

# Les indices de prix hédoniques

Principes et Illustration à partir du Prix des Terrains à Bâtir

Ronan Le Saout, Benjamin Vignolles

INSEE SMS, 24 Mars 2017

# Motivations

- La méthode des prix hédoniques : un large emploi pour les études économiques avec une théorie micro-économique établie ;
- Une réflexion plus statistique pour la construction des indices de prix ;
- Relier ces 2 approches, à travers l'exemple du prix des terrains à bâtir ;
- Manski (2015) : nécessité d'ouvrir "la boîte noire" des statistiques publiques pour limiter les incertitudes conceptuelles.

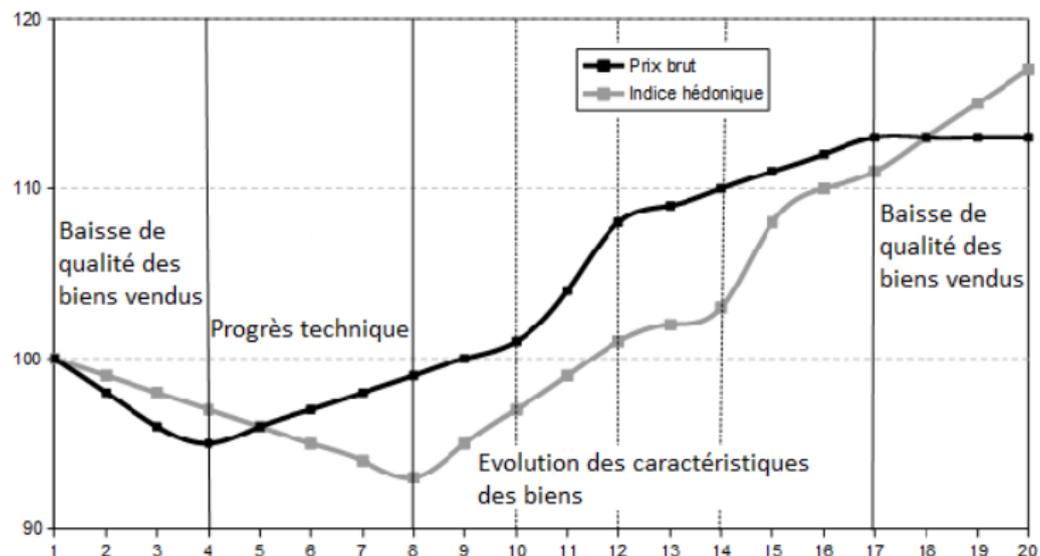
# Les indices hédoniques

- Pour les indices de prix, Moulton (2001) détaille les pratiques aux Etats-Unis

Prix de référence
Evolution technique
Qualité
Caractéristiques (surface, nombre de pièces...)

- Construire des indices de prix pour des biens hétérogènes ;
- Prendre en compte les différences et les évolutions de la qualité des biens, ainsi que le progrès technique ;
- Effectuer des ajustements de qualité ou de mode lorsque des produits sont substitués dans un panier de biens.

# Un exemple d'évolution des prix



# En France

- Usages principaux en France
  - Indice de prix des logements anciens ;
  - Indice de prix des logements neufs ;
  - Indice des terres et prés agricoles ;
  - Indice des coûts de la construction ;
  - Ajustement de la qualité pour l'indice des prix à la consommation.

# Questions et analyses

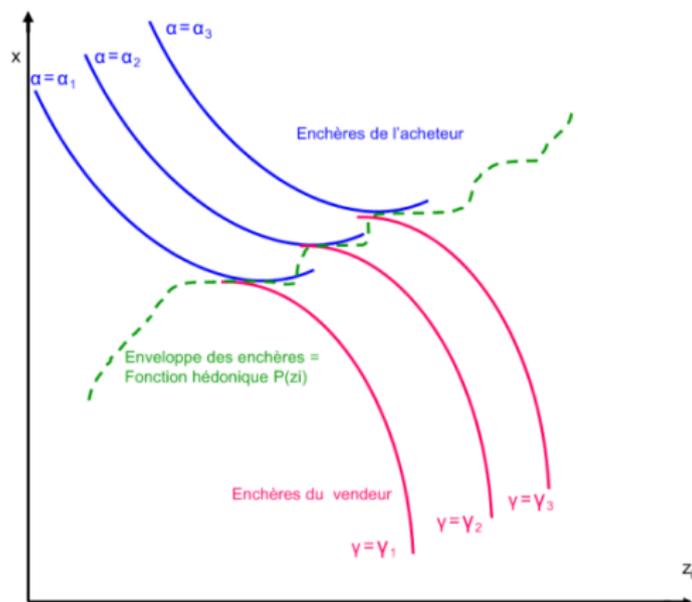
- Que nous apprend la théorie économique pour la construction des indices ? Quel est le rôle du panier de biens de référence ?
- Analyse hédonique des prix des terrains à bâtir sur la période 2006-2012 à l'aide de près de 400.000 observations individuelles ;
- Quantification du biais associé aux variables omises : contraintes réglementaires de la construction, aménités locales et accessibilité aux équipements ;
- Conclusions en matière de bonnes pratiques pour la construction d'indices hédoniques.

