

Projet Econometrie

DJOKO FOKOU Keyssel

Contents

1	Introduction	1
2	Choix et description des variables	2
3	Statistiques descriptives	4
3.1	Gestion des NA	4
3.2	Analyse univariée	4
3.3	Analyse bivariée	5
4	Régressions	9
4.1	Première approche	9
4.2	Gestion de l'hétéroscédasticité	12
5	Conclusion	18

1 Introduction

Le 27 juillet 2023, le projet de loi dite “Kasbarian” a été approuvé par le conseil Constitutionnel, dans un but très clair : étouffer toutes tentatives d’ouverture de “squat”, ces lieux de vie occupés de manière illicites par des gens précaires cherchant des solutions pour se loger. Ce phénomène principalement observé dans les métropoles vient en réaction à une crise du logement. Cette crise est tellement présente qu des réponses d’organisation collectives face à cette crise apparaissent d’ailleurs partout en France, dont à Rennes avec l’Assemblée Logement. Si on peut se demander d’où vient cette crise, il vient assez naturellement qu’elle vient du coût du logement qui prend de plus en plus de place dans le pouvoir d’achat, du fait de son augmentation constante : “Le logement est une marchandise, il a un coût, une valeur. Conséquemment, la crise est une crise des prix, en lien avec un pouvoir d’achat logement insuffisant”(Le logement : une crise globale et individuelle, Sabrina Daine et al.). Si ce coût est de plus en plus important, on peut se demander s’il l’est de plus en plus partout. S’il y

a bien des zones où les prix des loyers sont plus importants, quelles sont leur caractéristiques territoriales?

Dans ce contexte, notre étude porte donc sur le prix du loyer au mètre carré des appartements dans les villes constituant les métropoles. On appelle loyer le prix mensuel du logement que l'on loue, que l'on rapporte au mètre carré. Limiter l'analyse aux métropoles permet une meilleure compréhension du phénomène urbain. Inclure tous les territoires pourrait biaiser l'analyse, car les différences se marqueraient surtout sur des caractéristiques de ruralité. De plus, nous n'étudions ici que les appartements, un type de logement plus courant en milieu urbain. Après avoir étudié la littérature, notre choix de variable détaillé par la suite se concentre autour de 2 axes : l'activité socio-économique de la commune et sa disposition géographique.

2 Choix et description des variables

Variable	Description	Type	Source	Unité
loyer_m2	Loyer au mètre carré estimé à partir d'annonces immobilières	Quantitative continue	INSEE / SeLoger.com / Leboncoin (T3 2023)	€/m ²
med_revenu	Médiane du niveau de vie annuel des individus (2021)	Quantitative continue	Statistiques locales de l'INSEE	€
logements_vacants	Part des logements vacants (2021)	Quantitative continue	Observatoire des Territoires (2021)	Pourcentage (%)
densite_pop	Densité de population (2021)	Quantitative continue	Observatoire des Territoires (2021)	Habitant/km ²
si_paris	Indicateur : 1 si commune appartient au Grand Paris, sinon 0	Qualitative binaire	Construction interne	1/0
dist_metropole	Longueur de l'itinéraire le plus proche en voiture	Quantitative continue	Construction interne	Kilomètres (km)

Variable	Description	Type	Source	Unité
jeunes	Proportion des 18-24 ans dans la commune (2021)	Quantitative continue	Observatoire des Territoires (2021)	Pourcentage (%)
temps_trajet	Temps de trajet le plus court de la ville à la capitale de la métropole	Quantitative continue	Construction interne	Minutes
temps_tgv_paris	Temps en minutes en TGV entre Paris et la ville principale de la métropole	Quantitative continue	Construction interne	Minutes
nouveaux_menages	Part des ménages arrivés dans la commune il y a moins de 2 ans (2021)	Quantitative continue	Observatoire des Territoires (2021)	Pourcentage (%)

Le choix des variables a été guidé par des considérations économiques, sociales et géographiques pour expliquer les variations des loyers au mètre carré (**loyer_m2**). Cette variable cible, issue d’annonces immobilières sur SeLoger.com et Leboncoin, reflète les loyers “de flux”. Bien que légèrement surévalués par rapport aux loyers en place, ces flux offrent une base uniforme pour comparer les territoires. Les données, datées du troisième trimestre 2023, incluent des approximations pour les communes sans annonces, basées sur celles de communes voisines aux caractéristiques similaires.

Parmi les déterminants principaux, le revenu médian des habitants (**med_revenu**) a été intégré, car il constitue un facteur clé selon la bibliographie. Ce revenu reflète directement la capacité des habitants à payer des loyers, influençant donc les prix. En complément, la part des logements vacants (**logements_vacants**) permet d’évaluer l’offre potentielle de logements disponibles, tandis que la densité de population (**densite_pop**) capture l’impact de la pression démographique sur les prix, une population dense étant souvent associée à des loyers plus élevés.

Pour isoler des particularités géographiques, une variable indicatrice a été construite. **si_paris** distingue les communes du Grand Paris où les loyers sont notoirement plus élevés. Pour évaluer les effets de la distance à la ville principale de la métropole, on a créé **dist_metropole** et ‘**temps_trajet**’.

Enfin, certaines variables traduisent des aspects spécifiques de la demande. La proportion de jeunes âgés de 18 à 24 ans (**jeunes**) sert de proxy pour l'influence des étudiants, dont la présence peut accroître la demande locative, soutenue par des aides comme les APL. Dans un autre registre, la proportion de bénéficiaires du RSA (**beneficiaires_RSA**) explore si une population à faibles revenus exerce une pression à la baisse sur les loyers, en adaptant l'offre à cette demande spécifique.

