

Logement et discrimination en économie : une approche empirique mêlant expérience naturelle et Big Data

Résumé de la thèse de Jean-Benoît Eyméoud

Octobre 2018

1 L'élasticité de l'offre de logements et les coûts d'agglomération

L'élasticité de l'offre de logements est depuis longtemps un paramètre fondamental pour les économistes car elle détermine le succès des politiques du logement et joue un rôle important dans la compréhension des cycles immobiliers. Pourtant, elle reste largement méconnue, surtout en comparaison de l'élasticité de la demande de logements. Cette méconnaissance s'explique tant par la difficulté de documenter les décisions de construction de logements que le manque de données relatives au secteur de la construction. Néanmoins, les maigres contributions tentant d'estimer l'élasticité de l'offre de logements ont permis de souligner son caractère hétérogène. Cet article comble ce manque pour la France. En rassemblant des données microéconomiques inédites et en comparant les méthodes utilisées par les macroéconomistes et économistes urbains, nous calculons des élasticités du marché de l'immobilier locales ainsi que leurs déterminants.

La littérature montre en effet que la macroéconomie et l'économie urbaine ont tenté de mesurer l'élasticité de l'offre de logements. Cependant, les deux littératures restent mal reliées: si les deux courants tentent d'estimer un même objet, leur approche semble différente. L'un des objectifs de cet article est de clarifier la distinction entre les deux approches. Pour nous, la macroéconomie estime deux quantités: une élasticité de très court terme qui correspond à l'ajustement des prix suite à un choc de demande et une élasticité de moyen terme qui correspond à la réaction du marché de la construction. Parce que la construction d'un logement prend du temps, l'élasticité de très court terme est plus faible que celle de long terme. De plus, dans le cadre macroéconomique, les entreprises sont preneuses de prix et ajustent leur niveau de production (mises en chantier) de manière à ce que le coût marginal soit égal au prix. Les mises en chantier sont ainsi une conséquence des variations de prix des maisons. Cette littérature peut facilement faire des comparaisons entre pays comme dans Meen (2002) ou dans Andrews, Sánchez, and Johansson (2011) et met en évidence l'existence d'un écart transatlantique: les pays européens sont beaucoup plus inélastiques que les États-Unis. L'une des limites de ces contributions est néanmoins le fait que l'espace et la croissance urbaine sont des impensés.

Les économistes urbains prennent la consommation d'espace comme point de départ de leur analyse. Pour eux, la croissance urbaine est le résultat d'externalités spatiales positives qui poussent les entreprises et les ménages à se concentrer sur une partie du territoire. Cependant, la croissance des villes n'est pas infinie car leur expansion est freinée par des coûts d'agglomération. L'exemple traditionnel développé dans la littérature est celui du modèle monocentrique (voir Combes, Duranton, and Gobillon (2016) ou Saiz (2010) pour une illustration) où la croissance urbaine pousse le prix de la terre et donc le prix des logements à la hausse. Les économistes urbains adoptent une perspective de long terme où les prix sont le résultat de la croissance urbaine. Jusqu'à présent, cette littérature s'est surtout concentrée sur les États-Unis et a montré la forte hétérogénéité de l'élasticité de l'offre de logements entre

les différentes zones métropolitaines¹ américaines. En Europe, de rares travaux ont tenté d'estimer l'élasticité de l'offre de logements: Hilber and Vermeulen (2016) ont estimé l'élasticité-prix du revenu pour le Royaume-Uni tandis que Combes, Duranton, and Gobillon (2016) ont estimé l'élasticité du prix du terrain par rapport à la taille des villes en France. Toutefois, ces études utilisent des spécifications et des variables différentes qui fournissent des estimations difficilement comparables aux estimations américaines.

Dans cet article, nous essayons de faire le lien entre ces deux courants de littérature. En considérant l'absence d'espace et de croissance urbaine, les macroéconomistes s'intéressent à ce que l'on choisit d'appeler la marge intensive de l'offre de logements, c'est à dire l'offre de logements en considérant que la taille de la ville reste fixe. Ils étudient l'élasticité liée aux coûts marginaux de production des promoteurs et s'intéressent aux mises en chantier. Par conséquent, ils doivent utiliser des chocs de demande de court terme et estimer une équation d'offre de court terme en utilisant les prix des logements neufs. A l'inverse, en considérant la croissance urbaine comme un phénomène décennal, les économistes urbains s'intéressent à ce que l'on choisit d'appeler la marge extensive de l'offre de logements, c'est à dire l'offre de logements lorsque la taille des villes change. Comme les prix des logements sont la conséquence de la croissance de la ville, ils doivent estimer l'équation inverse de l'offre en utilisant des chocs de long terme comme dans Combes, Duranton, and Gobillon (2016) et Saiz (2010).

