



FaSEST

M2 Economie Appliquée

Econometrie Spatiale

Devoir

Encadré par:

Mme Aurore Lavigne

Réalisé par:

MAHAMADOU ISSOUFOU Khadija

REN Yue

Parcours:

Econométrie Appliquée

2025 - 2026

Introduction.....	2
1. Méthodologie.....	2
1.1. Sources et définition des variables.....	2
1.2. Traitement des données.....	3
1.3. Approche analytique.....	3
1.4. Outils techniques.....	4
2. Résultats.....	4
2.1. Analyse descriptive.....	4
2.2. Résultats des estimations.....	6
2.2.1. Le modèle linéaire simple comme benchmark et ses limites.....	6
2.2.2. Diagnostic de la dépendance spatiale.....	7
2.2.3. Sélection et validation du modèle final.....	7
2.2.4. Détail du modèle choisi et interprétation.....	8
2.2.5. Interprétation des effets – Modèle SDM.....	8
Annexes.....	10

Introduction

Le marché locatif francilien présente des disparités spatiales majeures, souvent corrélées à la structure socio-économique du territoire. L'objectif de ce rapport est de modéliser les déterminants du loyer au mètre carré à l'échelle communale en intégrant explicitement la dimension géographique. Nous cherchons à tester si le loyer d'une commune subit l'influence des caractéristiques de son voisinage, justifiant ainsi le passage d'une modélisation linéaire classique (OLS) à une modélisation d'économétrie spatiale.

Le choix des variables explicatives repose sur les fondements de l'économie urbaine. Le **pouvoir d'achat local** (`revenu_median`) est retenu comme principal moteur de la demande, les zones à hauts revenus exerçant une pression mécanique sur les loyers. La **tension foncière** (`densite_pop`) capture l'arbitrage entre une offre limitée et une demande concentrée, caractéristique des zones urbaines denses. La **qualité de vie et l'attractivité** sont modélisées par les aménités urbaines (`nb_sport` et `nb_commerces`), l'accès aux services étant un facteur déterminant du consentement à payer des locataires. Enfin, les **dynamiques sociales** (`tx_chomage` et `pct_etrangers`) permettent d'analyser les phénomènes de ségrégation spatiale et leur impact sur la valorisation immobilière.

Nous cherchons ainsi à tester si le loyer d'une commune subit l'influence des caractéristiques de son voisinage, justifiant le recours à une modélisation d'économétrie spatiale plutôt qu'à une approche linéaire classique (OLS).

1. Méthodologie

1.1. Sources et définition des variables

L'étude repose sur la combinaison de données socio-économiques et de données spatiales. L'échantillon final comprend **1 250 communes** d'Île-de-France.

- **Données socio-économiques** : Issues de l'Observatoire des Territoires, elles décrivent les caractéristiques économiques et démographiques des communes.

- **Données spatiales** : Les contours géographiques proviennent du shapefile officiel diffusé par cadastre.data.gouv.fr.

Tableau 1: Dictionnaire des variables

Variable	Description	Source
loyer_m2	Loyer moyen au m ² (en €)	Observatoire des Territoires
revenu_median	Revenu médian par unité de consommation (en €)	Observatoire des Territoires
densite_pop	Densité de population (hab./km ²)	Observatoire des Territoires
tx_chomage	Taux de chômage des 15–64 ans (%)	Observatoire des Territoires
pct_etrangers	Part de la population étrangère (%)	Observatoire des Territoires
nb_sport	Nombre d'équipements sportifs	Observatoire des Territoires
nb_commerces	Nombre de commerces de proximité	Observatoire des Territoires
geometry	Contours géographiques des communes	cadastre.data.gouv.fr

1.2. Traitement des données

Après appariement des bases à l'aide du code INSEE et nettoyage des données (conversion des valeurs manquantes, suppression des valeurs aberrantes), l'échantillon final comprend **1 250 communes** disposant de l'ensemble des informations nécessaires à l'analyse.

Les variables `revenu_median` et `densite_pop` ont été transformées en **logarithmes** afin de stabiliser les variances et faciliter l'interprétation des coefficients.

