

# **Indice de prix du foncier: réflexions théoriques et première approche empirique**

Gowtam Jinnuri (INSEE) et Anne Laferrère (CREST, INSEE)

30 juin 2009

version 1 (préliminaire et incomplète)  
préparée pour les Troisièmes Rencontres du Logement,  
Marseille 22-23 octobre 2009

## **Résumé**

**Le coût du foncier n'est observé directement qu'au moment des transactions de terrains. Il est ensuite intrinsèquement incorporé au prix des logements. Nous cherchons ici si, moyennant des hypothèses à expliciter, un prix relatif du foncier peut être extrait à partir d'un modèle hédonique sur les données de transactions immobilières à une date donnée, dite date de base. Un tel modèle permet de construire un indice spatial de prix comparant le coût de se loger dans le parc existant d'une zone à l'autre. On estime aussi un tel indice spatial à une autre date pour étudier l'évolution temporelle de ce coût relatif des localisations. Nous utilisons les données notariales de transactions de logement pour tester ce genre de modèles. Comme il existe par ailleurs des indices temporels de prix par zone (les indices sous-jacents aux indices agrégés Notaires-INSEE), on peut utiliser ces indices pour faire évoluer l'indice spatial de la date de base. Le résultat permet d'évaluer la transivité des évolutions spatiales et temporelles de prix. Cette comparaison avec les indices temporels Notaires INSEE permet d'explicitier davantage les hypothèses sous-jacentes aux uns et aux autres types d'indices.**

## **1. Introduction**

Le coût du foncier n'est observé directement qu'au moment des transactions portant sur les terrains nus. Il est ensuite intrinsèquement incorporé au prix des logements au moment de chaque transaction successive. Pourtant on sent bien qu'en achetant un logement, on achète à la fois une construction et une localisation. Moyennant des hypothèses (fortes) à expliciter, on pourrait envisager de l'extraire, à partir des données sur les transactions immobilières, en particulier celles des notaires. Cela permettrait, non pas d'avoir le prix absolu du foncier, mais de comparer le coût de se loger dans le parc existant d'un endroit à l'autre (indice spatial) et de suivre l'évolution temporelle de ce coût relatif.

Indices spatiaux, prix relatifs des différentes localisations, séparation du prix d'un logement entre prix du terrain et prix de la structure, sont différentes facettes d'une même question complexe. Des indices spatiaux à l'échelle d'un pays sont peu fréquemment calculés. Cheshire et Sheppard (1995) offrent une première réflexion et une application empirique sur un marché local anglais. Laferrère et Dubujet (2000) raisonnent sur données agrégées, les prix des biens de référence (maison et appartement anciens) des indices Notaires des 104 agglomérations de province en 1996. Ils les relient à des variables agrégées locales (taux de chômage, taux de propriétaires, de résidences secondaires, structure par âge, taille de la population ...), mettant en évidence un effet régional (ZEAT). Ensuite ils étudient l'évolution

des prix entre 1994 et 1999, mettant là encore en évidence un effet régional. Mais ce travail restait encore préliminaire. Davis et Heathcote (2006, 2007) estiment un indice de prix des terrains bâtis (essentiellement par différence entre un indice de prix des logements, et un indice officiel de prix de construction, en lien avec des indicateurs de quantité, par la méthode de l'inventaire perpétuel), et ils le comparent à celui des logements anciens aux États-Unis. Ils trouvent au niveau agrégé que le prix des terrains résidentiels évolue plus vite que celui des logements, et est plus volatil, que le terrain compte pour 38% de la valeur du stock de logement en début de période (1970) et que sa part augmente. Ils ne s'intéressent pas à des comparaisons spatiales.

Hill (2004) expose clairement la différence entre un indice de prix temporel, qui compare un prix dans le temps, en un lieu donné, et un indice spatial qui compare un prix entre localisations, à un instant donné. Il le fait dans le cadre de comparaison de prix à la consommation entre 15 pays de l'Union Européenne. Il se propose de construire ce qu'il appelle un panel à partir de comparaison spatiale de prix, répétées dans le temps. Il pose la question de la transitivité des indices spatiaux et temporels, et celle de la convergence des prix spatiaux au fil du temps.