En rassemblant une quantité importante de données de prix des logements neufs et anciens ainsi que les mises en chantier, les contraintes géographiques et réglementaires pour les principales zones urbaines françaises, nous calculons et décomposons les élasticités des offres extensives et intensives de logements.

Pour traiter le biais de simultanéité inhérent au calcul d'élasticités, nous utilisons des chocs de demande et plus précisément des chocs de type Bartik. Nous nous servons également du climat et d'un instrument utilisé en macroéconomie (voir Monnet and Wolf (2016)): le nombre de naissances pris vingt ans auparavant. L'impact de la réglementation sur les prix des logements est appréhendé de manière causale en utilisant des lois nationales permettant d'instrumenter un indice de réglementation local de l'immobilier défini dans l'article.

Au final, nous montrons que des séries de prix différentes (prix de l'ancien et prix du neuf) donnent des élasticités différentes qui ont des déterminants différents. L'élasticité intensive de l'offre de logements, basée sur les prix du neuf est principalement expliquée par la réglementation. L'élasticité extensive de l'offre de logements, basée sur les prix de l'ancien est expliquée par la part des terrains déjà aménagés et le niveau de réglementation.

Nos résultats confirment l'existence d'un écart transatlantique, la France disposant

1. metropolitan statistical area

d'une élasticité de long terme par rapport au prix de l'ancien proche de 0,3 contre 1,5 pour les Etats-Unis (voir Saiz (2010)). Compte tenu du faible niveau de contrainte géographique des villes européennes, nos conclusions laissent penser que les différences de réglementation expliquent en grande partie les différences d'élasticités observées.

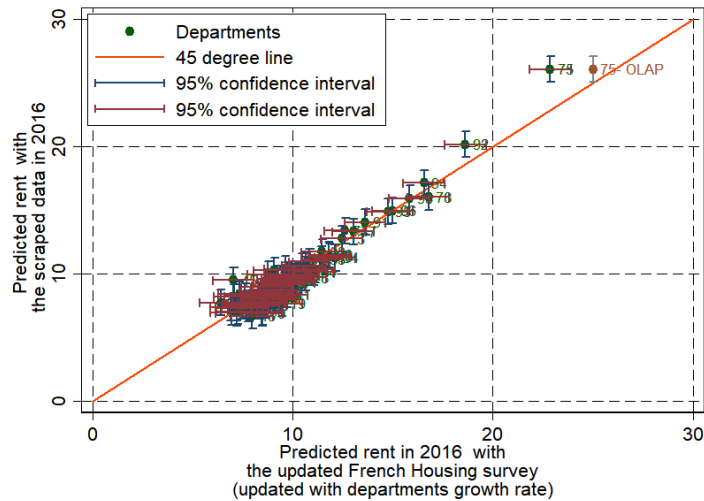
2 Les apports du Big Data peuvent-ils nous aider à mieux comprendre le marché de l'immobilier ?

Le secteur locatif privé joue un rôle important dans l'économie et concentre de nombreuses politiques publiques. En France, 33% des 40 milliards d'euros dépensés chaque année en aides au logement sont destinés au secteur locatif privé sous forme d'allocations logement ou d'aides directes aux propriétaires. Si plusieurs contributions récentes ont souligné les limites de ces programmes, on peut regretter que l'accès aux statistiques sur le marché locatif privé reste limité par rapport au secteur du logement social ou aux transactions immobilières. En effet, si les prix de transactions sont systématiquement enregistrés par l'administration fiscale ou les notaires, l'accès des chercheurs aux données locatives reste limité, en particulier au niveau local. Dans cet article, nous proposons de combler cette lacune en tirant parti de la pertinence des sites web immobiliers qui permettent d'apparier propriétaires et locataires.

L'objectif de notre travail est triple. Premièrement, fournir une base de données fiable du marché locatif local français aux universitaires et autres personnes intéressées par le marché du logement. Deuxièmement, souligner la forte complémentarité entre données administratives et données en ligne : les données en ligne ne révèlent leur potentiel que si elles sont utilisées avec des données externes. Troisièmement, dans une période où Internet est de plus en plus utilisé dans les transactions économiques, nous voulons partager une méthodologie pour extraire et utiliser des données en ligne.

