



# 1 Introduction

Les prix de l'immobilier sont déterminés par de nombreux facteurs économiques, démographiques et environnementaux. Dans le cadre de cette étude sur données de panel, nous allons déterminer quels sont ces facteurs et dans quelle mesure ceux-ci impactent le niveau de prix des biens immobiliers dans 49 États américains de 1976 à 2003.

Nous pouvons dès à présent poser des hypothèses pour comparer les résultats de la littérature en lien avec ce sujet et nos résultats empiriques personnels. Ces hypothèses sont les suivantes :

- Une hausse du revenu réel par habitant provoque une hausse des prix de l'immobilier.
- Une hausse de la population augmente en conséquence les prix de l'immobilier.
- Une hausse du taux d'emprunt réel diminue les prix de l'immobilier.

L'ensemble de ces propositions découlent directement de la loi de l'offre et de la demande. D'une part, une hausse du revenu et de la population vont entraîner une hausse de la demande et donc des prix. D'autre part, la hausse du taux d'intérêt réel des emprunts va diminuer le nombre d'acheteurs et peut aussi pousser des propriétaires à revendre leurs biens si le coût de l'emprunt devient trop important. En d'autres termes, la demande va diminuer, l'offre va augmenter et le prix va donc s'amoinrir.

De manière générale, ces hypothèses sont vérifiées empiriquement dans la littérature qui porte sur l'appréciation des biens immobiliers. Les sensibilités à ces facteurs sont variables, et dans de rares cas, on retrouve des résultats qui contredisent directement nos hypothèses. Ces différences sont expliquées par des caractéristiques locales qui varient selon les marchés (Reichert 1990). En effet, la population ainsi que les revenus varient au niveau des États mais peuvent aussi varier localement dans les villes ou les quartiers. Un État rural aura donc des résultats qui diffèrent d'un État urbanisé, et au sein même d'un État, les caractéristiques peuvent aussi varier localement. Cependant, pour la variable purement macroéconomique qu'est le taux d'intérêt réel des emprunts, la relation avec le prix est homogène peu importe la localisation des biens immobiliers.

Le modèle choisi est un modèle à effets fixes avec une spécification à deux effets spécifiques:

$$y_{it} = x_{it}\beta + \mu_i + \epsilon_t + v_{it} ; t = 1, \dots, T ; i = 1, \dots, N$$

avec  $\mu_i$ , un terme invariant dans le temps désigné par l'effet spécifique individuel,  $\epsilon_t$  un terme invariant entre individus (ici les États) désigné par l'effet spécifique temporel et  $v_{it}$ , une perturbation aléatoire.

Ce modèle permet de contrôler les effets spécifiques et individuels qui pourraient biaiser notre estimation. Le modèle à effets fixes considère les effets individuels de variables inobservées comme constants dans le temps. On considère donc que les comportements des agents sont fixes.

Dans un premier temps, nous allons explorer notre base de données et déterminer quelles sont les informations pertinentes que nous pourrions en tirer, puis nous détaillerons les estimations et leurs résultats, pour enfin conclure et déterminer quelles sont les améliorations que nous pourrions apporter à ce premier modèle.

## 2 Le cadre statistique

### 2.1 Données utilisées

Les données utilisées proviennent du package “pder” disponible sur CRAN. La base de données a pour identifiant “HousePricesUS” et comporte les observations annuelles de 4 variables pour 49 états américains de 1976 à 2003.

| Variables | Définitions   |
|-----------|---|
| price     | indice des prix de l’immobilier corrigé de l’inflation (1980 = 100) |
| income    | revenu horaire réel par habitant                                    |
| pop       | population totale   |
| intrate   | taux d’intérêt réel des emprunts                                    |

### 2.2 Statistiques Descriptives

Table 2: Statistiques Descriptives

|         | N     | Mean          | Std. dev.     | Min.        | Median        | Max.           |
|---------|-------|---------------|---------------|-------------|---------------|----------------|
| year    | 1,421 | 1,989.000     | 8.370         | 1,975.000   | 1,989.000     | 2,003.000      |
| price   | 1,421 | 99.894        | 19.657        | 58.092      | 96.867        | 224.118        |
| income  | 1,421 | 9.933         | 1.764         | 5.910       | 9.718         | 18.219         |
| pop     | 1,421 | 5,076,285.940 | 5,427,431.666 | 380,477.000 | 3,495,939.000 | 35,484,453.000 |
| intrate | 1,421 | 4.363         | 2.597         | -5.544      | 4.572         | 11.225         |