# Plan de la présentation

- Quelques rappels théoriques
- Synthèse sur les indices hédoniques
- Données EPTB et autres sources
- Estimation et résultats

# Le cadre de Rosen (1974)

- Bien composite  $\underline{z}$  de caractéristiques  $(z_1, \dots, z_n)$
- Un acheteur maximise son utilité  $U(x, \underline{z}, \alpha_i)$  avec  $x$  autres biens de prix unitaire et  $\alpha_i$  caractéristiques individuelles, le vendeur sa fonction de profit



## Le cadre de Rosen (1974)

- La fonction de prix  $p(\underline{z})$  peut être analysée comme l'enveloppe des fonctions d'enchère (des acheteurs et des vendeurs), elle n'est donc pas reliée directement à une fonction d'offre ou de demande.
- Elle doit être non linéaire et ne doit pas contenir de variables sur les caractéristiques des acheteurs ( $\alpha_i$ ) ou des vendeurs ( $\beta_k$ ).
- Le prix implicite de  $z_j$  est  $p_j = \frac{\partial P}{\partial z_j} = \frac{\partial U / \partial z_j}{\partial U / \partial x}$  et permet de calculer dans une seconde étape les fonctions de demande et les surplus du consommateur. Cette étape n'est pas utile pour le calcul d'un indice de prix.

# Théorie des indices

- Pas uniquement la simple application de la régression de première étape  $p(\underline{z})$ , dont l'estimation relève de choix statistiques.
- Hypothèse sur la composition du panier de biens de référence.
- Indice de Laspeyres : fixe le panier de biens de référence à la période 0 de caractéristiques  $X^0$ , i.e.  $\frac{p(X^0, \beta^t)}{p(X^0, \beta^0)}$ .
- Indice de Paasche : fixe le panier de biens de référence à la période  $t$ , i.e.  $\frac{p(X^t, \beta^t)}{p(X^t, \beta^0)}$ .
- Indice de Fisher et de Törnqvist, racine carrée des indices de Laspeyres et Paasche, arithmétique ou géométrique.

# La méthode sur périodes adjacentes

- Un indice de prix sur périodes adjacentes se rapproche d'un indice géométrique type Törnqvist.
- Une seule fonction de prix est estimée, qui inclut une (ou plusieurs) indicatrice d'évolution temporelle,  $I_{0,t}^{PA} = \frac{p(X^{0 \cup t}, \beta^t)}{p(X^{0 \cup t}, \beta^0)} = \frac{p(X^{0 \cup t}, (\beta_0^t, \beta_{-0}^t))}{p(X^{0 \cup t}, (\beta_0^0, \beta_{-0}^0))}$ .
- Si la fonction de prix explique le log du prix

$$\text{Ln}(\text{Prix}) = \text{Ln}(\text{Prix Ref}) + \beta \cdot X + \gamma \cdot \mathbf{1}\{\text{Année } t\} + \varepsilon$$

L'indice est  $\exp(\gamma)$ , indépendant de la composition du panier des biens de référence.