Diewert (2006) propose une méthode de modèle non linéaire pour décomposer la valeur d'un bâtiment entre celle de la terre et celle de la superstructure. Le calcul des indices Notaires-INSEE s'accompagne de l'estimation de quelques 300 modèles hédoniques dans autant de zones élémentaires couvrant tout le territoire français métropolitain. La plupart de ces modèles incorporent des variables indicatrices de quartiers, qui sous réserve des hypothèses sous-jacentes aux modèles pourraient fournir des prix relatifs locaux. Ces modèles ne sont pas publiés et ne servent qu'à estimer le prix du bien de référence local qu'entre dans le calcul des indices agrégés (Gouriéroux et Laferrère, 2004). Notre stratégie ici est d'établir au contraire un seul modèle national à une date donnée, et d'interpréter les coefficients des indicatrices de zones comme des prix relatifs de ces zones pour le bien de référence national défini par le modèle. Le modèle est réestimé à une seconde date, pour comparaison de l'évolution de ces prix relatifs. Dans un troisième temps nous faisons évoluer les prix de chaque zone entre la date 1 et la date 2 comme ont évolué les prix des biens de référence sous-jacents aux indices Notaires-INSEE, et nous comparons le résultat à ce que donne le modèle global.

Nous proposons donc ici une première réflexion théorique (partie 1), et un essai d'application empirique à partir des données des notaires<sup>1</sup>, à la fois pour produire un indice spatial à deux dates (2000 et 2008) (partie 2) et pour comparer la déformation des prix spatiaux relatifs à celle prédite par l'évolution des prix sous-jacents aux indices Notaires-INSEE (partie 3).

## 2. Réflexion théorique

Soient  $Pf_i$  le coût du foncier correspondant au logement  $i$ , et  $Pc_i$  le prix de la structure (construction), avec  $P_i = Pf_i + Pc_i$ . Supposons

$$Pc_i = \sum b_k X_{ki} + \varepsilon_i$$

$$Pf_i = \sum c_m Y_{mi} + \eta_i$$

On a alors  $P_i = \sum \alpha_{k,t} X_{ik}^t + \sum \beta_{m,t} Y_{im}^t + \varepsilon_i + \eta_i$

Les variables  $X$  sont liées à la qualité de la construction: surface, nombre de pièces, époque de construction, nombre de salle de bains, de garages, cave, état du logement...

---

<sup>1</sup> Nous remercions Perval et la CNIP de nous avoir autorisé à utiliser ces données.

Les variables Y sont liées au terrain: surface, quartier, distance aux commerces, au centre etc... Ce modèle s'applique facilement à des maisons.

Si on isole les variables de superficie du logement  $A_i^t$ , superficie du lot  $B_i^t$ , et  $L_s$ , indicatrice d'une zone géographique s, on obtient :

$$P_i^t = [\alpha_t A_i^t + \sum_k X_{ik}^t \alpha_{k,t}] + [\beta_t B_i^t + \sum_m Y_{im}^t \beta_{m,t} + \sum_s L_s^t \beta_s] + \varepsilon_i^t + \eta_i^t \quad (1)$$

$\alpha_t$  est le prix estimé d'un mètre carré de logement/structure pour un logement de référence,  $\beta_t$  le prix estimé d'un mètre carré de terrain de référence,  $\beta_s$  le prix de la zone s, le tout à la date t. Ceci n'a de sens que si les X déterminent bien un logement de référence (sans l'intervention des Y ?) et vice versa.

Donc stratifier ? petits lots/ moyens/ grands lots.

Le même modèle estimé à la date t' fournit les estimations de ces prix à cette date.

Un autre modèle est le suivant, où le prix est une fonction non linéaire de la surface. Si  $A_n^0$  et  $B_n^0$  sont respectivement la superficie du logement et celle du terrain à la date 0, et  $A_n^t$  et  $B_n^t$  sont les mêmes à la date t, Diewert (2006) propose la spécification suivante :

$$\begin{aligned} p_n^0 &= \{[\alpha_0 + \sum_{k=1}^K X_{nk}^0 \alpha_k] A_n^0 + [\beta_0 + \sum_{m=1}^M Y_{nm}^0 \beta_m] B_n^0\} \eta_n^0 \quad n = 1, \dots, N(0) \\ p_n^t &= \{\gamma^t [\alpha_0 + \sum_{k=1}^K X_{nk}^t \alpha_k] A_n^t + \delta^t [\beta_0 + \sum_{m=1}^M Y_{nm}^t \beta_m] B_n^t\} \eta_n^t \quad n = 1, \dots, N(t) \end{aligned}$$

