références de voisinage » collectées par l'OLAP.

Dans la section 2, nous réalisons une revue de la littérature académique sur les conséquences de l'encadrement des loyers. Dans la section 3, nous présentons les résultats de l'estimation de l'impact de l'encadrement parisien avec une méthode de différence-de-différences appliquée aux données du groupe SeLoger de Paris et d'une sélection de grandes villes. Dans la section 4, nous appliquons la même méthodologie sur les données des observatoires locaux des loyers accessibles sous CASD. Dans la section 5, nous mobilisons les données de la base « références de voisinage » de l'OLAP sur Paris et sa petite couronne avec une méthode de différence-de-différences qui tient compte des effets de débordement (*spillovers*) sur les communes proches de Paris. Enfin, nous concluons dans la dernière section.

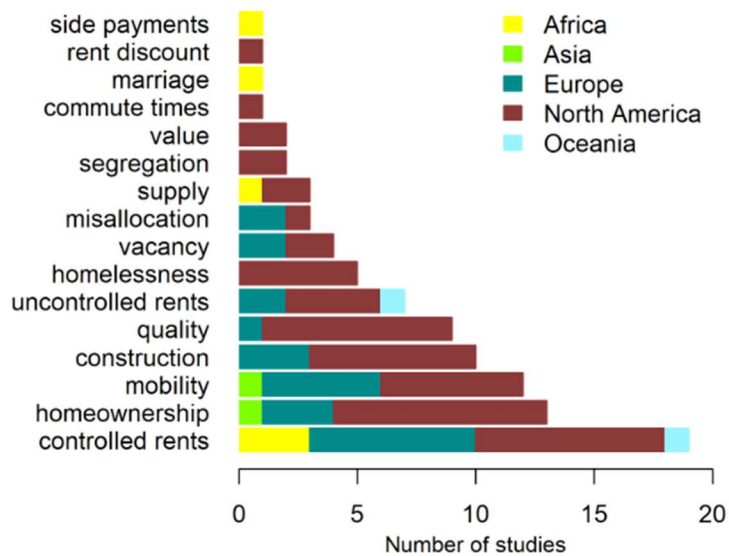
2. Revue de la littérature

Des encadrements de « première génération », qui consistent en un gel des loyers, ont été appliqués dès le début du 20^{ème} siècle, avec la loi britannique de 1915 sur l'augmentation des loyers et des intérêts hypothécaires, et reproduits dans plusieurs pays d'Europe durant l'entre-deux guerres. Certaines villes Etatsuniennes ont également mis en place différentes formes de contrôle des loyers, à l'instar de New York, Washington D.C., San Francisco, Boston, Cambridge ou Los Angeles. D'autres états en revanche interdisent entièrement les lois sur le contrôle des loyers. Ces expériences récentes relèvent de ce qu'on nomme le contrôle des loyers de « troisième génération », qui limite les augmentations de loyer dans les locations. Dans certains cas en Europe, les politiques mises en place relèvent du contrôle de « seconde génération » qui limitent les augmentations de loyers dans et entre les locations.

La mesure de l'efficacité de l'encadrement sur les loyers, et plus largement sur le marché immobilier voire la société, a suscité de nombreuses controverses (Arnott, 1995 ; Pastor et al., 2018). Mis en place pour favoriser l'accession à la location ou protéger les locataires d'augmentations excessives (préservant ainsi leur capital social en leur permettant de rester dans leur location), l'encadrement des loyers a également fait l'objet de nombreuses critiques relatives aux conséquences négatives à long terme. D'un point de vue théorique, en distordant les signaux-prix du marché, cette politique entraînerait une mauvaise allocation des ressources immobilières, une diminution de l'offre de logements et une réduction de la qualité des logements locatifs suite à la moindre incitation pour les propriétaires à entretenir et à investir dans des biens locatifs. D'autres conséquences ont également été pointées, qu'il s'agisse des effets sur la gentrification, la vacance, la construction, ou encore la part des sans domiciles fixes.

Les études empiriques sont nombreuses et s'intéressent à la fois aux effets voulus de la politique, i.e. l'impact sur le niveau des loyers, mais également sur toutes les autres dimensions affectées par cette politique. Le graphique suivant issu de Kholodilin (2018) comptabilise le nombre d'études empiriques selon le continent et la variable d'intérêt :

Figure 1 : Nombre d'études sur les effets de l'encadrement des loyers selon le continent et la variable d'intérêt



Source : Kholodilin, 2018

Nous focalisons notre revue de littérature sur les études empiriques évaluant l'impact des politiques d'encadrement des loyers sur les niveaux des loyers, en ciblant celles qui mobilisent des méthodes d'inférence causale, permettant une évaluation d'impact causale de ces politiques, comme les études mises en œuvre dans ce livrable.

La première étude empirique permettant de mettre en évidence un impact causal d'une politique d'encadrement des loyers sur différents segments du marché, dont les loyers, est celle de Sims (2007). Exploitant l'arrêt soudain de l'encadrement des loyers dans le Massachusetts en 1995, il montre, à l'instar de travaux subséquents, que si l'encadrement a permis de réduire substantiellement les loyers, elle a aussi incité les propriétaires à retirer leur bien du marché locatif et a contribué à détériorer la qualité des logements.

En France, une seule étude est à noter, celle de Bonneval et al. (2021) qui s'intéressent au cas de Lyon qui, sur la période 1890-1968 a connu différentes formes d'encadrement des loyers. Sur la base de données historiques de baux et en utilisant les appartements non soumis à l'encadrement comme groupe de contrôle, ils trouvent que l'impact de l'encadrement s'est renforcé dans le temps avec une baisse induite de 11% entre 1914 et 1929 pour atteindre 47% sur la période 1949-1968.

Monras et Montavo (2022) et Jofre-Monseny et al. (2022) s'intéressent au contrôle des loyers de seconde génération mis en place en Catalogne (Espagne) en septembre 2020, consistant à établir un loyer de référence pour chaque bien sur le marché, dépendant du loyer moyen des 25 biens les plus proches. Toutes les villes ne sont pas affectées par cette politique : seules les villes d'une taille supérieure à 20 000 habitants et affectées par des marchés locatifs en tension le sont. Les biens situés dans des villes non encadrées sont utilisés comme point de comparaison pour évaluer l'impact de l'encadrement sur les loyers et l'offre locative. Si l'encadrement a bien permis de diminuer les loyers encadrés de 5% relativement aux loyers non encadrés, les résultats de ces deux études diffèrent quant à son impact sur l'offre locative. En effet, alors que Monras et Montavo (2022) trouve qu'elle a également conduit à réduire l'offre de biens offerts à la location sur le marché, Jofre-Monseny et al. (2022) n'obtiennent pas ce résultat, la différence étant due à la différence de périmètre de l'étude, de période exacte étudiée et du nombre de données dans les échantillons. Kholodilin et al. (2022) trouvent en outre que l'arrêt de cette mesure en mars 2022 a causé une augmentation importante des loyers et des prix.

Le cas de l'Allemagne a été analysé par plusieurs articles. D'une part, Thomschke (2019) et Breidenbach et al. (2022) se focalisent sur l'encadrement des loyers de deuxième génération mis en place en 2015. D'autre part, Hahn et al. (2021) s'intéressent à la politique d'encadrement à Berlin mise en place en février 2020 et pouvant s'apparenter à un encadrement de première génération puisqu'elle détermine un loyer au m² maximum. Les trois articles s'appuient sur des données d'annonces. En comparant l'évolution des loyers dans 6 villes par rapport à une ville test (Leipzig), Thomschke (2019) montre que l'encadrement des loyers mis en place en 2015 a globalement permis une réduction des loyers d'annonce mais a eu aussi des effets spatialement différenciés. Si les loyers offerts à Hamburg, Berlin et Munich ont décru jusqu'à 5%, il n'y a pas eu d'effet significatif à Cologne et Düsseldorf et des effets en dessous de ce qui était attendu dans les autres villes de l'échantillon. Des résultats sont confirmés par Breidenbach et al. (2022) sur l'ensemble du pays mais la baisse des loyers constatés apparaît essentiellement temporaire. De leur côté, en se basant sur des données d'annonce parues entre 2018 et 2020 à Berlin, Hahn et al. (2021) trouvent un effet de réduction important entre 7 et 11% sur les loyers à la mise en place de la politique ainsi qu'un effet de substitution affectant quelques villes satellites autour de Berlin, comme Potsdam et une réduction du nombre d'offres.

Enfin, le cas de l'Irlande a été examiné par O'Toole et al. (2021) où un encadrement de 4% sur les changements de loyers est mis en place. Si en moyenne, le taux de croissance des loyers a diminué de 1 à 2% suivant la mise en place de cette régulation, les impacts sont différents le long de la distribution des loyers et des pressions à la hausse sur des loyers précédemment non affectés et que les propriétaires ont rapproché du seuil.

3. Mesure d'impact de l'encadrement des loyers à Paris sur données du groupe SeLoger

3.1. *Données du groupe SeLoger*

3.1.1. *Collecte et contenu des bases de données*

Notre premier jeu de données est constitué de l'ensemble des annonces diffusées au moins un jour sur un des **sites du Groupe SeLoger** (SeLoger.com, Logic-Immo et BellesDemeures.com) **entre le 1^{er} janvier 2018 et le 30 juin 2023**. En cas de mises à jour des annonces, nous retenons le loyer associé à l'annonce la plus récente. Ce jeu de données comprend, avant tout traitement, 5 015 485 annonces de locations d'appartements publiées par des professionnels de l'immobilier (agents immobiliers ou mandataires) sur la France métropolitaine.

Les annonces sont **géolocalisées** à partir de l'adresse textuelle renseignée par les agents immobiliers, qui comprend le numéro de la rue, le nom de la rue, le code postal et le nom de la commune. Une première étape consiste, pour chaque observation, à retrouver l'adresse dans une base d'adresses interne à Meilleurs Agents (qui appartient également au groupe AVIV). Cette dernière recoupe les

données des bases publiques BAN⁶ et BANO⁷. Dans le cas où l'adresse exacte de l'annonce n'est pas retrouvée, nous recherchons la rue et en dernier recours la commune. Les volumes obtenus sont reportés dans le Tableau 1 :

Tableau 1 : Géolocalisation des données

Niveau de géolocalisation le plus fin disponible	Nombre de données concernées	Proportion du jeu de données total
Adresse	2 025 311	40.4%
Rue	289 201	5.7%
Commune	2 677 256	53.4%
Géolocalisation impossible de la commune	23 717	0.5%

Les variables associées aux caractéristiques des logements sont reportées dans le Tableau 2 avec leur taux de complétion calculé sur le jeu de données pour l'ensemble de la France métropolitaine.

Tableau 2 : Liste des variables disponibles avec leur taux de complétion

Variables	Taux de complétion (%)
Date de début de l'annonce (première apparition sur le site)	100
Type de bien (appartement ou maison)	100
Sous-type de bien (loft, duplex, etc.)	100
Libellé de la commune	100
Code postal	100
Le bien est-il neuf ?	100
Nombre de pièces	99.97
Piscine	98.30
Surface	98.14
Nombre de chambres	96.13
Loyer charges comprises	95.30
Le bien est-il meublé ?	87.81
Loyer hors charges	87.36
Etage	84.92
Nombre de toilettes	81.98
L'immeuble a-t-il un ascenseur ?	74.02
Nombre de salles de bains	73.28
DPE	66.86
Type de cuisine (ouverte, équipée, ...)	59.44

⁶ <https://www.data.gouv.fr/fr/datasets/base-adresse-nationale/>

⁷ <https://www.data.gouv.fr/fr/datasets/base-d-adresses-nationale-ouverte-bano/>

Nombre de balcons	54.22
Y a-t-il une cave ?	50.86
Nombre d'étages de l'immeuble	50.38
Nombre de parking extérieurs	47.65
Année de construction	36.77
Nombre de terrasses	32.36
Y a-t-il une cheminée ?	14.74
Surface du balcon	4.73

3.1.2. Traitements réalisés

Nous avons tout d'abord **écarté les annonces avec une caractéristique majeure manquante**, à savoir la surface (93 449 annonces soit 1.9%), le nombre de pièces (1684 annonces soit 0.03%). Nous écartons aussi les annonces dont le libellé de la commune n'est pas reconnu dans notre référentiel géographique (23 717 annonces soit 0.5%).

Ensuite, les annonces étant potentiellement postées par un même agent sur plusieurs sites du groupe SeLoger, nous avons procédé à leur **déduplication**. Afin de réaliser ce traitement, nous nous sommes appuyés sur le texte et sur les caractéristiques des biens. Nous supprimons tout d'abord les doublons évidents qui présentent des textes identiques après conversion des textes en minuscule, sans ponctuation (remplacée par un espace) et sans numéro de référence d'annonce. Cette étape élimine 22.2% des annonces.

Nous procédons ensuite à une déduplication à partir des caractéristiques du bien, en ne conservant qu'une annonce lorsque deux annonces ont :

- Moins de 35 jours d'écart entre les dates de début
- Un écart de surface < 1% de la plus grande des surfaces
- Un écart des loyers < 5% du plus grand des loyers
- Le même étage
- La même meilleure maille de géolocalisation. Par exemple, les deux annonces sont géolocalisées au 7, boulevard Haussmann ou les deux annonces sont géolocalisées dans le quartier des Halles et leurs adresses sont manquantes, ou enfin, nous n'avons retrouvé ni l'adresse ni la rue de l'annonce mais elles sont toutes les deux géolocalisées dans la même commune.

Nous gardons toujours l'annonce qui a le moins de données manquantes et, s'il n'y a pas de différences sur ce point, nous retenons l'annonce publiée en dernier. Cette deuxième étape de déduplication réduit de 12,2% la taille du jeu de données restant, faisant passer le volume total des annonces postées sur les sites du groupe SeLoger sur la période considérée en France métropolitaine à 3 424 523 annonces. En définitive, la déduplication en deux étapes a réduit l'échantillon de 31.7% des annonces, dont la grande majorité des suppressions est imputable à la déduplication de textes identiques.

Enfin, le dernier traitement a consisté à **filtrer les valeurs extrêmes** des champs, pour ne retenir que les annonces avec les caractéristiques suivantes :

- Une surface comprise entre 9 et 250 m² (pas de valeurs manquantes)
- Un nombre de pièces compris entre 1 et 7 (pas de valeurs manquantes)
- Un nombre de chambres compris entre 0 et 6, avec possibilité de valeurs manquantes

- Un loyer hors charges compris entre 100 et 10 000 € (pas de valeurs manquantes)
- Un loyer hors charges au m² compris entre 2 et 200 €/m² (pas de valeurs manquantes)
- Des charges comprises entre 0 et 30% du loyer hors charges, avec possibilité de valeurs manquantes
- Un nombre de balcons (resp. terrasses) compris entre 0 et 5, avec possibilité de valeurs manquantes
- Un nombre de salles de bain (resp. salles d'eau ou toilettes) compris entre 0 et 5, avec possibilité de valeurs manquantes
- Un nombre d'étages entre 0 et 20, avec possibilité de valeurs manquantes
- Une année de construction comprise entre 1500 et 2022, avec possibilité de valeurs manquantes

Après l'ensemble de ces traitements, **notre échantillon d'annonces mises en ligne entre le 1^{er} janvier 2018 et le 30 juin 2023 est constitué de 2 424 056 observations pour l'ensemble de la France métropolitaine et de 284 221 annonces à Paris intra-muros.**

3.1.3. Analyse statistique des dépassements de loyers

Afin de mener une analyse statistique des dépassements de loyers, nous nous restreignons aux annonces de notre jeu de données qui sont localisées dans la ville de Paris (284 221 annonces). Sur ce jeu, nous devons calculer le loyer de référence majoré qui s'applique à chaque annonce en fonction des caractéristiques du bien (secteur géographique, nombre de pièces, époque de construction, meublé ou non) pour le comparer au loyer hors charges de l'annonce.

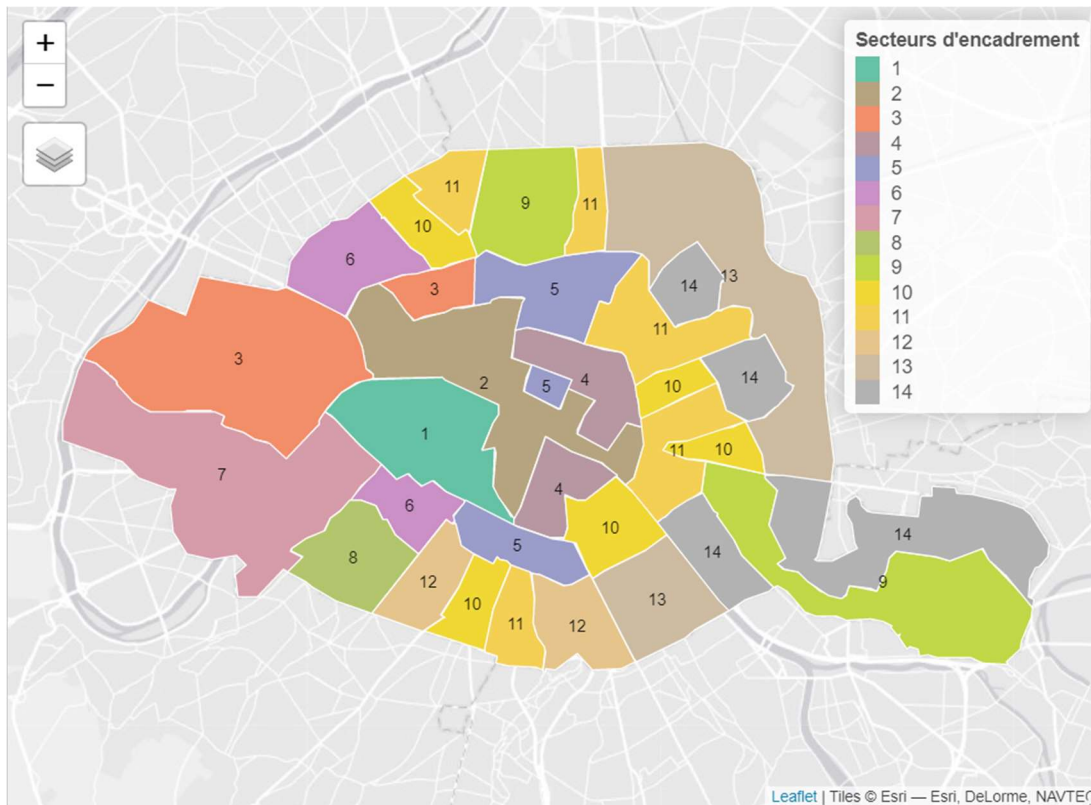
Calcul du loyer de référence majoré pour chaque annonce

Comme certaines de ces caractéristiques sont manquantes, nous avons dû, soit récupérer l'information manquante, soit écarter ces annonces.

1. Détermination du secteur géographique

Nous relierons tout d'abord chaque annonce avec une adresse complète à son secteur géographique (58% des annonces parisiennes). Nous cherchons ensuite à relier chaque annonce avec une adresse incomplète à un des 14 secteurs géographiques (voir Figure 2). Il est aisé d'affecter un secteur géographique à une annonce avec une adresse incomplète lorsque la rue est connue et complètement incluse dans un même secteur.

Figure 2 : Représentation graphique des secteurs d'encadrement des loyers à Paris



Cette affectation simple est en revanche impossible lorsqu'une rue appartient à plusieurs secteurs géographiques. A titre illustratif, la rue de Vaugirard est tronçonnée en 4 secteurs géographiques différents de l'encadrement des loyers. En conséquence, nous excluons de l'analyse statistique (mais pas de l'analyse économétrique, comme nous le verrons ultérieurement dans cette section) toutes les annonces qui ne peuvent pas être reliées avec certitude à un secteur géographique. Ces annonces exclues représentent 30.4% des annonces de la base de données après déduplication⁸.

2. Détermination de la période de construction

L'année de construction a un faible taux de complétion (42% sur le jeu de données à Paris) alors que c'est une variable clé qui conditionne la fixation du loyer de référence, et ce selon 4 modalités : < 1946, 1946 - 1970, 1971 - 1990, > 1990. Nous avons donc **complété les époques de construction**, dans les cas où nous connaissons précisément l'adresse du bien (56% des annonces sans époque de construction). La première étape consiste à lier l'adresse à la (ou les) parcelle(s) correspondante(s) grâce au lien adresse/parcelle(s) de la base Adresse Premium de l'IGN. La seconde étape s'appuie sur la Table Unifiée du Parcellaire des Fichiers Fonciers pour récupérer l'ensemble des locaux situés sur ces parcelles. La troisième étape consiste à ajouter l'information sur la date de construction de ces locaux. Dans le cas où toutes les dates de construction des locaux correspondent à la même période de construction (i.e., une des 4 modalités listées ci-dessus), nous pouvons imputer la période de construction du logement. La démarche est la suivante, lorsque l'année de construction n'est pas fournie par l'agence : 1/ nous recourrons aux informations collectées par l'APUR dans le cadre d'une enquête en 2007 visant à enrichir les fichiers fonciers (variable AN_CONST), sauf si la date fournie par les fichiers fonciers est postérieure, 2/ pour les cas restants, nous mobilisons les fichiers fonciers du CEREMA (variable [jannath](#)). A la suite de cette complétion, le taux de remplissage de l'époque de

⁸ 26.9% des annonces n'ont pas d'adresse ni de rue, 15.1% ont seulement une rue et cette rue est complètement contenue dans un secteur dans 76.7% des cas. Les autres 58% ont une adresse et donc un secteur.

construction est passée de 42% à 72.4%. Les observations pour lesquelles nous n'avons pu retrouver la date de construction sont écartées de cette analyse statistique.

3. *Suppression des annonces sans statut meublé/non meublé*

Enfin, nous avons supprimé de notre échantillon les annonces dont l'information sur le statut meublé/non meublé est manquante (6.2% des données).

Au total, le jeu de données utilisé pour l'analyse statistique comprend **164 384 annonces de location dans Paris intra-muros** (ce qui représente une perte totale de 42.2% des annonces).

Statistiques descriptives des dépassements du loyer de référence majoré

A partir de ce jeu de données, nous analysons tout d'abord **le pourcentage des annonces au-delà des seuils d'encadrement des loyers** sur la période considérée, en calculant la différence entre le loyer hors charges mentionné sur l'annonce et le loyer de référence majoré qui s'applique au bien sur la base de ses caractéristiques (nombre de pièces, période de construction, statut « meublé/non meublé »). Il est à noter que nous ne sommes **pas en mesure de distinguer les dépassements de loyers illégaux des dépassements justifiés par des caractéristiques exceptionnelles donnant lieu à un complément de loyer**. Les résultats sont reportés dans le Tableau 3. Pour les observations antérieures à la mise en place du second dispositif d'encadrement des loyers, soit entre le 1^{er} janvier 2018 et le 30 juin 2019, nous appliquons les seuils de l'encadrement applicables au 1^{er} juillet 2019, afin d'évaluer notamment dans quelle mesure l'écart des loyers à ce loyer de référence majoré a évolué avec l'entrée en vigueur du dispositif.

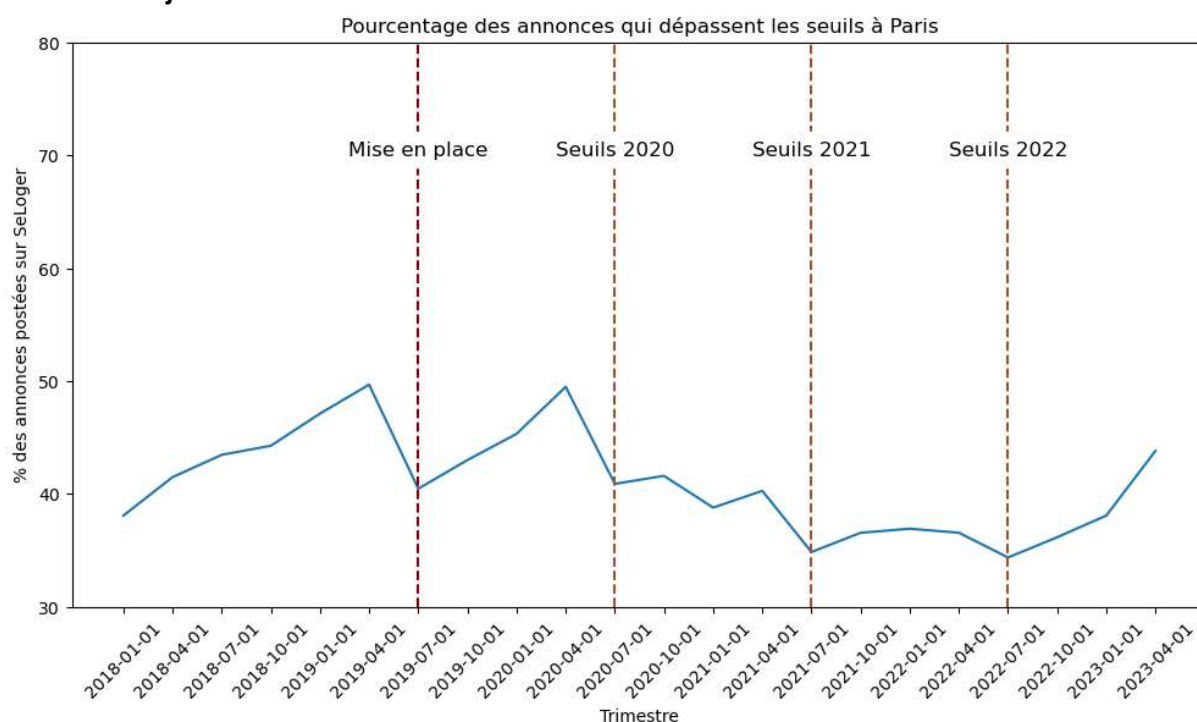
Tableau 3 : Pourcentage des annonces à Paris avec un loyer supérieur ou égal au loyer de référence majoré

Période	Logements meublés	Logements non meublés	Tous les logements
1 ^{er} janvier 2018 – 30 juin 2018	49.4 %	33.8%	39.6%
1 ^{er} juillet 2018 – 30 juin 2019	55.6 %	40.3%	46.2%
Mise en place de l'encadrement			
1 ^{er} juillet 2019 – 30 juin 2020	54.6%	36.4%	44.7%
1 ^{er} juillet 2020 – 30 juin 2021	51.1%	30.7%	40.3%
1 ^{er} juillet 2021 – 30 juin 2022	46.7%	27.9%	36.2%
1 ^{er} juillet 2022 – 30 juin 2023	48.2%	29.3%	38.0%

Sur le premier semestre 2018, soit quelques mois après l'annulation du 1^{er} dispositif d'encadrement des loyers par décision du tribunal administratif de Paris (en date du 28 novembre 2017), la part des annonces dont le loyer hors charges était supérieur au loyer de référence majoré (qui sera fixé à partir du 1^{er} juillet 2019) s'élève à 39,6%. Elle augmente de presque 7 points durant l'année qui précède l'application du premier dispositif, pour atteindre son niveau le plus haut à 46,2%, vraisemblablement sous l'effet des annonces de réglementation, le vote du rétablissement du dispositif par le Conseil de

Paris du 11 décembre 2018 et l'autorisation donnée par le gouvernement avec la publication du décret publié au Journal officiel le 13 avril 2019, qui ont conduit les bailleurs à signer des baux aux loyers élevés avant de tomber sous le joug de la réglementation. Depuis l'instauration du second dispositif, la part des annonces avec un loyer supérieur au loyer de référence majoré diminue progressivement, de 44.7 % lors de la première année jusqu'à 38% entre juillet 2022 et juillet 2023. Ainsi, près de 4 annonces sur 10 proposent un loyer supérieur au seuil fixé par la réglementation parisienne. Cette part est en progression sur la dernière année (entre juillet 2022 et juillet 2023) par rapport à l'année précédente (entre juillet 2021 et 2022) où elle était de 36.2%. Sur cette dernière période, nos chiffres (restreints au parc locatif des non meublés) sont cohérents avec le taux constaté par l'OLAP dans leur étude sur le parc locatif privé non meublé (OLAP, 2022) de 30% des loyers en dépassement.

