

Le Big Data peut-il améliorer notre connaissance du marché de la location ? Estimer le coût de la densité avec des données de loyers

Guillaume Chapelle² Jean-Benoît Eyméoud³

²LIEPP, Sciences Po

³Banque de France - ACPR

LIEPP, Sciences Po

Séminaire Fourgeaud

5 juin 2019

Feuille de route

- En exploitant les possibilités offertes par internet, on construit une base de données inédite rassemblant des millions d'offres de location immobilières
- En comparant données administratives et internet, on montre que les données en ligne ne souffrent d'aucun biais systématique
- On utilise les données internet pour estimer l'élasticité des coûts d'agglomération par rapport à la taille des villes
- On montre que cette élasticité est plus faible lorsqu'elle est estimée avec les loyers plutôt qu'avec les prix

Outline

- 1 Introduction
- 2 Utiliser internet pour étudier le marché de l'immobilier
 - Référencement des données disponibles
 - Collecter des données internet
 - Représentativité et biais statistiques
- 3 Estimer l'élasticité des coûts d'agglomération avec les loyers
 - Coût d'agglomération et coût du logement
 - Stratégie empirique
 - Résultats

Données disponibles pour étudier la location

- Peu de données sont disponibles pour étudier le parc locatif privé:
 - Les **observatoires des loyers** étudient quelques grandes villes mais ne communiquent pas leur données
 - L'**Enquête Logement** fournit des données détaillées tous les cinq ans mais uniquement représentatives à l'échelle nationale
 - L'**Enquête Loyers et Charges** fournit des données trimestrielles représentatives à l'échelle nationale
 - **CLAMEUR**, un regroupement de professionnels de l'immobilier, fournit des données annuelles pour quelques grandes villes

Méthode pour trouver un appartement

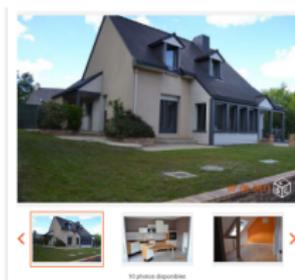
	Non meublé	Meublé	Total
Privé (internet ou journaux)	37	42	37
Agence de location	41	22	39
Réseau personnel	19	20	19
Employeur	1	3	2
Services sociaux	2	10	3
Autres	0	3	1
Total	100	100	100

Calculs des auteurs à partir de l'Enquête Logement 2013

Outline

- 1 Introduction
- 2 Utiliser internet pour étudier le marché de l'immobilier
 - Référencement des données disponibles
 - **Collecter des données internet**
 - Représentativité et biais statistiques
- 3 Estimer l'élasticité des coûts d'agglomération avec les loyers
 - Coût d'agglomération et coût du logement
 - Stratégie empirique
 - Résultats

Web scraping



11 photos disponibles

Maison en ligne le 19 décembre à 16:43

MÉTADONNÉES

Loyer mensuel	1 295 € charges comprises
Site	Repa (transaction sur fiche REPA)
Type de bien	Maison
Pièces	6
Mobilier / Non-mobilier	Non meublé
Surface	160 m ²
DPS	€ 106 11 à 205
Classe énergie	C (de 91 à 130)

Description :

A LOUER AU 15 décembre prochain

Maison (non meublée) située à NAINVAL, CHEFFELLES SUR YSCHE (à 10 kms de Rennes Sud) par 4 axes direction Nantes

Régime foncier en acte 2015.

Surface habitable 140 m² sur terrain de 300 M²

Chambre suite

Accès à un jardin avec piscine, séjour salon avec cheminée, cuisine aménagée, une chambre avec salle de bain (matras), douche et baignoire (avec supports et garage)

4 Hggs / 3 Chambres sur parquet + un grand salon sur garage (peuvent être utilisés en chambre, salle de jeu, lingerie ou bureau, café, bureau etc.)

Chauffage au gaz.

Niveau de confort excellent, haut.

Proximité de Rennes à 40 minutes.

Avec un grand TGV à 10 minutes.

Rue pour 1000€ en option avec 1000€ à 100 mètres.

Construite, habitée et protégée par un permis.

2 200 € TTC



Description de Appartement : Paris 16ème

Paris 16^{ème}, au pied du métro Solignac - A proximité du Bassin de la Vierge et du Paulsen St. Martin, rue de l'aparteur. En étage élevé avec vue panoramique sur les toits de Paris. Bien orienté, lumineux, aménagé. 4 pièces meublées comprenant une entrée générale, deux entrées un vaste double séjour sur rue, deux grandes chambres au calme sur toit, une suite avec dressing entièrement équipé, salle de bains avec baignoire et douche, WC séparés, Pergola, meubles et électroménagers. Prestations de qualité. Environnement agréable, bien desservi avec de nombreuses commerces à proximité.