1.3. Approche analytique

L'analyse s'articule en trois phases :

1. **Analyse descriptive exploratoire** : caractérisation de la structure des données (statistiques par département, corrélations, distributions, extrêmes) et visualisation via des cartes.
2. **Construction de la structure spatiale** : calcul des centroïdes et construction d'une matrice de voisinage \$W\$ selon la méthode des **k-plus proches voisins** (\$k = 5\$). La matrice est normalisée par ligne (style « W ») pour assurer un poids total identique pour chaque commune et garantir la connexité du graphe.
3. **Économétrie spatiale** :
 - Évaluation de l'autocorrélation spatiale via le **test de Moran global**, permettant de vérifier si les loyers des communes voisines présentent des similarités significatives.
 - Estimation d'un **modèle OLS de référence** pour servir de base et contrôler la dépendance spatiale dans les résidus.
 - Réalisation des **tests de Lagrange Multiplier (LM)** pour identifier la forme de la dépendance : dans la variable dépendante (LM-lag, modèle SAR) ou dans les erreurs (LM-error, modèle SEM).
 - Estimation d'un **modèle Spatial Durbin (SDM)** pour capturer les effets directs et indirects des variables explicatives.
 - Vérification via **LR test** que le SDM apporte une amélioration significative par rapport au SAR.
 - Choix final confirmé à l'aide du **critère d'information d'Akaike (AIC)**, le SDM présentant la valeur la plus faible.

1.4. Outils techniques

L'ensemble de l'analyse a été réalisé sous **R**, avec les packages suivants :

- **sf** pour la gestion des données spatiales
- **spdep** et **spatialreg** pour l'économétrie spatiale
- **dplyr** pour la préparation et la manipulation des données

2. Résultats

2.1. Analyse descriptive

L'analyse exploratoire des données des communes d'Île-de-France montre une forte variabilité spatiale des loyers et des caractéristiques socio-économiques. Les loyers moyens sont nettement plus élevés dans les départements centraux comme les Hauts-de-Seine (92) et le Val-de-Marne (94), avec des valeurs dépassant 20 €/m², tandis que les départements périphériques comme la Seine-et-Marne (77) présentent des loyers moyens plus faibles autour de 14 €/m². Cette cherté est mécaniquement liée à un revenu médian record (**31 355 €**) et une densité de population extrême (**11 106 hab./km²**).

À l'inverse, la **Seine-Saint-Denis (93)** présente un profil atypique : bien que les loyers y soient élevés (19,9 €/m²), le revenu médian est le plus bas de la région (20 306 €). Ce département affiche également les taux de chômage (15%) et d'étrangers (23,4%) les plus importants. Cela suggère que dans les zones très denses, la rareté du foncier maintient les loyers élevés malgré une précarité économique marquée.

Tableau 2: Statistiques descriptives agrégées par département

Dépt	Loyer moyen	Revenu moyen	Densité	Chômage %	Étrangers %	Sport	Commerces
92	25.0	31,355.0	11,106.9	9.9	13.0	8.7	2.2
94	21.2	25,693.6	7,216.8	11.0	15.9	5.9	1.8
93	19.9	20,306.2	8,142.3	15.0	23.4	8.0	2.3
95	16.9	27,201.4	1,215.2	8.9	7.5	1.4	0.6
78	16.8	30,333.7	781.3	7.9	6.8	1.1	0.6
91	16.2	27,935.2	994.5	7.8	7.8	1.3	0.7
77	14.4	26,076.9	334.2	8.9	5.7	0.6	0.4

• Top 10 des communes les plus chères

Le classement des loyers les plus élevés (Tableau 3) révèle une concentration géographique extrême, exclusivement située en **petite couronne** (92 et 94). **Neuilly-sur-Seine** domine ce segment avec **34,1 €/m²**, portée par un revenu médian record de 48 010 €. Ces pôles d'hypervalorisation se caractérisent par une **densité critique** (souvent > 20 000 hab./km²) et une proximité immédiate avec Paris, illustrant une saturation du marché où la rareté foncière et le haut pouvoir d'achat maintiennent des prix records.