En nous appuyant sur la littérature récente du webscraping, nous avons périodiquement collecté, nettoyé et analysé les annonces de location de logements en ligne provenant des deux plus grands sites français entre décembre 2015 et juin 2017. Chaque annonce fournit les caractéristiques du bien immobilier ainsi que son emplacement précis. Dans cet article, nous montrons que l'activité web fournit une image fidèle du marché de la location car les sites web rassemblent des millions d'annonces et tendent à l'exhaustivité. Aussi, le webscraping peut être utilisé pour suivre la dynamique du marché de l'immobilier.

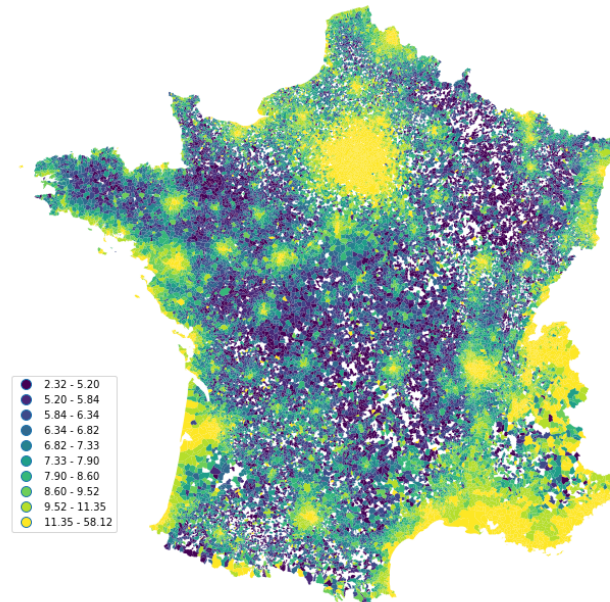
Pour montrer le caractère représentatif de nos données, nous calculons la valeur d'un bien comparable dans les régions et départements français grâce à nos données et confrontons nos résultats avec le même exercice basé sur les données de l'enquête logement de l'INSEE. La Figure 1 montre qu'aucun biais systématique n'apparaît lorsque l'on compare les deux bases de données.



Chaque point représente l'estimation du prix d'un bien représentatif au niveau départemental, l'axe horizontal fournit l'estimation en utilisant les données de l'enquête logement, l'axe vertical fournit l'estimation en utilisant les données en ligne.

Figure 1: Comparaison données en ligne et données administratives

Ces données en plus d'être représentatives, seront utiles pour les recherches futures traitant du marché de l'immobilier. Nous présentons deux applications potentielles. La première consiste à calculer la valeur d'un appartement similaire dans les zones urbaines françaises les plus importantes afin de comparer les tensions locales des marchés de l'immobilier. L'analyse des résultats en Figure 2 montre que le marché locatif local est hétérogène et que les marchés les plus tendus sont ceux de Paris et du Sud-Est de la France.



Chaque point représente l'estimation du prix d'un bien représentatif au niveau communal en utilisant les données en ligne.

Figure 2: Loyer moyen des communes de Françaises

La seconde application utilise nos données pour estimer la valeur locative marchande des logements sociaux. La précision de nos données nous permet de calculer la subvention implicite du logement social, également appelée avantage HLM, à une échelle très fine. Conformément à Trevien (2014) nous retrouvons une subvention implicite médiane proche de 42% de la valeur locative du logement mais nous la finesse de nos données nous permet d'aller plus loin en caractérisant la distribution complète de la subvention implicite.

3 La discrimination des femmes en politique

Les femmes sont-elles victimes de discrimination en politique? Alors que depuis des décennies les chercheurs en sciences sociales ont tenté d'étudier les raisons de la sous-représentation des femmes en politique, il s'est avéré difficile de démêler comportements discriminatoires de la part des électeurs et effets de sélection affectant les caractéristiques observées et non observées des femmes présentes dans l'arène politique.

Dans cet article, nous mettons en avant une discrimination causale vis à vis des femmes en politique. Pour ce faire, nous exploitons une caractéristique unique des élections départementales de 2015 qui nous permet de démêler, sans ambiguïté, les effets de sélections parmi les candidats se présentant à l'élection des effets de discrimination de la part des électeurs.