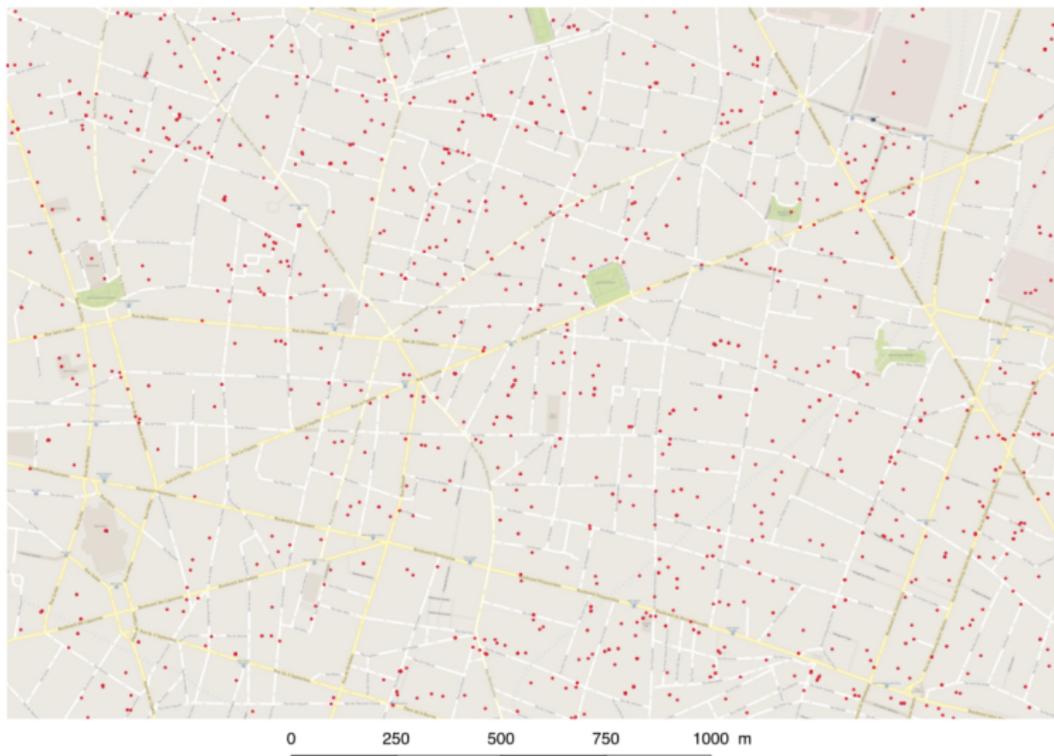
Surface de 64 m ²	4 pièces
4 Pièces	3 Chambres
1 Salle de bains	1 Terrasse
Nombre pièces	Meublé
Parquet	Armoire
Cheminée	Chauffage individuel gaz
Cuisine avec cuisine équipée	Interphone
Douche	Garage
Porte	Salle de bain
Couleur	DPE : D (176)
DPE : E (145)	Meublé ou Non : Meublé
Garantie : 3000 €	Charges : 150 €
Disponibilité : 01/11/2018	

Vous désirez en savoir plus ? Contacter l'agence
 01 1 45 15 81 82

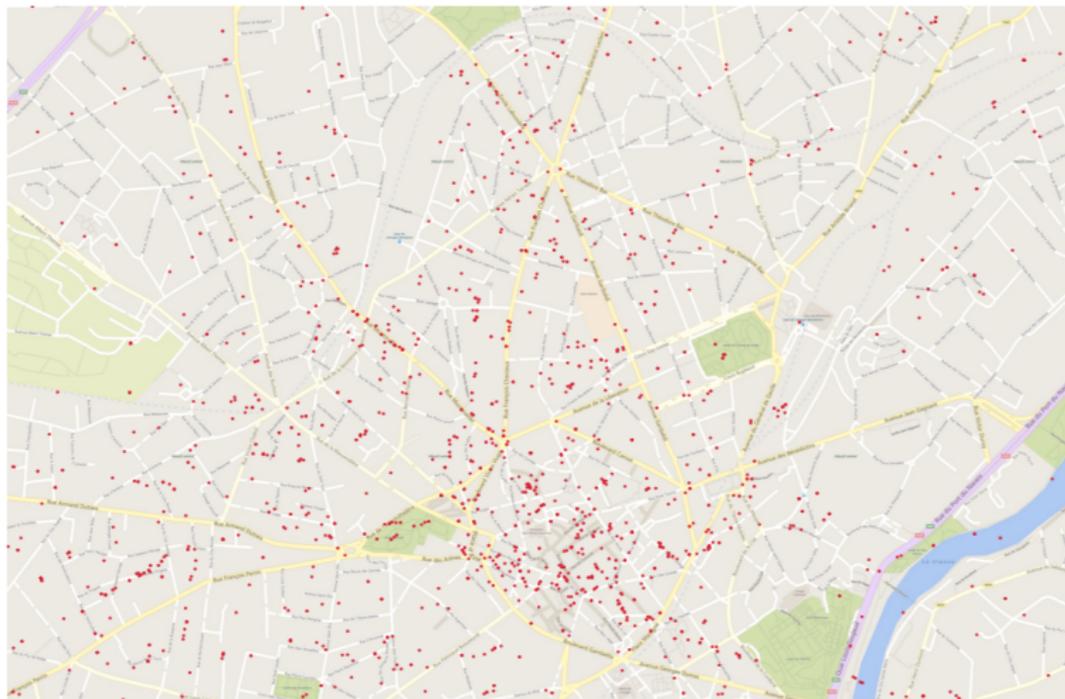
Collection des données: les variables

- La plupart des variables sont extraites directement dans la structure HTML: prix, surface, nombre de chambres, géolocalisation, indice énergétique, ...
- D'autres sont extraites en analysant les descriptions des offres: étage, parking, ...
- Description des variables
 - Surface and nombre de chambres
 - Prix et dépenses
 - Etage et autres aménités
 - Efficacité énergétique
 - Localisation