Figure 3 : Evolution du pourcentage des annonces à Paris avec un loyer supérieur ou égal au loyer de référence majoré



L'analyse graphique (Figure 3) avec un pas de temps plus fin (i.e., trimestriel) fait ressortir des fluctuations importantes sur la période, avec une hausse forte et continue du nombre d'annonces avec un loyer supérieur ou égal au loyer de référence majoré de près de 13 points en 18 mois, avant l'entrée en vigueur du second dispositif d'encadrement. L'effet du dispositif s'est apparemment fait ressentir immédiatement, avec près de 40% des annonces proposant un loyer supérieur ou égal au loyer de référence majoré, soit une diminution (la plus forte sur la période) de près de 9 points, qui est ensuite revenu au niveau initial au bout de 3 trimestres. Entre le 2^{ème} trimestre 2020, qui marque le début de la crise sanitaire, et la fin du 3^{ème} trimestre 2022, le trend de l'évolution des surloyers (justifiés par un complément ou non) est à la baisse. Depuis le 4^{ème} trimestre 2022, le pourcentage de loyers en dépassement repart en forte hausse.

Le Tableau 3 montre un écart important du pourcentage d'annonces avec un loyer excédant le loyer de référence majoré, selon que le logement soit meublé ou non meublé. Entre juillet 2022 et juillet 2023, 48,2% des annonces de logements meublés sont au-dessus la fourchette du dispositif contre seulement 29,3% pour les non meublés.

La plus forte hétérogénéité dans le dépassement du seuil de référence s’observe pour les surfaces, comme le montre le Tableau 4. Au deuxième trimestre 2023, la part des logements ayant un loyer hors charges supérieur à la limite maximum est de 60,6% pour les logements de moins de 18m², contre 41,4% pour les logements entre 18 et 24 m² et 32,4% pour les plus grandes surfaces. Nous retrouvons ainsi le résultat de Malard et Poulhès (2020) dans le cadre du premier dispositif, mais dans des proportions très différentes, car les auteurs pointaient des taux de dépassement de 44% pour les plus petites surfaces contre 8% pour les plus grandes. Avec l’encadrement de loyers à partir de juillet 2019, le taux de dépassement a surtout fortement chuté pour les plus petites surfaces, à savoir de 24,2 points jusqu’au 3^{ème} trimestre 2022, contre une faible baisse de 4,2 points pour les grandes surfaces. Sur la période la plus récente, le taux de dépassement des petites surfaces est à nouveau celui qui a subi la variation la plus forte, mais cette fois-ci à la hausse, avec une augmentation de 11,4 points en moins d’un an. Les évolutions des autres surfaces sont très proches, avec des hausses comprises entre 8,5 et 9 points.

Tableau 4 : Pourcentage des annonces à Paris avec un loyer supérieur ou égal au loyer de référence majoré, en fonction de la surface moyenne des pièces

Période	Surface moyenne des pièces inférieure à 18m ²	Surface moyenne des pièces entre 18m ² et 24m ²	Surface moyenne des pièces supérieure à 24m ²
2018-T1	62.6%	34.3%	22.8%
2018-T2	65.2%	36.8%	22.5%
2018-T3	67.3%	39.4%	23.9%
2018-T4	67.2%	41.6%	26.4%
2019-T1	70.6%	46.5%	26.9%
2019-T2	73.4%	48.2%	27.6%
Mise en place de l’encadrement			
2019-T3	58.9%	38.6%	27.1%
2019-T4	62.6%	41.7%	29.1%
2020-T1	64.1%	44.9%	30.6%
2020-T2	69.7%	45.7%	33.1%
2020-T3	60.0%	38.6%	26.8%
2020-T4	62.1%	39.2%	24.7%
2021-T1	58.1%	36.0%	23.2%
2021-T2	59.0%	36.2%	25.3%
2021-T3	54.3%	32.0%	19.7%
2021-T4	55.2%	33.7%	23.1%
2022-T1	55.6%	33.2%	24.6%
2022-T2	51.2%	35.6%	23.3%
2022-T3	49.2%	32.9%	23.4%
2022-T4	53.0%	33.5%	25.8%
2023-T1	52.5%	36.8%	28.4%
2023-T2	60.6%	41.4%	32.4%

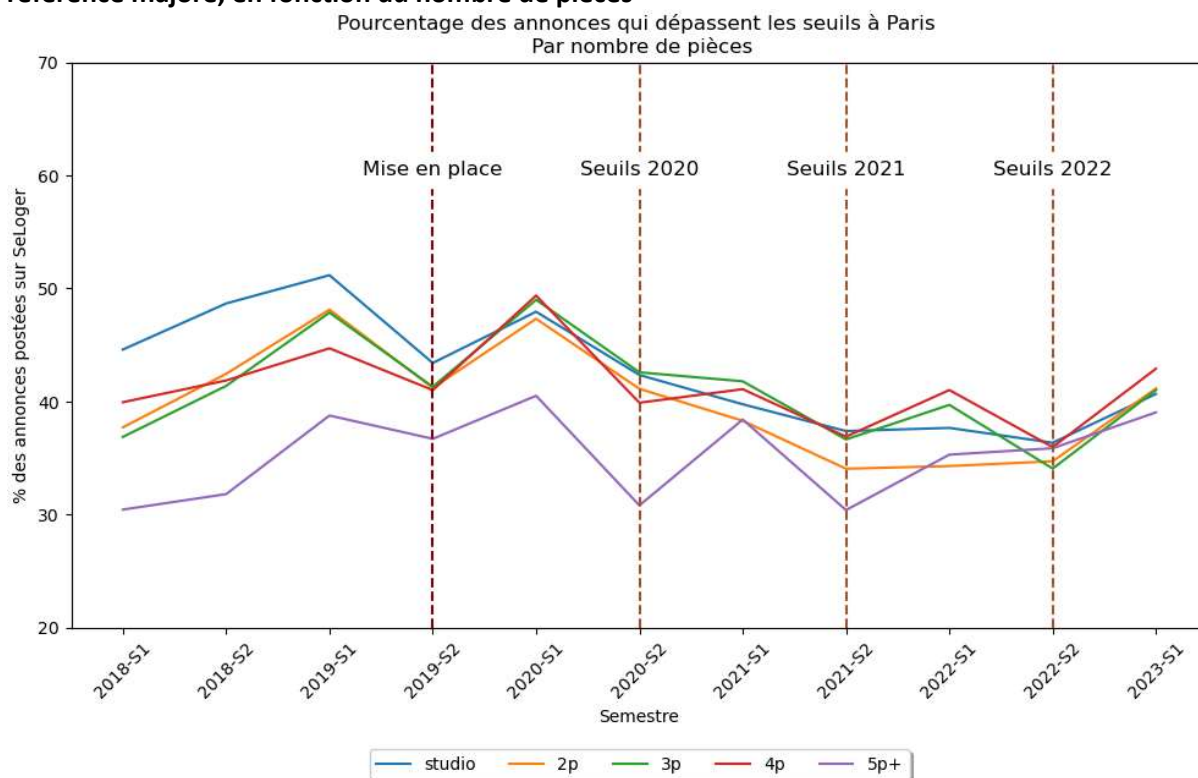
Les différences selon le nombre de pièces sont nettement moins marquées, comme le montre le Tableau 5.

Tableau 5 : Pourcentage des annonces à Paris avec un loyer supérieur ou égal au loyer de référence majoré, en fonction du nombre de pièces

Période	Studio	2 pièces	3 pièces	4 pièces	5 pièces et plus
2018-S1	44.6%	37.7%	36.9%	39.9%	30.4%
2018-S2	48.7%	42.4%	41.4%	41.8%	31.8%
2019-S1	51.2%	48.1%	47.9%	44.7%	38.7%
Mise en place de l'encadrement					
2019-S2	43.4%	41.2%	41.3%	41.0%	36.7%
2020-S1	47.9%	47.3%	49.0%	49.4%	40.5%
2020-S2	42.3%	41.1%	42.6%	39.9%	30.8%
2021-S1	39.7%	38.3%	41.8%	41.1%	38.4%
2021-S2	37.4%	34.0%	36.6%	36.9%	30.4%
2022-S1	37.7	34.3%	39.7%	41.0%	35.3%
2022-S2	36.3%	34.7%	34.1%	36.0%	35.9%
2023-S1	40.7%	41.4%	41.0%	42.9%	39.0%

La représentation graphique ci-dessous (Figure 4) montre plus nettement l'écart élevé dans le taux de dépassement entre les logements de moins de 5 pièces et les très grandes surfaces, avec toutefois une disparition de cet écart sur la fin de période. Sur le premier semestre 2023, nous retrouvons une hausse quasi-généralisée du taux d'annonces proposant un loyer supérieur au loyer de référence majoré.

Figure 4 : Evolution du pourcentage des annonces à Paris avec un loyer supérieur ou égal au loyer de référence majoré, en fonction du nombre de pièces



Enfin, le dispositif d'encadrement des loyers semble avoir eu un effet relativement homogène selon l'époque de construction, comme l'indique le tableau 6.

Tableau 6 : Pourcentage des annonces à Paris avec un loyer supérieur au loyer de référence majoré, en fonction de l'époque de construction

Période	Avant 1946	Entre 1946 et 1970	Entre 1971 et 1990	Après 1990
2018-T1	37.6%	42.7%	37.3%	35.4%
2018-T2	41.3%	44.8%	43.0%	34.4%
2018-T3	44.5%	41.2%	41.5%	36.4%
2018-T4	43.6%	47.8%	51.8%	33.9%
2019-T1	47.5%	49.5%	44.7%	40.2%
2019-T2	50.1%	51.6%	49.7%	38.0%
Mise en place de l'encadrement des loyers				
2019-T3	39.5%	46.1%	45.3%	33.3%
2019-T4	42.7%	46.2%	43.5%	38.6%
2020-T1	45.2%	48.4%	46.1%	38.9%
2020-T2	49.6%	54.1%	43.7%	44.4%
2020-T3	39.8%	48.4%	40.2%	39.9%

2020-T4	41.3%	47.3%	39.0%	36.8%
2021-T1	37.7%	46.5%	41.1%	34.1%
2021-T2	39.3%	49.8%	41.1%	33.0%
2021-T3	34.6%	40.6%	31.2%	31.6%
2021-T4	35.6%	42.6%	38.7%	32.1%
2022-T1	35.9%	45.0%	39.1%	32.5%
2022-T2	35.5%	43.7%	37.2%	34.7%
2022-T3	33.0%	40.1%	33.4%	40.4%
2022-T4	35.3%	44.1%	33.0%	35.1%
2023-T1	36.5%	45.3%	42.7%	34.9%
2023-T2	41.7%	54.1%	45.4%	44.0%

Nous réalisons également une analyse par secteur géographique. Le montant médian du dépassement, comme le montre la figure 5a, suit pour la plupart des secteurs géographiques un trend négatif, à l'exception notable des secteurs 2 et 3. Aux extrêmes, les écarts entre dépassements médians peuvent atteindre près de 15 points de pourcentages, entre le secteur géographique 2 qui se distingue par de fortes valeurs alors que le secteur 12 présente de faibles valeurs. Le centre de Paris présente les montants de dépassement les plus élevés (figure 5b).

Figure 5a : Evolution du montant médian de dépassement du loyer de référence majoré selon le secteur géographique

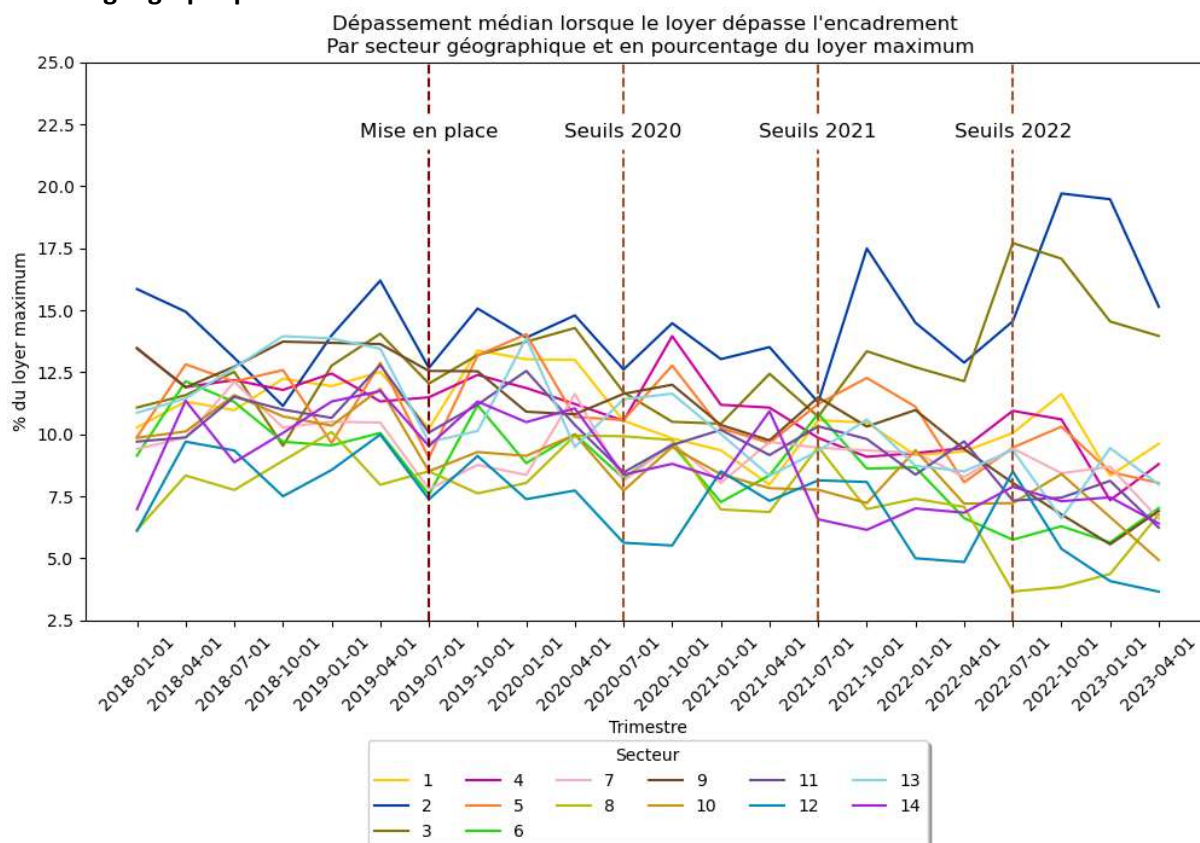


Figure 6 : Evolution du pourcentage des annonces à Paris avec un loyer supérieur ou égal au loyer de référence majoré, en fonction du secteur géographique

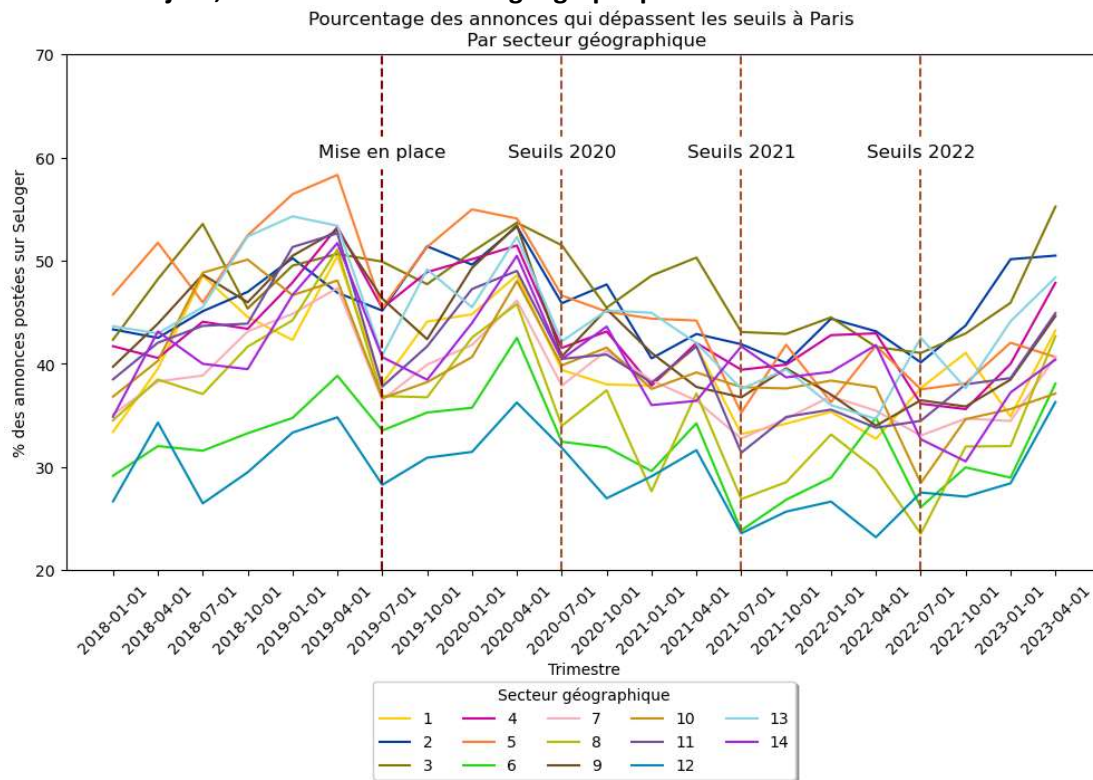
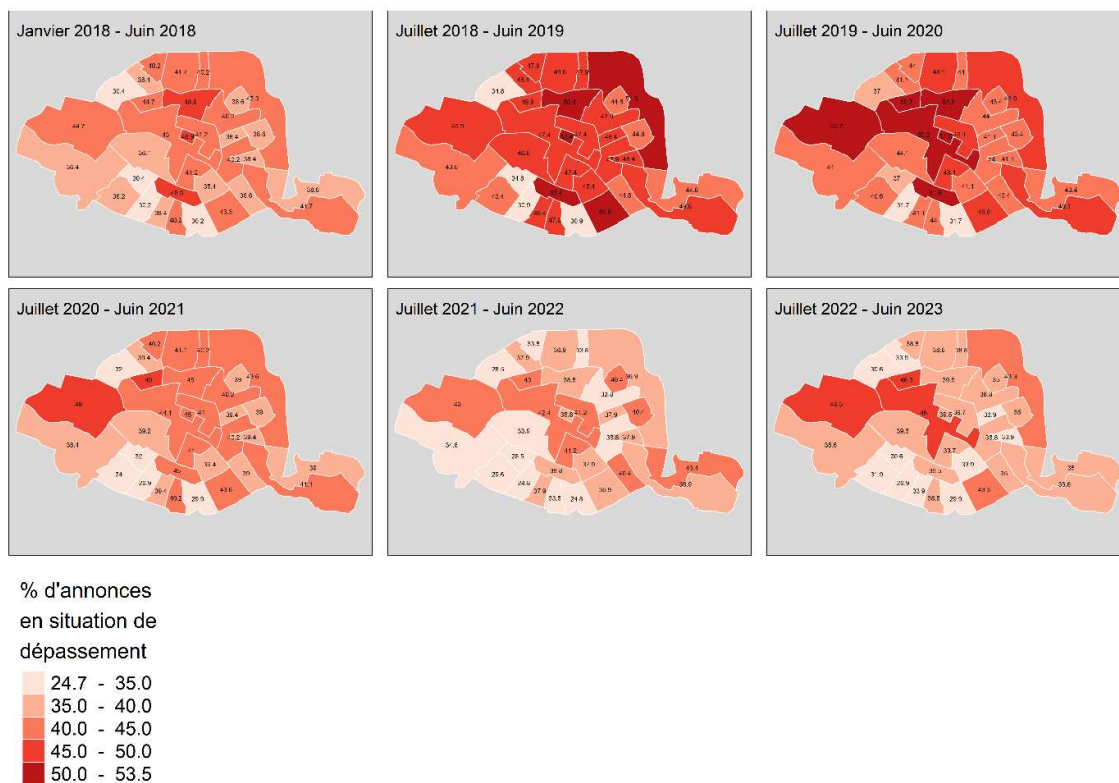
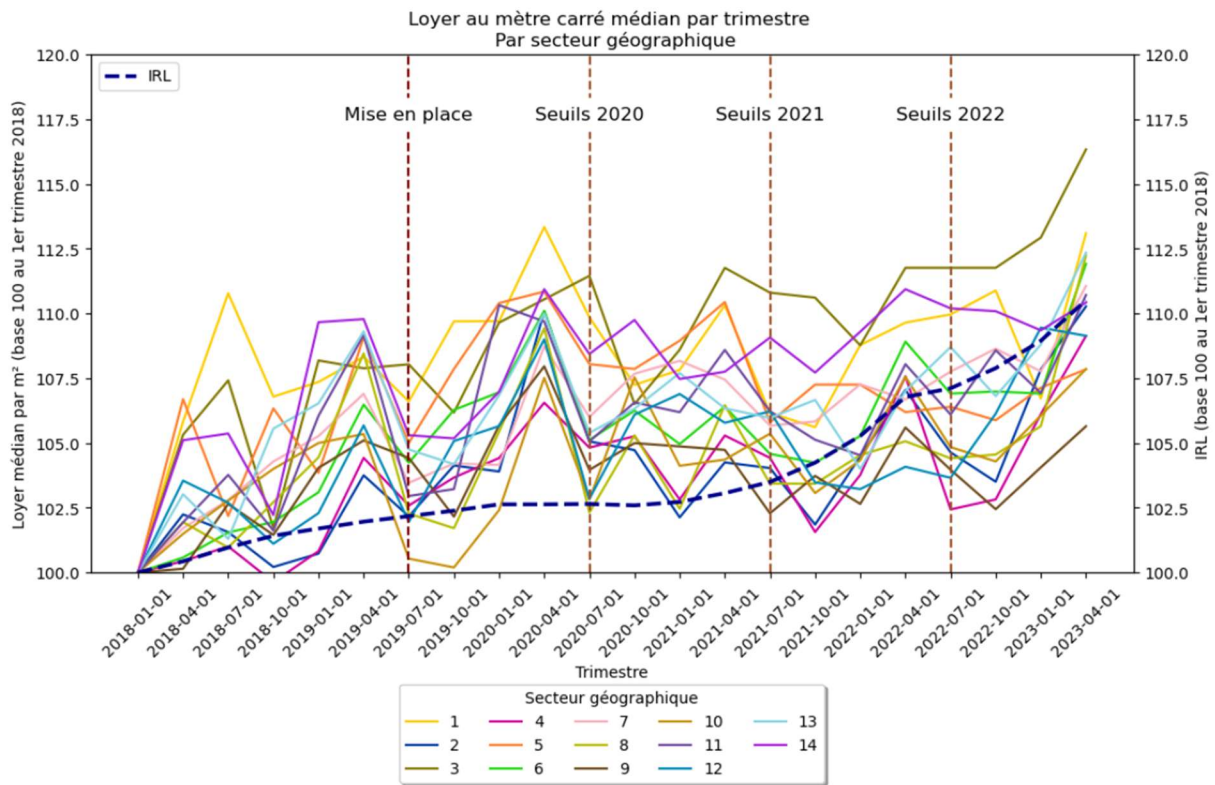


Figure 7 : Pourcentage des annonces à Paris avec un loyer supérieur ou égal au loyer de référence majoré, en fonction du secteur géographique



Enfin, la figure 8 montre que l'évolution du loyer médian suit un trend relativement similaire pour la plupart des 14 secteurs. Cette hausse est par ailleurs plus rapide que celle de l'IRL sur la période pré-encadrement et plus lente depuis mi-2021, notamment en raison de la forte évolution des prix à la consommation qui détermine le calcul de l'IRL.