Pour la première fois, les candidats des élections départementales devaient se présenter par binômes obligatoirement mixtes en terme de genre, de sorte qu'une paire de candidats comprenait nécessairement un homme et une femme ainsi qu'un remplaçant du même genre. Au moment du vote, les électeurs ne pouvaient choisir qu'une seule paire de candidats, de sorte que pour chaque paire, chaque candidat et chaque candidate recevaient exactement le même nombre de voix. Si la paire était élue, ses candidats étaient nommés au même siège du conseil départemental. Au sein de chaque paire, l'ordre d'apparition des candidats sur le bulletin de vote était déterminé par l'ordre alphabétique. Comme nous le soutenons dans l'article, ce mécanisme crée une allocation de la position du genre sur le bulletin de vote quasi-aléatoire qui nous permet d'examiner l'impact causal d'avoir une femme en première position sur le bulletin de vote et le nombre de voix reçues lors de l'élection.

Ce test de position est pertinent dans notre étude car, bien que l'ordre d'apparition des candidats sur le bulletin de vote n'ait aucune incidence sur les prérogatives subséquentes attribuées aux candidats, certains électeurs ont pu croire, à tort, que le premier candidat était le candidat "principal". En effet, comme les électeurs avaient l'habitude de voter que pour un seul candidat et un suppléant, il est peu probable que les nouvelles règles aient été pleinement comprises par tous. Comme l'institut français de statistique IFOP l'a reconnu quelques semaines avant les élections : "Ces élections ont été caractérisées par une information insuffisante" et "l'introduction de couples de candidats a ébranlé des repères établis de longue date" dans l'esprit des électeurs. Par conséquent, toute différence observée entre les paires avec un candidat ou une candidate figurant sur la première position signifierait que l'on observe deux phénomènes simultanément. Premièrement, un déficit d'attention de la part de certains électeurs, tel que définie par DellaVigna (2009). En effet, comme les destins des deux candidats sont liés, si tous les électeurs connaissaient parfaitement les règles des élections, nous ne trouverions aucun effet de traitement. Deuxièmement, un pur préjugé sexiste de la part des électeurs.

L'identification de la discrimination provient de plusieurs caractéristiques particulièrement intéressantes de l'élection. Tout d'abord, le nombre de candidats et de candidates est exactement identique, afin d'assurer une stricte parité au sein des conseils locaux. Deuxièmement, bien que les caractéristiques des candidats et candidates soient en moyenne différentes, les caractéristiques individuelles ne permettent pas de prédire si le candidat ou la candidate apparaît en premier sur le bulletin de vote. L'effet que nous mesurons est donc peu susceptible d'être affecté par un quelconque biais de sélection, puisqu'il consiste à comparer si des paires identiques ont en moyenne des performances différentes lorsqu'un homme ou une femme est en premier sur le bulletin de vote. De plus, notre stratégie d'identification est renforcée par le fait que les

partis n'ont pas semblé apparier stratégiquement les candidats masculins et féminins en fonction de leur nom de famille afin, par exemple, de placer le candidat masculin en premier sur le bulletin : en effet, la distribution de la première lettre des noms de famille masculins et féminins est identique.

En comparant des paires traitées et non traitées de candidats d'affiliations politiques identiques dans les circonscriptions, nous montrons que les paires de droite où le candidat féminin apparaît en première position sur le bulletin de vote perdent environ 1,5 point de pourcentage de vote au premier tour, alors qu'en moyenne ce n'est pas le cas pour les paires des autres partis. Ces effets ont significativement affecté le résultat de l'élection : les paires affectées avaient 4 points de pourcentage de moins de chance d'aller au second tour ou de remporter l'élection.

Notre cadre d'étude permet non seulement d'identifier la discrimination pure, mais aussi de caractériser le type de discrimination en cause. Théoriquement, la discrimination est soit fondée sur le goût (taste based discrimination) ou sur les stéréotypes (statistical based discrimination). Dans le premier cas, les électeurs n'aiment pas voter pour des femmes, quelles que soient leurs caractéristiques ou les informations dont ils disposent à leur sujet. Dans le second cas, les électeurs appliquent des stéréotypes sur les femmes candidates en raison d'un manque d'information sur les caractéristiques des candidats. Nous soutenons que, dans notre contexte, l'effet mesuré reflète une discrimination statistique plutôt qu'une discrimination fondée sur le goût. Pour l'identifier, nous suivons une méthodologie similaire à celle développée par Altonji and Pierret (2001) et exploitons une caractéristique unique de la loi électorale française, qui stipule que les candidats peuvent rapporter des informations supplémentaires les concernant sur le bulletin de vote - telles que leur expérience politique, leur âge, leur profession ou leur photo. En comparant les effets de traitement entre les bulletins de vote avec de l'information ajoutée et les bulletins sans information, nous montrons que, pour les couples de droite, la discrimination disparaît lorsque l'information sur les candidats est affichée.