# La méthode sur périodes adjacentes

- Choix retenu pour les indices des terres et prés agricoles (Lefebvre et Rouquette 2012) et des maisons neuves (Balcone 2013).
- Précision aisée à calculer par la Delta-Méthode
$$\sqrt{n} \left( \exp \left[ \widehat{\beta}_0 \right] - \exp \left[ \beta_0 \right] \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \mathbb{V} \left( \widehat{\beta}_0 \right) \exp \left[ 2\beta_0 \right] \right)$$

## Les autres méthodes

- Une régression hédonique est estimée à chaque période et la valeur du panier de référence est estimée à chaque date,

$$\text{Log}(\text{Prix})^{(i)} = \beta^{(i)} \cdot X^{(i)} + \varepsilon^{(i)}$$

- La méthode d'imputation est un indice arithmétique (L pour

$$\text{Laspeyres}), I_{0,t}^{\text{Imputation,L}} = \sum_{i \in 0} \text{Exp} \left( \left( \widehat{\beta}^{(t)} - \widehat{\beta}^{(0)} \right) \cdot X^{(i)} \right);$$

## Les autres méthodes

- La méthode par bien de référence est un indice géométrique,  
$$I_{0,t}^{\text{BienRef.,L}} = \text{Exp} \left( \left( \widehat{\beta}^{(t)} - \widehat{\beta}^{(0)} \right) \cdot \overline{X^{(0)}} \right);$$

- La méthode à qualité fixée est un indice géométrique qui recalcule un prix de référence

$$I_{0,t}^{\text{QualiteFixee,L}} = \text{Exp} \left( \left( \overline{\left( \text{Log} \left( \text{Prix}^{(t)} \right) - \widehat{\beta}^{-(0)} X^{(t)} \right)} - \widehat{\alpha}^{(0)} \right) \right);$$

## Les autres méthodes

- Ces 2 dernières méthodes sont très proches. L'avantage de la méthode à qualité fixée est de ne nécessiter, pour l'indice de Laspeyres, qu'une estimation de la fonction hédonique à la période 0.
- Potentiel biais car  $\mathbb{E}(\text{Log}(\text{Prix})/X) = \text{Log}(\text{Prix})$  mais  $\mathbb{E}(\text{Prix}/X) = (\text{Prix}) \cdot \text{Exp}(\sigma^2/2)$
- Précision complexe à calculer.

## Biais de variables omises

- Des signes contraires à l'intuition peuvent apparaître du fait de relations complexes entre la fonction de prix et les courbes d'offre et de demande (Pakes 2003) ;
- Mais cela peut aussi être le cas avec des variables omises (Benkard et Bajari 2005) :
  - Le rendement des caractéristiques inobservées peut différer selon les périodes ;
  - Le contenu moyen des caractéristiques inobservées diffère à chaque période, en cas de progrès technique non pris en compte par exemple.
- A relier au mode de prise en compte de l'espace et de l'hétérogénéité. Des analyses désagrégées réduisent ce biais.

# Les données EPTB

- Enquête sur le prix des terrains à bâtir, période 2006-2012.
- Environ 400.000 observations.
- Peu de caractéristiques physiques : surface et viabilisation.
- Pas de relation claire entre le prix et la surface.
- Appariement avec les aménités au niveau commune :
  - Distance à la commune de 50.000 habitants et zonage unité urbaine ;
  - Aménités touristiques ;
  - Distance aux équipements (train, aéroport, hôpital, supermarché) ;
  - Risques.

## Modèles estimés

- Modèle annuel liant le log du prix (ou prix au m<sup>2</sup>) aux variables explicatives.
- Goffette-Nagot (2000) pour un cadre théorique, Combes *et al.* (2011) pour une analyse empirique hédonique non centrée sur les indices.
- 4 types de variables : caractéristiques du terrain, de la localisation, contraintes réglementaires, aménités environnementales et autres variables de localisation.
- Analyse de la robustesse de la spécification (analyse semi-paramétrique).
- Différents modèles :
  - Moyenne simple ;
  - Indice sur périodes adjacentes ;
  - Indice par méthode d'imputation, bien de référence et à qualité fixée.