Figure 8 : Evolution du loyer médian des annonces à Paris en fonction du secteur géographique



3.1.4. Evolution de la probabilité de dépassement avec l'encadrement des loyers

Dans cette partie, nous reproduisons dans le cadre du second dispositif d'encadrement des loyers, la méthodologie retenue par Malard et Poulhès (2020) pour comparer l'évolution de la probabilité de dépassement du seuil de loyer avant et après la mise en place du (premier, dans leur cas) dispositif d'encadrement des loyers à Paris. Pour cela, nous estimons la probabilité de dépassement étant donné les caractéristiques des logements, avec un modèle du type :

$$P(D = 1|X) = G(X\beta)$$

où la variable $D = 1$ si le loyer demandé dans l'annonce est supérieur au seuil de l'encadrement, et 0 s'il est en dessous. $G(.)$ est la fonction de répartition logistique. La matrice X contient l'ensemble des variables de contrôle, à savoir la surface moyenne par pièce, le nombre de pièces, l'interaction de ces deux variables, l'époque de construction, un effet fixe par grand quartier, un effet fixe par mois ainsi qu'un effet fixe par année. Ce modèle est estimé sur la période avant la mise en place de l'encadrement, soit entre le 1^{er} Janvier 2018 et le 30 juin 2019.

A partir des paramètres estimés, nous prédisons la probabilité de dépassement pour chaque bien sur la période allant du 1^{er} juillet 2019 au 30 juin 2023. A l'instar de Malard et Poulhès (2020), nous appliquons l'effet fixe estimé pour l'année 2019 aux années suivantes, comme il n'est pas estimé pour les périodes postérieures à l'application du dispositif. Nous calculons ensuite le taux de dépassement agrégé à partir des prédictions. Il se définit comme la moyenne des probabilités de dépassement prédites sur l'ensemble de l'échantillon pour chacune des périodes.

Sur les graphiques suivants (Figures 9 à 13), la courbe bleue représente le dépassement moyen constaté alors que la courbe rouge représente le dépassement moyen estimé en l'absence d'encadrement des loyers. Les traits horizontaux représentent la valeur moyenne sur chaque période (avant et après la date de mise en place de l'encadrement des loyers à Paris). Globalement, en supposant une stabilité des effets temporels après le 1^{er} juillet 2019, le dispositif parisien aurait permis de réduire en moyenne de plus de 10 points le taux de dépassement du loyer de référence majoré, en le faisant passer de 49% à 38%, avec des variabilités importantes de ces chiffres selon le nombre de pièces, la surface, la taille des pièces et le caractère meublé ou non du logement.

Figure 9 : Evolution comparée du dépassement moyen constaté et du dépassement moyen prédit

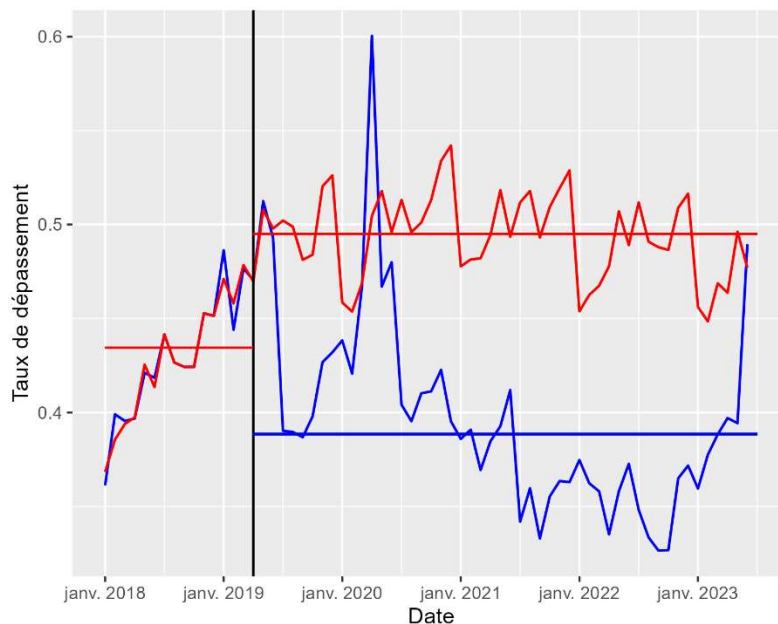


Figure 10 : Evolution comparée du dépassement moyen constaté et du dépassement moyen prédit selon le nombre de pièces

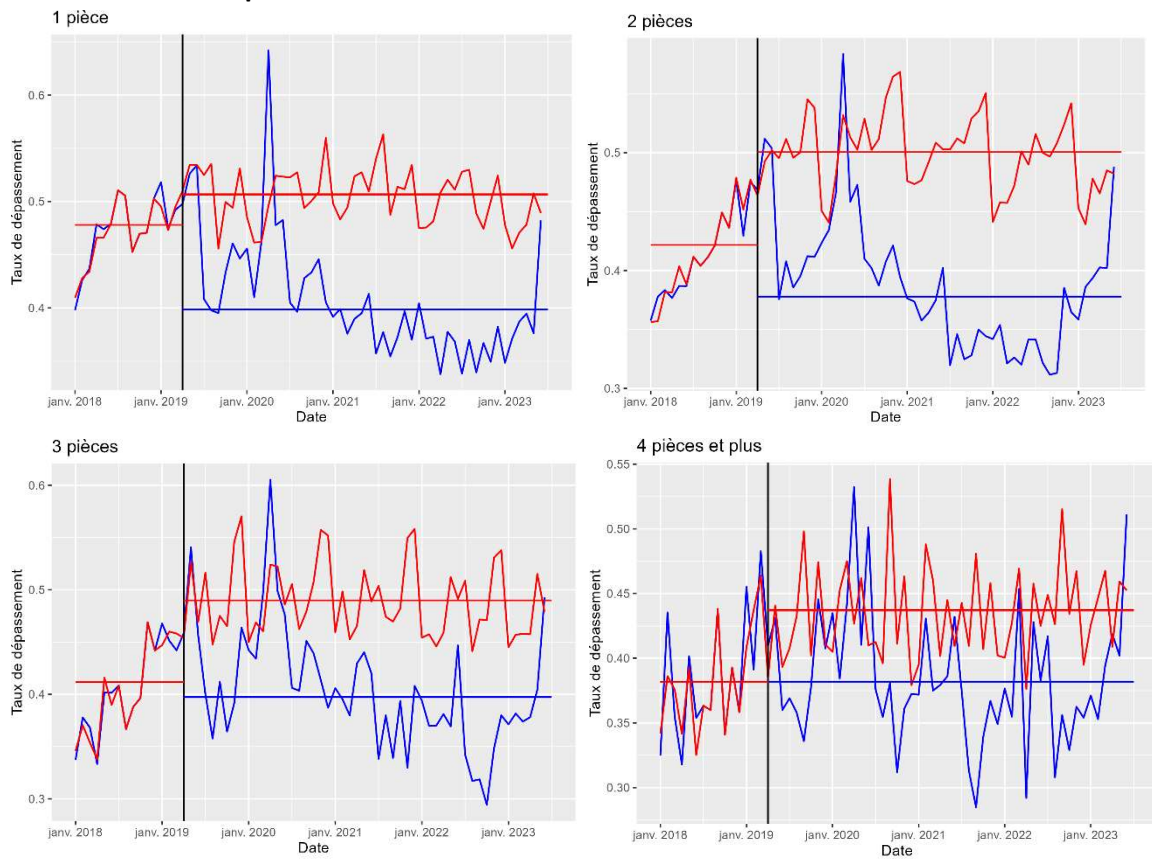
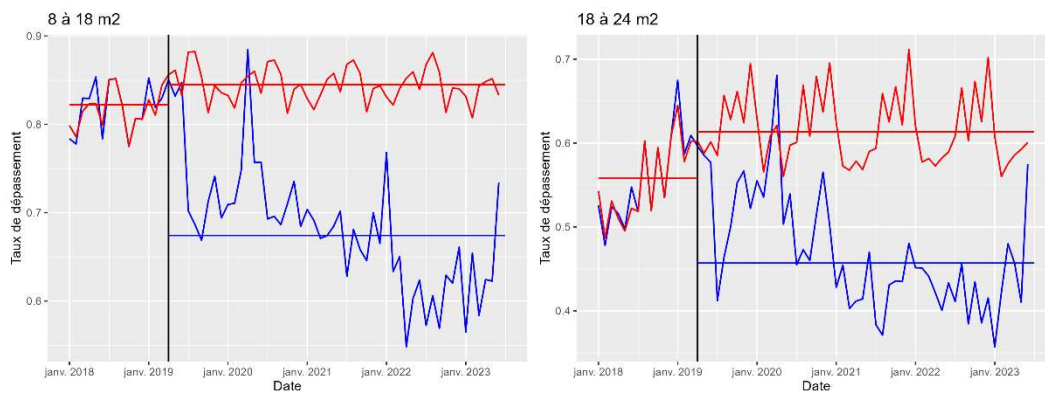


Figure 11 : Evolution comparée du dépassement moyen constaté et du dépassement moyen prédit selon la surface



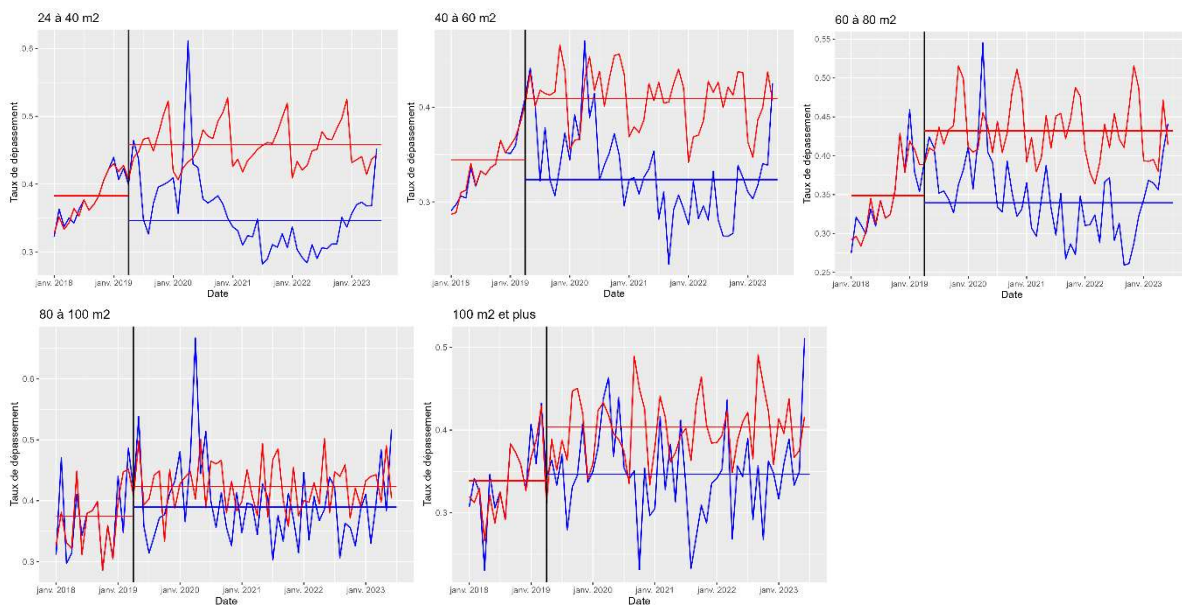


Figure 12 : Evolution comparée du dépassement moyen constaté et du dépassement moyen prédit selon la taille des pièces

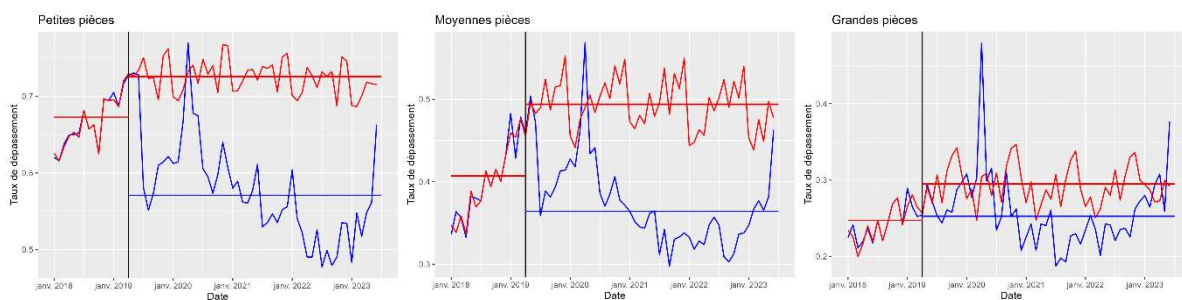
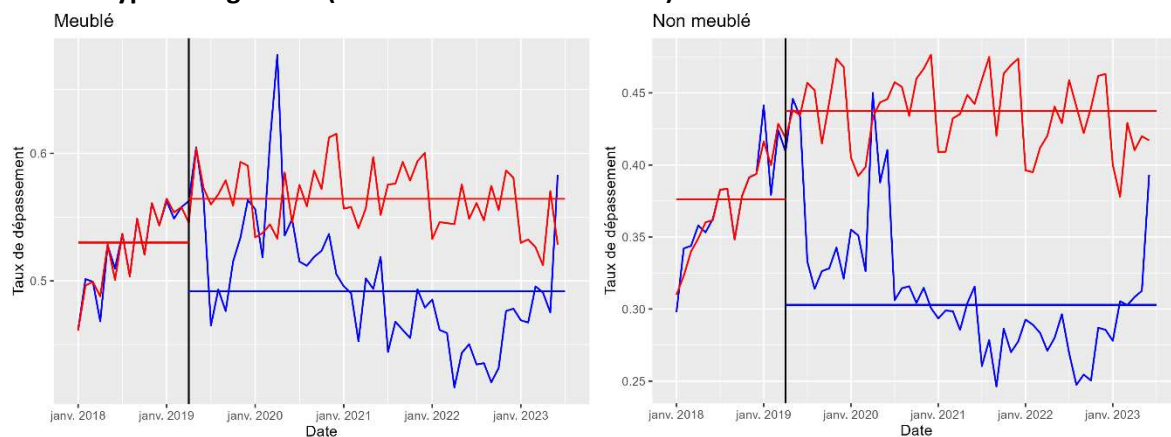


Figure 1 : Evolution comparée du dépassement moyen constaté et du dépassement moyen prédit selon le type de logement (meublé versus non meublé)



Cette reproduction de la méthodologie de Malard et Poulhès (2020) a permis de réaliser des prédictions sur le taux de dépassement des loyers en se basant sur les valeurs observées des caractéristiques des logements, à partir des données pré-encadrement. Ce modèle logistique présente

cependant plusieurs limites, à savoir l'hypothèse de relation linéaire et l'application de l'effet fixe de 2019 pour les années suivantes (a fortiori sur des années marquées par la crise de la Covid-19). Des techniques alternatives de prédiction comme des régressions pénalisées ou des *random forests* pourraient être mobilisées.

Plus fondamentalement, il convient de noter que ces prédictions sur le taux de dépassement des loyers ne peuvent pas s'interpréter de façon causale. Des différences post-encadrement entre le dépassement moyen constaté et le dépassement moyen prédit ne peuvent pas être expliquées intégralement par la mise en place de la politique d'encadrement en raison des variables confondantes inobservées et des biais de sélection que cette spécification logistique ne permet pas de contrôler. Afin de mener une évaluation causale isolant l'impact de l'encadrement, nous mettons en œuvre une méthode d'évaluation d'impact, détaillée dans la section 3.2.

3.2. Méthodologie

3.2.1. Avantages et limites des principales méthodes d'inférence causale pour évaluer l'impact de l'encadrement des loyers

Les méthodes d'inférence causale visent à évaluer l'impact d'un traitement (une politique publique, un programme, etc.) sur une variable d'intérêt pour un échantillon composé d'observations (individus, logements, etc.). Idéalement, il faudrait disposer de la valeur de la variable d'intérêt dans les deux situations : en l'absence et en présence du traitement. Dans notre cas, puisque l'objectif est d'évaluer l'effet de l'encadrement des loyers (le « traitement ») sur les montants des loyers (la variable d'intérêt) des logements situés dans Paris intra-muros, il faudrait pouvoir observer simultanément les montants des loyers avec et sans encadrement. Dans l'impossibilité de faire cette comparaison, puisqu'un seul cas est observé en réalité⁹, et également dans la mesure où réaliser une expérience contrôlée n'est pas envisageable, nous devons recourir à l'économétrie d'évaluation d'impact. Différentes méthodes permettent d'estimer le contrefactuel, c'est-à-dire ce qui se serait passé en l'absence de la mise en place de la politique, ici l'encadrement. Le respect d'hypothèses d'identification, sous lesquelles la comparaison du groupe des traités (les loyers observés à Paris) et d'un groupe de contrôle (les loyers observés dans une autre zone) permet d'identifier et d'interpréter de façon causale l'impact du traitement.

Une première catégorie de méthodes repose sur l'appariement. Sous l'hypothèse de disposer de l'ensemble des variables qui permettent de contrôler les différences entre les observations du groupe traité et les observations du groupe de contrôle (hypothèse d'*ignorabilité*), nous pouvons appairer chaque observation du groupe des traités à un ou plusieurs « jumeaux » en termes de caractéristiques observables. La moyenne de ces différences, éventuellement inversement pondérée par la distance entre les jumeaux permet d'estimer l'effet causal du traitement. Une variante est la méthode d'appariement par score de propension, où l'appariement entre les observations traitées et les observations du groupe de contrôle se fait sur la base de la proximité en termes du score de propension, lequel est défini comme la probabilité d'être traité, conditionnellement aux caractéristiques observables. On peut alors réduire la dimension du problème en appariant les observations sur une dimension, le score de propension, plutôt que sur de multiples dimensions (toutes les caractéristiques observables). Nous n'avons pas retenu cette méthode en raison de l'hypothèse forte d'identification qui est de pouvoir disposer de l'ensemble des caractéristiques différenciant les logements. Malgré la disponibilité de nombreuses variables de contrôle, il reste de

⁹ Holland (1986) indique qu'il s'agit du problème fondamental de l'évaluation.

l'hétérogénéité inobservable dans les déterminants des loyers (aménités très localisées, caractéristiques structurelles de logements non incluses dans les variables de contrôle) ce qui biaiserait l'estimation de telles différences. D'autres estimateurs basés sur des pondérations utilisant l'inverse du score de propension (*inverse probability weighting*) ne sont pas non plus applicables car nos estimations sur la base des caractéristiques structurelles présentes dans les bases renvoient des scores de propension très proches de 0 ou de 1, rendant la pondération par l'inverse de ces scores très instable.

D'autres méthodes n'imposent pas l'hypothèse d'ignorabilité mais s'appuient sur d'autres hypothèses d'identification. Il en est ainsi des régressions sur discontinuité qui exploitent la présence de seuils ou de discontinuité. Dans notre cas, cela revient à utiliser la frontière séparant Paris et les villes limitrophes comme discontinuité géographique et comparer les montants des loyers entre les logements parisiens situés à proximité des limites de Paris avec ceux des logements situés en Petite Couronne également situés à proximité de la frontière avec Paris. Cependant, outre la nécessité de supposer une forme de continuité sur toutes les variables, hormis le traitement et la variable d'intérêt, entre les traités et les contrôles, cette méthode est particulièrement sensible à une autre hypothèse, la *Stable Unit Treatment Value Assumption* (SUTVA). Cette hypothèse SUTVA, qui impose que le traitement d'une observation n'a pas d'effet sur le résultat des autres observations (ou autrement dit l'absence d'interactions, d'externalités ou d'effets de report), est hautement improbable dans le cas de l'encadrement. Il est à noter que l'hypothèse SUTVA n'est pas spécifique aux régressions sur discontinuité mais commune à la majeure partie des méthodes d'évaluation, cependant, elle est particulièrement problématique pour les régressions sur discontinuité géographique.

Pour toutes ces raisons, nous avons choisi une autre méthode d'évaluation, qui est la méthode de différence-en-différences, pour laquelle l'hypothèse d'identification est une hypothèse de tendances parallèles, que nous présentons dans la section 3.2.2.

3.2.2. La méthode de la différence-de-différences

La méthode de la différence de différences (ou DID) permet d'évaluer l'effet de l'encadrement des loyers (i.e., le « traitement ») sur les montants des loyers (i.e., la variable d'intérêt) des logements situés dans Paris intra-muros (i.e., les observations du « groupe traité ») en les comparant avec des logements situés dans d'autres villes n'appliquant pas l'encadrement des loyers (i.e., les observations du « groupe de contrôle »). Elle est basée sur une double différence, la première permettant d'éliminer les effets fixes qui sont spécifiques à chaque groupe alors que la seconde différence permet d'éliminer les effets temporels qui sont communs aux deux groupes. Formellement, l'effet de l'encadrement des loyers (i.e., « l'effet de traitement ») que nous notons β se définit comme :

$$\beta = [E(Y_{ict}|c = Tr, t = t_{post}) - E(Y_{ict}|c = Tr, t = t_{pre})] - [E(Y_{ict}|c = Co, t = t_{post}) - E(Y_{ict}|c = Co, t = t_{pre})]$$

où Y_{ict} est le niveau de loyer de l'annonce de location i qui est située dans une ville c , et observé à la période t , $E(.)$ est l'espérance, Co représente le groupe de contrôle, Tr le groupe traité (Paris), t_{pre} la période avant encadrement (du 1^{er} janvier 2018 au 30 juin 2019) et t_{post} la période après encadrement (du 1^{er} juillet 2019 au 30 juin 2023).

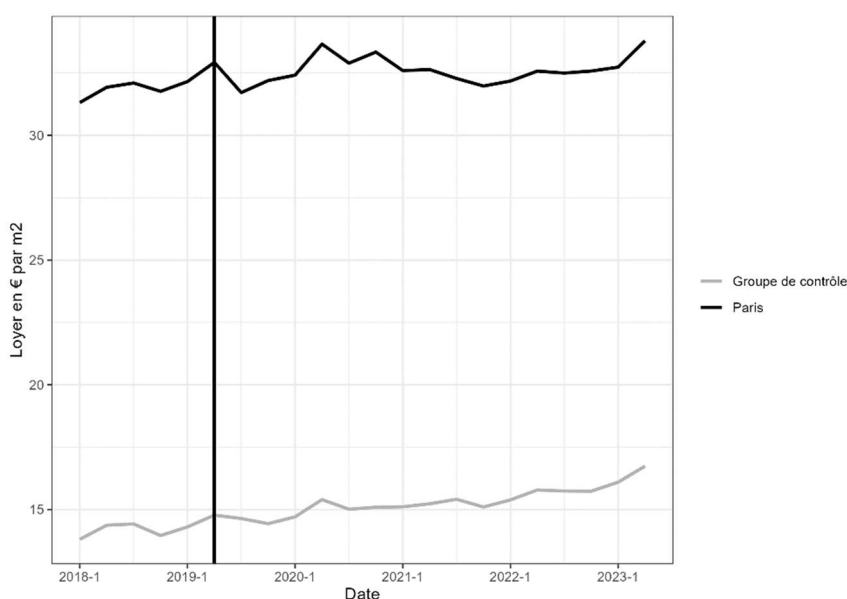
Un défi associé à cette méthode est de disposer d'un contrefactuel qui permet d'obtenir la valeur qui aurait été observée en l'absence de la politique publique. Le groupe de contrôle doit ainsi être construit pour que deux hypothèses fondamentales soient vérifiées :

- PTA (*Parallel Trend Assumption*), selon laquelle le groupe de contrôle et Paris doivent présenter une tendance d'évolution des loyers similaire (i.e., parallèle) ;

- SUTVA (*Stable Unit Treatment Value Assumption*), qui impose que le loyer demandé pour un bien dans le groupe de contrôle ne doit pas être affecté par l’encadrement des loyers de biens mis en location à Paris. Inversement, le fait que loyers ne soient pas encadrés dans les villes du groupe de contrôle ne doit pas avoir d’effet sur le niveau des loyers à Paris.

Nous définissons nos deux groupes à partir du jeu de données. D’une part, le **groupe traité est constitué des annonces de location de biens situés dans la ville de Paris**. D’autre part, le **groupe de contrôle est constitué de 8 villes, à savoir Aix-en-Provence, Grenoble, Marseille, Nantes, Nice, Strasbourg, Toulon et Toulouse**. Pour les sélectionner, nous avons tout d’abord considéré toutes les villes de plus de 100 000 habitants (d’après les données de la population communale INSEE) qui ne sont pas localisées en Ile-de-France. Cette dernière restriction éliminant toute proximité géographique entre groupe traité et groupe de contrôle vise à garantir le respect de l’hypothèse SUTVA. Nous avons ensuite écarté celles qui ont appliqué un encadrement des loyers entre 2018 et 2022 et/ou qui n’appartiennent pas à une « zone tendue » au sens de la loi ALUR. Cela nous permet d’estimer l’effet propre de cet encadrement des loyers en niveau, et également de considérer des zones caractérisées par des marchés immobiliers semblables à celui de Paris. En résumé, les villes du groupe de contrôle ne sont pas soumises à un encadrement des loyers en niveau mais les loyers sont encadrés en évolution par l’indice de référence des loyers (IRL) lors des relocations, comme l’était Paris avant la mise en place du double encadrement « IRL et seuils municipaux » le 1^{er} juillet 2019. Comme le montre la figure 14, la tendance de loyers agrégée dans ces villes est très similaire à celle de Paris avant que l’encadrement des loyers ne soit mis en place (avec une corrélation estimée de 0,94), ce qui semble valider l’hypothèse de tendances parallèles explicitée précédemment.