Annonces postées *Grands Boulevards*

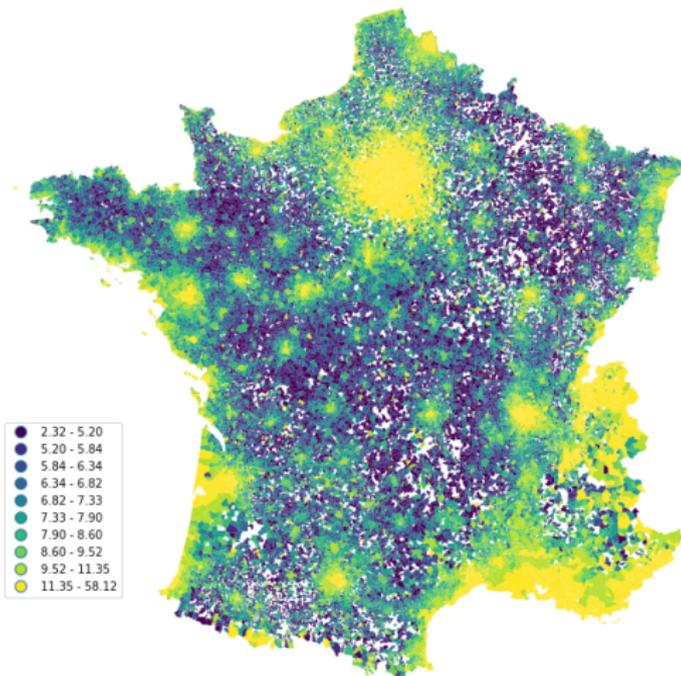


Annonces postées *Limoges*



0 250 500 750 1000 m

Loyer moyen des municipalités françaises (euros/m²)



Outline

- 1 Introduction
- 2 Utiliser internet pour étudier le marché de l'immobilier
 - Référencement des données disponibles
 - Collecter des données internet
 - Représentativité et biais statistiques
- 3 Estimer l'élasticité des coûts d'agglomération avec les loyers
 - Coût d'agglomération et coût du logement
 - Stratégie empirique
 - Résultats

Question méthodologique

- En théorie, les prix postés ne sont pas les prix signés:
 - Desgranges and Wasmer (2000) montrent que lorsque le pouvoir de négociation du locataire est proche de zéro, le loyer signé converge vers le loyer posté s'il existe une compétition en prix entre les propriétaires.
 - Les sites en ligne fournissant à chaque instant une information gratuite permettant de connaître l'état du marché, la concurrence en prix pousse les propriétaires à révéler leur prix de réserve.

Tester la qualité de la base

- L'idée générale consiste à comparer les données internet aux données administratives
- Deux tests sont mis en oeuvre pour tester la qualité des données:
 - Comparer les prix au mètre carré
 - Tester une différence statistique de distributions
 - Comparer des indices hédoniques
 - Comparer des indices de prix une fois les caractéristiques du parc de logement prises en compte

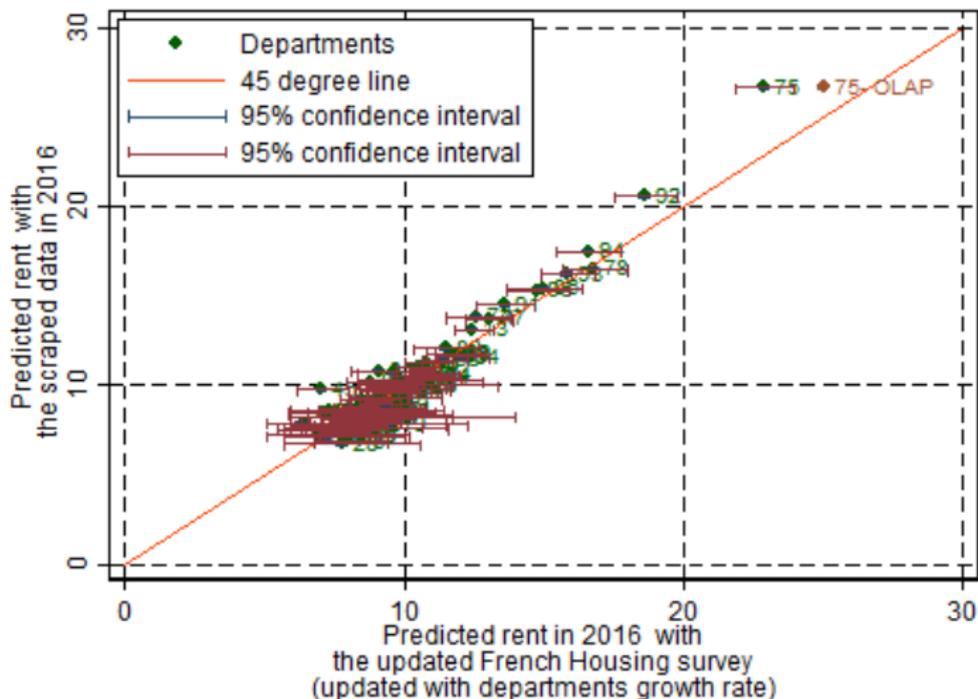
Tester la qualité de la base: l'approche hédonique