# Le choix des variables explicatives

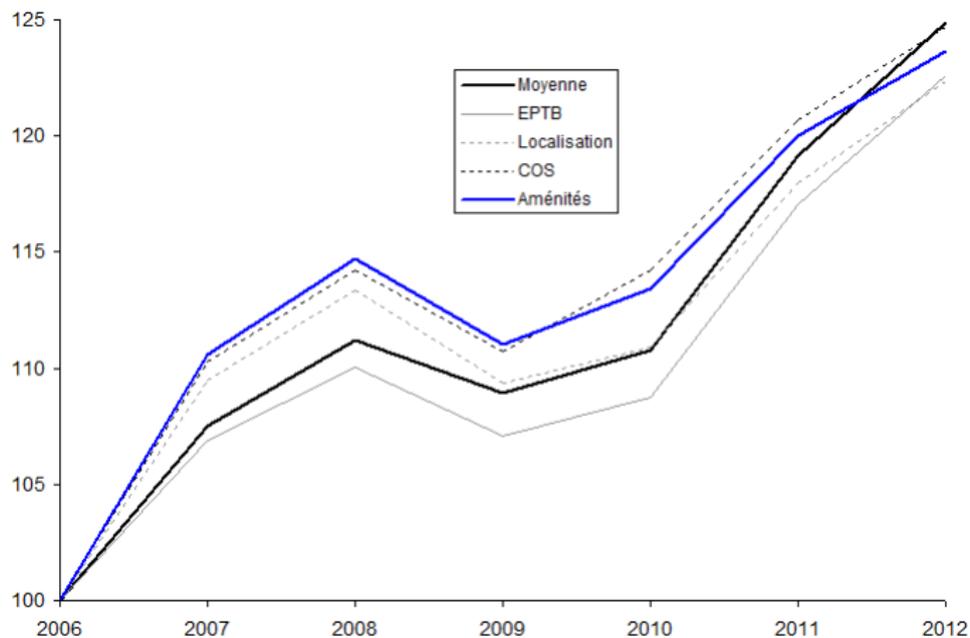
	(1) Log(Prix) EPTB	(2) Log(Prix) Loc.	(3) Log(Prix) COS	(4) Log(Prix) Aménités
Surface ( $\cdot 10^{-3}$ )	-0.068	0.041	0.115	0.119
Viabilisé	0.057	0.058	0.053	0.052
Distance commune 50.000 hab. Grand pôle (GP) (Ref. Couronne GP)		-0.009 0.248	-0.005 0.138	-0.005 0.058
Proche plusieurs GP		-0.131	-0.100	-0.089
Petit pôle (PP)		-0.167	-0.135	-0.206
Couronne PP		-0.479	-0.400	-0.408
Hors pôles		-0.406	-0.339	-0.328
COS			0.046	0.039
Haut débit				0.020
Haute montagne				0.452
Plage				0.094
Taux équ. touristiques				0.038
Dist. littoral <30 kms				-0.018
(Dist. littoral <30 kms) <sup>2</sup>				0.649
Dist. gare				-0.002
Dist. aéroport				-0.002
Dist. hôpital				-0.003
Dist. supermarchés				-0.014
Constante	10.810	11.037	10.258	10.487
Nb. Obs.	368.537	368.537	368.537	368.537
R2 ajusté	0.025	0.284	0.358	0.388

Lecture : Modèle liant le log du prix aux variables explicatives avec des indicatrices d'année pour la période 2006-2012.

# L'effet des aménités

- Un modèle très peu explicatif avec les seules variables issues de l'enquête ;
- L'effet des unités urbaines est différent avec ou sans contrôle des aménités ;
- Les aménités et l'accessibilité aux équipements sont valorisées dans le prix ;
- Mais inclure les aménités n'améliore que peu le pouvoir explicatif du modèle ;
- Les indicatrices géographiques ou d'unités urbaines captent une partie de l'hétérogénéité spatiale.