Tableau 3: Top 10 des communes les plus chères

INSEE	Commune	Loyer (€/m²)	revenu_median	densite_pop	tx_chomage
92051	NEUILLY SUR SEINE	34.1	48,010.0	16,000.0	8.5
94067	SAINT MANDE	31.4	36,180.0	23,581.1	8.0
92044	LEVALLOIS PERRET	30.7	34,500.0	28,505.0	8.4
92012	BOULOGNE BILLANCOURT	29.7	35,040.0	19,387.9	8.7
94080	VINCENNES	29.4	34,880.0	25,456.8	7.7
92040	ISSY LES MOULINEAUX	27.9	33,650.0	16,117.9	7.5
92062	PUTEAUX	27.5	30,980.0	13,811.9	9.4
92026	COURBEVOIE	27.3	32,080.0	19,510.7	8.9
94018	CHARENTON LE PONT	27.2	30,250.0	15,134.7	8.8
92073	SURESNES	26.7	31,810.0	12,876.8	9.6

Le classement des loyers les plus élevés (Tableau 3) révèle une concentration géographique extrême, exclusivement située en **petite couronne** (92 et 94). **Neuilly-sur-Seine** domine ce segment avec **34,1 €/m²**, portée par un revenu médian record de 48 010 €. Ces pôles d'hypervalorisation se caractérisent par une **densité critique** (souvent > 20 000 hab./km²) et une proximité immédiate avec Paris, illustrant une saturation du marché où la rareté foncière et le haut pouvoir d'achat maintiennent des prix records

Tableau 4: Matrice de corrélation entre les variables

Variable	loyer m2	revenu_median	densite_pop	tx_chomage	pct_etrangers	nb_sport	nb_commerces
loyer_m2	1.000	0.295	0.724	0.087	0.491	0.621	0.451
revenu_median	0.295	1.000	-0.128	-0.646	-0.467	-0.230	-0.239
densite_pop	0.724	-0.128	1.000	0.351	0.590	0.711	0.444
tx_chomage	0.087	-0.646	0.351	1.000	0.640	0.432	0.383
pct_etrangers	0.491	-0.467	0.590	0.640	1.000	0.677	0.586
nb_sport	0.621	-0.230	0.711	0.432	0.677	1.000	0.702
nb_commerces	0.451	-0.239	0.444	0.383	0.586	0.702	1.000

L'analyse des corrélations démontre que la **densité de population** ($r = 0,72\$$) et la dotation en **équipements sportifs** ($r = 0,62\$$) sont les principaux leviers de valorisation des loyers, surpassant l'effet brut du revenu et confirmant l'importance des aménités urbaines dans la formation des prix immobiliers.

Les cartes ci-dessous **complètent et illustrent visuellement les statistiques descriptives** en rendant visible la dimension spatiale des phénomènes étudiés. Elles montrent que les loyers les plus élevés se concentrent au centre, autour de Paris et dans certaines communes de la petite couronne, tandis que les loyers diminuent progressivement vers la périphérie. La carte du revenu médian met en évidence une distribution similaire, avec des valeurs élevées localisées principalement dans l'ouest et le centre, et des revenus plus faibles dans l'est et les zones périphériques. Enfin, la carte de densité de population illustre clairement la forte concentration de population dans les zones centrales et une baisse rapide vers la périphérie.

Ainsi, ces représentations cartographiques **renforcent l'interprétation des données statistiques**, permettant de visualiser la corrélation entre loyers, revenus et densité de population à l'échelle régionale.