Les variables géographiques, comme le temps de trajet minimal en TGV jusqu'à Paris (**temps_tgv_paris**), mesurent l'influence de la connectivité ferroviaire sur les loyers, tandis que la proportion de ménages récents (**nouveaux_menages**) reflète l'attractivité d'une commune et la dynamique de renouvellement de sa population. Ces éléments permettent de capturer les effets conjugués de la gentrification et des migrations périurbaines, souvent observés autour des grandes villes.

3 Statistiques descriptives

3.1 Gestion des NA

La première étape est de gérer le cas des valeurs manquantes.

ville_principale	Code	Libellé	loyer_m2
0	0	0	1
logements_vacants	med_revenu	jeunes	densite_pop
0	7	0	0
nouveaux_menages	metropole	si_paris	temps_tgv_paris
0	0	0	0
dist_metropole	temps_trajet		
1	1		

On constate une valeur manquante de la variable cible dans notre base de données, qui est également manquante dans notre variable **temps_trajet**, que nous allons donc écarter. C'est la ville de Oullins-Pierre-Bénite, dans la métropole lyonnaise qui est une fusion de deux villes qui date de 2024, sachant que nos données datent souvent de 2023. Comme il n'y en a qu'une, on peut se permettre de l'enlever. On décide également de supprimer les 7 villes présentant une valeur manquante de **med_revenu**.

3.2 Analyse univariée

On regarde d'abord une distribution globale des variables.

Tableau 1 : Statistiques descriptives des variables quantitatives

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
loyer_m2	13.705	3.310	6.600	11.600	13.000	14.800	31.200
logements_vacants	5.947	2.466	0.400	4.300	5.500	7.175	22.800
med_revenu	25,822.420	4,182.618	14,580	23,295	25,600	28,160	48,010
jeunes	7.536	2.866	2.700	6.000	7.000	8.200	34.500
densite_pop	2,002.549	3,574.429	4.250	228.762	630.800	1,822.878	28,337.080
nouveaux_menages	10.994	3.676	2.400	8.600	10.600	12.800	31.300
temps_tgv_paris	122.008	83.133	0	70	106	185	345
dist_metropole	17.632	12.471	0.000	9.672	15.062	21.423	88.295
temps_trajet	25.213	11.821	0.000	17.387	23.083	31.446	92.583

On voit dans le tableau 1 que le loyer au mètre carré est en moyenne de 13.705 euros à l'échelle des villes qui constituent les métropoles. L'écart-type assez faible nous indique qu'une bonne partie des territoires possède le même loyer au mètre carré moyen à 3 euros près environ.

3.3 Analyse bivariable

On peut regarder une distribution des variables en fonction de si on se trouve dans la métropole du grand Paris ou non.

\$`Autre métropole`

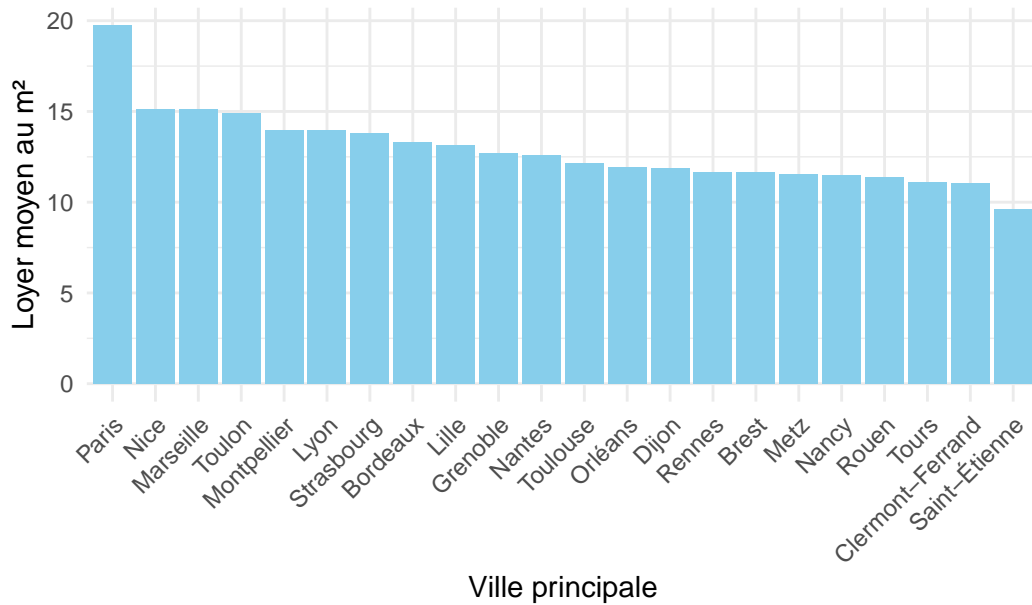
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
6.60	11.40	12.60	12.75	13.80	27.80

\$`Métropole du Grand Paris`

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15.50	17.35	18.60	19.73	21.60	31.20