Figure 14 : Comparaison des tendances de loyers



Le tableau 7 présente quelques caractéristiques des biens en fonction de leur localisation (Paris versus groupe de contrôle). Une version plus complète se trouve en annexe A.3.1.

Tableau 7 : Statistiques descriptives pour le groupe de contrôle et le groupe traité

Variables	Groupe de contrôle	Groupe traité
Nombre d'observations	275 079	284 221
Loyer moyen hors charge au m2	15	32.5
Loyer moyen hors charge	664.5	1471
Surface moyenne (écart-type)	48.9 (24.8)	48 (32.8)
Proportion de meublés	23.1%	42.1%
Proportion de 1 pièce	28%	33.5%
Proportion de 2 pièces	36.1%	37.1%
Proportion de 3 pièces	25.6%	17.5%
Proportion de 4 pièces et plus	10.3%	11.9%
Taille moyenne des pièces	22.7	22.2
Construction avant 1945	7.5%	53%
Construction 1946-1970	8%	9.3%
Construction 1971-1990	5.4%	6%
Construction 1991-2023	17.8%	4.1%
Construction manquante	61.3%	27.6%

Le loyer moyen parisien hors charges est, comme attendu, plus du double de celui du groupe de contrôle. La proportion d'appartements meublés est plus importante à Paris et les appartements y sont plus anciens. Par ailleurs, malgré une surface moyenne relativement similaire entre les deux groupes, Paris est composé de davantage de petits et de très grands logements avec moins d'appartements de taille moyenne (entre 3 et 4 pièces).

Modèle standard de différence-de-différences

Sous ces hypothèses, nous estimons l'effet de l'encadrement des loyers à Paris avec une spécification classique dans un premier temps, grâce au modèle de régression suivant :

$$Y_{ict} = \beta T_{ct} + X_{ict}\theta + \mu_c + \mu_t + \varepsilon_{ict} \quad (1)$$

où Y_{ict} est le logarithme du loyer au m² pour le bien i dans la ville c à la période t , T_{ct} est une variable binaire qui vaut 1 si l'encadrement des loyers est appliqué dans la ville c à la période t , β est l'effet du traitement, X_{ict} est un ensemble de variables de contrôle représentant les caractéristiques du bien i , θ est le vecteur de paramètres estimés associés, μ_c est un effet fixe par ville (ou par IRIS), μ_t est un effet fixe par période et ε_{ict} est un bruit blanc iid. L'effet fixe IRIS permet de capter l'hétérogénéité locale invariante dans le temps (notamment les aménités locales qui peuvent se capitaliser dans les loyers). Le modèle est estimé grâce à la méthode des Moindres Carrés Ordinaires avec inférence robuste.

Pour interpréter le coefficient β , étant donné que le loyer est exprimé en logarithme, il est nécessaire de prendre l'exponentiel du coefficient, ce qui donne le facteur par lequel la variable de loyer est multipliée en cas de traitement, puis de soustraire 1 pour obtenir la variation, i.e., $\exp(\beta)-1$. Ce coefficient transformé, qui est directement interprétable, sera indiqué à la ligne « Traitement (%) » dans les tableaux de résultats. Par ailleurs, l'effet du traitement est mesuré par comparaison à un

contrefactuel qui n'a pas bénéficié du traitement et s'interprète donc comme une variation par rapport à ce contrefactuel. Supposons par exemple que le coefficient transformé s'élève à -0.042. Cela signifie que l'encadrement des loyers a permis une baisse de 4,2% par rapport à une situation dans laquelle l'encadrement des loyers n'aurait pas existé. La hausse des loyers évitée atteint ainsi $[1/(1-\text{coefficient transformé})-1]*100$, soit 4,38%.

Modèle de différence-de-différences avec coefficients hétérogènes et effets fixes temporels différenciés

Nous interagissons dans un second temps et estimons le modèle suivant :

$$Y_{ict} = \beta T_{ct} + Z_{tr}\beta_{tr} + Z_{co}\beta_{co} + X_{ict}\theta + \mu_c + \mu_t + Z_{ict} \times \mu_t + \varepsilon_{ict} \quad (2)$$

où Z représente un sous-ensemble de variables pour lesquelles on associe des coefficients hétérogènes selon le traitement (groupe traité ou groupe de contrôle) et des effets fixes temporels pour chacun des traitements (ces variables ne sont donc plus dans le sous ensemble de variables représenté par la matrice X). Si l'observation est dans le groupe de contrôle, Z_{co} prend pour valeur les variables associées, sinon 0, et inversement pour Z_{tr} . Ces variables sont les meublés/non meublés, l'année de construction et le nombre de pièces. β_{tr} représente l'effet des variables de contrôle Z_{tr} pour le groupe traité et β_{co} leur effet sur le groupe de contrôle. En outre, nous incluons une interaction : $Z_{ict} \times \mu_t$ qui représente les effets fixes temporels spécifiques à chaque modalité des variables dans Z_{ict} . Par exemple, relativement à la variable nombre de pièces, cette interaction permet d'estimer un effet fixe temporel pour les 1 pièce, un autre pour les 2 pièces etc. La spécification inclut également un effet fixe temporel pour les meublés et non meublés, puis un effet fixe temporel pour chaque période de construction. D'une part, ces coefficients hétérogènes permettent d'estimer des chocs différenciés par catégorie de bien. D'autre part, l'ajout de ces effets fixes temporels par catégorie et de ces coefficients hétérogènes par groupe permettent de s'approcher d'une spécification où l'on estimerait le modèle indépendamment pour chaque catégorie de bien (une estimation pour les biens avec 1 pièce, une pour les biens avec 2 pièces etc.) sans pour autant procéder à une spécification séparée par catégorie de bien, cette stratégie réduirait en effet trop fortement le nombre d'observations pour l'estimation des effets fixes par IRIS.

Modèle de différence-de-différences avec effet de traitement différencié selon les caractéristiques des logements

Nous ajoutons enfin à la spécification précédente (équation 2) une interaction de la variable de traitement avec des caractéristiques clés des logements, à savoir le nombre de pièces, la surface, l'époque de construction, le statut meublé/non meublé, le secteur géographique de l'encadrement, l'arrondissement et l'année de l'encadrement (du 1^{er} juillet au 30 juin). Cette spécification permet d'estimer un effet de traitement hétérogène selon les dimensions voulues. Nous estimons le modèle suivant :

$$Y_{ict} = (T_{ct} \times M_{ict}) \beta + Z_{ict}\beta_{tr} + Z_{ict}\beta_{co} + X_{ict}\theta + \mu_c + \mu_t + Z_{ict} \times \mu_t + \varepsilon_{ict} \quad (3)$$

où M_{ict} est la variable avec laquelle on interagit le traitement, et β est un vecteur représentant l'effet de l'encadrement sur les loyers pour chaque modalité de la variable M_{ict} . Nous réalisons une estimation pour chaque variable pour laquelle nous désirons analyser l'hétérogénéité de l'effet de l'encadrement (soit une estimation où M_{ict} correspond au nombre de pièces, une estimation où M_{ict} correspond à la période de construction, etc.).

3.3. Résultats

3.3.1. Evaluation de l'impact de l'encadrement des loyers

Le Tableau 8 contient les résultats de l'estimation du modèle de différence-de-différences. La 1^{ère} colonne contient les résultats d'estimation avec des coefficients homogènes (modèle 1) alors que la 2^{ème} colonne contient les résultats d'estimation incluant des coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque modalité des variables mentionnées dans la partie 3.2.2. (modèle 2). Les variables de contrôle concernant les caractéristiques des biens sont incluses et les coefficients estimés associés sont reportés en annexe A.3.2. Les écarts-types sont robustes et groupés par ville.

Tableau 8 : Résultats d'estimation du modèle de différence-de-différences

	Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
Traitement (%)	-3,7%	-4,2%
Traitement	-0.038*** (0.005)	-0.043*** (0.006)
Num.Obs.	559300	559300
R2	0.870	0.879
R2 Adj.	0.870	0.879
R2 Within	0.510	0.545
Écarts-types	Par ville	Par ville
Effet-fixe : trimestre	X	X
Effet-fixe : IRIS	X	X
Variables de contrôle	X	X

La variable *Traitement* indique l'impact de l'encadrement des loyers sur le niveau des loyers au m2 exprimés en logarithme. Le niveau de significativité est indiqué par : + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 et la standard error apparait entre parenthèses. Pour interpréter le coefficient, il convient de calculer son exponentielle. Par exemple pour la colonne 1 : $\exp(-0.038)-1 = -0.0372$, soit une baisse de 3.7%. Le coefficient transformé, qui est directement interprétable, est indiqué à la ligne *Traitement (%)*.

La variable *Traitement (%)* indique l'impact de l'encadrement des loyers sur le niveau des loyers au m2. Les effets de traitement issus des deux spécifications sont relativement proches : **sur la période allant du 1^{er} juillet 2019 au 30 juin 2023, l'encadrement des loyers a conduit à une baisse des loyers parisiens de 3,7% pour le modèle coefficients homogènes et de 4,2% pour le modèle avec coefficients hétérogènes.** Sur la période, sachant que le loyer moyen constaté dans les annonces parues à Paris est de 1 469€, le loyer moyen en l'absence du dispositif d'encadrement aurait atteint une fourchette comprise entre 1 525€ et 1 533€. **L'encadrement des loyers a donc permis de réduire le niveau moyen des loyers entre 56€/mois (672€ par an) et 64€/mois (768€ par an).**

3.3.2. Analyse d'hétérogénéité par rapport aux caractéristiques

L'analyse d'hétérogénéité (équation 3) permet de comprendre l'effet différencié de l'encadrement des loyers sur les loyers pour différentes caractéristiques des logements. Nous les représentons graphiquement, avec en abscisse les modalités de la variable avec laquelle nous interagissons le traitement, et en ordonnée l'effet du traitement pour cette modalité. Si l'intervalle de confiance à 95% croise la valeur 0 en ordonnée, cela signifie que l'encadrement n'a pas d'effet pour la modalité étudiée. Les tableaux de résultat associés à ces graphiques sont dans les annexes A.3.5 à A.3.11.

La première dimension d'hétérogénéité considérée est temporelle (Figure. Voir figure A.3.3 pour le modèle à coefficients homogènes). Nous n'utilisons pas des années calendaires, mais des années commençant au 1^{er} juillet, pour la correspondance avec la mise en œuvre des nouveaux seuils d'encadrement. L'année 2017 correspond donc au début du jeu de données du 1^{er} janvier 2018 au 30 juin 2018, l'année 2018 correspond aux données du 1^{er} juillet 2018 au 30 juin 2019, ..., et l'année 2022 correspond aux données du 1^{er} juillet 2022 au 30 juin 2023. Le traitement n'est pas significatif avant la mise en place de l'encadrement : préalablement au 1^{er} juillet 2019, les bailleurs n'ont donc ni baissé ni augmenté leur loyer de façon significative. Plus largement sur la période, nous constatons une forte hétérogénéité de l'encadrement d'une période à l'autre. La capacité de l'encadrement à faire baisser les montants des loyers augmente progressivement dans le temps, et ce jusque fin juin 2023.

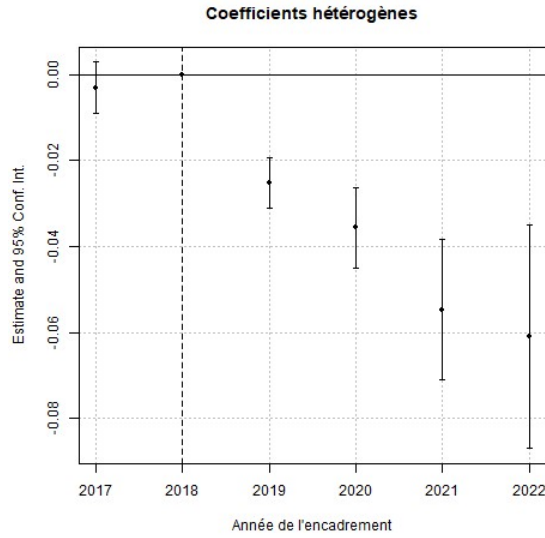
Sur la période allant du 1^{er} juillet 2019 au 30 juin 2020, sachant que le loyer moyen constaté dans les annonces parues à Paris est de 1 504€ et que l'effet de l'encadrement est de -2,5%, le loyer moyen en l'absence du dispositif d'encadrement aurait atteint 1 543€. L'encadrement des loyers a donc permis de réduire le niveau moyen des loyers de 39€ par mois (de 468€ par an). Sur la période plus récente allant du 1^{er} juillet 2022 au 30 juin 2023, sachant que le loyer moyen constaté dans les annonces parues à Paris est de 1 549€ et que l'effet de l'encadrement est de -5,9%, le loyer moyen en l'absence du dispositif d'encadrement aurait atteint 1 646€. L'encadrement des loyers a donc permis de réduire le niveau moyen des loyers de 97€ par mois (1 164€ par an). Le tableau 9 fournit ces calculs pour l'ensemble des périodes.

Tableau 9 : variations de loyer imputables à l'encadrement par année

	2019	2020	2021	2022
Loyer observé (€)	1504	1429	1427	1549
Effet du traitement (%)	-2,5	-3,5	-5,3	-5,9
Loyer contrefactuel (€)	1543	1481	1507	1646
Variation de loyer par mois (€)	-39	-52	-80	-97
Variation de loyer par an (€)	-468	-624	-960	-1164

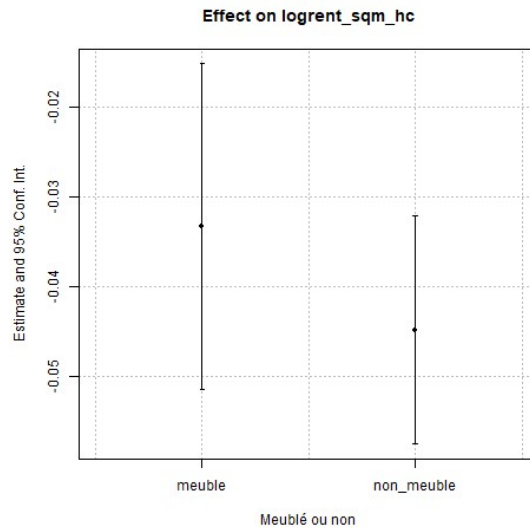
Il est à noter que l'intervalle de confiance augmente au fil des périodes d'encadrement. Cela pourrait signifier qu'avec le temps, l'écart se creuse entre les biens respectant l'encadrement et ceux qui ne le respectent pas. D'un côté les biens qui le respectent voient leur loyer diminuer progressivement par rapport au groupe de contrôle, de l'autre les biens qui ne le respectent pas font diminuer l'efficacité de l'encadrement et donc augmentent la variance de l'estimation.

Figure 15 : Evolution de l'effet du traitement



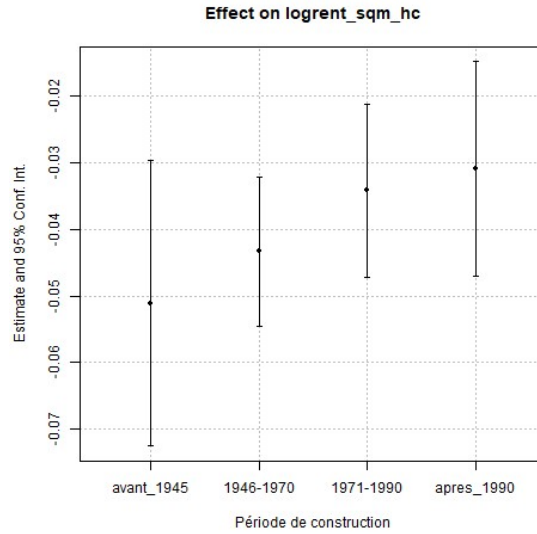
Nous différencions ensuite l'effet de l'encadrement selon que les appartements en location sont meublés ou non (Figure 16). L'effet de l'encadrement est plus important pour les non-meublés que pour les meublés, mais le fait que les intervalles de confiance se recoupent largement nous conduit à nuancer la différence entre les deux.

Figure 16 : Effet de l'encadrement en distinguant les logements meublés et non meublés



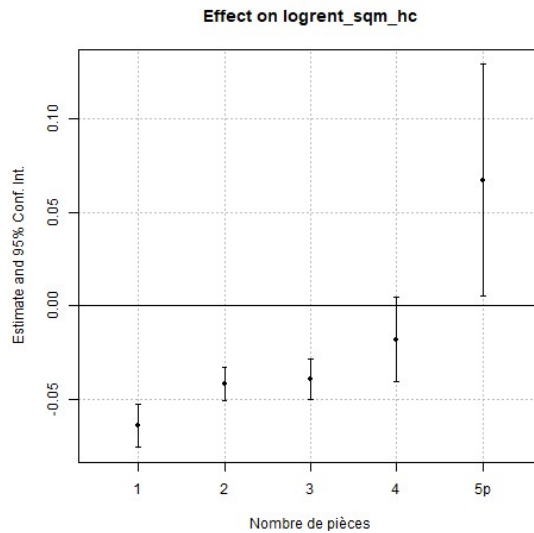
Concernant les différences de l'effet de l'encadrement selon la période de construction du logement (Figure), l'effet semble relativement similaire pour toutes les périodes de construction. Nous constatons toutefois que la variance associée aux biens construits avant 1945 est plus élevée que pour les autres groupes.

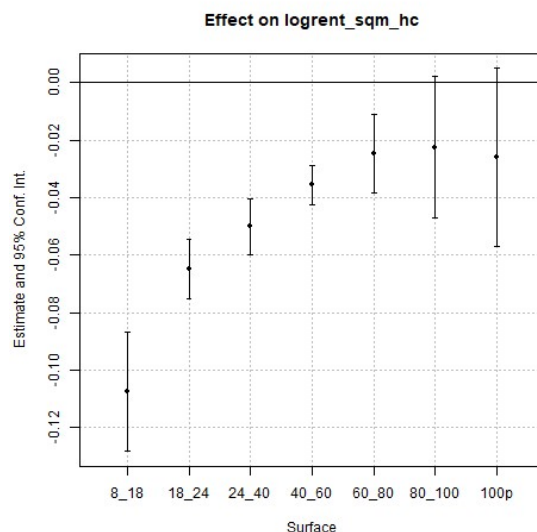
Figure 17 : Effet de l'encadrement selon la période de construction



Nous estimons ensuite l'effet de l'encadrement selon le nombre de pièces puis selon la surface des logements (Figure 18 18). Les résultats sont cohérents entre ces deux estimations : l'encadrement fait davantage baisser les loyers des petits appartements (1 pièce ou moins de 18m²). La capacité de l'encadrement à réduire les montants des loyers diminue à mesure que les surfaces/nombre de pièces augmente. L'effet de l'encadrement a une très forte variance illustrée par un intervalle de confiance très large pour les biens de 5 pièces et plus ou de 100m² et plus. Cet effet peut notamment être dû à la rareté de ce type de bien qui peut provoquer des estimations moins précises. L'impact de l'encadrement devient non significatif à partir de 4 pièces et 80 m². L'intervalle de confiance des biens de 5 pièces et plus est également trop élevé par rapport aux autres types de biens pour conclure à un effet significatif.

Figure 18 : Effet de l'encadrement selon le nombre de pièces et selon la surface





Sur la période allant du 1^{er} juillet 2019 au 30 juin 2023, sachant que le loyer moyen constaté dans les annonces parues à Paris est de 808€ pour les studios et que l'effet de l'encadrement est de -6,2%, le loyer moyen en l'absence du dispositif d'encadrement aurait atteint 861€. L'encadrement des loyers a donc permis de réduire le niveau moyen des loyers de 53€ par mois (636€ par an). Les résultats sont très proches pour les logements avec une surface comprise entre 18 et 24 m², avec une réduction qui atteint 636€ par an. Pour un logement de 3 pièces, sachant que le loyer moyen constaté dans les annonces parues à Paris est de 1922€ et que l'effet de l'encadrement est de -3,8%, le loyer moyen en l'absence du dispositif d'encadrement aurait atteint 1998€. L'encadrement des loyers a donc permis de réduire le niveau moyen des loyers de 76€ par mois (912€ par an). Les tableaux 10 et 11 fournissent ces calculs pour toutes les catégories de pièces et de surface.

Tableau 10 : variations de loyer imputables à l'encadrement par nombre de pièces

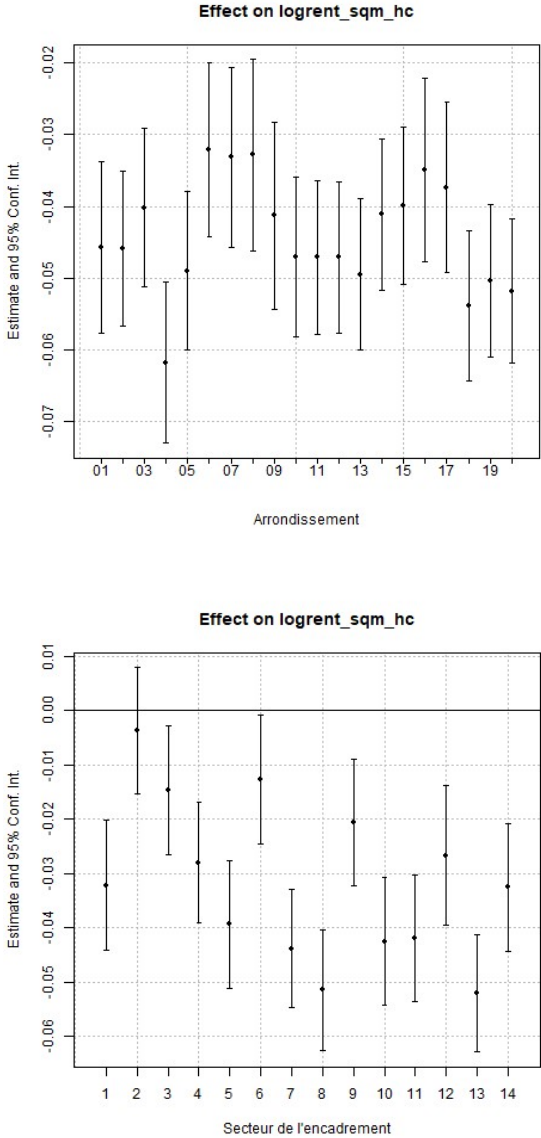
	1 pièce	2 pièces	3 pièces	4 pièces	5 pièces et +
Loyer observé (€)	808	1282	1922	2783	4100
Effet du traitement (%)	-6,2	-4,1	-3,8	-1,8	6,9
Loyer contrefactuel (€)	861	1337	1998	2834	3835
Variation de loyer par mois (€)	-53	-55	-76	-51	265
Variation de loyer par an (€)	-636	-660	-912	-612	3180

Tableau 11 : variations de loyer imputables à l'encadrement par surface

	8 à 18m ²	18 à 24m ²	24 à 40m ²	40 à 60m ²	60 à 80m ²	80 à 100m ²	>100m ²
Loyer observé (€)	612	790	1032	1471	1992	2624	4029
Effet du traitement (%)	-10,2	-6,3	-4,9	-3,5	-2,4	-2,2	-2,6
Loyer contrefactuel (€)	682	843	1085	1524	2041	2683	4137
Variation de loyer par mois (€)	-70	-53	-53	-53	-49	-59	-108
Variation de loyer par an (€)	-840	-636	-636	-636	-588	-708	-1296

Nous examinons enfin l'hétérogénéité de l'impact de l'encadrement selon les arrondissements et selon les secteurs (Figure 19) de l'encadrement. Peu de différences significatives apparaissent selon le secteur de l'encadrement. Ce résultat peut s'expliquer par le mode de délimitation des périmètres de ces secteurs, qui ont précisément été définis comme des zones relativement homogènes. Les différences existantes sont donc captées par les effets fixes, rendant l'effet de la politique relativement homogène entre ces secteurs. A l'exception du 4^{ème} arrondissement, l'effet est également relativement homogène entre les arrondissements : les intervalles de confiance se croisent pour la plupart des arrondissements.

Figure 19 : Effet de l'encadrement selon l'arrondissement et le secteur géographique



3.3.3. Analyse d'hétérogénéité par niveau de contrainte des loyers

Plusieurs études ont exploité la co-existence, sur un même territoire, d'un marché contraint (où les biens sont encadrés) et d'un marché non contraint où les prix pratiqués sont libres. Elles ont montré que si l'encadrement des loyers permettait une baisse des loyers dans le secteur contraint, elle engendrait une hausse des loyers dans le secteur non contraint (Skak et Bloze, 2013 ; Chapelle, Wasmer, et Bono, 2019 ; Mense, Michelsen, et Kholodilin, 2023). L'encadrement des loyers à Paris concerne toutes les observations dans notre base de données, ce qui ne permet pas une analyse similaire. Cependant, nous pouvons définir plusieurs niveaux de contrainte des loyers en utilisant la différence entre le seuil de loyer maximum autorisé par l'encadrement et le loyer contrefactuel (loyer qui aurait été effectif à Paris si l'encadrement des loyers n'avait pas été mis en place).

Nous commençons par estimer le loyer contrefactuel en utilisant une méthode par imputation (Borusyak, Jaravel, et Spiess, 2022) qui consiste à estimer le loyer contrefactuel sous les mêmes hypothèses qu'une différence-de-différences (SUTVA et hypothèse de tendances parallèles). Plus précisément, nous estimons le modèle sur les observations non traitées et nous nous servons des coefficients estimés pour imputer le loyer sur les observations à Paris pendant la période d'encadrement. Nous obtenons ainsi le loyer contrefactuel noté $\hat{Y}_{ict}(0)$ qui correspond au loyer à Paris si l'encadrement n'avait pas été mis en place. Nous pouvons ainsi calculer une mesure du niveau de contrainte des loyers :

$$RC_{ict} = \frac{\hat{Y}_{ict}(0) - RT_{ict}}{\hat{Y}_{ict}(0)}$$

où RT_{ict} est le loyer de référence majoré appliqué au bien i dans la commune c (en l'occurrence, Paris) à la période t . Nous divisons cette mesure par le loyer contrefactuel afin de limiter les effets taille (les biens avec un loyer plus élevé sont aussi ceux qui dépassent le plus le seuil en valeur). Cette mesure prend des valeurs négatives pour les biens qui auraient été en dessous du seuil de loyer sans la mise en place de l'encadrement, et des valeurs positives dans l'autre cas. Nous divisons cette mesure de contrainte en 20 quantiles afin de l'utiliser dans une analyse d'hétérogénéité. Nous pouvons ainsi étudier l'effet de l'encadrement des loyers en fonction du niveau de contrainte des loyers. Autrement dit, cela nous permet de savoir si les biens qui étaient relativement bas par rapport au seuil d'encadrement annoncé en ont profité pour augmenter leur loyer.