- On estime un modèle hédonique pour chaque régions et départements français:

$$\ln(r_{i,s}) = \ln(r_{m(i)}^{ref}) + X_i\beta + u_i \quad (1)$$

- $\ln(r_{i,s})$ est le loyer par mètre carré de l'observation i dans la strate s
- X_i est un vecteur de caractéristiques hédoniques commun aux deux bases de données (surface, nombre de chambre, autres aménités (meublé, jardin, balcon, garage, ascenseur))
- β est estimé pour chaque strate s
- On teste l'absence de biais en comparant $\ln(r_{m(i)}^{ref})$ pour chaque départements et régions lorsqu'il est calculé avec les données internet et administratives.

Tester la qualité de la base: les effets fixes départements



Champs d'utilisation de la base

- Recherche fondamentale
 - Etude des mécanismes urbains à l'échelle micro
 - Etude des dynamiques de prix et formation des bulles
- Evaluation des politiques publiques
 - Contrôle des loyers
 - Impact de Airbnb
- Mise en oeuvre des politiques publiques
 - Indicateurs de stabilité financière
 - Calibration des réformes du marché de l'immobilier (APL, taxe foncière, ...)

Outline

- 1 Introduction
- 2 Utiliser internet pour étudier le marché de l'immobilier
 - Référencement des données disponibles
 - Collecter des données internet
 - Représentativité et biais statistiques
- 3 Estimer l'élasticité des coûts d'agglomération avec les loyers
 - Coût d'agglomération et coût du logement
 - Stratégie empirique
 - Résultats

L'économie urbaine et l'apparition des villes

- Fujita and Thisse (2002) montrent que les **villes** résultent d'un **équilibre** entre:
 - **Economies d'agglomération** (gains de productivité, externalités Marshalliennes)
 - **Coûts d'agglomération** (congestion, pollution, augmentation du prix de la terre)
- Les mécanismes d'économies d'agglomération ont été largement étudiés dans la littérature (Combes and Gobillon (2015)).
- Dans un article précurseur Combes, Duranton et Gobillon (2018) estiment l'élasticité des coûts urbains à la population
 - ⇒ Pour estimer cette élasticité, les auteurs utilisent les prix de l'immobilier

Le coût d'agglomération

- Combes, Duranton, and Gobillon (2018) montrent que l'élasticité des coûts urbains à la population ϵ_N^{UC} peut s'écrire de la manière suivante:

$$\epsilon_N^{UC} = s_E^h \times \epsilon_N^{\bar{P}} \quad (2)$$

- Où s_E^h est la part des dépenses de logement dans le budget des ménages et $\epsilon_N^{\bar{P}}$ est l'élasticité du coût du logement au centre de l'aire urbaine par rapport à la taille de la population:

$$\epsilon_N^{\bar{P}} = \frac{\partial \ln(C)}{\partial \ln(Pop)} \quad (3)$$

- Empiriquement, les auteurs utilisent les prix du logement (P) pour mesurer l'élasticité et mesurent ainsi $\frac{\partial \ln(P)}{\partial \ln(Pop)}$

Le coût du logement avec les prix et les loyers

- Le logement est un bien spécifique car il remplit une double fonction:
 - bien de consommation
 - capital économique
- Son prix ne reflète pas seulement son coût mais incorpore les anticipations des agents et dépend de paramètres économiques potentiellement locaux (Himmelberg, Mayer, and Sinai (2005))
- A l'inverse, les loyers reflètent exclusivement le coût du logement
- Estimer l'élasticité des coûts urbains en utilisant les prix ou les loyers change la perspective des coûts d'agglomération

Outline

- 1 Introduction
- 2 Utiliser internet pour étudier le marché de l'immobilier
 - Référencement des données disponibles
 - Collecter des données internet
 - Représentativité et biais statistiques
- 3 Estimer l'élasticité des coûts d'agglomération avec les loyers
 - Coût d'agglomération et coût du logement
 - **Stratégie empirique**
 - Résultats

Mesurer le loyer au coeur de la ville

- On estime un modèle hédonique spatial pour chacune des municipalités françaises
- Les indices hédoniques sont ensuite utilisés pour estimer des loyers comparables au centre de chaque aire urbaine ($C_{c(m)}^r$):

$$\ln(r_m) = C_{c(m)}^r - \delta_{c(m)}^r \times \ln(D_m) + X_m \alpha^r + \mu_m \quad (4)$$