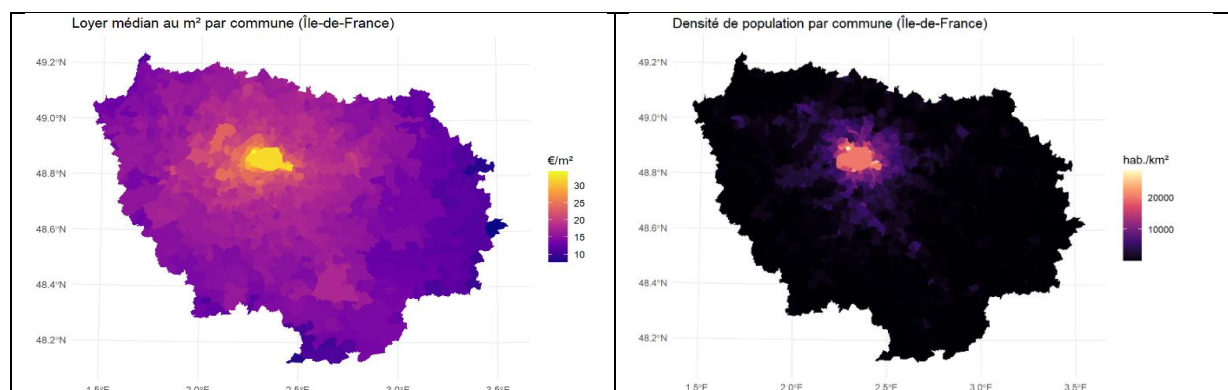


Figure 1: : Loyer médian au par m2 commune en Île-de-France

Figure 2: Densité de population par commune en Île-de-France

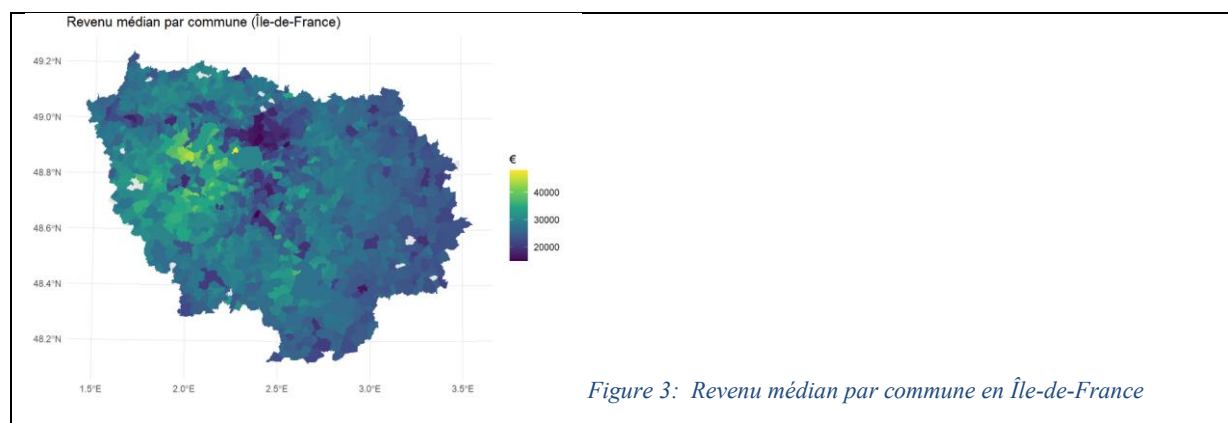


Figure 3: Revenu médian par commune en Île-de-France

2.2. Résultats des estimations

2.2.1. Le modèle linéaire simple comme benchmark et ses limites

Avant d'introduire la dimension spatiale, un modèle de Moindres Carrés Ordinaires (OLS) a été estimé. Ce modèle explique **81 % de la variance** ($R^2 = 0,81$), identifiant le revenu et la densité comme les piliers du loyer.

Le modèle OLS identifie le **revenu** et la **densité** comme les piliers du loyer ($R^2 = 0,81$), l'attractivité étant renforcée par les **équipements sportifs**. Toutefois, l'effet négatif des commerces et la non-significativité du chômage soulignent des limites structurelles. Surtout, le test de Moran ($I = 0,89$) prouve que l'OLS échoue à capter la réalité géographique : les loyers sont portés par une **dynamique de voisinage** massive. Le passage aux modèles spatiaux est donc primordial.