Nous ne sommes pas en mesure d'identifier le seuil de loyer pour tous les biens dans notre base de données, en raison des données manquantes sur la période de construction et/ou l'adresse. Nous parvenons à identifier le seuil de loyer pour 124 688 observations de nos 210 544 observations à Paris en période de post-traitement¹⁰. Le Tableau 12 compare les estimations de l'effet de l'encadrement entre cet échantillon et la base complète utilisée dans les estimations précédentes. Les colonnes 1 et 2 indiquent les résultats sur toutes les données et les colonnes 3 et 4 les résultats sur les données où nous avons pu identifier le seuil de l'encadrement des loyers applicable au bien. Les effets de l'encadrement estimés sont très proches pour les 4 colonnes. Nous en déduisons que cet échantillon ne devrait pas biaiser les résultats de notre analyse d'hétérogénéité par niveau de contrainte des loyers.

¹⁰ Nous n'avons pas besoin du seuil de loyer avant la mise en place de l'encadrement car l'analyse se concentre ici sur l'effet de l'encadrement après sa mise en place et non sur les dépassements avant/après comme dans la section 3.1.

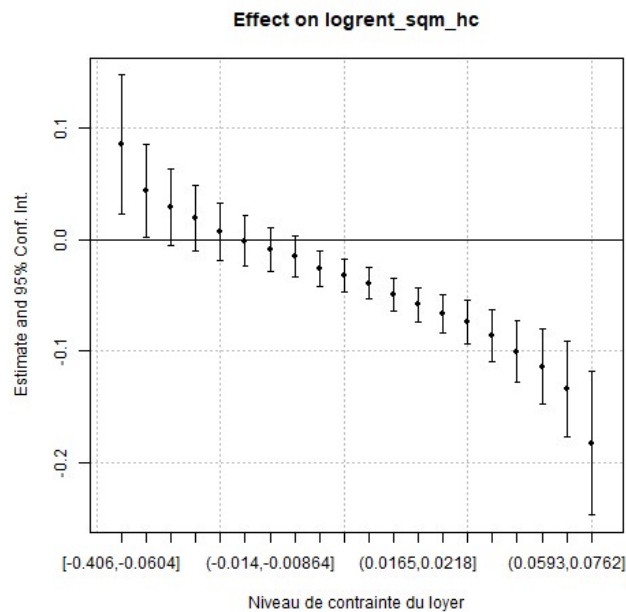
Tableau 12 : Résultats d'estimation du modèle de différence-de-différences

	Toutes les observations / Coefficients homogènes	Toutes les observations / Coefficients hétérogènes	Observations avec seuil / Coefficients homogènes	Observations avec seuil / Coefficients hétérogènes
Traitement (%)	-3,7%	-4,2%	-3,0%	-3,5%
Traitement	-0.038*** (0.005)	-0.043*** (0.006)	-0.031*** (0.006)	-0.035** (0.008)
Num.Obs.	559300	559300	473444	473444
R2	0.870	0.879	0.872	0.881
R2 Within	0.510	0.545	0.530	0.565
Écart-types	Par ville	Par ville	Par ville	Par ville
Effet fixe trimestre	X	X	X	X
Effet fixe IRIS	X	X	X	X
Variables de contrôle	X	X	X	X

La variable *Traitement* indique l'impact de l'encadrement des loyers sur le niveau des loyers au m2 exprimés en logarithme. Le niveau de significativité est indiqué par : + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 et la standard error apparait entre parenthèses. Pour interpréter le coefficient, il convient de calculer son exponentielle. Par exemple pour la colonne 1 : $\exp(-0.038)-1 = -0.037$, soit une baisse de 3.7%. Le coefficient transformé, qui est directement interprétable, est indiqué à la ligne *Traitement (%)*.

Les résultats de l'analyse d'hétérogénéité par niveau de contrainte des loyers sont indiqués dans la Figure 20 (voir annexe A.3.12 pour le tableau avec la valeur des coefficients). Dans l'ensemble, l'encadrement des loyers a été plus efficace pour réduire les loyers qui auraient été au-dessus du seuil d'encadrement. Plus le loyer contrefactuel aurait été élevé (sans mise en place de l'encadrement), plus l'effet de l'encadrement a été élevé. Les loyers qui ont été les moins contraints par la politique (ceux du 1^{er} quantile), ont vu leur loyer augmenter de 2% malgré l'encadrement. Cet effet est proche de 0 et non significatif pour les biens du 2^{ème} et du 3^{ème} quantile. L'encadrement a permis une réduction significative du loyer des biens à partir du 4^{ème} quartile. Nous pouvons en déduire qu'en présence d'un encadrement des loyers, les loyers n'augmentent pas uniquement dans le marché non contraint, mais également pour les biens qui sont fortement en-dessous du seuil d'encadrement imposé par la loi.

Figure 20 : Effet de l'encadrement selon le niveau de contrainte des loyers



3.3.4. Effet de l'encadrement si tous les biens avaient un loyer inférieur ou égal au seuil maximum autorisé

Notre analyse statistique a révélé que le taux de dépassement du loyer de référence majoré se situait entre 34.36 et 49.48% sur la période allant de juillet 2019 à juin 2023 (figure 3 dans la partie 3.1.3.). Ce dépassement peut être légal si un complément de loyer est justifié par des caractéristiques exceptionnelles du bien qui le distinguent des biens équivalents à proximité ou, dans les autres cas, être illégal. Nous cherchons dans cette partie à quantifier quel serait l'effet de l'encadrement des loyers à Paris si tous les bailleurs le respectaient et si aucun complément de loyer n'était pratiqué. En d'autres termes, cela revient à étudier une borne haute théorique de l'effet de la politique d'encadrement des loyers.

Pour estimer cet effet, nous utilisons également l'échantillon d'observations pour lequel nous pouvons identifier le seuil de l'encadrement applicable à chaque bien à Paris pendant la période d'encadrement. Nous remplaçons ensuite tous les loyers qui dépassent le seuil majoré par la valeur de ce seuil majoré. Ainsi aucun des loyers dans notre échantillon ne dépasse ce seuil. Nous réestimons ensuite le modèle en différence-de-différences pour obtenir la borne haute de l'effet de la politique. Les écarts-types sont obtenus à partir d'une méthode de rééchantillonnage de type bootstrap avec 1000 répliques. Les résultats sont présentés dans le Tableau 13. La 1^{ère} colonne indique les résultats pour un modèle avec coefficients homogènes et la 2nde pour un modèle avec coefficients hétérogènes. Dans chacun des cas, l'effet de l'encadrement est significatif. D'après ces estimations, la borne haute de l'effet de l'encadrement des loyers à Paris se situe entre 8,2% (colonne 2) et 8,7% (colonne 1). Cet effet théorique est plus de deux fois plus élevé que l'effet actuel de la politique d'encadrement des loyers. Le loyer moyen constaté sur ce sous-échantillon est de 1420€ sur la période, et en l'absence d'encadrement le loyer contrefactuel aurait été compris entre 1547€ et 1555€. L'encadrement a donc permis de réduire le niveau moyen des loyers de 127€ à 135€ par mois (ou de 1524€ à 1620€ par an).

Tableau 13 : Résultats d'estimation du modèle si l'encadrement était respecté

	Coefficients homogènes	Coefficients hétérogènes
Traitement (%)	-8,7%	-8,2%
Traitement	-0.091*** (0.001)	-0.085*** (0.002)
Num.Obs.	473444	473444
R2	0.870	0.881
R2 Within	0.528	0.566
Écarts-types	Par ville	Par ville
Effet fixe trimestre	X	X
Effet fixe IRIS	X	X
Variables de contrôle	X	X

La variable *Traitement* indique l'impact de l'encadrement des loyers sur le niveau des loyers au m2 exprimés en logarithme. Le niveau de significativité est indiqué par : + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 et la standard error apparait entre parenthèses. Pour interpréter le coefficient, il convient de calculer son exponentielle. Par exemple pour la colonne 1 : $\exp(-0.091)-1 = -0.087$, soit une baisse de 8.7%. Le coefficient transformé, qui est directement interprétable, est indiqué à la ligne *Traitement (%)*.

3.3.5. Analyse de robustesse

Cette section s'attache à évaluer la sensibilité de nos résultats aux hypothèses faites pour les estimations. Le modèle estimé précédemment repose sur l'hypothèse d'une tendance de loyers parallèle entre le groupe traité et le groupe de contrôle. Afin de vérifier si notre estimation est robuste à cette hypothèse, nous estimons une différence-de-différence synthétique (SDID) qui ne repose pas sur l'hypothèse de tendances parallèles (Arkhangelsky et al., 2021). Cette méthode est présentée en détail dans l'annexe A.3.13.

Le tableau 14 présente les résultats de ces estimations pour les modèles à coefficient homogènes et hétérogènes. Dans chacun des cas, nous comparons le résultat obtenu avec la différence-de-différences et la différence-de-différences synthétique. Les écarts-types des colonnes 2 et 4 sont obtenus à partir d'un bootstrap utilisant 1000 répétitions. Pour ces deux modèles, l'effet estimé de l'encadrement est très proche : il est exactement le même dans le cas des coefficients hétérogènes (une baisse de 4.2%), et est très proche pour le cas avec des coefficients homogènes (baisse de 3.7% pour le modèle DID contre 3.9% pour le modèle SDID). L'effet de l'encadrement estimé est donc robuste puisque nous trouvons un résultat très similaire lorsque nous relâchons l'hypothèse de tendances parallèles avec une autre méthode d'estimation.

Tableau 14 : Résultats d'estimation du modèle avec différence-de-différence synthétique

	DID coefficients homogènes	SDID coefficients homogènes	DID coefficients hétérogènes	SDID coefficients hétérogènes
Traitement (%)	-3,7%	-3,9%	-4,2%	-4,2%
Traitement	-0.038*** (0.005)	-0.040*** (0.003)	-0.043*** (0.006)	-0.043*** (0.003)
Num.Obs.	559300	425044	559300	450575
Effets fixes : trimestre	X	X	X	X
Effets fixes : IRIS	X	X	X	X
Variables de contrôle	X	X	X	X

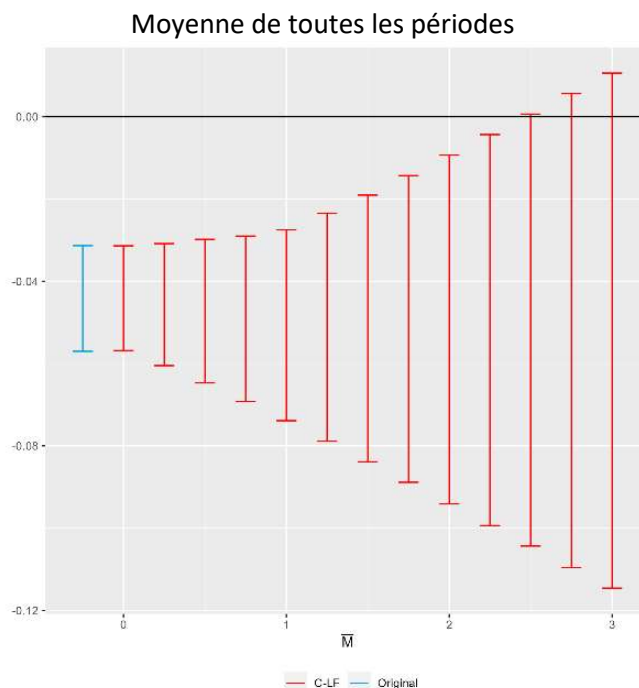
La variable *Traitement* indique l'impact de l'encadrement des loyers sur le niveau des loyers au m2 exprimés en logarithme. Le niveau de significativité est indiqué par : + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ et la standard error apparait entre parenthèses. Pour interpréter le coefficient, il convient de calculer son exponentielle. Par exemple pour la colonne 1 : $\exp(-0.038)-1 = -0.037$, soit une baisse de 3.7%. Le coefficient transformé, qui est directement interprétable, est indiqué à la ligne *Traitement (%)*.

Un deuxième test de robustesse a consisté à appliquer la méthode développée par Rambachan et Roth (2022) afin de tester la sensibilité de l'effet de traitement estimé à l'hypothèse PTA. La Figure montre que si les tendances de loyer étaient proches entre Paris et le groupe de contrôle avec une corrélation estimée de 0.94, elles ne sont pas exactement identiques. Nous mettons en œuvre un test statistique pour l'effet de cette hypothèse sur nos résultats pour différents niveaux de violation de l'hypothèse. Deux tests sont possibles. Le premier suppose que l'existence de chocs temporaires affecte différemment le groupe de contrôle et le groupe traité, alors que le deuxième suppose une différence

de tendance qui change avec le temps. Dans notre cas, le plus gros risque pour l'identification est la crise sanitaire de la covid-19 qui pourrait avoir affecté différemment le marché immobilier parisien par rapport au groupe de contrôle. C'est donc le premier test qui semble le plus pertinent dans notre cas.

Nous réalisons le test pour l'effet moyen du traitement estimé précédemment en utilisant pour cela le modèle à coefficient hétérogènes. Les résultats sont présentés dans la figure 21. Le graphique pour le modèle à coefficients homogènes est en annexe A.3.4. L'effet du traitement est robuste (significatif) jusqu'à une violation de la PTA deux fois plus élevée en période de post-traitement qu'en période de pré-traitement. Cela confirme que l'encadrement des loyers a bien pour effet de baisser le montant des loyers sur le marché Parisien par rapport au groupe de contrôle et que cet effet est robuste à l'hypothèse de tendances parallèles.

Figure 21 : Test de robustesse



4. Mesure d'impact de l'encadrement des loyers à Paris sur données des observatoires locaux des loyers accessibles sous CASD

Notre second jeu de données est constitué des données de baux versées au Centre d'Accès Sécurisé aux Données (CASD) par les observatoires locaux des loyers (OLL) pour les mettre à disposition des chercheurs dans un environnement sécurisé. Ces données proviennent de collectes réalisées par les observatoires et de données transmises par les professionnels de l'immobilier dans le cadre de la loi ELAN. Les OLL ont été créés pour améliorer la connaissance des loyers pratiqués dans le parc locatif privé et disposer de données fiables, homogènes et comparables d'un territoire à l'autre. Les informations collectées sont une aide à la décision pour les politiques publiques locales et nationales (PLH, zonages, plafonds Anah, encadrement des loyers...) et aident les professionnels (administrateurs de biens, promoteurs...) et les particuliers (bailleurs et locataires) dans leurs arbitrages.

4.1. Données des OLL

Nous recourons à la même stratégie d'identification que précédemment. En conséquence, nous gardons uniquement les données de baux de Paris intra-muros et des 8 villes de province qui appartiennent au groupe de contrôle (Aix-en-Provence, Grenoble, Marseille, Nantes, Nice, Strasbourg, Toulon et Toulouse). Nous décrivons le contenu de ce jeu de données disponibles sur la période allant de 2015 à 2022, à l'aide de quelques statistiques descriptives. Le tableau 15 fournit le nombre d'observations disponibles chaque année pour chacune des 9 villes.

Tableau 15 : Nombre d'observations par année et par ville

	2015	2016	2017	2018	2019	2020	2021	2022
AIX-EN-PROVENCE	0	2061	2318	2504	4241	4313	3806	0
GRENOBLE	0	6215	4411	6657	4279	6321	5355	0
MARSEILLE	0	10156	11670	13087	20252	20621	20095	0
NANTES	0	9129	7550	8570	7885	10622	11904	0
NICE	0	9384	11386	11940	9677	9744	0	0
PARIS	46842	44133	48075	44600	46761	56673	0	0
STRASBOURG	0	5570	9898	9724	9060	8488	8228	0
TOULON	0	1605	2778	1645	3803	0	5308	0
TOULOUSE	0	13073	14340	11800	10662	18462	19686	23433

D'une part, à l'exception de l'OLAP pour la région parisienne, les observatoires n'ont pas collecté ou pas versé au CASD des données de baux pour l'année 2015. En conséquence, nous ne pouvons pas les mobiliser pour évaluer l'impact du premier dispositif d'encadrement des loyers. D'autre part, la forte proportion d'observations manquantes en $t - 2$ (i.e., en 2021 actuellement) pour Nice, et surtout pour Paris, ne permet pas d'évaluer l'impact récent du second dispositif d'encadrement des loyers. Il est

également à noter l'absence de données à Toulon en 2020 ce qui est limitant pour notre groupe de contrôle.

La géolocalisation des logements au niveau infra-communal, comme le montre le tableau 16, est par ailleurs limitée. En effet, l'information sur l'IRIS d'appartenance est manquante pour 42,5% des observations.

Tableau 16 : nombre d'observations avec un IRIS manquant

	IRIS RENSEIGNE	IRIS MANQUANT
AIX-EN-PROVENCE	19157	86
GRENOBLE	32926	312
MARSEILLE	95881	50 ou moins
NANTES	54335	1325
NICE	52131	50 ou moins
PARIS	50 ou moins	287084
STRASBOURG	3422	47546
TOULON	2721	12418
TOULOUSE	11496	99960

L'IRIS est notamment systématiquement manquant à Paris, ce qui est problématique car cette information nous permettrait d'introduire un effet fixe IRIS afin de capter l'hétérogénéité locale invariante dans le temps de façon plus fine. Cet effet permettrait de capter des aménités locales comme la densité d'arrêts de bus et de métro, la part d'espaces verts, l'offre culturelle ou encore la densité de bars et restaurants. Les options pour les évaluations ultérieures seraient 1) l'ajout de cette variable directement par les observatoires, et en particulier par l'OLAP, et 2) la possibilité accordée à l'équipe de géocoder les observations à partir des adresses dans la bulle CASD, mais cette solution dépend des autorisations accordées par l'ANIL, est très coûteuse en temps à mettre en œuvre et comporte le risque de ne pas pouvoir tout géocoder. Dans le pas de temps du contrat, nous n'avons pas pu mettre en œuvre cette option, par ailleurs trop coûteuse.

Enfin, le taux de complétion est inégal et trop faible sur certaines caractéristiques clés des logements. Si le nombre de pièces, la surface, la période de construction et le code postal sont systématiquement renseignés, les variables suivantes ont un fort pourcentage de valeurs manquantes : l'étage (60%), le nombre de salle de bains (89%), le nombre de toilettes (90%), la présence d'un jardin (89%), la présence d'un balcon (87%), la présence d'un ascenseur (78%) et la présence d'un parking (85%). Ce taux de complétion faible sur des variables structurelles clés comme l'étage ne permet pas d'assurer la fiabilité de l'analyse économétrique et de l'évaluation d'impact en raison du biais de variables omises qu'il génère. Le détail du taux de complétion par ville est disponible en annexe (annexes A.4.1 à A.4.6) Nous retenons donc dans l'analyse les seules variables suivantes : le loyer au m2 hors charges (introduit en log), la surface (introduite en log), le nombre de pièces, la période de construction, le trimestre et l'année d'enquête, la ville et la date d'emménagement.

Afin de ne garder que les loyers de marché, nous menons les analyses économétriques sur l'échantillon restreint aux logements avec une date d'emménagement < 365 jours à la date d'enquête. Le tableau 17 fournit le nombre d'observations par ville et par année d'enquête.

Tableau 17 : nombre d'observations par ville et année d'emménagement

	2018	2019	2020
AIX-EN-PROVENCE	952	1111	820
GRENOBLE	1150	1861	1474
MARSEILLE	3668	3938	2821
NANTES	2810	3500	3000
NICE	1316	1306	78
PARIS	5235	8484	2012
STRASBOURG	2172	2446	2145
TOULON	521	135	569
TOULOUSE	3222	4371	5537

4.2. Méthodologie

Afin d'évaluer l'impact de l'encadrement des loyers sur les loyers de baux à Paris, nous avons répliqué la même stratégie d'identification que celle menée sur les données d'annonces du groupe SeLoger, à savoir une méthode de différence-en-différences avec un groupe traité (les baux à Paris) et un groupe de contrôle (les baux dans 8 villes de province).

4.3. Résultats d'évaluation de l'impact de l'encadrement des loyers

Nous avons estimé le modèle sur l'échantillon restreint aux logements avec une date d'emménagement < 365 jours à la date d'enquête, avec deux modalités d'introduction des effets fixes spatiaux : d'une part en introduisant des effets fixes par ville (colonne 1 du tableau 18) et d'autre part en introduisant des effets fixes à l'IRIS lorsque l'IRIS est disponible et à la ville sinon (colonne 2 du tableau 18). Avec cette dernière spécification, les effets fixes ne sont pas inclus dans la régression au même niveau géographique, ce qui rend les résultats très peu fiables. Nous les présentons uniquement à titre d'information et centrerons nos commentaires sur la première spécification avec effets fixes par ville.

Tableau 18 : Résultats d'estimation du modèle de différence-de-différences sur l'échantillon restreint aux logements avec une date d'emménagement < 365 jours à la date d'enquête, avec effets fixes par ville et par iris (ou par ville en cas de donnée manquante)

	Effets fixes par ville	Effets fixes par IRIS
Traitement (%)	-3,6%	-3,3%
Traitement	-0.037*** (0.007)	-0.034** (0.007)
Log de la surface	-0.417*** (0.076)	-0.442*** (0.075)
2 pièces	0.023 (0.024)	0.039+ (0.018)
3 pièces	0.058 (0.038)	0.083* (0.033)
4 pièces	0.148** (0.038)	0.175** (0.038)
5 pièces et plus	0.357*** (0.043)	0.358*** (0.046)
Construction : avant 1946	-0.073* (0.023)	-0.047** (0.013)
Construction : 1971-1990	-0.046 (0.034)	-0.013 (0.021)
Construction : après 1991	-0.038 (0.037)	-0.005 (0.024)
Num.Obs.	66654	66654
R2	0.809	0.814
R2 Within	0.471	0.471
Log.Lik.	22347.872	23353.067
Ecarts-types	Par ville	Par ville
Effets fixes : trimestre	X	X
Effets fixes : ville	X	X

La variable *Traitement* indique l'impact de l'encadrement des loyers sur le niveau des loyers au m2 exprimés en logarithme. Le niveau de significativité est indiqué par : + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 et la standard error apparait entre parenthèses. Pour interpréter le coefficient, il convient de calculer son exponentielle. Par exemple pour la colonne 1 : $\exp(-0.037) - 1 = -0.036$, soit une baisse de 3.6%. Le coefficient transformé, qui est directement interprétable, est indiqué à la ligne *Traitement (%)*.

Le coefficient associé à l'effet de traitement est négatif et significatif, pointant un impact modérant sur les loyers de l'encadrement. En termes d'ampleur, nous trouvons une baisse des loyers à Paris de 3,6%. Afin d'avoir un point de comparaison avec les résultats obtenus sur les données du groupe SeLoger (qui vont jusqu'au 30 juin 2023), nous avons réestimé l'équation de différences-en-différences sur les données d'annonces en nous arrêtant fin 2020. Nous trouvons que la baisse des loyers entre juillet 2019 et décembre 2020 est de 1,8% avec les données du groupe SeLoger. Cette différence dans l'ampleur de l'impact de l'encadrement peut s'expliquer d'une part par le biais lié à l'absence de

contrôle des aménités locales, l'effet fixe étant introduit au niveau de la ville, et d'autre part par le manque de variables de contrôle dans les données de baux.

A l'instar des traitements effectués sur les données d'annonces, nous avons également effectué une analyse d'hétérogénéité du traitement selon la date. Les résultats sont reportés ci-dessous dans les figures 22 et 23.