⇒ Se concentrer sur le centre de l'aire urbaine permet d'évacuer la question des coûts de transport

- X_m est un vecteur d'aménités et de caractéristiques municipales

Stratégie empirique

- On veut estimer l'élasticité suivante $\epsilon_{N}^{\bar{P}} = \frac{\partial \ln(R)}{\partial \ln(Pop)}$
- On se base sur la méthodologie développée par Combes, Duranton, and Gobillon (2018) et on estime :

$$\hat{C}_c^r = \epsilon_{N}^{\bar{P}} \times \ln(Population_c) + \gamma \times \ln(Land_Area_c) + Z_c \beta + \eta_c \quad (5)$$

- Où \hat{C}_c^r est le logarithme du loyer au centre de l'agglomération
- Le coefficient de $\ln(Population)$ est $\epsilon_{N}^{\bar{P}}$
- Z_c est un vecteur de variables de contrôle

Endogénéité

- *Population* and *Land Area* sont potentiellement sujettes à un biais de causalité inverse
- On utilise plusieurs instruments:
 - Aménités naturelles: température en janvier, nombre d'étoiles des hotels, budget des hotels
 - Données historiques de population et de densités: population au XIXème siècle (Motte, Séguy, and Théré (2003))
 - La complémentarité des instruments permet de tester leur puissance et leur exogénéité

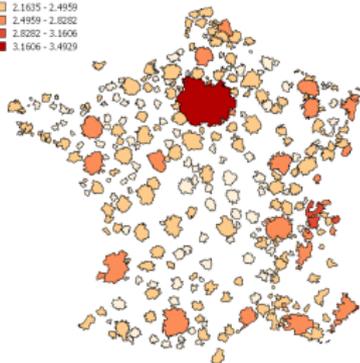
Outline

- 1 Introduction
- 2 Utiliser internet pour étudier le marché de l'immobilier
 - Référencement des données disponibles
 - Collecter des données internet
 - Représentativité et biais statistiques
- 3 Estimer l'élasticité des coûts d'agglomération avec les loyers
 - Coût d'agglomération et coût du logement
 - Stratégie empirique
 - Résultats

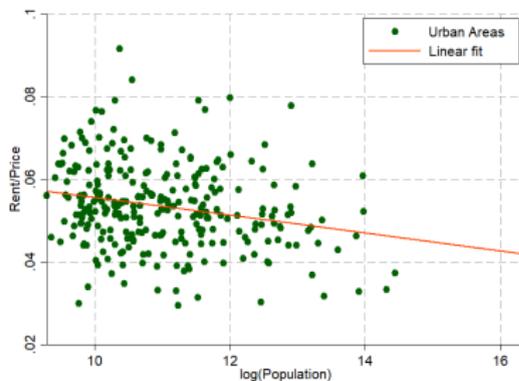
Loyers dans le centre des aires urbaines françaises

Figure: Loyer et ratio loyer / prix au centre des aires urbaines

Log(Rent per square meter in CBD)



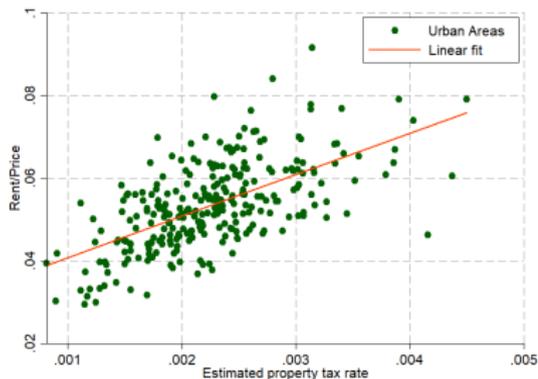
(a) Log du loyer au mètre carré par aires urbaines



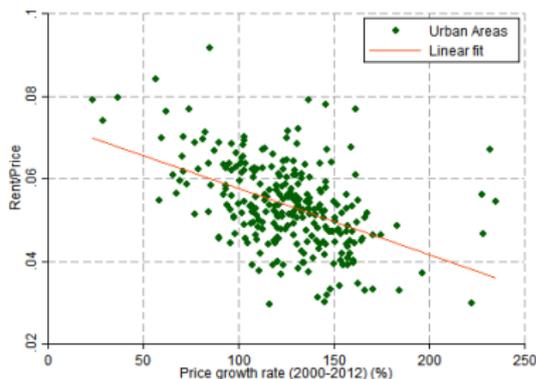
(b) Ratio loyer / prix et taille des aires urbaines

Ratio prix / loyer et paramètres économiques locaux

Figure: Loyer et ratio loyer / prix au centre des aires urbaines



(a) Ratio loyer / prix et taux de taxation



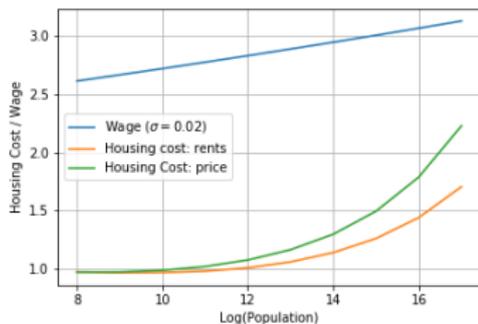
(b) Ratio loyer / prix et taux de croissance des prix de l'immobilier