Tableau 5: : Résultats des OLS

Variable	Estimate	Std. Error	t value	p-value	Significativité
(Intercept)	-94.076	3.932	-23.924	< 2e-16	***
log revenu	10.173	0.379	26.817	< 2e-16	***
log densité	0.966	0.038	25.357	< 2e-16	***

Variable	Estimate	Std. Error	t value	p-value	Significativité
tx_chomage	-0.003	0.020	-0.150	0.881	
pct_etrangers	0.103	0.013	7.990	3.07e-15	***
nb_sport	0.259	0.021	12.208	< 2e-16	***
nb_commerces	-0.104	0.046	-2.259	0.024	*

2.2.2. Diagnostic de la dépendance spatiale

Le test de Moran global révèle une autocorrélation spatiale extrêmement forte et positive des loyers en Île-de-France, avec un indice $I = 0,894$ et une p-value quasi nulle ($p < 2,2e-16$). Cette valeur, très proche de 1, permet de rejeter sans équivoque l'hypothèse d'une distribution aléatoire : elle démontre que le marché locatif est structuré en « clusters », où les communes à loyers élevés sont systématiquement entourées de voisins aux prix similaires. Ce résultat justifie scientifiquement l'abandon du modèle OLS classique — qui suppose l'indépendance des observations — au profit d'un modèle d'économétrie spatiale capable de capturer ces effets de voisinage et de contagion géographique.

Pour identifier la forme exacte de cette dépendance, nous avons réalisé les **tests de Lagrange Multiplier (LM)**.

La significativité des tests robustes indique la nécessité d'un modèle capable de traiter simultanément le retard spatial et les variables explicatives spatialisées. Le passage au modèle SDM (Spatial Durbin Model) est donc requis.

Tableau 6: Résultats des tests de Lagrange

Test LM	Statistique	df	p-value	Interprétation
LM lag (RSlag)	842.99	1	< 2.2e-16	Très significatif → effet spatial lag
LM error (RSerr)	611.53	1	< 2.2e-16	Très significatif → autocorr. erreur
Robust LM lag (adjRSlag)	280.30	1	< 2.2e-16	Effet lag significatif même après correction de l'erreur
Robust LM error (adjRSerr)	48.85	1	2.77e-12	Effet erreur significatif même après correction du lag

Les résultats des tests de dépendance spatiale (LM et LM robustes) montrent que la variable d'intérêt, le loyer au m², présente une **autocorrélation spatiale très significative**. Autrement dit, les valeurs observées dans une commune sont fortement influencées par celles de ses voisines. Dans ce contexte, les hypothèses classiques de l'OLS — notamment l'indépendance des résidus — sont **violées**, ce qui rend les estimations potentiellement biaisées et inefficaces. Pour tenir compte de cette dimension spatiale, il est donc nécessaire d'utiliser des modèles économétriques spatiaux tels que le **SAR (Spatial Autoregressive Model)** ou le **SDM (Spatial Durbin Model)**, qui permettent d'intégrer explicitement les interactions entre communes et d'obtenir des coefficients fiables et interprétables.

2.2.3. Sélection et validation du modèle final

La comparaison des modèles confirme la supériorité du SDM pour capturer la réalité du terrain francilien.

Tableau 7: Choix du modèle final

Critère	OLS (Classique)	SAR	SDM (Retenu)
AIC	4354.3	3307.8	3286.4

Critère	OLS (Classique)	SAR	SDM (Retenu)
Moran des résidus (p-value)	< 0.001 (Biaisé)	0.689	0.934
ρ (Rho)	-	0.745	0.754

Après comparaison des critères d'information (AIC) et analyse de la **distribution des résidus**, le modèle **SDM** s'impose comme le plus performant. Il permet de distinguer les **impacts directs** (propres à la commune) des **impacts indirects** (effets de débordement des voisins), offrant ainsi une lecture plus fine des dynamiques immobilières en Île-de-France et le test de rapport de vraisemblance confirme cela.

Le paramètre $\rho = 0,75$ est fondamental : il indique que **75 % de la variation du loyer d'une commune est expliquée par la dynamique de son voisinage**.

2.2.4. Détail du modèle choisi et interprétation

Forme générale du SDM :

$$Y = \rho WY + X\beta + WX\theta + \varepsilon$$

où :

- Y : loyer_m2
- X : variables explicatives locales (log revenu, log densité, tx_chomage, pct_etrangers, nb_sport, nb_commerces)
- W : matrice de poids spatiaux (5 voisins, normalisée)
- ρ : paramètre de dépendance spatiale
- θ : effets des variables explicatives dans les communes voisines
- ε : erreur classique

2.2.5. Interprétation des effets – Modèle SDM

L'intérêt principal du modèle Spatial Durbin (SDM) réside dans la décomposition des effets des variables explicatives en **effets directs**, **indirects (spillovers)** et **totaux**, permettant de tenir compte des interactions spatiales et des mécanismes de rétroaction entre communes.