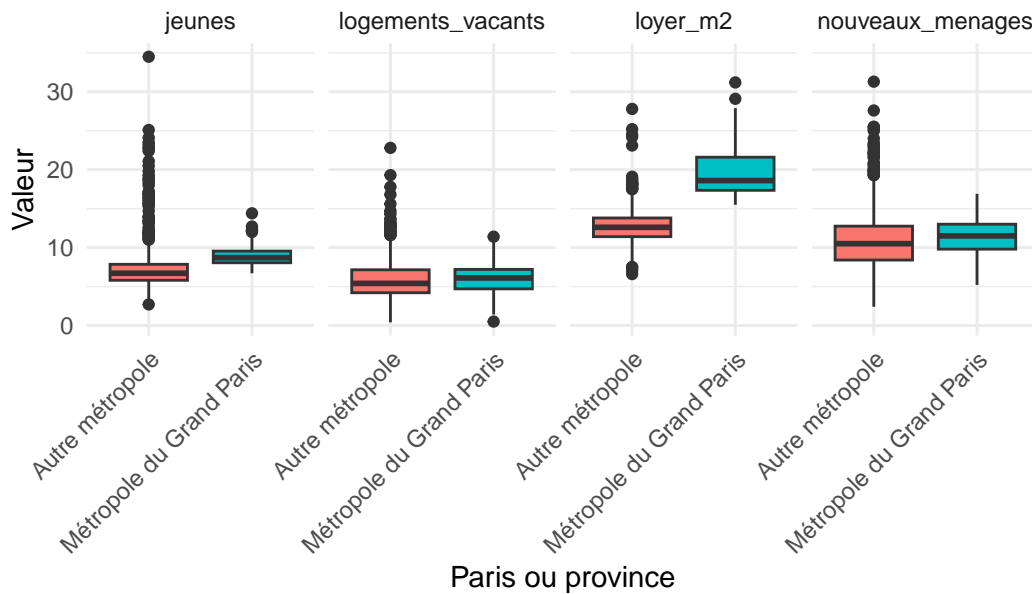
Il vient assez nettement que le loyer au m² moyen est bien plus élevé dans les villes de la capitale, avec une différence d'environ 6 euros entre le loyer au m² moyen dans la métropole Parisienne et le loyer au m² moyen dans les autres métropoles. On peut illustrer ce résultat par un graphique :

Graphique 1 : Classement des villes principales par loyer moyen



Le graphique 1 montre bien à quel point les loyers sont plus élevés à Paris qu'ailleurs. Il convient d'avoir une représentation graphique, dans laquelle nous allons regarder la distribution d'autres variables en fonction de si nous sommes dans la métropole du Grand Paris ou non

Graphique 2 : Comparaison de variables en fonction de Paris ou

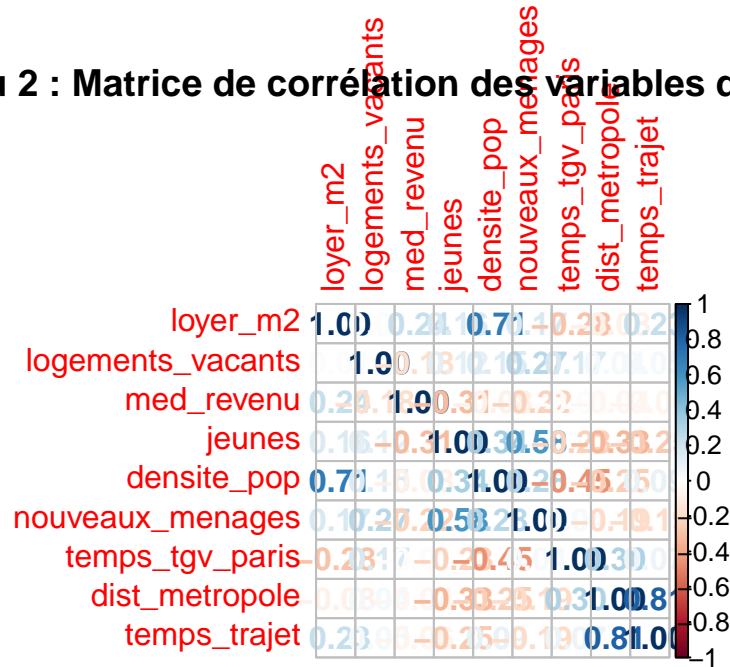


Il est clair que le loyer au mètre carré est bien plus élevé dans la métropole du Grand Paris que dans les autres métropoles (Graphique 2). On note également que les jeunes habitent en moyenne un peu plus à Paris que dans les autres métropoles.

Il apparait que parmi les liens de corrélation avec la variable `si_paris`, le plus élevé est de loin celui avec le loyer au mètre carré, ce qui nous évite un problème de colinéarité avec cette variable.

Il faut à présent regarder les corrélations entre les autres variables entre elles.

Tableau 2 : Matrice de corrélation des variables quantitatives

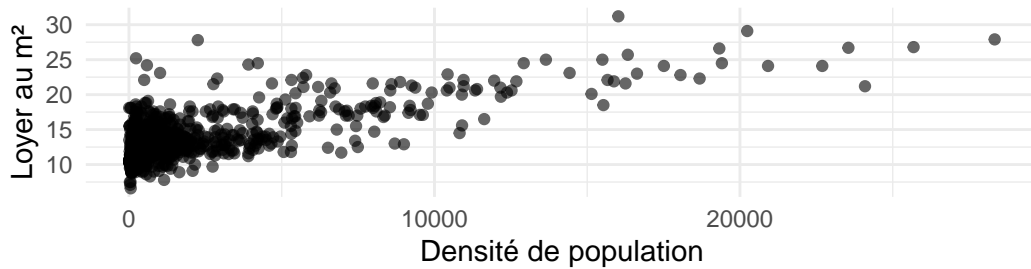


D'après le tableau 2, le loyer au mètre carré est corrélé positivement très fortement avec la densité de population. La relation légèrement positive entre le loyer au mètre carré et le temps de trajet depuis la capitale de la métropole interroge, car on pourrait s'attendre à ce que plus on s'éloigne de la ville, plus les prix baissent.