Résultats, OLS

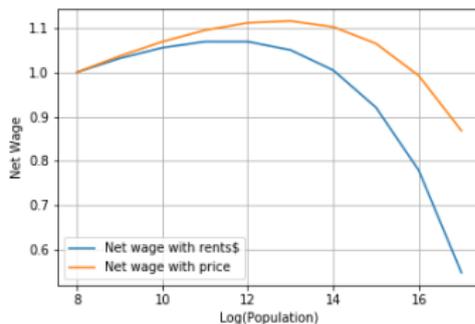
- Les coefficients estimés avec les données de loyers sont stables, robustes et systématiquement inférieurs à ceux estimés avec les données de prix
 - Résultats OLS
- L'élasticité retenue par Combes, Duranton, and Gobillon (2018) est de 0,208, la nôtre est de 0,125
- Cette différence est statistiquement significative
- Les coefficients estimés par les variables instrumentales sont du même ordre de grandeur
 - Results 2SLS

Calibration sur données françaises

Figure: Salaire, coût du logement et salaire net en France



(a) Salaire et coût du logement



(b) Salaire net

Conclusion

- Internet fournit des données de qualité à faible coût
- L'élasticité des coûts urbains par rapport à la taille des villes est plus faible lorsqu'elle est mesurée avec les loyers plutôt qu'avec les prix
- Le coût de l'écart à la taille optimale est plus faible en termes de bien-être en considérant les loyers plutôt que les prix
- Cette différence s'explique par la prise en compte d'autres variables économiques dans les prix de l'immobilier
- Les deux élasticités ne sont pas mutuellement exclusives mais répondent à deux exercices de pensée différents

The cost of agglomeration, OLS Back

Table: OLS regression

	(1)	(2)	(3)	(4)	(5)	(6)
First-step	Only Fixed Effects			Full set of controls		
Controls	N	Y	Ext.	N	Y	Ext.
Panel A: Original specification using price fixed effects						
Log(Population)	0.217*** (0.0160)	0.176*** (0.0111)	0.224*** (0.0223)	0.252*** (0.0198)	0.208*** (0.0137)	0.304*** (0.0284)
Log(Land area)	-0.151*** (0.0168)	-0.153*** (0.0107)	-0.224*** (0.0231)	-0.143*** (0.0185)	-0.152*** (0.0135)	-0.276*** (0.0294)
N	1937	1937	1937	1937	1937	1937
R ²	0.352	0.645	0.720	0.403	0.659	0.726
Panel B: substituting price fixed effect with rent fixed effect						
Log(Population)	0.177*** (0.0164)	0.146*** (0.0132)	0.193*** (0.0245)	0.170*** (0.0152)	0.147*** (0.0137)	0.223*** (0.0250)
Log(Land area)	-0.0314** (0.0135)	-0.0419*** (0.0112)	-0.104*** (0.0231)	-0.0719*** (0.0133)	-0.0818*** (0.0122)	-0.171*** (0.0234)
N	277	277	277	277	277	277
R ²	0.625	0.750	0.790	0.544	0.668	0.748
Panel C: and updating population and control variables						
Log(Population)	0.172*** (0.0124)	0.138*** (0.0128)	0.0920*** (0.0290)	0.167*** (0.0117)	0.138*** (0.0126)	0.101*** (0.0285)
Log(Land area)	-0.0289*** (0.0145)	-0.0385*** (0.0133)	-0.0133 (0.0287)	-0.0713*** (0.0137)	-0.0787*** (0.0130)	-0.0586** (0.0282)
N	277	277	277	277	277	277
R ²	0.619	0.709	0.783	0.541	0.621	0.717
Panel D: and extending the sample						
Log(Population)	0.169*** (0.0135)	0.131*** (0.0125)	0.109*** (0.0213)	0.159*** (0.0129)	0.125*** (0.0127)	0.117*** (0.0214)
Log(Land Area)	-0.0511*** (0.0137)	-0.0527*** (0.0116)	-0.0363 ^c (0.0210)	-0.0817*** (0.0132)	-0.0837*** (0.0117)	-0.0801*** (0.0211)
N	352	352	352	352	352	352
R ²	0.446	0.640	0.778	0.355	0.535	0.719