Les statistiques descriptives du tableau 2 indiquent que le revenu réel par habitant médian est de 9,72\$ de l’heure. L’écart-type est assez faible mais la valeur maximum est élevée. La Figure 1 permet de visualiser graphiquement cette répartition asymétrique du revenu réel par habitant avec des valeurs élevées à droite. Concernant la population, les valeurs sont très hétérogènes avec des États dont la population est très faible (Wyoming) et des États dont la population est très élevée (Californie). Le tableau 4 souligne les valeurs extrêmes de ces variations. Le taux d’intérêt réel moyen de 1976 à 2003 est égal à 4,36%, la Figure 2 nous montre l’évolution de celui-ci sur cette période. On constate aussi que ce taux d’intérêt corrigé de l’inflation a été négatif vers les années 80. Cette période correspond au deuxième choc pétrolier caractérisée par une forte hausse du niveau général des prix. Enfin, la variation pour chaque année entre États, représentée par les traits sur chaque points est très faible. Ceci confirme encore une fois le caractère macroéconomique du taux d’intérêt.

La Figure 3 est une matrice de corrélation des variables présentes dans notre base de données. Nous pouvons en tirer quelques informations pertinentes : le prix est corrélé faiblement mais positivement avec le temps, ainsi en retirant l’effet de l’inflation, les prix de l’immobilier sont en hausse depuis 1976. La figure 4 confirme ce premier constat et contraste avec l’évolution du taux d’intérêt par les variations de prix entre chaque État représentées par les traits au niveau de chaque point. Les prix sont donc très hétérogènes. De plus, le prix est corrélé positivement avec le revenu et dans une moindre mesure avec la population. Enfin, à l’image de l’hypothèse que nous avons posé précédemment, le taux d’intérêt est corrélé négativement avec le prix. Cette première analyse des corrélations confirment donc nos prédictions, mais il faut confirmer ces effets en utilisant des modèles plus précis.

## 2.3 Estimations

Table 3: Estimations Pooled OLS, WITHIN et FGLS

|                         | <i>Variables dépendantes : revenu, log(population) et taux d'intérêt</i> |                             |                      |
|-------------------------|--|-----------------------------|----------------------|
|                         | (1)  | (2)                         | (3)                  |
| income                  | 5.290***<br>(0.260)  | 13.282***<br>(0.648)        | 8.355***<br>(0.450)  |
| log(pop)                | 1.725***<br>(0.452)  | 22.333***<br>(3.761)        | 1.892<br>(1.613)     |
| intrate                 | -1.869***<br>(0.172)   | -2.972***<br>(0.315)        | -2.624***<br>(0.231) |
| Constant                | 29.692***<br>(6.695)   |                             | 0.034<br>(24.119)    |
| Observations            | 1,421  | 1,421                       | 1,421                |
| R <sup>2</sup>          | 0.280  | 0.300                       | 0.249                |
| Adjusted R <sup>2</sup> | 0.278  | 0.258                       | 0.247                |
| Residual Std. Error     | 16.699 (df = 1417)   |                             |                      |
| F Statistic             | 183.514*** (df = 3; 1417)  | 191.256*** (df = 3; 1341)   | 469.558***           |
| <i>Note:</i>            |  | *p<0.1; **p<0.05; ***p<0.01 |                      |

La spécification (1) est une régression multiple OLS, la spécification (2) correspond au modèle de choix que nous avons sélectionné et détaillé dans l'introduction. Enfin, la spécification (3) est un modèle à effets aléatoires qui prend lui aussi l'effet individuel ainsi que l'effet temporel en compte. Dans la suite de notre développement, nous expliquerons notre choix de modèle et pourquoi les modèles (1) et (3) sont moins pertinents. L'utilisation du logarithme pour la population est justifiée par la croissance exponentielle de la population au fil du temps (Figure 5).