Nous montrons que les votes manquants ne reflètent pas une abstention différentielle et ne se sont pas traduits par des votes blancs ou nuls ; ils se sont traduits par des parts de votes plus élevées pour les candidats concurrents. Cependant, les paires concurrentes avec une candidate en première position n'ont pas obtenu plus de votes que les autres. Un tel résultat conforte notre interprétation des résultats comme reflétant une discrimination statistique plutôt qu'une discrimination fondée sur le goût. En effet, si notre résultat avait été motivé par une discrimination fondée sur le goût, nous aurions dû constater que les couples de candidats avec un candidat masculin en première position bénéficiaient du report de voix.

Enfin, nous exploitons deux types d'hétérogénéité. Premièrement, nous montrons que la discrimination ne dépend pas directement des caractéristiques observées chez les candidats, à savoir leur expérience politique antérieure et leur âge. Si l'on suppose que les caractéristiques sont une approximation de la qualité des candidats, nos résultats ne sont pas expliqués par les différences de qualité entre candidats mas-

culins et féminins. Deuxièmement, nous regardons si la discrimination dépend des caractéristiques locales. Nous montrons que, si la discrimination électorale ne varie pas en fonction de l'âge, du taux de chômage et du niveau d'éducation de la population, elle est plus importante dans les zones où la discrimination des femmes est élevée sur le marché du travail.

Nos résultats ont des implications importantes pour le débat public sur la discrimination des femmes en politique. Tout d'abord, si nos résultats montrent une discrimination à l'encontre des candidates de droite, cela ne signifie pas que les électeurs de droite ont plus de préjugés contre les femmes que les électeurs des autres partis. En effet, la présence d'un déficit d'attention est nécessaire pour identifier une quelconque discrimination. Le fait de ne pas observer de discrimination à l'égard des candidates des autres partis peut simplement indiquer que leurs électeurs sont moins sujets au biais d'attention limité. Deuxièmement, comme la quantité d'informations disponibles à propos des candidats sur le bulletin de vote semble jouer un rôle important sur le résultat de l'élection, il conviendrait d'ouvrir une réflexion plus générale sur la standardisation de la présentation des bulletins de vote. Enfin, étant donné que la discrimination électorale semble être plus élevée dans les endroits où les femmes sont discriminées sur le marché du travail, les politiques visant à réduire la discrimination des femmes en politique seront probablement plus efficaces si elles sont coordonnées avec les politiques antidiscriminatoires des autres marchés.

References

- Altonji, Joseph G, and Charles R Pierret. 2001. "Employer learning and statistical discrimination." *The Quarterly Journal of Economics* 116 (1): 313–350.
- Andrews, Dan, Aida Caldera Sánchez, and Åsa Johansson. 2011. "Housing markets and structural policies in OECD countries." *OECD Technical Report*.
- Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2016. "The costs of agglomeration: House and Land prices in French cities." *CEPR Discussion Paper No.9240 (2012 revised in 2016)* .
- DellaVigna, Stefano. 2009. "Psychology and economics: Evidence from the field." *Journal of Economic literature* 47 (2): 315–372.
- Hilber, Christian AL, and Wouter Vermeulen. 2016. "The impact of supply constraints on house prices in England." *The Economic Journal* 126 (591): 358–405.
- Meen, Geoffrey. 2002. "The time-series behavior of house prices: a transatlantic divide?" *Journal of Housing Economics* 11 (1): 1–23.
- Monnet, Eric, and Clara Wolf. 2016. "Demographic Cycle, Migration and Housing Investment: a Causal Examination." *Banque de France - Working Paper*.
- Saiz, Albert. 2010. "The geographic determinants of housing supply." *The Quarterly Journal of Economics* 125 (3): 1253–1296.
- Trevien, Corentin. 2014. "Habiter en HLM: quel avantage monétaire et quel impact sur les conditions de logement?" *Economie et statistique* 471 (1): 33–64.