en supposant les prix relatifs des caractéristiques X et Y fixes dans le temps. Soit, en prenant le log des deux côtés :

$$\begin{aligned} \ln p_n^0 &= \ln \{[\alpha_0 + \sum_{k=1}^K X_{nk}^0 \alpha_k] A_n^0 + [\beta_0 + \sum_{m=1}^M Y_{nm}^0 \beta_m] B_n^0\} + \varepsilon_n^0 \quad n = 1, \dots, N(0) \\ \ln p_n^t &= \ln \{\gamma^t [\alpha_0 + \sum_{k=1}^K X_{nk}^t \alpha_k] A_n^t + \delta^t [\beta_0 + \sum_{m=1}^M Y_{nm}^t \beta_m] B_n^t\} + \varepsilon_n^t \quad n = 1, \dots, N(t) \end{aligned} \quad (2)$$

Si on suppose que les X agissent sur le prix de la structure et les Y sur le prix du terrain, on obtient un indice de l'évolution temporelle du prix de la structure par  $\gamma^t$  et un indice de l'évolution du prix du terrain par  $\delta^t$ .

En fait pour Diewert l'indice de prix du mètre carré de structure est  $[\alpha_t + \sum_k X_{ik}^t \alpha_{k,t}] / [\alpha_0 + \sum_k X_{nk}^0 \alpha_{k,0}]$  et de même pour le terrain.

Un troisième modèle est plus proche de la spécification des indices Notaires-INSEE des appartements en ce qu'il utilise le prix au mètre carré.

On part partir du modèle suivant qui relie  $p_i$  le prix au mètre carré d'un logement i, à ses caractéristiques:

$$\log p_i = \log p_0 + \sum \beta_k X_{k,i} + \sum \lambda_m L_{_m,i} + \varepsilon_i \quad (3)$$

où  $X_{_k,i}$  sont les caractéristiques de qualité k du logement i et  $L_{_m,i}$  des variables indicatrices de la localisation (par exemple les zones de l'indice notaire-Insee, un quartier dans une zone, le climat, la présence de bruit extérieur (?)) m du logement i. On estimerait un seul modèle sur la France entière.

On suppose que les  $p_i$ , X et L sont observés à une date donnée t dont l'indice est omis. Un tel modèle fait plusieurs hypothèses qui seraient à étudier (log linéarité, prix au m<sup>2</sup> plutôt que prix total et surface en variable explicative). Il suppose aussi (contrairement à ce que fait l'indice Notaires-Insee) que le prix relatif des caractéristiques des logements (les  $\beta_k$ ) est

indépendant du quartier (par exemple l'avantage relatif à disposer d'un ascenseur est le même dans tous les quartiers). En d'autres termes, il n'y a pas d'interactions entre les  $X_k$  et les  $L_m$ . On sait que cette hypothèse n'est pas forcément vérifiée. Les valeurs omises des caractéristiques  $X_k$  et des localisations  $L_m$ , définissent alors automatiquement un bien de référence localisé, dont le prix est  $p_0$ . C'est, si on raisonne par exemple sur les appartements, le prix (au mètre carré) d'un deux pièces, avec salle de bain, au 1er étage, d'un immeuble de 5 étages, construit entre 1981 et 1991, sans garage, de surface par pièce moyenne, dans la localisation de référence,  $m=0$ . Alors  $\lambda_m$  représente le prix de la localisation  $m$  relativement à la localisation de référence 0. On peut bien évidemment choisir n'importe quelle zone pour la référence.

Le prix relatif  $\lambda_m$  incorpore trois éléments:

- (1) la différence du coût du foncier,
- (2) les différences non observées ou non observables de qualité des logements entre les quartiers, i.e. les variables  $X$  qui "manquent" à l'observateur,
- (3) les différences possibles de coût de construction entre localisations.

La différence du coût du foncier (1) reflète la qualité de vie au sens large d'un endroit à l'autre, en anglais la "site value". C'est ce qu'on recherche en principe. Les différences non observées de qualité (2) sont par exemple le fait que dans la localisation  $m$  les logements sont en béton, alors que les logements du quartier  $m'$ , de même période de construction, même taille d'immeuble etc (donc de même qualité observée) sont en pierre de taille (le type de matériau n'est pas observé). Les différences possibles de coût de construction entre localisation (3) proviennent par exemple du fait que le salaire du manoeuvre est plus élevé dans un endroit que dans l'autre ou que le coût d'y transporter une quantité donnée de béton varie.