### 2.3.1 Interprétations

Pour la spécification (1), les interprétations sont les suivantes : une hausse des revenus de 1\$ correspond à une hausse des prix de l'immobilier de 5.29 points d'indice. Sachant que la différence entre le minimum (58.092) et le maximum (224.188) qu'a connu cet indice est égale à environ 166 (voir Tableau 2), et en faisant un produit en croix, ceci correspond à une hausse des prix de 3%. Une hausse de la population de 100% provoque une hausse des prix de 1.73 points sur l'indice des prix de l'immobilier (lin-log). Enfin, une hausse du taux d'intérêt de 1% cause une baisse des prix de l'immobilier de 1.87 points sur l'indice des prix. L'ensemble de ces interprétations sont effectuées toutes choses égales par ailleurs.

Pour la spécification (2), les interprétations sont similaires : une hausse du revenu de 1\$ a pour effet d'augmenter le prix de l'immobilier de 13.28 points sur l'indice des prix. Une hausse de la population de 100% occasionne une hausse des prix de 22.33 points sur l'indice des prix. Pour finir, une hausse du taux d'intérêt de 1% cause une baisse de 2.97 points sur l'indice des prix toutes choses égales par ailleurs.

Les interprétations sont strictement identiques pour la troisième spécification.

## 2.4 Commentaires

Dans cette partie, nous discuterons de la validité économique et statistique de nos trois modèles et pourquoi nous avons sélectionné la deuxième spécification comme notre modèle de référence.

### 2.4.1 Validité économique

La validité économique de notre modèle semble conforme à nos hypothèses et aux résultats de la littérature de ce domaine. La hausse de la population et des revenus provoque effectivement une appréciation des biens immobiliers. Une hausse des taux d'intérêt des emprunts impactent à la baisse les prix de l'immobilier.

### 2.4.2 Validité statistique

#### 2.4.2.1 Pooled OLS

La régression Pooled OLS correspond à une régression multiple avec les Moindres Carrés Ordinaires (MCO) appliquée aux données de panel. Cette méthode ignore les effets individuels et temporels et contrôle uniquement les effets entre les individus. Celle-ci est donc inadaptée lorsqu'elle est utilisée avec des données de panel. Cette technique n'est donc pas valide statistiquement et c'est pourquoi elle n'a pas été sélectionnée.

#### 2.4.2.2 Effets fixes (WITHIN)

Le modèle à effets fixes détermine les effets individuels et temporels de variables inobservés considérées comme fixe dans le temps. On suppose donc que les comportements des agents sont rationnels face aux variations des variables explicatives que nous avons à disposition. De plus l'intégralité de nos variables sont significatives au seuil de 5% et la F-Stat est significative aussi.

Les différents F-tests et Tests de Lagrange (Tableaux 5 à 10 en annexe), nous indiquent que les effets individuels, temporels et les deux effets sont des effets significatifs puisqu'on rejette l'hypothèse nulle à chaque fois. L'utilisation des deux effets spécifiques est donc valide statistiquement.

#### 2.4.2.3 Effets aléatoires (FGLS)

Pour le modèle aléatoire, on considère que les comportements des agents sont aléatoires. Il s'agit là d'un premier problème pour l'utilisation de ce modèle. Dans nos hypothèses, il est supposé que les agents connaissent la loi de l'offre et de la demande et que les prix sont donc fixés de manière rationnelle.

De plus dans le Tableau 3, la population n'est pas une variable significative dans l'estimation. Il n'est donc pas possible de l'interpréter correctement.

Enfin, le dernier facteur qui nous a poussé à ne pas considérer ce modèle est le test d'Hausman (Tableau 11 en annexe). Le test d'Hausman est un test d'endogénéité. L'hypothèse nulle est que la covariance entre les variables explicatives et  $\alpha$ , le facteur qui contient l'hétérogénéité des variables omises est nulle. Dans ce cas, le modèle à effets aléatoires est préféré au modèle à effets fixes. Dans notre cas, l'hypothèse nulle est rejetée, il faut donc utiliser le modèle à effets fixes.