Standard errors in parentheses

^{*} p<0.10, ^{**} p<0.05, ^{***} p<0.01

The cost of agglomeration, 2SLS Back

Table: 2SLS regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Log price per m2, with first step and second step controls								
log(population)	0.204*** (0.0180)	0.207*** (0.0177)	0.204*** (0.0180)	0.186*** (0.0157)	0.187*** (0.0156)	0.287*** (0.0472)	0.201*** (0.0141)	0.287*** (0.0474)
log(surface)	-0.140*** (0.0162)	-0.142*** (0.0159)	-0.140*** (0.0162)	-0.114*** (0.0145)	-0.116*** (0.0144)	-0.218*** (0.0573)	-0.110*** (0.0160)	-0.218*** (0.0576)
N	1937	1937	1937	1937	1937	1937	1937	1937
Weak	215.9	187.2	145.4	239.0	240.4	16.3	346.4	11.0
Overld	.	0.49	0.67	0.72	0.10	.	0.05	0.72
Panel B: Log rent per m2, with first step controls and second step controls								
log(population)	0.153*** (0.0247)	0.157*** (0.0246)	0.156*** (0.0252)	0.143*** (0.0219)	0.142*** (0.0219)	0.191*** (0.0674)	0.137*** (0.0184)	0.211** (0.0952)
log(surface)	-0.0725*** (0.0159)	-0.0755*** (0.0157)	-0.0749*** (0.0161)	-0.0565*** (0.0142)	-0.0572*** (0.0142)	-0.126 (0.0802)	-0.0599*** (0.0154)	-0.150 (0.114)
N	277	277	277	277	277	277	277	277
Weak	93.7	84.2	61.7	97.2	98.5	7.5	61.7	5.0
Overld	.	0.21	0.07	0.34	.	.	0.38	0.13
Panel C: Log rent per m2, updated population, surface and control variables								
log(population)	0.128*** (0.0264)	0.134*** (0.0261)	0.128*** (0.0266)	0.118*** (0.0229)	0.119*** (0.0230)	0.203*** (0.0772)	0.134*** (0.0190)	0.191*** (0.0928)
log(surface)	-0.0652*** (0.0162)	-0.0698*** (0.0161)	-0.0654*** (0.0164)	-0.0524*** (0.0143)	-0.0541*** (0.0143)	-0.143 (0.0882)	-0.0526*** (0.0158)	-0.131 (0.106)
N	277	277	277	277	277	277	277	277
Weak	76.30	67.88	50.47	73.65	73.79	6.21	35.69	4.39
Overld	.	0.12	0.24	0.25	0.30	.	0.27	0.20
Panel D: Log rent per m2, extended sample								
log(Population)	0.0908 ^a (0.0318)	0.108 ^a (0.0297)	0.0823 ^b (0.0334)	0.0769 ^a (0.0291)	0.0798 ^a (0.0283)	0.275 ^b (0.119)	0.115 ^a (0.0245)	0.0279 (0.178)
log(Surface)	-0.0628 ^a (0.0162)	-0.0748 ^a (0.0161)	-0.0570 ^b (0.0164)	-0.0501 ^a (0.0143)	-0.0522 ^a (0.0143)	-0.210 ^b (0.0882)	-0.0539 ^a (0.0158)	0.0296 (0.106)
N	352	352	352	352	352	352	352	352
Weak	51.91	56.79	37.71	56.96	52.97	1.79	30.54	2.97
Overld	.	0.03	0.25	0.27	0.24	.	0.13	0.09
Urban population in 1831	Y	Y	Y	Y	Y	N	N	N
Urban population in 1851	Y	Y	Y	Y	N	N	N	N
Urban srca in 1881	N	Y	N	N	N	N	N	N
Urban pop. density in 1881	N	Y	N	N	N	N	N	N
January temperature	N	N	Y	Y	N	N	N	Y
Number of hotel rooms	N	N	N	N	N	Y	Y	Y
Share of one-star hotel rooms	N	N	N	N	Y	Y	Y	Y

Standard errors in parentheses
^a p<0.10, ^{**} p<0.05, ^{***} p<0.01

Type of unit, surface and number of rooms [Back](#)

Table: Type of units , number of rooms and surface

	count	mean	std	min	25%	50%	75%	max
surface	4225940.0	55.9	31.1	1.0	34.0	50.0	70.0	1080.0
single unit (%)	4225940.0	15.8	36.4	0.0	0.0	0.0	0.0	100.0
Rooms (%): 01	4225940.0	20.7	40.5	0.0	0.0	0.0	0.0	100.0
Rooms (%): 02	4225940.0	32.5	46.9	0.0	0.0	0.0	100.0	100.0
Rooms (%): 03	4225940.0	26.2	44.0	0.0	0.0	0.0	100.0	100.0
Rooms (%): 04	4225940.0	12.6	33.2	0.0	0.0	0.0	0.0	100.0
Rooms (%): 05	4225940.0	5.5	22.8	0.0	0.0	0.0	0.0	100.0
Rooms (%): 6+	4225940.0	2.5	15.7	0.0	0.0	0.0	0.0	100.0

Price, expenditures and type of lease [Back](#)

Table: Price, expenditures and type of lease

	count	mean	std	min	25%	50%	75%	max
Rent	4225940.0	646.2	396.7	8.0	445.0	561.0	730.0	65000.0
Rent per square meter	4225940.0	13.3	6.7	2.0	8.9	11.7	16.0	100.0
Expenditures : Included	4225940.0	72.3	44.8	0.0	0.0	100.0	100.0	100.0
Expenditures : Not Included	4225940.0	5.9	23.6	0.0	0.0	0.0	0.0	100.0
Expenditures : Unknown	4225940.0	21.8	41.3	0.0	0.0	0.0	0.0	100.0
Expenditures	1594691.0	58.3	51.9	0.0	30.0	45.0	72.0	3705.0
Collective heating (%)	4225940.0	3.5	18.4	0.0	0.0	0.0	0.0	100.0
Hot water (%)	4225940.0	0.2	4.1	0.0	0.0	0.0	0.0	100.0
Trash collection (%)	4225940.0	4.7	21.1	0.0	0.0	0.0	0.0	100.0
Furnished (%): No	4225940.0	81.1	39.2	0.0	100.0	100.0	100.0	100.0
Furnished (%): Yes	4225940.0	18.9	39.2	0.0	0.0	0.0	0.0	100.0