Il semble difficile de séparer ces trois éléments. Par exemple les coûts de transport (accessibilité) font bien partie du coût du foncier, et sont donc dans (1) autant que (3). Dans une zone d'emploi donnée, à réseau de transport donné, par exemple en Ile-de-France, on pourrait supposer que (3) est en première approximation identique dans toutes les localisations  $L_m$  (hypothèse sans doute non tenable au niveau de la France entière). Si on fait l'hypothèse additionnelle que la partie non observable (2) est négligeable, on obtient ce qu'on cherche (1). Si (2) n'est pas négligeable, on estime la somme (1) + (2), qui n'est peut être pas sans intérêt, si qualité des logements et qualité des quartiers sont assez mêlées: ainsi le plaisir d'habiter le quartier  $m$  et non  $m'$  provient non seulement des biens publics qu'il procure, mais du fait que le logement  $y$  est (intrinsèquement) de meilleure qualité (on devrait pouvoir mesurer le biais?). Quant à l'estimation de (1) + (2) + (3) ne peut-on la voir comme celle du coût relatif d'un foncier bâti, résultant de la confrontation offre/demande?

Dans quelle mesure l'ajout de variables  $Y$  dans le modèle (1) modifie l'estimation de  $\beta_s$  ?

Que donne l'estimation du modèle (2) ? De même dans ce modèle, quelle est la sensibilité de l'estimation à la présence de variables  $Y$ , et comment interpréter les coefficients de ces variables ?

Toujours en restant à une date donnée, on peut donc estimer le modèle à partir des transactions observées et l'indice spatial est  $I_{s,m}=100 \lambda_m$  de base 100 dans la localisation 0 (avec une spécification en logarithmes).

Ajoutons maintenant une dimension temporelle. On peut réestimer le même modèle à chaque date  $t$ .

$$\log p_{i,t} = \log p_{0,t} + \sum \beta_{k,t} X_{k,i,t} + \sum \lambda_{m,t} L_{m,i,t} + \varepsilon_{i,t}$$

On a alors un nouvel indice spatial pour chaque date. On peut aussi interpréter  $\lambda_{m,t+1} / \lambda_{m,t}$  comme l'évolution de t à t+1 du prix relatif du quartier m.

Il reste à écrire comment cet indice se compare aux indices temporels Notaires-Insee (NI dans ce qui suit) et surtout à tester l'effet de l'hypothèse d'indépendance du prix relatif des caractéristiques des logements (les  $\beta_k$ ). Pour cela nous faisons évoluer les prix de chaque zone entre la date 1 et la date 2 comme ont évolué les prix des biens de référence sous-jacents aux indices Notaires-INSEE, et nous comparons le résultat à ce que donne le modèle global.

### 3. Estimation des deux modèles globaux en 2000 et 2008

à compléter

### 4 Comparaisons avec les prix sous-jacents aux indices Notaires-INSEE

à compléter

### Références (à compléter)

Cheshire et Sheppard, 1995, On the Price of land and the value of Amenities.

Davis M.A. et J. Heathcote, 2007, The Price and Quantity of Residential Land in the United States, Journal of Monetary Economics, 54, 2595-2620.

Diewert, W. E., 2006. The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions. Published now in Diewert, W.E., B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (2009), PRICE AND PRODUCTIVITY MEASUREMENT: Volume 1 -- Housing. Trafford Press.

<http://www.vancouvervolumes.com/V1%20FCh6%2004.03.09%20Diewert%20OECD.pdf>

Gouriéroux, C., Laferrère, A., 2006. Managing Hedonic Housing Price Indexes: the French Experience, paper presented at the OECD-IMF Workshop on Real Estate Price Indexes, Paris, November 6-7. <http://www.oecd.org/dataoecd/2/24/37583497.pdf>

Hayes, P. 2005, Estimating UK Regional price Indices, 1974-96, Regional Studies, 93, 3, 333-344.

Laferrère A. et F. Dubujet (2003), Niveau et évolution des prix des logements anciens en province, Données urbaines, 293-306.

Laferrère A. et D. le Blanc (2003), Gone with the windfall, how do housing allowances affect student co-residence?, Document de travail du CREST.

Frigitt, .