Figure 22 : Evolution de l'effet de l'encadrement des loyers sur les montants des baux parisiens au cours du temps avec effets fixes par ville

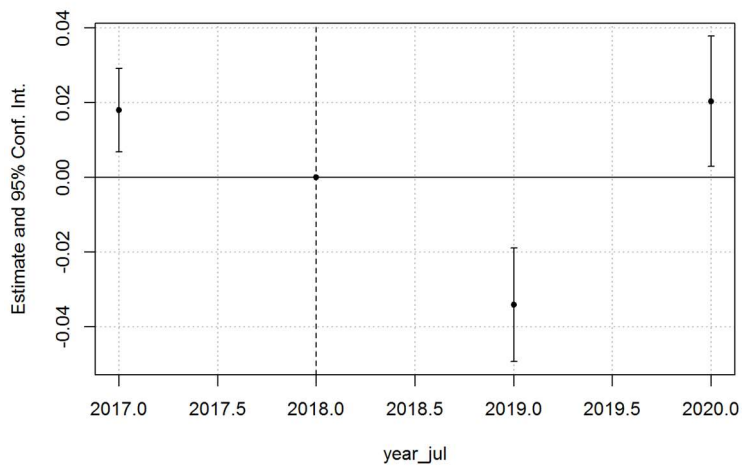
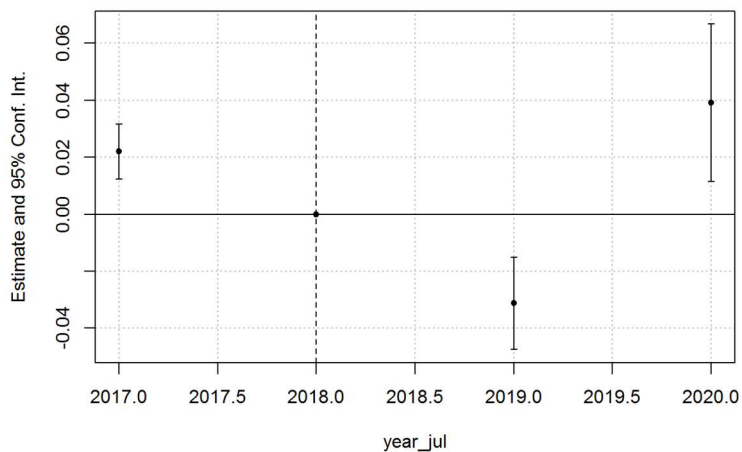


Figure 23 : Evolution de l'effet de l'encadrement des loyers sur les montants des baux parisiens au cours du temps avec effets fixes par IRIS



Les intervalles de confiance sont très larges (beaucoup plus larges qu'avec les données du groupe SeLoger), ce qui fait peser de doutes sérieux sur la fiabilité des résultats. Cela peut notamment s'expliquer par le faible nombre d'observations et par le faible nombre de variables de contrôle dans le modèle à cause des valeurs manquantes. La précision des estimations est dès lors très affectée. Par ailleurs, nous notons que l'hypothèse de tendances parallèles n'est pas respectée en raison de la pré-tendance significative en 2017. Par ailleurs, l'effet de l'encadrement en 2020 est positif, ce qui est peu cohérent. Là encore, ces problèmes sont probablement dus au manque de variables de contrôle et l'absence d'IRIS dans les données parisiennes notamment.

En conclusion, la très faible profondeur temporelle et le nombre réduit d'observations ne permettent pas de mener une analyse fiable de l'impact de l'encadrement. Pour une évaluation ultérieure s'appuyant sur ces données, il serait souhaitable que :

- l'observatoire des loyers du département du Var complète son jeu de données avec l'année 2020 et que l'ensemble des observatoires versent dans un délai beaucoup plus court les jeux de données les plus récents, pour évaluer les effets récents de l'encadrement des loyers.
- l'IRIS soit systématiquement renseigné pour chaque logement, pour capter l'hétérogénéité locale invariante dans le temps.
- le taux de complétion sur des variables structurelles clés comme l'étage soit de 100% pour réduire le biais de variables omises et améliorer la fiabilité des résultats.
- le nombre d'observations de baux de loyers soit significativement augmenté

5. Mesure d'impact de l'encadrement des loyers à Paris sur données « Références de voisinage » de l'OLAP

5.1. Données Références de voisinage

Nous exploitons la base de données « Références de voisinage » constituée par l'OLAP, qui contient des informations sur les baux de locations situées dans Paris et sa petite couronne. Cette base représente bien toutes les catégories de logements (le panel en est extrait) mais pas de façon uniforme sur le territoire parisien car elle n'a pas de vocation statistique. Son objectif est de fournir des listes de références exploitables. En conséquence, les logements peu fréquents sont sur-échantillonnés et les très fréquents au contraire sous-enquêtés. Au prix d'une bonne pondération, elle pourrait être une bonne représentation du parc locatif privé non meublé. Elle a été transmise à l'équipe de recherche après brouillage des coordonnées géographiques. Nous disposons de ces données jusqu'au millésime 2021. Un encadrement des loyers est appliqué à Plaine Commune depuis le 1er juin 2021 et à Est Ensemble depuis le 1^{er} décembre 2021. Nous retirons donc de l'échantillon toutes les observations après le 31 mai 2021¹¹. Par ailleurs, dans les estimations réalisées, nous gardons les observations à partir du 1^{er} janvier 2018. En effet, cette date de début correspond à celle des données du groupe SeLoger, ce qui nous permet de comparer les résultats obtenus sur les deux bases. Nous évitons également ainsi que nos estimations soient affectées par les effets résiduels du précédent encadrement des loyers.

Comme cette base de données contient des données de panel (les logements pouvant être suivis dans le temps), nous conservons comme Malard et Poulhès (2020) l'observation la plus récente pour chaque paire logement-date d'installation. Le tableau 19 indique le nombre d'observations à Paris intra-muros et dans les zones voisines qui sont contenues dans la base finale.

Tableau 19 : nombre d'observations de la base finale des références de voisinage sur la période allant du 1er janvier 2018 au 31 mai 2021

Paris	2km et moins	2 à 4km	4 à 6km	6km et plus
7771	2652	1955	1189	1078

5.2. Méthodologie

La méthode de la différence-de-différences appliquée sur les données du groupe SeLoger repose sur l'hypothèse SUTVA que le traitement n'affecte que les unités traitées, i.e., qu'il n'y a pas d'effet sur le groupe de contrôle. Lorsque le groupe de contrôle est composé d'observations dans d'autres villes éloignées géographiquement de la ville encadrée, cette hypothèse est plausible. Cependant, le périmètre géographique de la base de données « Références de voisinage » de l'OLAP nous impose de

¹¹ Il serait possible d'utiliser une méthode de différence-de-différences « staggered » qui prend en compte le fait que le traitement commence à des moments différents selon les zones. Toutefois, cela nécessiterait davantage de profondeur temporelle pour la zone Plaine Commune et Est Ensemble.

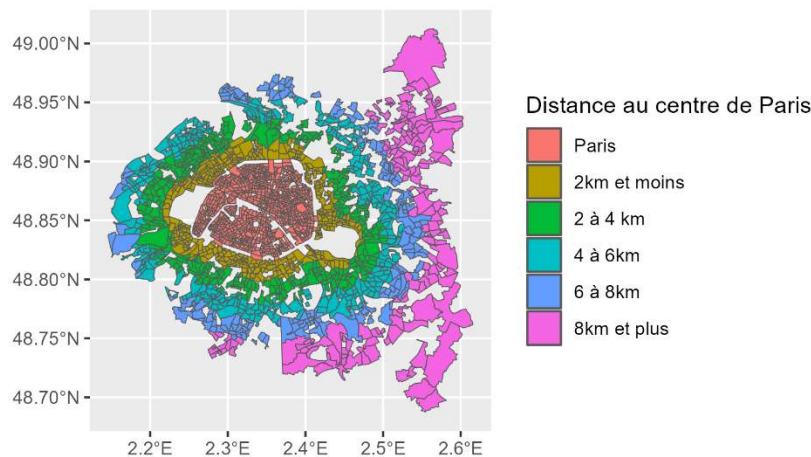
trouver un groupe de contrôle situé dans la petite couronne de Paris, avec le risque qu'il soit affecté par l'encadrement des loyers mis en place à Paris.

Le défi consiste donc à trouver un ensemble de logements situés dans la petite couronne de Paris qui ont des caractéristiques similaires aux logements à Paris, mais qui ne sont pas trop proches géographiquement (afin que les externalités n'affectent pas ce groupe). Nous faisons l'hypothèse qu'*en moyenne*, plus les logements utilisés dans le groupe de contrôle sont loin de Paris, moins ils sont affectés par les externalités, mais moins ils sont semblables à ceux du marché Parisien. A l'inverse, plus on se rapproche de Paris, plus les logements sont semblables *en moyenne*, mais plus également les externalités sont susceptibles de biaiser les résultats de nos estimations.

Nous calculons tout d'abord des anneaux basés sur la distance entre le centroïde de chaque IRIS non concerné par l'encadrement des loyers et le centroïde de l'IRIS traité le plus proche. Cette distance est calculée pour l'ensemble des IRIS de Paris et de la petite couronne, et non seulement pour ceux disponibles dans la base de données pour éviter qu'un IRIS n'appartenant pas à notre base ne biaise la définition des anneaux de distance.

La carte ci-dessous représente les IRIS de Paris et de sa petite couronne, avec des anneaux de distance tous les 2km jusqu'à 8 km et plus. Les zones blanches correspondent à une absence de données de baux.

Figure 24 : anneaux de distance à Paris



A partir de ces anneaux de distance, nous réalisons deux catégories d'estimations.

5.2.1. Modèle de différence-de-différences sans prise en compte des externalités

Nous estimons en premier un modèle de différence-de-différences classique, pour différents contrefactuels composés de logements situés dans des anneaux de distance. Plus précisément, les

groupes de contrôle considérés correspondent aux observations situées alternativement dans les anneaux (ou groupes d’anneaux) suivants : entre 2 et 4km, entre 2 et 6km, entre 2 et 8km, entre 4 et 6km et entre 4 et 8km. Cette étape permet de retenir les zones de la petite couronne qui pourront jouer le rôle de contrefactuel lorsque les externalités seront intégrées dans le modèle.

La spécification est la suivante :

$$Y_{ict} = \beta T_{ct} + X_{ict}\theta + \mu_c + \mu_t + \varepsilon_{ict} \quad (4)$$

où Y_{ict} est le logarithme du loyer hors charges au m², T_{ct} est une variable binaire qui vaut 1 si l’encadrement des loyers est appliqué dans la ville c à la période t , X_{ict} est un ensemble de variables de contrôle représentant les caractéristiques du bien i dans la ville c à la période t , θ est le vecteur de paramètres estimés associés, μ_c est un effet fixe par IRIS, μ_t est un effet fixe par période et ε_{ict} est un bruit blanc iid. Le modèle est estimé grâce à la méthode des Moindres Carrés Ordinaires (MCO) avec inférence robuste.

Les variables de contrôle utilisées sont le nombre de pièces, le logarithme de la surface, la période de construction, l’étage, le nombre de salles de bains et de toilettes, une variable binaire pour le statut meublé, la présence d’une cave ou d’un grenier, la présence d’un balcon, la présence d’une loggia ou d’une terrasse, la présence d’un jardin, et enfin la présence d’un parking inclus dans le loyer.

5.2.2. Modèle de différence-de-différences avec prise en compte des externalités

Nous tenons compte dans un second temps des potentielles externalités induites par l’encadrement des loyers, i.e., nous intégrons dans la spécification précédente la possibilité d’un effet de l’encadrement des loyers à Paris sur les biens localisés à proximité de Paris. Nous utilisons pour cela la spécification de différence-de-différences intégrant des externalités, proposée par Butts (2023) pour quantifier :

- l’impact de l’encadrement (traitement) sur les logements situés à Paris (groupe traité)
- l’impact de l’encadrement (traitement) sur les logements non encadrés mais situés à proximité et potentiellement affectés par des externalités liées à l’encadrement des loyers à Paris (groupes potentiellement affectés par des externalités, à savoir un sous-groupe d’anneaux qui n’a pas le statut de groupe de contrôle).

Cette stratégie repose sur l’hypothèse qu’à partir d’une certaine distance, il existe un groupe de contrôle (composé des observations dans un anneau donné) non affecté par les externalités tout en ayant des caractéristiques proches de Paris, ce qui permet la plausibilité de l’hypothèse de tendances parallèles entre ces groupes. Comme nous cherchons à mesurer les externalités, nous devons également faire une hypothèse de tendances parallèles entre les groupes potentiellement affectés par les externalités et le groupe supposé non affecté et non traité (groupe de contrôle). Nous estimons ce modèle pour un groupe de contrôle à différentes distances de Paris et pour des groupes potentiellement affectés par des externalités à différentes distances de Paris, i.e., pour des anneaux de distance alternatifs.

Formellement, le modèle de différence-de-différences est étendu pour inclure directement les externalités comme paramètres du modèle à estimer :

$$Y_{ict} = \beta T_{ct} + \sum_{j \in dist} Anneau_{ij} \delta_j + X_{ict}\theta + \mu_c + \mu_t + \varepsilon_{ict} \quad (5)$$

où δ_j est l'effet de l'encadrement des loyers sur les observations non traitées proches géographiquement et $dist$ correspond aux anneaux de distance, exprimée en km, (0, 2], (2, 4] et (4, 6]. Les variables de contrôle sont les mêmes que dans le modèle précédent.

5.3. Résultats

5.3.1. Evaluation de l'impact de l'encadrement des loyers sans prise en compte des externalités

Le tableau 20 présente les résultats de l'estimation de l'équation (4) pour un contrefactuel situé à différentes distances de Paris, i.e., pour un anneau ou groupe d'anneaux de distance, mais sans prise en compte des externalités. Concernant l'inférence robuste, nous supposons que les résidus sont i) corrélés dans l'espace à une distance *maximum* correspondant à la borne inférieure de la distance du contrefactuel utilisé et ii) corrélés au sein de chaque IRIS pendant 4 périodes (soit 1 an comme notre unité de temps est le trimestre). Nous ne reportons que l'effet de l'encadrement (variable « Traitement ») pour la lisibilité du tableau, mais les variables de contrôle présentées dans la partie précédente ont bien été incluses dans le modèle.

Tableau 20 : Résultats d'estimation du modèle de différence-de-différences sans prise en compte des externalités

	Entre 2 et 4km	Entre 2 et 6km	Entre 2 et 8km	Entre 4 et 6km	Entre 4 et 8km
Traitement (%)	-3,0%	-2,9%	-2,9%	-2,8%	-3,1%
Traitement	-0.030*** (0.009)	-0.029*** (0.007)	-0.030*** (0.007)	-0.029** (0.011)	-0.031*** (0.009)
Num.Obs.	9726	10915	11516	8960	9561
R2	0.698	0.743	0.758	0.717	0.747
R2 Within	0.438	0.433	0.435	0.422	0.424
FE : trimestre	X	X	X	X	X
FE : IRIS	X	X	X	X	X
Variables de contrôle	X	X	X	X	X

La variable *Traitement* indique l'impact de l'encadrement des loyers sur le niveau des loyers au m2 exprimés en logarithme. Le niveau de significativité est indiqué par : + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 et la standard error apparait entre parenthèses. Pour interpréter le coefficient, il convient de calculer son exponentielle. Par exemple pour la colonne 1 : $\exp(-0.03)-1 = -0.02955$, soit une baisse de 2.9%. Le coefficient transformé, qui est directement interprétable, est indiqué à la ligne *Traitement (%)*.

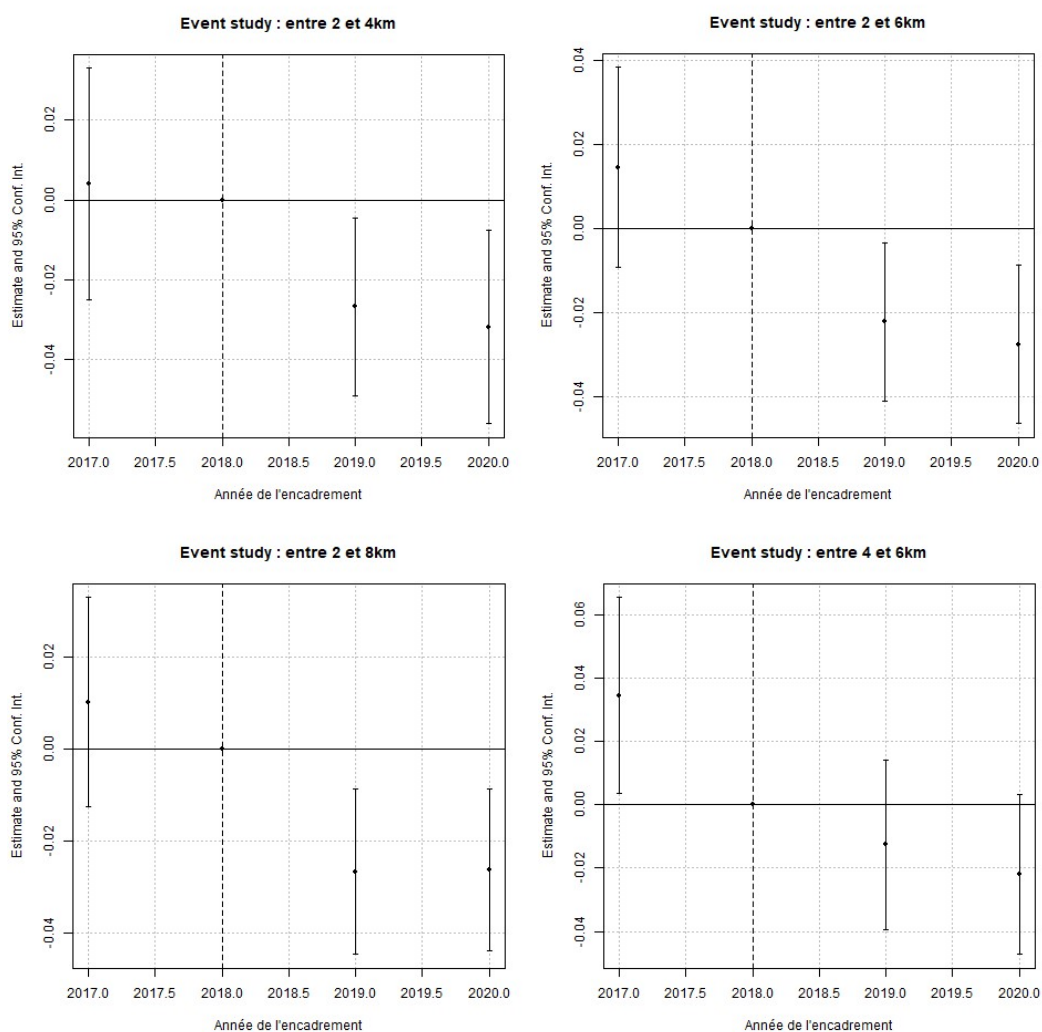
Nous obtenons des valeurs très proches des effets de traitement, quel que soit l'anneau (ou groupes d'anneaux) de distance sélectionné(s) comme contrefactuel. L'encadrement des loyers parisiens implique une baisse des loyers comprise entre 2,8 et 3,1% à Paris. Le fait que les résultats des spécifications pour différents anneaux de distance retenus comme contrefactuels soient proches laisse penser que les externalités doivent être faibles au-delà de 2km.

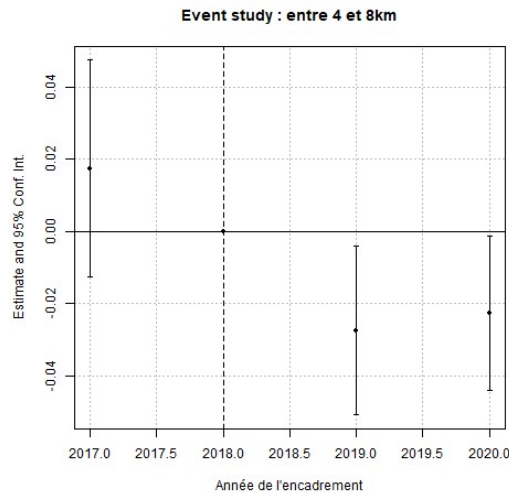
Afin d'avoir un point de comparaison avec les données du groupe SeLoger, nous avons également réestimé le modèle sur un échantillon restreint des données du groupe SeLoger qui s'arrête fin mai 2021. Sur données d'annonces sur la période janvier 2018-mai 2021, la politique d'encadrement a permis une baisse des loyers d'annonces de 2,6%. Les résultats sont donc proches pour ces deux sources de données.

L'estimation en différence-de-différences repose sur l'hypothèse que, en l'absence de politique d'encadrement, la tendance de loyers aurait été la même entre Paris et le groupe de contrôle. Afin de vérifier la plausibilité de cette hypothèse, nous réalisons une « event study », qui consiste à estimer l'effet du traitement avant qu'il ne soit mis en place. En l'absence de tendance de prétraitement significative, nous ne devons pas trouver d'effet du traitement avant la mise en place.

D'après les résultats représentés sur les graphiques de la figure 25, il n'existe de pré-tendance pour aucun des modèles que nous avons testés, à l'exception de celui avec un contrefactuel compris entre 4 et 6km, quoique l'intervalle de confiance soit proche de 0. L'hypothèse de tendance parallèle entre Paris et le groupe de contrôle semble donc être valable, même si le contrefactuel entre 4 et 6km semble moins fiable.

Figure 25 : Evolution de l'effet de l'encadrement des loyers sur les montants des baux parisiens au cours du temps en fonction des anneaux





5.3.2. Evaluation de l'impact de l'encadrement des loyers avec prise en compte des externalités

Nous estimons cette fois les externalités de l'encadrement des loyers à Paris sur les zones qui sont les plus proches. Le tableau 21 présente les résultats pour les différentes spécifications. Dans la colonne 1 par exemple, nous gardons uniquement les observations situées à moins de 8km d'une unité traitée, et nous estimons des externalités jusqu'à 6km. Cela signifie que nous définissons le contrefactuel comme étant les observations situées entre 6 et 8km.

Tableau 21 : Résultats d'estimation du modèle de différence-de-différences avec prise en compte des externalités

	Contrefactuel 6km à 8km \ Externalités de 0 à 6km	Contrefactuel 4 à 8km \ Externalités de 0 à 4km	Contrefactuel 4 à 6km \ Externalités de 0 à 4km	Contrefactuel 2 à 8km \Externalités de 0 à 2km	Contrefactuel 2 à 6km \ Externalités de 0 à 2km
Traitement (%)	-3,0%	-3,0%	-2,8%	-3,0%	-2,9%
Traitement	-0.030** (0.010)	-0.030*** (0.008)	-0.029*** (0.009)	-0.030*** (0.007)	-0.030*** (0.007)
treat_within_2	-0.011 (0.011)	-0.010 (0.09)	-0.009 (0.010)	-0.010 (0.008)	-0.010 (0.009)
treat_within_2 _4	0.000 (0.013)	0.000 (0.012)	0.003 (0.012)		
treat_within_4 _6	-0.001 (0.014)				

	Contrefactuel 6km à 8km \ Externalités de 0 à 6km	Contrefactuel 4 à 8km \ Externalités de 0 à 4km	Contrefactuel 4 à 6km \ Externalités de 0 à 4km	Contrefactuel 2 à 8km \Externalités de 0 à 2km	Contrefactuel 2 à 6km \ Externalités de 0 à 2km
Num.Obs.	14168	14168	13567	14168	13567
R2	0.747	0.747	0.735	0.747	0.735
R2 Within	0.434	0.434	0.432	0.434	0.432
FE : Trimestre	X	X	X	X	X
FE: IRIS	X	X	X	X	X
Variables de contrôle	X	X	X	X	X

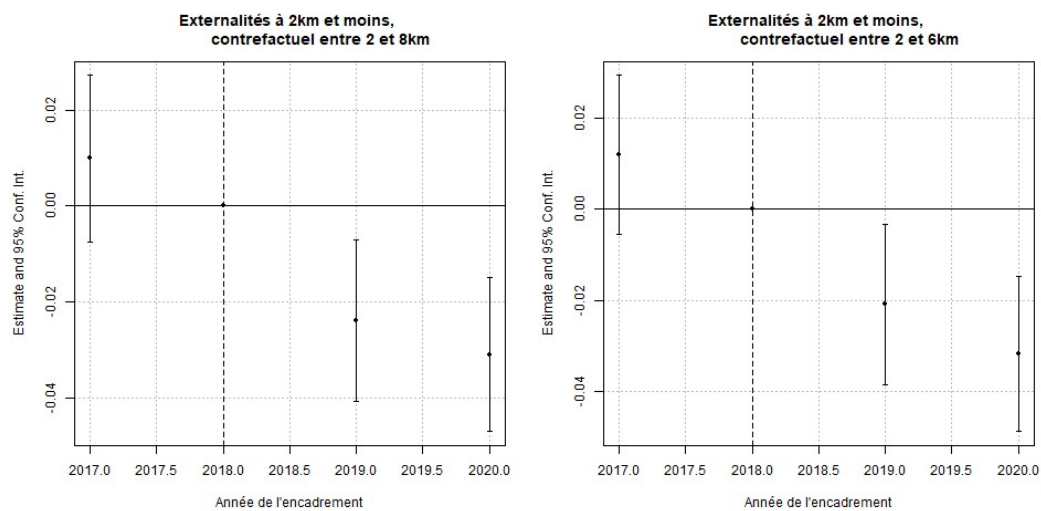
La variable *Traitement* indique l'impact de l'encadrement des loyers sur le niveau des loyers au m² exprimés en logarithme. Le niveau de significativité est indiqué par : + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 et la standard error apparait entre parenthèses. Pour interpréter le coefficient, il convient de calculer son exponentielle. Par exemple pour la colonne 1 : $\exp(-0.030)-1 = -0.02956$, soit une baisse de 2.96%. Le coefficient transformé, qui est directement interprétable, est indiqué à la ligne *Traitement (%)*.

L'effet de l'encadrement sur les loyers parisiens est significatif quels que soient les distances des externalités et les contrefactuels, avec une baisse des loyers comprise entre 2,8 et 3%, ce qui est très proche des coefficients obtenus en l'absence de prise en compte des externalités. Pour les logements situés à moins de 2km de Paris, l'effet de traitement est également négatif, mais n'est significatif dans aucune des spécifications. Cette non-significativité semble principalement attribuable à la faible valeur de l'effet estimé (-0,9%), puisque les écarts-types des coefficients associés aux externalités sont très proches de ceux associés au traitement, qui lui est significatif. Une autre explication possible est le nombre relativement faible d'observations dans la base de données.

Cette estimation repose sur deux hypothèses de tendances parallèles : entre Paris et le contrefactuel d'une part, et entre les zones potentiellement affectées par les externalités et le contrefactuel d'autre part. Afin de vérifier la plausibilité de ces hypothèses, nous procédons comme précédemment à une *event study*.