Table: Floors and other amenities

	count	mean	std	min	25%	50%	75%	max
Floor								
Floor (%): 0.0	4225940.0	8.9	28.5	0.0	0.0	0.0	0.0	100.0
Floor (%): 1.0	4225940.0	10.8	31.1	0.0	0.0	0.0	0.0	100.0
Floor (%): 2.0	4225940.0	7.9	27.0	0.0	0.0	0.0	0.0	100.0
Floor (%): 3.0	4225940.0	4.1	19.9	0.0	0.0	0.0	0.0	100.0
Floor (%): 4.0	4225940.0	2.0	13.9	0.0	0.0	0.0	0.0	100.0
Floor (%): 5.0	4225940.0	0.9	9.7	0.0	0.0	0.0	0.0	100.0
Floor (%): 6+	4225940.0	1.0	10.0	0.0	0.0	0.0	0.0	100.0
Floor (%): Unknown floor	4225940.0	59.6	49.1	0.0	0.0	100.0	100.0	100.0
Floor (%): last floor	4225940.0	4.7	21.1	0.0	0.0	0.0	0.0	100.0
Amenities								
Elevator (%)	4225940.0	14.0	34.7	0.0	0.0	0.0	0.0	100.0
Double glazing (%)	4225940.0	9.5	29.3	0.0	0.0	0.0	0.0	100.0
Kitchen with equipment (%)	4225940.0	35.5	47.9	0.0	0.0	0.0	100.0	100.0
garage (%)	4225940.0	46.1	49.8	0.0	0.0	0.0	100.0	100.0
Garden (%)	4225940.0	17.4	37.9	0.0	0.0	0.0	0.0	100.0
Balcony (%)	4225940.0	35.6	47.9	0.0	0.0	0.0	100.0	100.0

Table: Energy consumption and greenhouse gas emission

	count	mean	std	min	25%	50%	75%	max
Energy (%):A	4225940.0	3.6	18.5	0.0	0.0	0.0	0.0	100.0
Energy (%):B	4225940.0	5.8	23.4	0.0	0.0	0.0	0.0	100.0
Energy (%):C	4225940.0	13.2	33.8	0.0	0.0	0.0	0.0	100.0
Energy (%):D	4225940.0	25.6	43.6	0.0	0.0	0.0	100.0	100.0
Energy (%):E	4225940.0	17.8	38.3	0.0	0.0	0.0	0.0	100.0
Energy (%):F	4225940.0	6.2	24.2	0.0	0.0	0.0	0.0	100.0
Energy (%):G	4225940.0	2.4	15.4	0.0	0.0	0.0	0.0	100.0
Energy (%):H	4225940.0	0.0	1.8	0.0	0.0	0.0	0.0	100.0
Energy (%):I	4225940.0	0.0	1.5	0.0	0.0	0.0	0.0	100.0
Energy (%):None	4225940.0	8.9	28.5	0.0	0.0	0.0	0.0	100.0
Energy (%):V	4225940.0	4.8	21.4	0.0	0.0	0.0	0.0	100.0
GES (%):A	4225940.0	4.3	20.4	0.0	0.0	0.0	0.0	100.0
GES (%):B	4225940.0	13.6	34.3	0.0	0.0	0.0	0.0	100.0
GES (%):C	4225940.0	21.3	40.9	0.0	0.0	0.0	0.0	100.0
GES (%):D	4225940.0	13.9	34.6	0.0	0.0	0.0	0.0	100.0
GES (%):E	4225940.0	10.4	30.6	0.0	0.0	0.0	0.0	100.0
GES (%):F	4225940.0	4.8	21.5	0.0	0.0	0.0	0.0	100.0
GES (%):G	4225940.0	1.9	13.8	0.0	0.0	0.0	0.0	100.0
GES (%):H	4225940.0	0.0	2.1	0.0	0.0	0.0	0.0	100.0
GES (%):I	4225940.0	0.0	1.7	0.0	0.0	0.0	0.0	100.0
GES (%):None	4225940.0	11.2	31.5	0.0	0.0	0.0	0.0	100.0
GES (%):V	4225940.0	4.4	20.4	0.0	0.0	0.0	0.0	100.0

Table: Precision of the geocoding

Geocoding (%) Unknown:	0.0
Geocoding (%): address	17.9
Geocoding (%): browser	0.0
Geocoding (%): city	59.0
Geocoding (%): device	0.2
Geocoding (%): neighborhood	12.8
Geocoding (%): user	10.0
Geocoding (%): zipcode	0.0