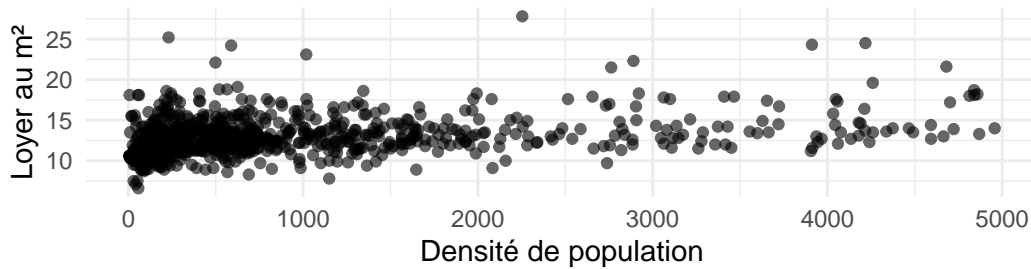
On remarque une corrélation négative entre la densité de population et le temps en TGV pour aller à paris, qui s'explique car pour toutes les communes de la métropole du Grand Paris la valeur de cette variable est égale à 0.

Comme on pouvait s'en douter du fait que les jeunes changent souvent de lieu de vie, on détecte une corrélation assez forte entre la part de nouveaux ménages et la part de jeunes, ce qui pourrait entraîner un problème de colinéarité dans notre régression.

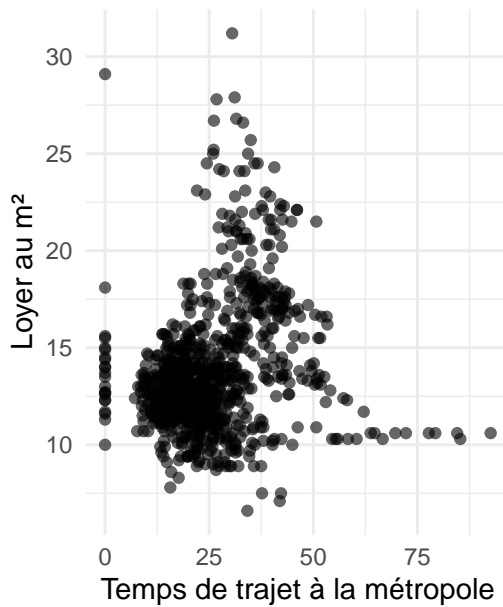
Graphique 3 : Relation entre densité de population et loyer au m²



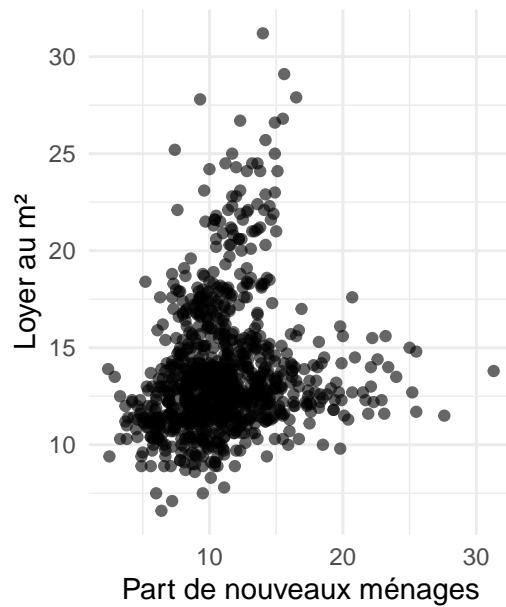
Graphique 4 : Relation entre densité de population et loyer au m²



Graphique 5



Graphique 6



Si la relation entre densité de population semble linéaire et positive, on note un clair problème d'hétéroscédasticité, car la plupart des individus ont une densité de population inférieure à 5 000. Ainsi en regardant de plus près, on constate que cette relation positive n'est vraie que pour les individus ayant une densité de population supérieure à 5000 (graphiques 3 et 4). Les

graphiques 5 et 6 montrent également des relations non linéaires entre le temps de trajet et le loyer au mètre carré, et entre la part de nouveaux ménages et le loyer au mètre carré.

4 Régressions

4.1 Première approche

Nous allons commencer par une régression simple avec toutes les variables, avec pour formule:

$$\text{loyer}_{m2} = \beta_0 + \beta_1 \cdot \text{logements_vacants} + \beta_2 \cdot \text{med_revenu} + \beta_3 \cdot \text{jeunes} + \beta_4 \cdot \text{nouveaux_menages} + \beta_5 \cdot \text{densite_pop} + \beta_6 \cdot \text{si_paris} + \beta_7 \cdot \text{temps_tgv_paris} + \beta_8 \cdot \text{dist_metropole} + \beta_9 \cdot \text{temps_trajet}.$$

La deuxième a pour formule : $\text{loyer}_{m2} = \beta_0 + \beta_1 \cdot \text{logements_vacants} + \beta_2 \cdot \text{med_revenu} + \beta_3 \cdot \text{nouveaux_menages} + \beta_4 \cdot \text{densite_pop} + \beta_5 \cdot \text{si_paris} + \beta_6 \cdot \text{temps_tgv_paris} + \beta_7 \cdot \text{dist_metropole} + \beta_8 \cdot \text{temps_trajet}.$

La troisième a pour formule : $\text{loyer}_{m2} = \beta_0 + \beta_1 \cdot \text{logements_vacants} + \beta_2 \cdot \text{med_revenu} + \beta_3 \cdot \text{nouveaux_menages} + \beta_4 \cdot \text{densite_pop} + \beta_5 \cdot \text{si_paris} + \beta_6 \cdot \text{temps_tgv_paris} + \beta_7 \cdot \text{dist_metropole} + \beta_8 \cdot \text{temps_trajet} + \beta_9 \cdot \text{nouveaux_menages} \cdot \text{metropole}.$

Pour notre régression, nous allons modifier les variables pour une meilleure lisibilité.