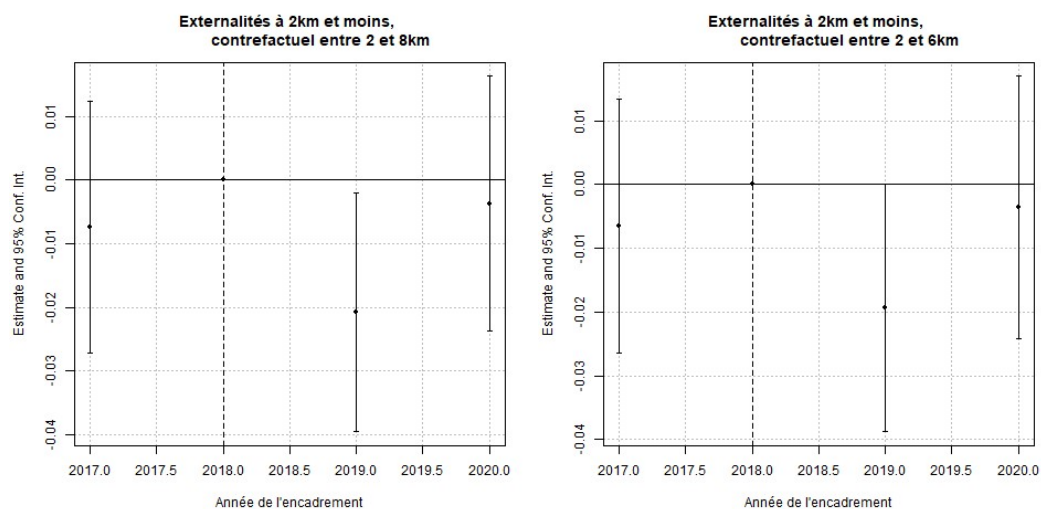
D'après les graphiques de la figure 26, il n'y a pas de pré-tendance entre le contrefactuel et Paris, l'hypothèse de tendances parallèles semble donc valide entre ces deux groupes.

Figure 26 : Evolution de l'effet de l'encadrement des loyers sur les montants des baux parisiens au cours du temps en fonction des anneaux (modèle avec externalités)



Les deux graphiques suivants (figure 27) montrent qu'il n'y a pas de pré-tendance non plus pour les deux contrefactuels. L'hypothèse de tendances parallèles est donc respectée entre le groupe affecté par les externalités et le contrefactuel. Les externalités n'ont un effet significatif sur les loyers qu'en 2019 pour le contrefactuel situé entre 2 et 8km. Toutefois, l'intervalle de confiance reste très proche de 0, et cet effet n'est plus significatif à partir de 2020. Cela signifie que l'encadrement n'a pas permis de baisser les loyers des zones non encadrées proche de Paris.

Figure 27 : Evolution de l'effet de l'encadrement des loyers sur les montants des baux parisiens des zones voisines non encadrées au cours du temps en fonction des anneaux (modèle avec externalités)



6. Conclusion

Cette étude a permis de mesurer l'impact du dispositif expérimental d'encadrement des loyers appliqué par la ville de Paris depuis juillet 2019. A cette fin, trois jeux de données complémentaires ont été mobilisés :

- 1/ les données d'annonces de location publiées sur les plateformes immobilières du groupe SeLoger entre janvier 2018 et juin 2023 sur France entière,
- 2/ les données de baux collectées entre 2017 et 2020 sur le périmètre des observatoires locaux des loyers et déposées sous le CASD
- 3/ les données brouillées des baux collectées par l'OLAP dans le cadre de sa base de « références de voisinage » pour Paris et sa petite couronne, entre janvier 2018 et mai 2021.

En termes de méthodologie, nous appliquons une méthode d'inférence causale qui isole l'effet de l'encadrement des loyers (i.e., le traitement) sur les montants des loyers pratiqués, en neutralisant les autres chocs susceptibles d'avoir affecté le marché de la location traditionnelle de façon homogène sur le territoire. Cette méthode de différence-de-différences s'appuie sur la comparaison entre deux groupes : un groupe traité constitué des logements situés à Paris et un groupe de contrôle constitué de logements situés en dehors. Plus précisément, le groupe de contrôle est soit constitué des logements de 8 villes en zone tendue qui n'ont pas appliqué de politique d'encadrement des loyers, dans le cadre de l'estimation sur données du groupe SeLoger ou données des OLL sous CASD, soit constitué de communes de petite couronne appartenant à des anneaux de distance sur la base des références de voisinage de l'OLAP. La double différence permet à la fois d'éliminer les effets fixes spécifiques à chaque groupe et les effets temporels communs aux deux groupes. Elle repose sur les deux hypothèses fondamentales de tendance parallèle d'évolution des loyers et d'absence d'externalités de l'encadrement des loyers à Paris sur les autres villes du groupe de contrôle. En raison de cette seconde hypothèse, l'analyse des données de l'OLAP sur Paris et sa petite couronne a nécessité de recourir à une méthodologie très récente qui permet d'intégrer et d'estimer les externalités dans le modèle de différence-de-différences.

L'analyse des données d'annonces du groupe SeLoger révèle que la part des annonces avec un loyer supérieur au loyer de référence majoré s'élève à 38% entre juillet 2022 et juin 2023, soit une réduction de plus de 8 points par rapport à l'année pré-introduction du dispositif. Les dépassements sont plus élevés pour les logements meublés que pour les non meublés, avec un écart de presque 20 points. Des écarts de même ampleur peuvent s'observer entre secteurs géographiques. Les écarts de dépassement sont en revanche nettement moins marqués selon le nombre de pièces et l'époque de construction. Sur ces données d'annonces, nous montrons que l'encadrement des loyers parisiens a conduit à une baisse des loyers comprises entre 3,7 et 4,2% entre juillet 2019 et juin 2023. L'efficacité de ce dispositif s'est accrue au fil des années. Si cet effet est beaucoup plus prononcé pour les petites surfaces/les studios, il est en revanche relativement homogène selon les époques de construction, le type de location (meublé/non meublé) ou le secteur géographique. Nous retrouvons un impact de l'encadrement de loyers d'ampleur moindre en estimant le modèle de différence-de-différences avec estimations des externalités sur les données OLAP de Paris et sa petite couronne, avec une baisse des loyers comprise entre 2,8 et 3,1% entre juillet 2019 et mai 2021. Ces différences de l'ordre de plus d'un point s'expliquent largement par le fait que notre période d'analyse s'arrête fin mai 2021 pour ne pas biaiser l'analyse du fait de la mise en œuvre du dispositif d'encadrement dans deux EPT voisins, alors même que la capacité du dispositif à réduire les loyers a été plus forte sur la période récente. Les résultats obtenus sur données des observatoires locaux des loyers sont quant à eux à considérer avec beaucoup de précautions en raison de quatre limites principales (impossibilité d'introduire un effet fixe à l'IRIS, faible taux de complétion des variables de contrôle, faible taille de l'échantillon et données peu récentes).

Références

- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12), 4088-4118
- Arnott, R. (1995). Time for revisionism on rent control?. *Journal of economic perspectives*, 9(1), 99-120.
- Bonneval, L., Goffette-Nagot, F., & Zhao, Z. (2022). The impact of rent control: Investigations on historical data in the city of Lyon. *Growth and Change*, 53(4), 1646-1667.
- Breidenbach, P., Eilers, L., & Fries, J. (2022). Temporal dynamics of rent regulations—The case of the German rent control. *Regional Science and Urban Economics*, 92, 103737.
- Hahn, A. M., Kholodilin, K. A., Walth, S. R., & Fongoni, M. (2023). Forward to the past: Short-term effects of the rent freeze in Berlin. *Management Science*.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396), 945-960
- Kholodilin, K. A. (2018). Measuring stick-style housing policies: A multi-country longitudinal database of governmental regulations.
- Kholodilin, K. A., López, F. A., Rey Blanco, D., & Gonzalez Arbués, P. (2022). Lessons from an aborted second-generation rent control in Catalonia.
- Kranz, S. (2021), Synthetic Difference-in-Differences with Time-Varying Covariates, mimeo
- Monràs, J., & Montalvo, J. G. (2022). *The effect of second generation rent controls: New evidence from Catalonia*. Universitat Pompeu Fabra, Department of Economics and Business.
- O'Toole, C., Martinez-Cillero, M., & Ahrens, A. (2021). Price regulation, inflation, and nominal rigidity in housing rents. *Journal of Housing Economics*, 52, 101769.
- Pastor, M., Carter, V., & Abood, M. (2018). Rent Matters: What are the Impacts of Rent Stabilization Measures?. *Los Angeles: USC Dornsife Program for Environmental and Regional Equity*.
- Segu, M., Jofre-Monseny, J., & Martínez-Mazza, R. (2022). Effectiveness and Supply Effects of High-Coverage Rent Control Policies. *Available at SSRN 4182928*.
- Sims, D. P. (2007). Out of control: What can we learn from the end of Massachusetts rent control? *Journal of Urban Economics*, 61(1), 129-151.
- Thomschke, L. (2019). Regional impact of the German rent brake. *German Economic Review*, 20(4), e892-e912.

Annexes

ANNEXES PARTIE 3

A.3.1 : STATISTIQUES DESCRIPTIVES PARIS VS CONTREFACTUEL

		Groupe de contrôle (N = 275079) / Moyenne	Groupe de contrôle (N = 275079) / Ecart-type	Paris (N = 284221) / Moyenne	Paris (N = 284221) / Ecart-type
Surface		22.7	6.0	22.2	6.9
Surface		48.9	24.8	48.0	32.8
Loyer au m2		15.0	5.3	32.5	8.6
		N	Pct.	N	Pct.
Nombre de pièces	1	77017	28.0	95278	33.5
	2	99330	36.1	105553	37.1
	3	70523	25.6	49662	17.5
	4	22443	8.2	19581	6.9
	5p	5766	2.1	14147	5.0
Meublé	meuble	63571	23.1	119571	42.1
	meuble_na	20352	7.4	17617	6.2
	non_meuble	191156	69.5	147033	51.7
Étage	0	36884	13.4	25356	8.9
	1	60217	21.9	44327	15.6
	2	54108	19.7	43711	15.4
	3	44184	16.1	41854	14.7
	4	29604	10.8	38037	13.4
	5	15912	5.8	32008	11.3
	6_8	14903	5.4	44807	15.8

		Groupe de contrôle (N = 275079) / Moyenne	de Groupe de contrôle (N = 275079) / Ecart-type	Paris (N = 284221) / Moyenne	Paris (N = 284221) / Ecart-type
	9_12	2548	0.9	3851	1.4
	13p	853	0.3	830	0.3
	Missing	15866	5.8	9440	3.3
Nombre de salles de bain	0	74752	27.2	53493	18.8
	1	139005	50.5	126208	44.4
	2	2384	0.9	4739	1.7
	3p	264	0.1	529	0.2
	Missing	58674	21.3	99252	34.9
Nombre de toilettes	0	33817	12.3	23968	8.4
	1	193555	70.4	218884	77.0
	2	5138	1.9	15204	5.3
	3p	340	0.1	2243	0.8
	Missing	42229	15.4	23922	8.4
Nombre de balcons	0	91399	33.2	73265	25.8
	1	78036	28.4	38457	13.5
	2	5408	2.0	3338	1.2
	3p	505	0.2	735	0.3
	Missing	99731	36.3	168426	59.3
Nombre de terrasses	0	52262	19.0	37392	13.2
	1	59270	21.5	11121	3.9
	2	1863	0.7	511	0.2
	3p	174	0.1	107	0.0
	Missing	161510	58.7	235090	82.7
Cave	False	90950	33.1	103562	36.4

		Groupe de contrôle (N = 275079) / Moyenne	Groupe de contrôle (N = 275079) / Ecart-type	Paris (N = 284221) / Moyenne	Paris (N = 284221) / Ecart-type
	Missing	131865	47.9	116543	41.0
	True	52264	19.0	64116	22.6
Jardin	False	53	0.0	79	0.0
	Missing	275012	100.0	284112	100.0
	True	14	0.0	30	0.0
Piscine	False	263434	95.8	279349	98.3
	Missing	4394	1.6	4263	1.5
	True	7251	2.6	609	0.2
Duplex	False	265798	96.6	277444	97.6
	True	9281	3.4	6777	2.4
Période de construction	avant_1945	20501	7.5	150618	53.0
	1946-1970	22012	8.0	26368	9.3
	1971-1990	14899	5.4	17082	6.0
	apres_1990	48963	17.8	11764	4.1
	Missing	168704	61.3	78389	27.6

A.3.2 : TABLEAU DE REGRESSION AVEC VARIABLES DE CONTROLE

	Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
Traitement (%)	-3,7%	-4,2%
nb_pieces2	0.034** (0.007)	
nb_pieces3	0.075*** (0.013)	
nb_pieces4	0.119*** (0.017)	
nb_pieces5p	0.180*** (0.020)	
log_surface	-0.365*** (0.063)	-0.383*** (0.067)
meublemeuble_na	-0.152*** (0.007)	
meublenon_meuble	-0.146*** (0.008)	
etage1	0.017*** (0.003)	0.015*** (0.003)
etage2	0.026*** (0.004)	0.024*** (0.004)
etage3	0.024*** (0.004)	0.022*** (0.004)
etage4	0.024** (0.005)	0.021** (0.004)

	Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
etage5	0.027*** (0.004)	0.022*** (0.004)
etage6_8	0.007 (0.008)	0.014* (0.005)
etage9_12	0.017* (0.007)	0.016* (0.006)
etage13p	0.006 (0.011)	0.009 (0.008)
etageMissing	0.046*** (0.003)	0.044*** (0.003)
nb_salles_de_bain1	-0.018** (0.004)	-0.017** (0.005)
nb_salles_de_bain2	0.004 (0.021)	0.000 (0.021)
nb_salles_de_bain3 p	-0.057 (0.087)	-0.051 (0.080)
nb_salles_de_bain Missing	-0.005 (0.004)	0.000 (0.003)
nb_toilettes1	-0.008 (0.007)	-0.008 (0.008)
nb_toilettes2	0.142** (0.033)	0.108** (0.022)
nb_toilettes3p	0.279** (0.078)	0.224** (0.049)

	Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
nb_toilettesMissing	-0.005 (0.007)	-0.007 (0.008)
nb_balcons1	-0.004 (0.014)	0.001 (0.014)
nb_balcons2	-0.005 (0.016)	0.008 (0.016)
nb_balcons3p	0.033+ (0.016)	0.035* (0.014)
nb_balconsMissing	0.000 (0.009)	0.001 (0.010)
nb_terrasses1	0.033+ (0.016)	0.044* (0.016)
nb_terrasses2	0.072* (0.024)	0.104** (0.026)
nb_terrasses3p	0.086* (0.028)	0.119** (0.032)
nb_terrassesMissin g	-0.026*** (0.003)	-0.025*** (0.002)
si_caveMissing	-0.001 (0.004)	-0.003 (0.004)
si_caveTrue	0.008+ (0.004)	0.004 (0.002)
si_jardinMissing	0.008 (0.019)	0.005 (0.021)
si_jardinTrue	0.047***	0.051***

	Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
	(0.007)	(0.005)
si_piscineMissing	-0.017***	-0.017***
	(0.003)	(0.003)
si_piscineTrue	0.035	0.041
	(0.021)	(0.022)
si_duplexTrue	0.037**	0.041**
	(0.008)	(0.009)
construct_cat1946- 1970	-0.030***	
	(0.005)	
construct_cat1971- 1990	-0.021***	
	(0.002)	
construct_catapres_ 1990	-0.014	
	(0.012)	
construct_catMissing	-0.010**	
	(0.003)	
treatment	-0.038***	-0.043***
	(0.005)	(0.006)
nb_pieces = 1 × Group = Control_group		0.016
		(0.070)
nb_pieces = 1 × Group = Paris		-0.191*
		(0.074)
nb_pieces = 2 × Group = Control_group		-0.008

Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
	(0.030)
nb_pieces = 2 × Group = Paris	-0.104**
	(0.029)
nb_pieces = 4 × Group = Control_group	0.038
	(0.023)
nb_pieces = 4 × Group = Paris	0.098**
	(0.024)
nb_pieces = 5p × Group = Control_group	0.039
	(0.064)
nb_pieces = 5p × Group = Paris	0.194*
	(0.061)
construct_cat = avant_1945 × Group = Control_group	0.053*
	(0.016)
construct_cat = avant_1945 × Group = Paris	0.011
	(0.014)
construct_cat = 1971-1990 × Group = Control_group	0.014
	(0.008)
construct_cat = 1971-1990 × Group = Paris	0.008

Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
	(0.005)
contract_cat = apres_1990 × Group = Control_group	0.033**
	(0.009)
contract_cat = apres_1990 × Group = Paris	0.050***
	(0.004)
contract_cat = Missing × Group = Control_group	0.041**
	(0.010)
contract_cat = Missing × Group = Paris	0.015
	(0.010)
nb_pieces = 1 × fact_yearjul = 2018	-0.005*
	(0.002)
nb_pieces = 1 × fact_yearjul = 2019	-0.008
	(0.005)
nb_pieces = 1 × fact_yearjul = 2020	-0.007
	(0.007)
nb_pieces = 1 × fact_yearjul = 2021	-0.013
	(0.008)
nb_pieces = 1 × fact_yearjul = 2022	-0.009
	(0.011)

Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
nb_pieces = 2 × fact_yearjul = 2018	0.000 (0.002)
nb_pieces = 2 × fact_yearjul = 2019	-0.001 (0.001)
nb_pieces = 2 × fact_yearjul = 2020	-0.001 (0.002)
nb_pieces = 2 × fact_yearjul = 2021	-0.005* (0.002)
nb_pieces = 2 × fact_yearjul = 2022	-0.001 (0.002)
nb_pieces = 4 × fact_yearjul = 2018	-0.003 (0.004)
nb_pieces = 4 × fact_yearjul = 2019	-0.005 (0.004)
nb_pieces = 4 × fact_yearjul = 2020	-0.018** (0.005)
nb_pieces = 4 × fact_yearjul = 2021	-0.022* (0.010)
nb_pieces = 4 × fact_yearjul = 2022	-0.037+ (0.017)

Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
nb_pieces = 5p × fact_yearjul = 2018	-0.010 (0.011)
nb_pieces = 5p × fact_yearjul = 2019	0.004 (0.005)
nb_pieces = 5p × fact_yearjul = 2020	-0.053 (0.030)
nb_pieces = 5p × fact_yearjul = 2021	-0.058 (0.038)
nb_pieces = 5p × fact_yearjul = 2022	-0.081 (0.059)
construct_cat = avant_1945 × fact_yearjul = 2018	0.005 (0.003)
construct_cat = avant_1945 × fact_yearjul = 2019	0.011 (0.008)
construct_cat = avant_1945 × fact_yearjul = 2020	0.002 (0.005)
construct_cat = avant_1945 × fact_yearjul = 2021	-0.006 (0.004)

	Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
construct_cat avant_1945 fact_yearjul = 2022	= ×	-0.011* (0.003)
construct_cat 1971-1990 fact_yearjul = 2018	= ×	0.000 (0.002)
construct_cat 1971-1990 fact_yearjul = 2019	= ×	-0.003 (0.005)
construct_cat 1971-1990 fact_yearjul = 2020	= ×	-0.001 (0.004)
construct_cat 1971-1990 fact_yearjul = 2021	= ×	-0.009 (0.005)
construct_cat 1971-1990 fact_yearjul = 2022	= ×	-0.017* (0.007)
construct_cat apres_1990 fact_yearjul = 2018	= ×	-0.011** (0.003)
construct_cat apres_1990 fact_yearjul = 2019	= ×	-0.019** (0.005)

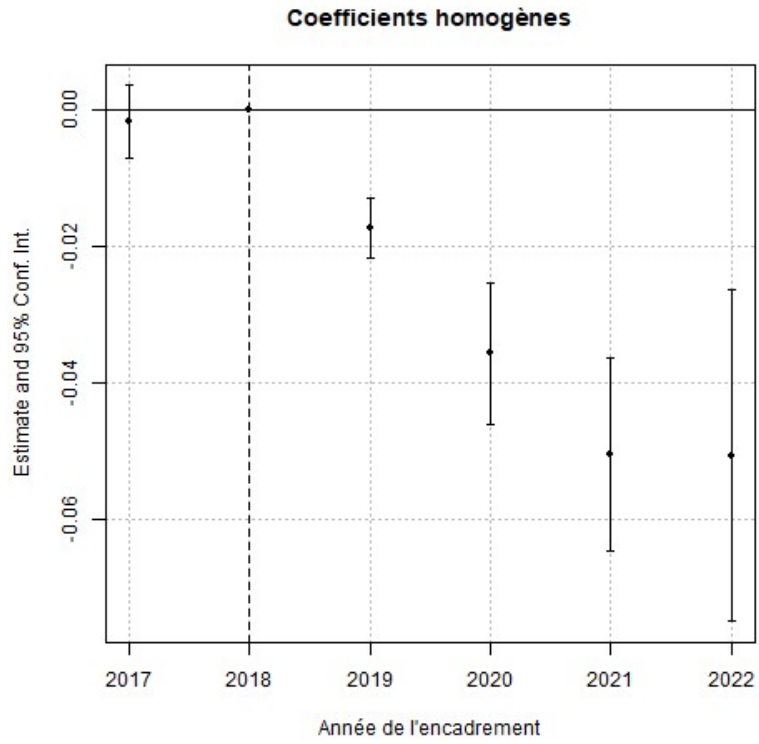
	Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
contract_cat = apres_1990 × fact_yearjul = 2020		-0.019** (0.004)
contract_cat = apres_1990 × fact_yearjul = 2021		-0.029*** (0.005)
contract_cat = apres_1990 × fact_yearjul = 2022		-0.035** (0.009)
contract_cat = Missing × fact_yearjul = 2018		-0.002 (0.003)
contract_cat = Missing × fact_yearjul = 2019		-0.015+ (0.007)
contract_cat = Missing × fact_yearjul = 2020		-0.015** (0.004)
contract_cat = Missing × fact_yearjul = 2021		-0.019** (0.005)
contract_cat = Missing × fact_yearjul = 2022		-0.027** (0.008)

Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
meuble = meuble × Group = Control_group	0.127*** (0.017)
meuble = meuble × Group = Paris	0.113*** (0.011)
meuble = non_meuble × Group = Control_group	-0.017 (0.011)
meuble = non_meuble × Group = Paris	-0.040** (0.011)
meuble = meuble × fact_yearjul = 2018	0.033+ (0.017)
meuble = meuble × fact_yearjul = 2019	0.059* (0.020)
meuble = meuble × fact_yearjul = 2020	0.029 (0.021)
meuble = meuble × fact_yearjul = 2021	0.023 (0.017)
meuble = meuble × fact_yearjul = 2022	0.022 (0.028)

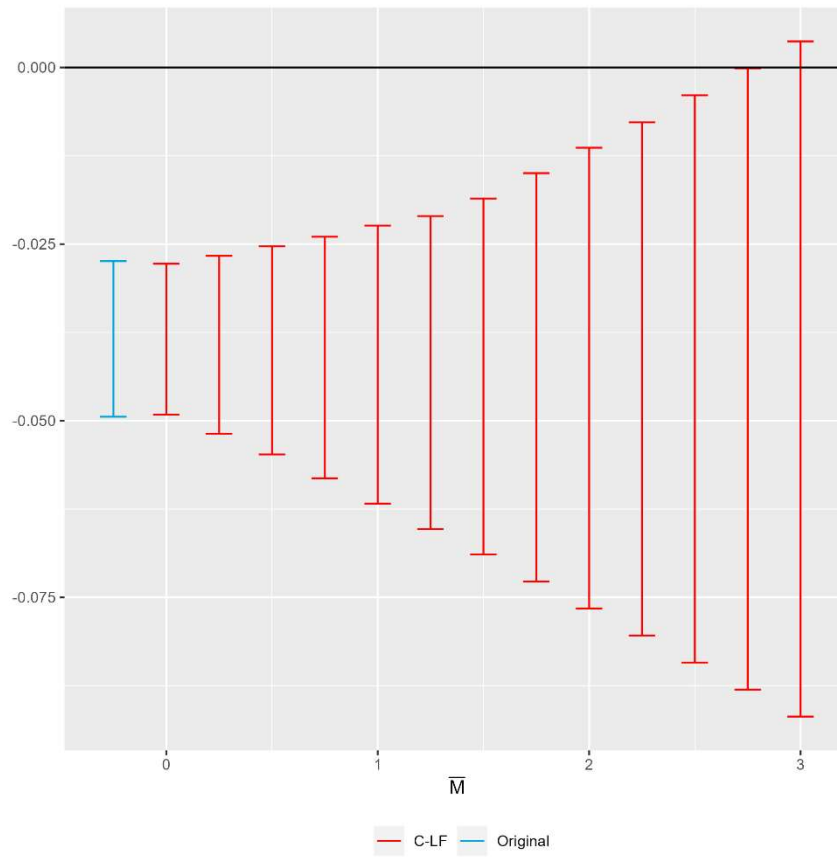
	Coefficients homogènes	Coefficients hétérogènes par groupe traité/contrôle et effet fixe temps pour chaque catégorie de beta hétérogène
meuble =		
non_meuble ×		0.040+
fact_yearjul = 2018		(0.017)
meuble =		
non_meuble ×		0.048+
fact_yearjul = 2019		(0.021)
meuble =		
non_meuble ×		0.031
fact_yearjul = 2020		(0.017)
meuble =		
non_meuble ×		0.029+
fact_yearjul = 2021		(0.014)
meuble =		
non_meuble ×		0.037*
fact_yearjul = 2022		(0.015)
Num.Obs.	559300	559300
R2	0.870	0.879
R2 Within	0.510	0.545
Std.Errors	by: city_title	by: city_title
FE: quarterdate	X	X
FE: area_fe	X	X

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.3.3 : EVENT STUDY : COEFFICIENTS HOMOGÈNES