Ainsi, c'est la validité statistique du deuxième modèle qui nous a incité à le privilégier. Le modèle pooled OLS étant directement éliminé et le modèle à effets aléatoires étant invalidé statistiquement par le test d'Hausman.

### 3 Conclusion

Les résultats de notre étude sont donc les suivants : une hausse de la population et des revenus conduisent à une hausse des prix de l'immobilier. Les prix diminuent à la suite d'une hausse du taux d'intérêt des emprunts. Ces résultats sont conformes à nos hypothèses et à la littérature que nous avons citée.

La hausse des revenus est le facteur qui influence le plus les prix de l'immobilier aux États-Unis, suivi par le taux d'intérêt et enfin la croissance démographique. La faiblesse de l'influence de la croissance démographique est avant tout explicable par le fait qu'il faut une hausse massive de la population pour voir un réel changement de la demande sur le marché de l'immobilier.

La première critique qui peut être émise envers notre étude est le manque de variables explicatives. En effet, les taxes, les aides du gouvernement, ou encore le nombre de nouvelles constructions sont des facteurs qui pourraient impacter le prix de l'immobilier dans une certaine mesure.

Ensuite, il serait plus pertinent d'avoir des observations plus locales pour pouvoir comparer les différences au sein d'un État et comparer les différences entre une zone rurale et une zone urbaine. Que ce soit au niveau de la population et des revenus qui impactent directement l'appréciation des biens immobiliers.

Enfin, effectuer cette même analyse sur une autre zone géographique permettrait d'étudier si les sensibilités et les effets de ces facteurs sont identiques ou non.

## 4 Annexe : tableaux

### 4.1 Statistiques Descriptives

Table 4: Valeurs extrêmes

| names      | year | pop        |
|------------|------|------------|
| California | 2003 | 35,484,453 |
| California | 2002 | 35,001,986 |
| California | 2001 | 34,533,054 |
| California | 2000 | 33,999,879 |
| California | 1999 | 33,499,204 |
| ...        | ...  | ...        |
| Wyoming    | 1979 | 451,850    |
| Wyoming    | 1978 | 430,897    |
| Wyoming    | 1977 | 411,530    |
| Wyoming    | 1976 | 395,446    |
| Wyoming    | 1975 | 380,477    |

### 4.2 F-Test

Table 5: F test for individual effects: price ~ income + log(pop) + intrate

| Test statistic | df1 | df2  | P value        | Alternative hypothesis |
|----------------|-----|------|----------------|------------------------|
| 26.57          | 48  | 1369 | 9.9e-161 * * * | significant effects    |

Table 6: F test for time effects: price ~ income + log(pop) + intrate

| Test statistic | df1 | df2  | P value        | Alternative hypothesis |
|----------------|-----|------|----------------|------------------------|
| 4.235          | 28  | 1389 | 1.75e-12 * * * | significant effects    |

Table 7: F test for twoways effects: price ~ income + log(pop) + intrate

| Test statistic | df1 | df2  | P value          | Alternative hypothesis |
|----------------|-----|------|------------------|------------------------|
| 26.66          | 76  | 1341 | 4.765e-215 * * * | significant effects    |

### 4.3 Lagrange Multiplier Test

Table 8: Lagrange Multiplier Test - individual effects :  $\text{price} \sim \text{income} + \log(\text{pop}) + \text{intrate}$

| Test statistic | df | P value                  | Alternative hypothesis |
|----------------|----|--------------------------|------------------------|
| 4233           | 1  | $< 2.2\text{e-}16$ * * * | significant effects    |

Table 9: Lagrange Multiplier Test - time effects :  $\text{price} \sim \text{income} + \log(\text{pop}) + \text{intrate}$

| Test statistic | df | P value                  | Alternative hypothesis |
|----------------|----|--------------------------|------------------------|
| 76.48          | 1  | $2.229\text{e-}18$ * * * | significant effects    |

Table 10: Lagrange Multiplier Test - two-ways effects :  $\text{price} \sim \text{income} + \log(\text{pop}) + \text{intrate}$

| Test statistic | df | P value                  | Alternative hypothesis |
|----------------|----|--------------------------|------------------------|
| 4309           | 1  | $< 2.2\text{e-}16$ * * * | significant effects    |