Régressions avec toutes les variables | sans variable non significative | avec effet croisés

Dependent variable:		
	Toutes les variables (1)	loyer_m2 Seulement variables significatives (2)
logements_vacants	-0.103*** (0.024)	-0.104*** (0.024)
med_revenu_m	0.237*** (0.014)	0.235*** (0.014)
jeunes	0.011 (0.027)	
nouveaux_menages	0.079*** (0.020)	0.083*** (0.017)

densite_pop_m	0.037*** (0.003)	0.037*** (0.003)
si_parisMétropole du Grand Paris	6.459*** (0.366)	6.473*** (0.364)
temps_tgv_paris	0.672*** (0.057)	0.669*** (0.057)
dist_metropole_m	0.327*** (0.106)	0.326*** (0.105)
metropole1		
nouveaux_menages:metropole1		
temps_trajet_m	-0.449*** (0.128)	-0.454*** (0.127)
Constant	4.817*** (0.552)	4.920*** (0.496)

Observations	962	962
R2	0.730	0.730
Adjusted R2	0.728	0.728
Residual Std. Error	1.727 (df = 952)	1.726 (df = 953)
F Statistic	286.550*** (df = 9; 952)	322.623*** (df = 8; 953)
=====		

Note:

Notre premier modèle explique 72,8% de la variance du loyer au m². On peut dire que ce modèle est globalement valide avec le test de Fisher. Simplement, toutes les variables ne sont pas significatives. Nous pouvons enlever *jeunes*, qui n'est pas représentative dans cette régression.

Le revenu médian des habitants (*med_revenu_m*), exprimé en milliers d'euros, conserve un effet positif significatif. Une augmentation de 1 000 € du revenu médian est associée à une hausse de 0,235 € du loyer au m². Cela confirme que les loyers sont directement influencés par le niveau de vie des habitants, qui détermine leur capacité à payer.

La proportion de ménages arrivés récemment (**nouveaux_ménages**), indicatrice de l'attractivité d'une commune, a également un effet significatif : une augmentation de 1 % de ménages récents est liée à une hausse de 0,083 € du loyer. Ces résultats suggèrent que les territoires attractifs, accueillant de nouveaux arrivants, subissent une pression accrue sur leur marché locatif.

Avec une transformation en centaines d'habitants par km², la densité de population (**densite_pop_m**) montre qu'une augmentation de 100 habitants par km² est associée à une hausse de 0,037 € du loyer. Ce résultat met en évidence l'effet des dynamiques urbaines sur la demande locative.

Du côté de l'accessibilité géographique, le temps de trajet en TGV vers Paris (**temps_tgv_paris**), maintenant exprimé en heures, a un effet significatif. Une réduction d'une heure de trajet est associée à une hausse de 0,672 € du loyer, soulignant l'importance de la connexion à la capitale dans la valorisation des biens immobiliers.

Pour la distance au centre métropolitain (**dist_metropole_m**), exprimée en dizaines de kilomètres, une augmentation de 10 km est associée à une hausse de 0,327 € du loyer. En parallèle, le temps de trajet en voiture vers la capitale métropolitaine (**temps_trajet_m**), réduit à des variations de 10 minutes, montre qu'une augmentation de 10 minutes est associée à une baisse de 0,449 € du loyer. Cela reflète une perte d'attractivité pour les communes plus éloignées des centres métropolitains. Il est à noter que ces variables ne sont significatives que si on les inclut de paire dans la régression. Leur effet est donc plutôt à interpréter ensemble. Il y a une dimension d'efficacité des infrastructures dans l'interprétation : plus je mets de temps à parcourir la distance, moins le loyer au mètre carré va être cher dans la commune (coefficient **temps_de_trajet_m** négatif).

Enfin, la part des logements vacants (**logements_vacants**) continue d'avoir un effet négatif significatif. Une augmentation de 1 % de logements vacants est associée à une baisse de 0,103 € du loyer, traduisant l'effet d'une sur-offre sur le marché.