A.3.4 : ROBUSTESSE A LA VIOLATION DE LA PTA : COEFFICIENTS HOMOGENES



A.3.5 : TABLEAU DE REGRESSION : HETEROGENEITE PAR PERIODE

	Coefficients homogènes	Coefficients hétérogènes
Traitement 2017 (%)	-0.2%	-0.3%
Traitement 2019 (%)	-1,7%	-02,5%
Traitement 2020 (%)	-3,5%	-3,5%
Traitement 2021 (%)	-4,9%	-5,3%
Traitement 2022 (%)	-4,9%	-5,9%
year_jul = 2017 × D	-0.002 (0.003)	-0.003 (0.003)
year_jul = 2019 × D	-0.017*** (0.002)	-0.025*** (0.003)
year_jul = 2020 × D	-0.036*** (0.005)	-0.036*** (0.005)
year_jul = 2021 × D	-0.050*** (0.007)	-0.055*** (0.008)
year_jul = 2022 × D	-0.051** (0.012)	-0.061** (0.013)
Num.Obs.	559300	559300
R2	0.870	0.880
R2 Within	0.510	0.545
Std.Errors	by: city_title	by: city_title
FE: quarterdate	X	X
FE: area_fe	X	X
Variables de contrôle	X	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

A.3.6 : TABLEAU DE REGRESSION : HETEROGENEITE PAR NOMBRE DE PIECES

	Coefficients hétérogènes
Traitement 1 pièce (%)	-6,2%
Traitement 2 pièces (%)	-4,1%
Traitement 3 pièces (%)	-3,8%
Traitement 4 pièces (%)	-1,8%
Traitement 5 pièces et plus (%)	6,9%
nb_pieces = 1 × traitement	-0.064***
	(0.006)
nb_pieces = 2 × traitement	-0.041***
	(0.005)
nb_pieces = 3 × traitement	-0.039***
	(0.005)
nb_pieces = 4 × traitement	-0.018
	(0.012)
nb_pieces = 5p × traitement	0.067+
	(0.032)
Num.Obs.	559300
R2	0.880
R2 Within	0.545
Std.Errors	by: city_title
FE: quarterdate	X
FE: area_fe	X
Variables de contrôle	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

A.3.7 : TABLEAU DE REGRESSION : HETEROGENEITE PAR SURFACE

	Coefficients hétérogènes
Traitement 8 à 18m2 (%)	-10,2%
Traitement 18 à 24m2 (%)	-6,3%
Traitement 24 à 40m2 (%)	-4,9%
Traitement 40 à 60m2 (%)	-3,5%
Traitement 60 à 80m2 (%)	-2,4%
Traitement 80 à 100m2 (%)	-2,2%
Traitement 100m2 et plus (%)	-2,6%
surface_cats2 = 8_18 × traitement	-0.107*** (0.011)
surface_cats2 = 18_24 × traitement	-0.065*** (0.005)
surface_cats2 = 24_40 × traitement	-0.050*** (0.005)
surface_cats2 = 40_60 × traitement	-0.036*** (0.003)
surface_cats2 = 60_80 × traitement	-0.025** (0.007)
surface_cats2 = 80_100 × traitement	-0.022 (0.013)
surface_cats2 = 100p × traitement	-0.026 (0.016)
Num.Obs.	559300
R2	0.879
R2 Within	0.542
Std.Errors	by: city_title
FE: quarterdate	X
FE: area_fe	X
Variables de contrôle	X

Coefficients hétérogènes

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

A.3.8 : TABLEAU DE REGRESSION : HETEROGENEITE PAR STATUT MEUBLE

Coefficients hétérogènes

Traitement Meublé (%)	-3,3%
Traitement Non meublé (%)	-4,4%
meuble = meuble × traitement	-0.033** (0.009)
meuble = meuble_na × traitement	-0.062*** (0.008)
meuble = non_meuble × traitement	-0.045*** (0.006)
Num.Obs.	559300
R2	0.879
R2 Within	0.545
Std.Errors	by: city_title
FE: quarterdate	X
FE: area_fe	X
Variables de contrôle	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

A.3.9 : TABLEAU DE REGRESSION : HETEROGENEITE PAR PERIODE DE CONSTRUCTION

	Coefficients hétérogènes
Traitement Avant 1945 (%)	-5,0%
Traitement Entre 1946 et 1970 (%)	-4,2%
Traitement Entre 1971 et 1990 (%)	-3,4%
Traitement Après 1990 (%)	-3,0%
construct_cat = avant_1945 × traitement	-0.051** (0.011)
construct_cat = 1946-1970 × traitement	-0.043*** (0.006)
construct_cat = 1971-1990 × traitement	-0.034*** (0.007)
construct_cat = apres_1990 × traitement	-0.031** (0.008)
construct_cat = Missing × traitement	-0.043*** (0.005)
Num.Obs.	559300
R2	0.879
R2 Within	0.545
Std.Errors	by: city_title
FE: quarterdate	X
FE: area_fe	X
Variables de contrôle	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

A.3.10 : TABLEAU DE REGRESSION : HETEROGENEITE PAR ARRONDISSEMENT

	Coefficients hétérogènes
Traitement Arrondissement 1 (%)	-4,5%
Traitement Arrondissement 2 (%)	-4,5%
Traitement Arrondissement 3 (%)	-3,9%
Traitement Arrondissement 4 (%)	-6,0%
Traitement Arrondissement 5 (%)	-4,8%
Traitement Arrondissement 6 (%)	-3,2%
Traitement Arrondissement 7 (%)	-3,3%
Traitement Arrondissement 8 (%)	-3,2%
Traitement Arrondissement 9 (%)	-4,0%
Traitement Arrondissement 10 (%)	-4,6%
Traitement Arrondissement 11 (%)	-4,6%
Traitement Arrondissement 12 (%)	-4,6%
Traitement Arrondissement 13 (%)	-4,8%
Traitement Arrondissement 14 (%)	-4,0%
Traitement Arrondissement 15 (%)	-3,9%
Traitement Arrondissement 16 (%)	-3,4%
Traitement Arrondissement 17 (%)	-3,7%
Traitement Arrondissement 18 (%)	-5,2%
Traitement Arrondissement 19 (%)	-4,9%
Traitement Arrondissement 20 (%)	-5,0%
arrond_agent = 01 × traitement	-0.046***
	(0.006)
arrond_agent = 02 × traitement	-0.046***
	(0.005)
arrond_agent = 03 × traitement	-0.040***
	(0.006)
arrond_agent = 04 × traitement	-0.062***
	(0.006)

	Coefficients hétérogènes
arrond_agent = 05 × traitement	-0.049*** (0.006)
arrond_agent = 06 × traitement	-0.032*** (0.006)
arrond_agent = 07 × traitement	-0.033*** (0.006)
arrond_agent = 08 × traitement	-0.033** (0.007)
arrond_agent = 09 × traitement	-0.041*** (0.007)
arrond_agent = 10 × traitement	-0.047*** (0.006)
arrond_agent = 11 × traitement	-0.047*** (0.005)
arrond_agent = 12 × traitement	-0.047*** (0.005)
arrond_agent = 13 × traitement	-0.050*** (0.005)
arrond_agent = 14 × traitement	-0.041*** (0.005)
arrond_agent = 15 × traitement	-0.040*** (0.006)
arrond_agent = 16 × traitement	-0.035*** (0.007)
arrond_agent = 17 × traitement	-0.037*** (0.006)
arrond_agent = 18 × traitement	-0.054*** (0.005)

	Coefficients hétérogènes
arrond_agent = 19 × traitement	-0.050*** (0.005)
arrond_agent = 20 × traitement	-0.052*** (0.005)
arrond_agent = NA × traitement	-0.041** (0.011)
Num.Obs.	559300
R2	0.879
R2 Within	0.545
Std.Errors	by: city_title
FE: quarterdate	X
FE: area_fe	X
Variables de contrôle	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

A.3.11 : TABLEAU DE REGRESSION : HETEROGENEITE PAR SECTEUR

	Coefficients hétérogènes
Traitement Secteur 1 (%)	-3,2%
Traitement Secteur 2 (%)	-0,4%
Traitement Secteur 3 (%)	-1,5%
Traitement Secteur 4 (%)	-2,8%
Traitement Secteur 5 (%)	-3,9%
Traitement Secteur 6 (%)	-1,3%
Traitement Secteur 7 (%)	-4,3%
Traitement Secteur 8 (%)	-5,0%
Traitement Secteur 9 (%)	-2,0%
Traitement Secteur 10 (%)	-4,2%

	Coefficients hétérogènes
Traitement Secteur 11 (%)	-4,1%
Traitement Secteur 12 (%)	-2,6%
Traitement Secteur 13 (%)	-5,1%
Traitement Secteur 14 (%)	-3,2%
sect_encad = 1 × traitement	-0.032***
	(0.006)
sect_encad = 2 × traitement	-0.004
	(0.006)
sect_encad = 3 × traitement	-0.015*
	(0.006)
sect_encad = 4 × traitement	-0.028**
	(0.006)
sect_encad = 5 × traitement	-0.039***
	(0.006)
sect_encad = 6 × traitement	-0.013+
	(0.006)
sect_encad = 7 × traitement	-0.044***
	(0.006)
sect_encad = 8 × traitement	-0.051***
	(0.006)
sect_encad = 9 × traitement	-0.021**
	(0.006)
sect_encad = 10 × traitement	-0.043***
	(0.006)
sect_encad = 11 × traitement	-0.042***
	(0.006)
sect_encad = 12 × traitement	-0.027**
	(0.007)

	Coefficients hétérogènes
sect_encad = 13 × traitement	-0.052*** (0.005)
sect_encad = 14 × traitement	-0.033*** (0.006)
sect_encad = Missing × traitement	-0.055*** (0.006)
Num.Obs.	559300
R2	0.880
R2 Within	0.546
Std.Errors	by: city_title
FE: quarterdate	X
FE: area_fe	X
Variables de contrôle	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

A.3.12 : TABLEAU DE REGRESSION : HETEROGENEITE PAR NIVEAU DE CONTRAINTE

	Coefficients	hétérogènes
Traitement quantile [-0.406,-0.0604] (%)	8,9%	
Traitement quantile (-0.0604,-0.0446] (%)	4,5%	
Traitement quantile (-0.0446,-0.0345] (%)	2,9%	
Traitement quantile (-0.0345,-0.0266] (%)	1,9%	
Traitement quantile (-0.0266,-0.02] (%)	0,7%	
Traitement quantile (-0.02,-0.014] (%)	-0,1%	
Traitement quantile (-0.014,-0.00864] (%)	-0,9%	
Traitement quantile (-0.00864,-0.00353] (%)	-1,5%	
Traitement quantile (-0.00353,0.00152] (%)	-2,6%	
Traitement quantile (0.00152,0.0065] (%)	-3,1%	
Traitement quantile (0.0065,0.0114] (%)	-3,8%	
Traitement quantile (0.0114,0.0165] (%)	-4,8%	
Traitement quantile (0.0165,0.0218] (%)	-5,6%	
Traitement quantile (0.0218,0.0274] (%)	-6,5%	
Traitement quantile (0.0274,0.0335] (%)	-7,1%	
Traitement quantile (0.0335,0.0405] (%)	-8,2%	
Traitement quantile (0.0405,0.0487] (%)	-9,5%	
Traitement quantile (0.0487,0.0593] (%)	-10,8%	
Traitement quantile (0.0593,0.0762] (%)	-12,5%	
Traitement quantile (0.0762,0.201] (%)	-16,7%	
cons_het_20d = [-0.406,-0.0604] × traitement	0.085*	(0.032)
cons_het_20d = (-0.0604,-0.0446] × traitement	0.044+	(0.021)
cons_het_20d = (-0.0446,-0.0345] × traitement	0.029	(0.018)

	Coefficients	hétérogènes
cons_het_20d = (-0.0345,-0.0266] × traitement	0.019	(0.015)
cons_het_20d = (-0.0266,-0.02] × traitement	0.007	(0.013)
cons_het_20d = (-0.02,-0.014] × traitement	-0.001	(0.012)
cons_het_20d = (-0.014,-0.00864] × traitement	-0.009	(0.010)
cons_het_20d = (-0.00864,-0.00353] × traitement	-0.015	(0.009)
cons_het_20d = (-0.00353,0.00152] × traitement	-0.026*	(0.008)
cons_het_20d = (0.00152,0.0065] × traitement	-0.032**	(0.008)
cons_het_20d = (0.0065,0.0114] × traitement	-0.039***	(0.007)
cons_het_20d = (0.0114,0.0165] × traitement	-0.050***	(0.007)
cons_het_20d = (0.0165,0.0218] × traitement	-0.058***	(0.008)
cons_het_20d = (0.0218,0.0274] × traitement	-0.067***	(0.009)
cons_het_20d = (0.0274,0.0335] × traitement	-0.074***	(0.010)
cons_het_20d = (0.0335,0.0405] × traitement	-0.086***	(0.012)
cons_het_20d = (0.0405,0.0487] × traitement	-0.100***	(0.014)

	Coefficients	hétérogènes
cons_het_20d = (0.0487,0.0593] × traitement	-0.114***	
	(0.017)	
cons_het_20d = (0.0593,0.0762] × traitement	-0.133***	
	(0.022)	
cons_het_20d = (0.0762,0.201] × traitement	-0.182***	
	(0.033)	
Num.Obs.	473411	
R2	0.884	
R2 Within	0.575	
Std.Errors	by: city_title	
FE: quarterdate	X	
FE: area_fe	X	
Variables de contrôle	X	

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

A.3.13 : METHODE : DIFFERENCE-DE-DIFFERENCES SYNTHETIQUE

L'objectif de cette annexe est de tester la robustesse de nos estimations à l'hypothèse de tendances parallèles. Pour ce faire, nous utilisons l'estimateur de différence-de-différences synthétique de Arkhangelsky et al. (2021), développé pour des données de panel et que nous adaptons pour les données en coupes transversales répétées.

La méthode de différence-de-différences synthétique combine les caractéristiques de la méthode de la différence-de-différences et de celle du contrôle synthétique. D'une part, comme dans le cas du contrôle synthétique, nous estimons des poids afin de faire correspondre la tendance de la variable expliquée entre le groupe traité et le groupe de contrôle en période de pré-traitement. D'autre part, comme dans les méthodes de différence-de-différences, nous pouvons contrôler des effets fixes individuels et ajouter des variables de contrôle dans le modèle. La méthode de différence-de-différences synthétique se met en œuvre en deux temps. Dans un premier temps, nous calculons les poids pour chaque observation dans l'échantillon, puis dans un 2^{ème} temps nous estimons une régression en différence-de-différences pondérée par les poids calculés dans l'étape précédente.

Dans cette note méthodologique, nous commençons par présenter le calcul des poids pour des données de panel (Arkhangelsky et al., 2021), nous décrivons ensuite la méthode pour inclure des variables de contrôle dans le modèle (Kranz, 2022), et enfin nous présentons la méthode que nous proposons pour utiliser ce modèle sur des données en coupes transversales répétées. La méthode est appliquée aux données SeLoger.

Calcul des poids

Nous considérons un panel comportant N unités observées pendant T périodes. Y_{it} est la variable expliquée pour l'unité i en t . W_{it} est une variable binaire qui représente l'exposition au traitement. L'échantillon est divisé en N_{co} unités de contrôle qui ne sont pas exposées au traitement et $N_{tr} = N - N_{co}$ unités qui sont exposées au traitement à partir de la période $t \geq T_{post}$.

Le premier type de poids que nous estimons a pour objectif de faire correspondre la tendance de la variable expliquée entre le groupe traité et le groupe de contrôle. Nous calculons ces poids de la façon suivante :

$$\begin{aligned} (\hat{\omega}_0, \hat{\omega}^{sdid}) &= \arg \min_{\omega_0 \in \mathbb{R}, \omega \in \Omega} \ell_{unit}(\omega_0, \omega) \quad \text{where} \\ \ell_{unit}(\omega_0, \omega) &= \sum_{t=1}^{T_{pre}} \left(\omega_0 + \sum_{i=1}^{N_{co}} \omega_i Y_{it} - \frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N Y_{it} \right)^2 + \zeta^2 T_{pre} \|\omega\|_2^2, \\ \Omega &= \left\{ \omega \in \mathbb{R}_+^N : \sum_{i=1}^{N_{co}} \omega_i = 1, \omega_i = N_{tr}^{-1} \text{ for all } i = N_{co} + 1, \dots, N \right\}, \end{aligned}$$

où le paramètre de régularisation, qui a pour objectif d'augmenter la dispersion des poids et d'assurer leur unicité, prend la forme :

$$\begin{aligned} \zeta &= (N_{tr} T_{post})^{1/4} \hat{\sigma} \quad \text{with} \quad \hat{\sigma}^2 = \frac{1}{N_{co}(T_{pre} - 1)} \sum_{i=1}^{N_{co}} \sum_{t=1}^{T_{pre}-1} (\Delta_{it} - \bar{\Delta})^2, \\ \text{where } \Delta_{it} &= Y_{i(t+1)} - Y_{it}, \quad \text{and} \quad \bar{\Delta} = \frac{1}{N_{co}(T_{pre} - 1)} \sum_{i=1}^{N_{co}} \sum_{t=1}^{T_{pre}-1} \Delta_{it}. \end{aligned}$$

Le second type de poids que nous calculons a pour objectif de pondérer davantage les périodes qui sont les plus similaires dans le groupe de contrôle en période de pré-traitement par rapport à la période de post-traitement. Ils sont calculés tel que :

$$\begin{aligned} (\hat{\lambda}_0, \hat{\lambda}^{sdid}) &= \arg \min_{\lambda_0 \in \mathbb{R}, \lambda \in \Lambda} \ell_{time}(\lambda_0, \lambda) \quad \text{where} \\ \ell_{time}(\lambda_0, \lambda) &= \sum_{i=1}^{N_{co}} \left(\lambda_0 + \sum_{t=1}^{T_{pre}} \lambda_t Y_{it} - \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^T Y_{it} \right)^2 \zeta^2 N_{co} \|\lambda\|^2, \\ \Lambda &= \left\{ \lambda \in \mathbb{R}_+^T : \sum_{t=1}^{T_{pre}} \lambda_t = 1, \lambda_t = T_{post}^{-1} \text{ for all } t = T_{pre} + 1, \dots, T \right\}. \end{aligned}$$

Le paramètre de régularisation est :

$$\zeta = 10^{-6} \hat{\sigma} \quad \text{with} \quad \hat{\sigma}^2 = \frac{1}{N_{co}(T_{pre} - 1)} \sum_{i=1}^{N_{co}} \sum_{t=1}^{T_{pre}-1} (\Delta_{it} - \bar{\Delta})^2,$$

$$\text{where } \Delta_{it} = Y_{i(t+1)} - Y_{it}, \quad \text{and} \quad \bar{\Delta} = \frac{1}{N_{co}(T_{pre} - 1)} \sum_{i=1}^{N_{co}} \sum_{t=1}^{T_{pre}-1} \Delta_{it}.$$

Une fois ces poids calculés, nous estimons une régression en différence-de-différences pondérée avec ces poids tel que :

$$\left(\hat{\tau}^{sdid}, \hat{\mu}, \hat{\alpha}, \hat{\beta} \right) = \arg \min_{\tau, \mu, \alpha, \beta} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau)^2 \hat{\omega}_i^{sdid} \hat{\lambda}_t^{sdid} \right\}$$

où μ est une constante, α_i un effet fixe individuel, β_t un effet fixe individuel et τ est l'effet causal de l'exposition au traitement.

Ajout de variables de contrôle

Afin d'ajouter des variables de contrôle dans l'analyse, nous utilisons la méthode de Krantz (2022). Pour cela, nous purgeons la variable expliquée Y_{it} de l'effet des variables de contrôle.

Nous commençons par estimer la régression suivante sur les observations non traitées, qui incluent les observations du groupe de contrôle et celles du groupe traité en période de pré-traitement) :

$$Y_{it} = X_{it}\theta + \alpha_i + \beta_t + \varepsilon_{it}$$

Ensuite, nous calculons la variable dépendante ajustée pour l'effet des variables de contrôle tel que :

$$Y_{it}^{adj} = Y_{it} - X_{it}\hat{\theta}$$

Une fois que Y_{it}^{adj} est calculé, nous appliquons la procédure décrit précédemment pour calculer les poids de la différence-de-différences synthétique.

Adaptation pour données en coupes transversales répétées

La méthode que nous avons présentée est valable pour des données de panel, nous proposons donc une adaptation de celle-ci pour les données en coupes transversales répétées. Nous considérons un ensemble de données composé de T coupes transversales indépendantes. Chacune de ces coupes transversales t est un échantillon aléatoire de la population mère. Cette population est composée de K groupes dont les premiers K^{co} sont le groupe de contrôle et les derniers $K^{tr} = K - K^{co}$ sont les groupes traités. Comme il s'agit de données en coupe transversale, les individus de ces groupes ne sont pas observés à chaque période t . L'indice associé à un individu dépend donc de la période d'observation tel que $i(k,t)$ est un individu observé en t dans le groupe k . La variable expliquée est notée $Y_{i(k,t)t}$. Le nombre d'observations peut varier à chaque période, on note ainsi N_{kt} le nombre d'observations dans le groupe k à la période t .

Nous considérons que le traitement est assigné aux groupes K^{tr} à partir de la période $t \geq T_{post}$. On utilise la méthode de la différence-de-différences synthétique pour sur les données agrégées par groupe afin de calculer les poids à appliquer. Les données sont agrégées pour chaque groupe k et chaque période t tel que :

$$\bar{Y}_{k,t} = \frac{1}{N_{k,t}} \sum_i Y_{i(k,t),t}$$

où \bar{Y}_{kt} est la moyenne de la variable Y du groupe k à la période t .

Une fois cette agrégation effectuée, on peut calculer les poids comme présenté précédemment. Afin d'ajouter des variables de contrôle dans le modèle, on commence par purger Y de l'effet des variables de contrôle (Krantz, 2022), et ensuite nous agrégeons les données par groupe. Nous obtenons ainsi les poids ω^{did} et λ^{did} . Cependant, on ne peut les utiliser directement pour estimer le modèle puisque le nombre d'observations par groupe est différent pour chaque période. Les poids ω^{did} ne permettraient donc pas d'assurer une tendance parallèle entre le groupe traité et le groupe de contrôle.

Nous proposons par conséquent de calculer un 3^{ème} type de poids pour chaque groupe-période :

$$\nu_{k,t}^{RC} = \frac{1}{N_{k,t}}$$

Ces poids permettent d'assurer que la fréquence de chaque groupe dans l'échantillon pondéré soit égale à $1/K$. Cela permet donc aux poids ω^{did} de garantir une tendance parallèle entre les groupes traités et le groupes de contrôle. Par construction, la somme de ces poids est égale à 1 pour chaque période t :

$$\sum_{\substack{k \\ k \in co}} \nu_{k,t}^{RC} = 1$$

Ainsi chaque période a une fréquence égale dans l'échantillon pondéré, comme cela doit être le cas pour calculer les poids temporels λ^{did} .

Une fois que ces poids sont obtenus, on peut finalement estimer l'effet du traitement avec la régression suivante :

$$\left(\hat{\tau}^{rc-sdid}, \hat{\mu}, \hat{\alpha}, \hat{\beta} \right) = \arg \min_{\tau, \mu, \alpha, \beta} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau)^2 \hat{\omega}_i^{sdid} \hat{\lambda}_t^{sdid} \nu_{k,t}^{RC} \right\}$$

Application de la méthode aux données du groupe SeLoger

Nous appliquons cette méthode aux données du groupe SeLoger afin d'estimer l'effet de l'encadrement des loyers.

ANNEXES PARTIE 4

A.4.1 : VARIABLE SALLE DE BAIN MANQUANTE PAR VILLE

	RENSEIGNE	MANQUANT
AIX-EN-PROVENCE	7686	11557
GRENOBLE	14851	18387
MARSEILLE	30069	65812
NANTES	4248	51412
NICE	5133	46998
PARIS	50 ou moins	287084
STRASBOURG	3422	47546
TOULON	2721	12418
TOULOUSE	11496	99960

A.4.2 : VARIABLE TOILETTES MANQUANTE PAR VILLE

	RENSEIGNE	MANQUANT
AIX-EN-PROVENCE	8423	10820
GRENOBLE	11238	22000
MARSEILLE	32061	63820
NANTES	2013	53647
NICE	3417	48714
PARIS	50 ou moins	287084
STRASBOURG	3318	47650
TOULON	2855	12284
TOULOUSE	12061	99395

A.4.3 : VARIABLE JARDIN MANQUANTE PAR VILLE

	RENSEIGNE	MANQUANT
AIX-EN-PROVENCE	1670	17573
GRENOBLE	19179	14059
MARSEILLE	7209	88672
NANTES	5558	50102
NICE	7365	44766
PARIS	50 ou moins	287084
STRASBOURG	7440	43528
TOULON	3823	11316
TOULOUSE	25796	85660

A.4.4 : VARIABLE BALCON MANQUANTE PAR VILLE

	RENSEIGNE	MANQUANT
AIX-EN-PROVENCE	2674	16569
GRENOBLE	19942	13296
MARSEILLE	10342	85539
NANTES	5289	50371
NICE	16360	35771
PARIS	50 ou moins	287084
STRASBOURG	6934	44034
TOULON	5442	9697
TOULOUSE	26282	85174

A.4.5 : VARIABLE ASCENSEUR MANQUANTE PAR VILLE

	RENSEIGNE	MANQUANT
AIX-EN-PROVENCE	6458	12785
GRENOBLE	26334	6904
MARSEILLE	30018	65863
NANTES	13855	41805
NICE	34192	17939
PARIS	50 ou moins	287084
STRASBOURG	13266	37702
TOULON	5705	9434
TOULOUSE	31222	80234

A.4.6 : VARIABLE PARKING MANQUANTE PAR VILLE

	RENSEIGNE	MANQUANT
AIX-EN-PROVENCE	2287	16956
GRENOBLE	24526	8712
MARSEILLE	11104	84777
NANTES	40212	15448
NICE	6478	45653
PARIS	50 ou moins	287084
STRASBOURG	11358	39610
TOULON	3536	11603