Tableau 8: Effets directs & indirects basés sur le modèle SDM

Variable	Effet direct	Effet indirect	Effet total
log revenu	4.765	6.341	11.105
Log densité	0.339	0.821	1.161
Tx chômage	0.016	-0.003	0.013
pct_etrangers	0.006	0.029	0.035
nb_sport	0.146	0.491	0.637
nb_commerces	-0.041	-0.709	-0.750

Note de lecture : L'effet total du revenu (11,10) se décompose en un effet direct (4,76) et un effet indirect (6,34). Autrement dit, une augmentation du revenu dans la commune fait monter le loyer local (effet direct), mais l'effet le plus fort vient de l'enrichissement des communes voisines (6,34) : le marché capte un effet de « standing » de la zone, où la richesse alentour tire les prix vers le haut.

- **Le revenu : un effet de standing spatial**

Le revenu apparaît comme un déterminant majeur du loyer au m², avec un **impact total élevé (11,10)**. L'effet indirect (6,34) est supérieur à l'effet direct (4,76), ce qui suggère que le niveau

de revenu des communes voisines joue un rôle important dans la formation des loyers locaux. En Île-de-France, les prix semblent ainsi sensibles au **contexte socio-économique du bassin de vie élargi**, et pas uniquement aux caractéristiques propres à la commune.

- **Densité et équipements sportifs : des externalités positives**

La densité urbaine présente un impact total positif (1,16), traduisant une pression immobilière accrue dans les zones denses. Cet effet est principalement local mais génère également des retombées positives sur les communes voisines.

Les équipements sportifs ont également un impact total positif (0,64), avec un effet indirect (0,49) plus important que l'effet direct (0,15). Cela indique que ces infrastructures constituent des aménités territoriales dont les bénéfices dépassent les frontières communales et renforcent l'attractivité du voisinage.

- **Les commerces : un effet spatial négatif**

Le nombre de commerces présente un **impact total négatif (-0,75)**, principalement porté par l'effet indirect (-0,71). Ce résultat suggère qu'une forte concentration commerciale dans les communes voisines est associée à une baisse des loyers locaux. Cette relation peut refléter des externalités négatives potentielles (congestion, nuisances, intensité des flux), sans pour autant permettre d'établir un lien causal strict.

- **Chômage et part de population étrangère**

Les impacts associés au taux de chômage et à la part de population étrangère sont faibles et non significatifs. Une fois pris en compte le revenu et la densité urbaine, ces variables ne semblent pas jouer un rôle déterminant dans la formation des loyers en Île-de-France.

Conclusion

Les résultats confirment que le loyer au m² est fortement influencé par les interactions spatiales entre communes. La valeur élevée du paramètre spatial ($\rho \approx 0,75$) met en évidence une **forte dépendance spatiale des prix**, justifiant pleinement l'utilisation d'un modèle SDM. Les politiques locales d'aménagement et d'équipement génèrent ainsi des effets qui dépassent largement le cadre communal, soulignant l'importance d'une approche territoriale coordonnée

Annexes

Tableau 9: Estimation du modèle Spatial Durbin (SDM)

Variable	Coefficient estimé	Erreur standard	z-statistique	p-value
Constante	-25.599	4.256	-6.015	< 0.001
log_revenu	4.341	0.348	12.488	< 0.001
log_densite	0.284	0.035	8.221	< 0.001
tx_chomage	0.016	0.012	1.317	0.188
pct_etrangers	0.004	0.009	0.396	0.692
nb_sport	0.113	0.014	8.244	< 0.001
nb_commerces	0.006	0.028	0.212	0.832
lag.log_revenu	-1.620	0.507	-3.194	0.001
lag.log_densite	0.000	0.053	0.000	1.000
lag.tx_chomage	-0.013	0.022	-0.592	0.554
lag.pct_etrangers	0.005	0.015	0.338	0.735
lag.nb_sport	0.043	0.027	1.610	0.107
lag.nb_commerces	-0.190	0.063	-2.992	0.003