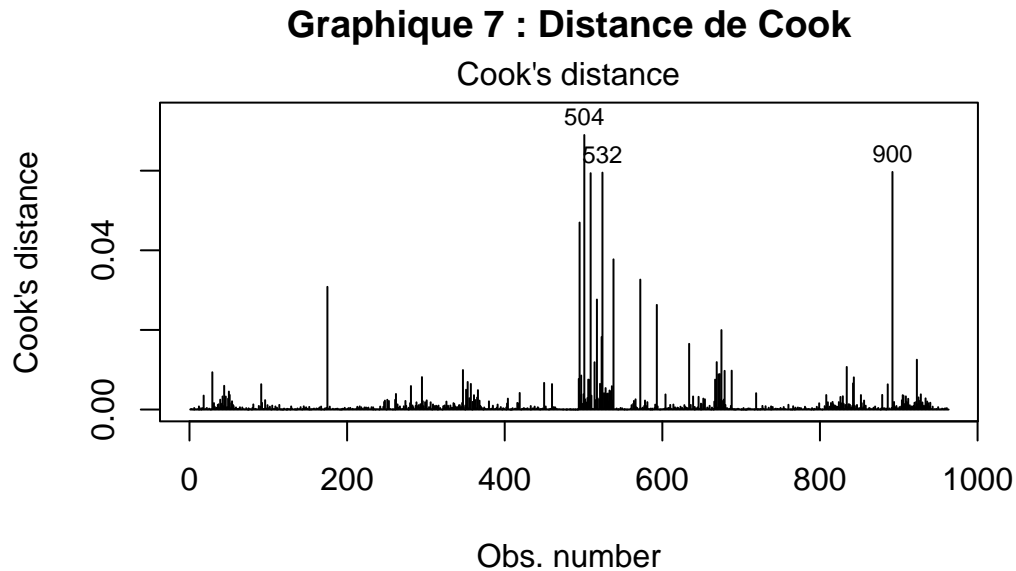
L'effet le plus fort reste celui de la variable indicatrice **si_paris**. Être situé dans la métropole du Grand Paris augmente le loyer moyen de 6,46 € par m², toutes choses égales par ailleurs. Ce résultat reflète l'attractivité unique de la région parisienne, où la demande de logements dépasse largement l'offre, entraînant des prix nettement supérieurs à ceux des autres métropoles.

Nous ajoutons une interaction entre la part de nouveaux ménages et la variable métropole afin d'examiner si leur impact sur les loyers varie selon le contexte, entre les métropoles et les zones non métropolitaines. Dans les zones non métropolitaines, une augmentation d'une unité de nouveaux ménages entraîne une hausse moyenne des loyers de 0,102 €/m², tandis que dans les métropoles, cet effet diminue de 0,272 €/m², devenant ainsi négatif à -0,17 €/m² (0,102 - 0,272). Cela révèle des dynamiques distinctes : dans les métropoles, la part de nouveaux ménages indique la taille du flux d'entrée et de sortie (souvent des étudiants et des jeunes), limitant l'effet sur les loyers, alors que dans les zones non métropolitaines, elle indique l'attractivité d'un territoire. Ainsi, dans les métropoles, plus il y a de flux, ou plus la ville est étudiante, moins les loyers sont onéreux. Tandis que dans les zones non métropolitaines, plus

la zone est attractive, plus les loyers sont onéreux. Une étude plus poussée avec la part de bénéficiaires des APL pourrait affiner cette analyse.

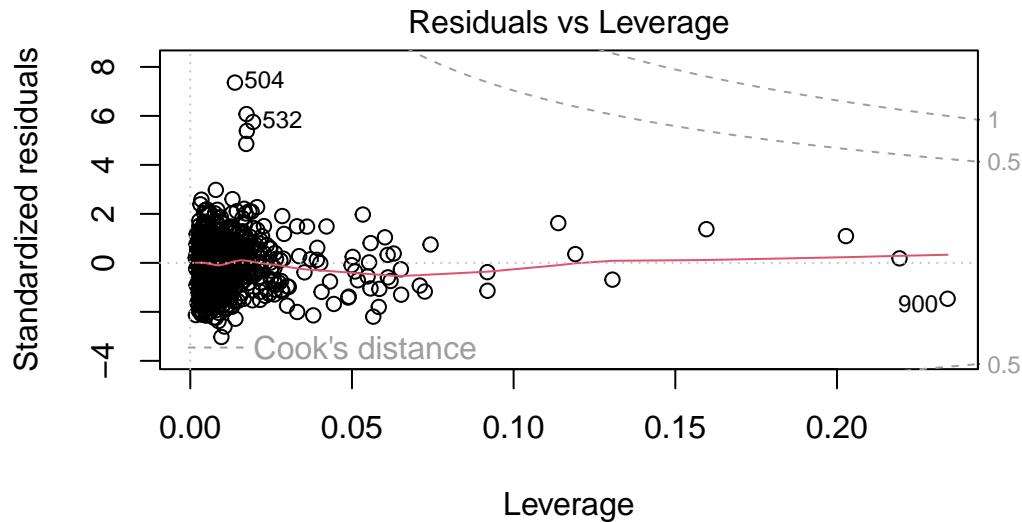
4.2 Gestion de l'hétéroscédasticité

Il est important de vérifier l'hétéroscédasticité dans les données que nous avons pressenti dans l'analyse descriptive précédemment.