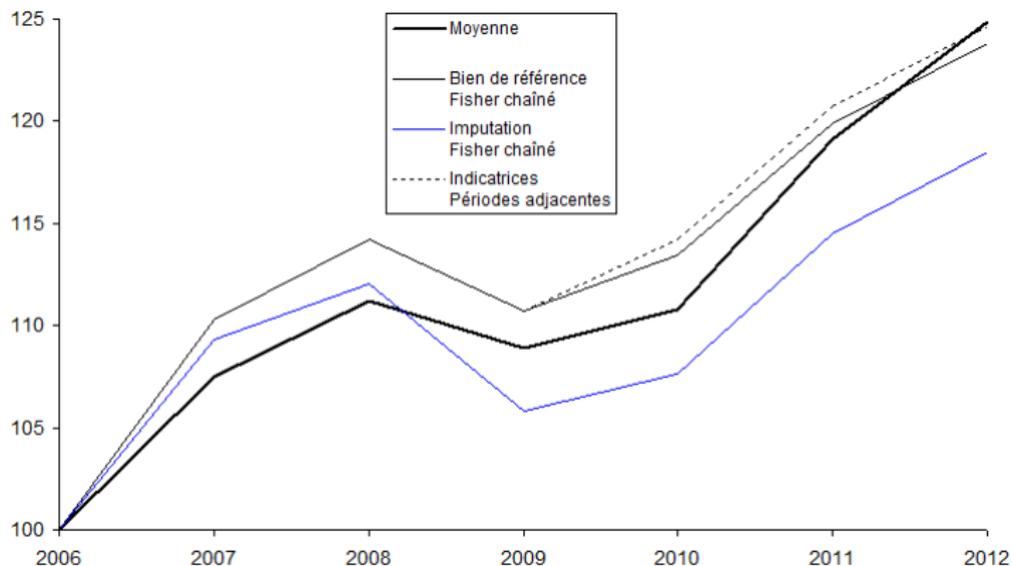
# Biais de variable omise



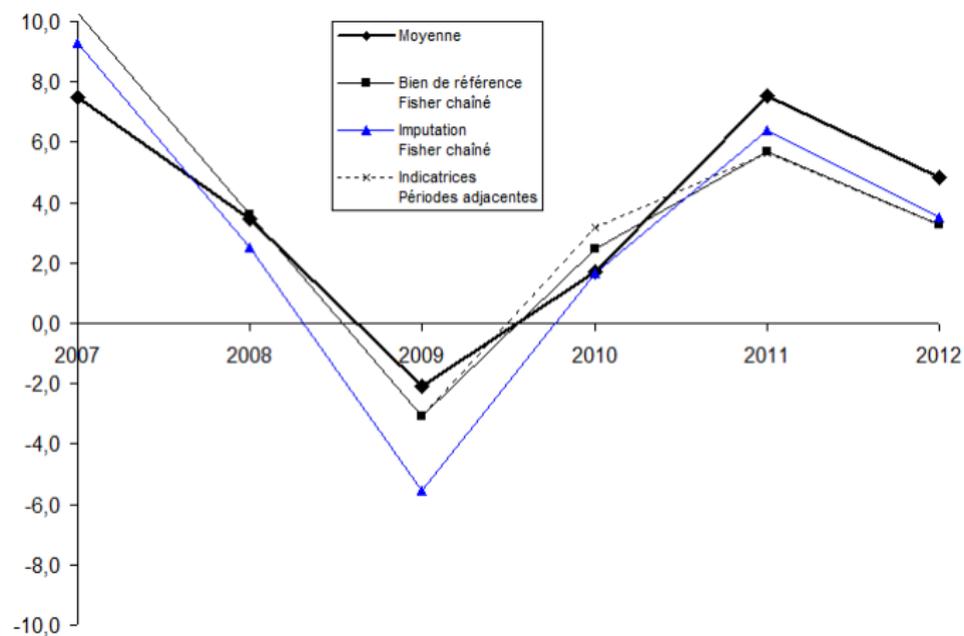
# Biais de variable omise

- Des évolutions cohérentes mais une interprétation économique qui peut être différente ;
- Des résultats proches en incluant les contraintes réglementaires, avec ou sans les aménités ;
- De fortes divergences de taux de croissance entre 2006 et 2007 et, entre 2008 et 2009 ;

# Comparaison des méthodes



# Taux de croissance



# Conclusions

- Des indices sensibles à la méthode choisie et au biais de variables omises, avec une forte divergence de la méthode d'imputation en 2009 ;
- Les indices construits, quelle que soit la méthode retenue, présentent des évolutions différentes de celle des prix moyens ;
- Les aménités et l'accessibilité aux équipements sont valorisées dans le prix, mais influencent peu les indices ;

# Conclusions

- L'indice de prix des terrains semble refléter des hausses plus élevées en début de période mais moins élevées par la suite.
- Avant la crise, les derniers terrains mis en vente pouvant être de qualité décroissante, il y a une correction à la hausse de la qualité. Après la crise, cela peut traduire un effet volume, seuls les terrains de meilleur qualité trouvant acheteurs ;
- La prise en compte de l'espace est effectuée à travers les variables explicatives ou une hypothèse d'hétérogénéité spatiale. La dépendance spatiale reste non prise en compte dans la construction des indices de prix.