oyer_m2 ~ logements_vacants + med_revenu_m + nouveaux_menages + c

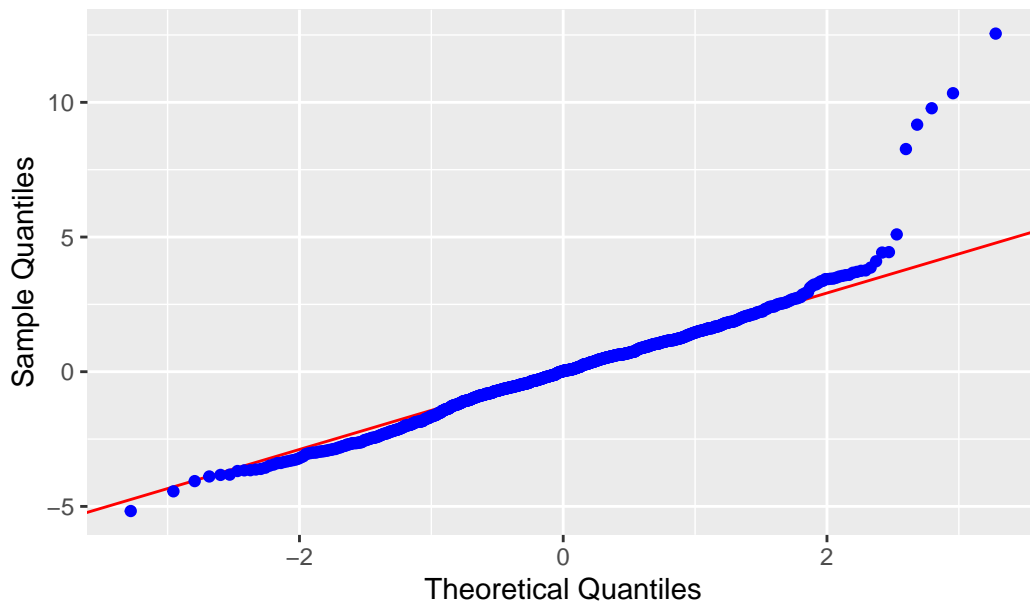
Graphique 8 : Résidus standardisés en fonction leur levie



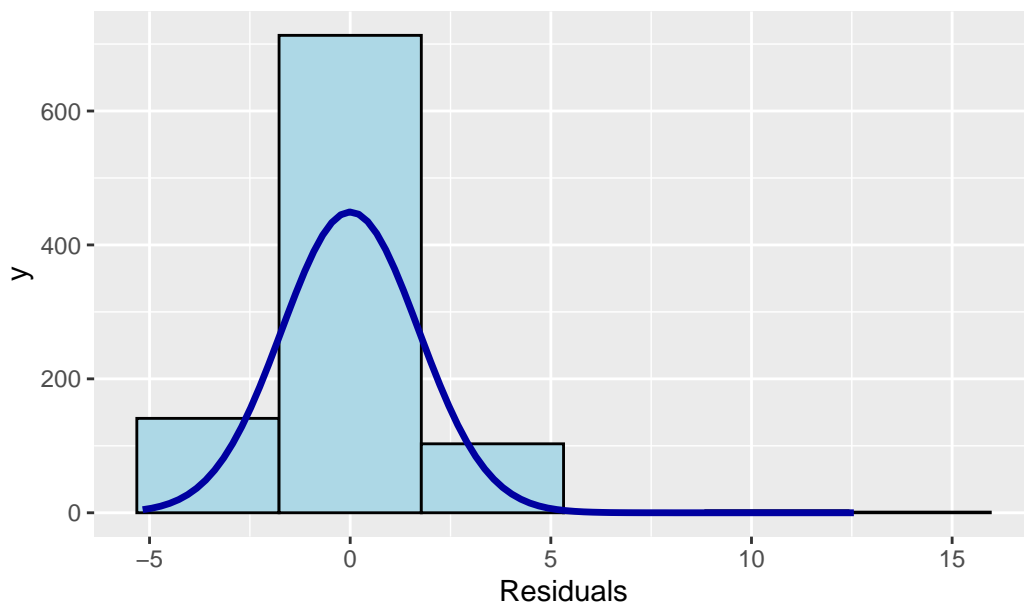
$\text{loyer_m2} \sim \text{logements_vacants} + \text{med_revenu_m} + \text{nouveaux_menages} + \epsilon$

On n'observe pas de points influents dans notre régression. Si certaines valeurs dénotent, comme les villes de Ouges, Toulon (communes avec un loyer au m² faible comparativement aux autres communes de leurs métropoles) qui ont un fort point de levier, et Cap-d'Ail (commune de bord de mer, très riche, frontalière de Monaco, avec un loyer au m² de 10 euros plus cher que les autres villes de la métropole de Nice) qui a un résidu standardisé très élevé, elles ne sont pas suffisamment influentes pour être retirées de notre régression, comme le montre les faibles distances de Cook. Nous pourrions cependant lancer une régression robuste pour vérifier si ces points ne sont pas influents.

Graphique 9 : Quantile–quantile des résidus



Graphique 10 : Histogramme des résidus



One Sample t-test

```
data: reg3$residuals
t = -4.7216e-15, df = 961, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.1080847  0.1080847
sample estimates:
 mean of x
-2.600499e-16
```

Shapiro-Wilk normality test

```
data: reg3$residuals
W = 0.94339, p-value < 2.2e-16
```

La moyenne des résidus est statistiquement égale à 0, d'après le test de moyenne nulle. On ne rejette pas H_0 car la p-value est égale à 1. C'est ce que l'on attend et que l'on pouvait observer dans le graphique . Cependant, le test de Shapiro-Wilk rejette l'hypothèse de normalité des résidus. L'hétéroscédasticité des résidus est bien visible dans les graphiques 7 et 8, avec une dispersion inégale des résidus dans l'histogramme et avec les queues du qqplot qui ne suivent pas la droite de normalité. Nous pouvons donc effectuer un test de White par la suite pour confirmer notre hypothèse.

studentized Breusch-Pagan test

```
data: reg3
BP = 133.66, df = 10, p-value < 2.2e-16
```

```
[1] "Test de White avec carrés des variables"
```

studentized Breusch-Pagan test

```
data: reg3
BP = 168.67, df = 15, p-value < 2.2e-16
```

```
[1] "Test de Goldfeld et Quandt sur le temps de trajet depuis Paris en TGV"
```

Goldfeld-Quandt test

```
data: reg3
GQ = 4.5158, df1 = 467, df2 = 467, p-value < 2.2e-16
alternative hypothesis: variance increases from segment 1 to 2
```

[1] "Test de Goldfeld et Quandt sur la temps de trajet en voiture à la métropole"

Goldfeld-Quandt test

```
data: reg3
GQ = 1.6361, df1 = 467, df2 = 467, p-value = 6.118e-08
alternative hypothesis: variance increases from segment 1 to 2
```

Le test de Breusch-Pagan rejette l'hypothèse nulle d'homoscédasticité des résidus. Le test de White, pour lequel nous n'avons pas utilisé les effets croisés mais simplement les carrés des variables, confirme ce résultat. Nous avons donc une hétéroscédasticité dans notre régression. En utilisant le test de Goldfeld et Quandt, nous pouvons voir que l'hétéroscédasticité est plus forte pour les communes les plus éloignées de la capitale de la métropole, ou de Paris, et que donc c'est le `temps_trajet_m` et le `temps_tgv_paris` qui sont les variables qui expliquent le plus l'hétéroscédasticité de notre régression. Nos hypothèses faites durant la première partie de l'analyse descriptive ne sont donc pas toutes confirmées.

À présent, nous allons effectuer une régression robuste à l'hétéroscédasticité et une régression MCG pour comparer les résultats.

Comparaison de l'estimation MCO, MCG et MCO correction WHITE

=====			
Variable dépendante : 'Loyer au mètre carré'			
	MCO	MCG	MCOWhite

logements_vacants	-0.100*** (0.024)	-0.099*** (0.023)	-0.100*** (0.025)
med_revenu_m	0.233*** (0.014)	0.227*** (0.012)	0.233*** (0.014)
nouveaux_menages	0.102***	0.106***	0.102***

	(0.018)	(0.016)	(0.019)
densite_pop_m	0.038*** (0.003)	0.039*** (0.002)	0.038*** (0.003)
si_parisMétropole du Grand Paris	6.526*** (0.368)	6.202*** (0.340)	6.526*** (0.355)
temps_tgv_paris	0.667*** (0.057)	0.532*** (0.059)	0.667*** (0.099)
temps_trajet_m	-0.528*** (0.135)	-0.520*** (0.131)	-0.528*** (0.142)
dist_metropole_m	0.369*** (0.108)	0.465*** (0.105)	0.369*** (0.119)
metropole1	4.616** (2.134)	9.821*** (1.125)	4.616** (2.072)
nouveaux_menages:metropole1	-0.272*** (0.102)	-0.482*** (0.054)	-0.272*** (0.096)
Constant	4.856*** (0.496)	5.048*** (0.444)	4.856*** (0.503)

Observations	962	962
R2	0.734	0.955
Adjusted R2	0.731	0.954
Residual Std. Error (df = 951)	1.717	0.767
F Statistic (df = 10; 951)	261.861***	2,014.604***

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

L'analyse des résultats montre des variations notables entre les trois méthodes d'estimation utilisées (MCO, MCG et MCO avec correction de White). Tout d'abord, le coefficient de détermination passe de 73,4 % avec les MCO à 95,4 % avec les MCG, suggérant une très forte amélioration de l'ajustement du modèle. Cette augmentation importante peut être interprétée comme une meilleure prise en compte des spécificités des données, notamment grâce aux pondérations liées à l'hétéroscédasticité. Cependant, elle pourrait également indiquer un risque de sur-ajustement qui nécessiterait une validation externe, ce qui ici nous empêche de tirer de réelles conclusions.

Les coefficients des variables restent relativement stables entre les trois modèles, mais les erreurs standards sont notablement réduites avec les MCG, renforçant la précision des estimations. Par exemple, les erreurs associées à des variables comme `temps_tgv_paris` ou `metropole1` diminuent considérablement sous l'effet des pondérations. Enfin, la correction de White appliquée aux MCO n'altère pas les coefficients estimés mais ajuste légèrement les erreurs standards, ce qui modifie subtilement la significativité de certaines variables. Cependant, il faut tout de même signifier que toutes les variables restent bel et bien significatives dans les trois modèles.

5 Conclusion

Notre analyse des déterminants des loyers au mètre carré dans les métropoles françaises a permis de mettre en lumière plusieurs facteurs clés. Le revenu médian des habitants, la proportion de nouveaux ménages, la densité de population, l'accessibilité géographique et la part de logements vacants ont tous un impact significatif sur les loyers. En particulier, la localisation dans la métropole du Grand Paris est le facteur le plus déterminant, avec des loyers nettement plus élevés que dans les autres métropoles.

Cette analyse permet d'observer des dynamiques globales sur le loyer au mètre carré, mais il est important de garder à l'esprit les limites de notre travail. Les données utilisées, bien que récentes, sont des estimations basées sur des annonces immobilières, qui peuvent ne pas refléter exactement la réalité des loyers en place, du fait de l'estimation mais surtout du fait de l'effet d'annonce (les villes ayant peu d'annonces ont un estimateur beaucoup moins robustes, et les autres sont calculés uniquement sur les annonces en vigueur et pas de tous les biens loués). De plus, certaines variables auraient pu apporter beaucoup à notre modèle, comme la part de bénéficiaires des APL : est-ce que cette allocation fait monter les loyers et à quel point? Il est nécessaire de considérer ces limites pour permettre l'avancée des travaux de recherches économétriques sur le sujet, afin de donner les clefs d'amélioration.