### 4.4 Hausman Test

Table 11: Hausman Test:  $\text{price} \sim \text{income} + \log(\text{pop}) + \text{intrate}$

| Test statistic | df | P value                  | Alternative hypothesis    |
|----------------|----|--------------------------|---------------------------|
| 139.4          | 3  | $5.077\text{e-}30$ * * * | one model is inconsistent |



## 5 Annexe : Figures

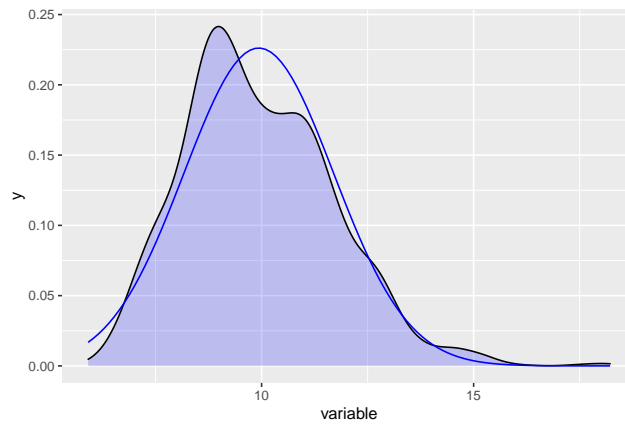


Figure 1: Courbe de densité du revenu réel par habitant

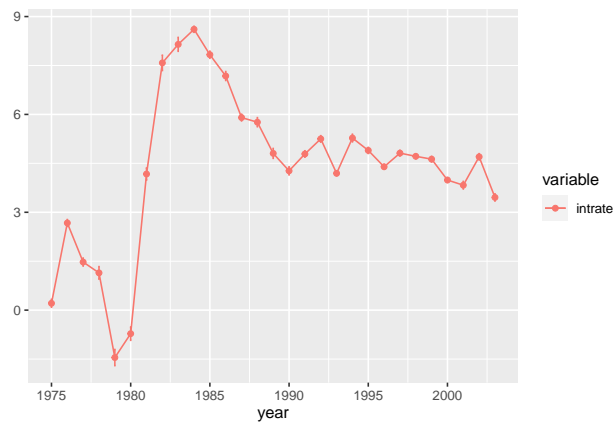


Figure 2: Variation du taux d'intérêt réel d'emprunt en fonction du temps

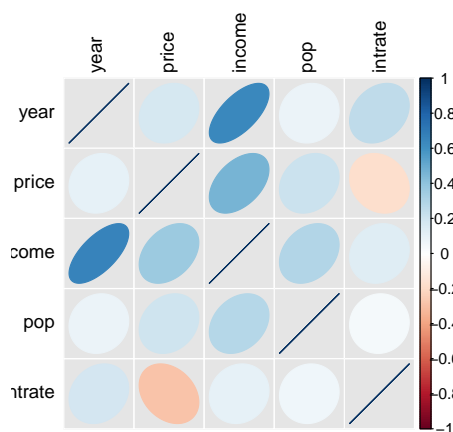


Figure 3: Matrice de corrélation des variables

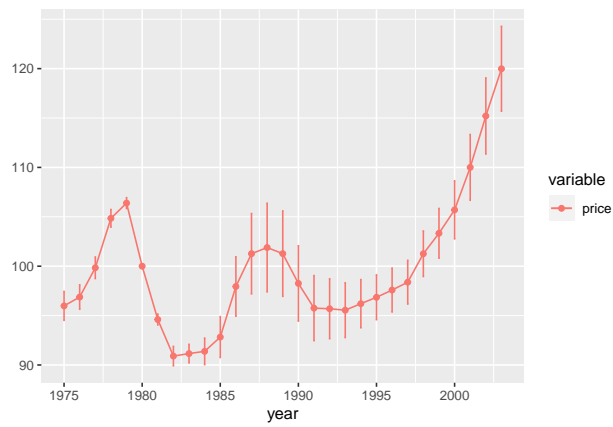


Figure 4: Evolution des prix de l'immobilier en fonction du temps

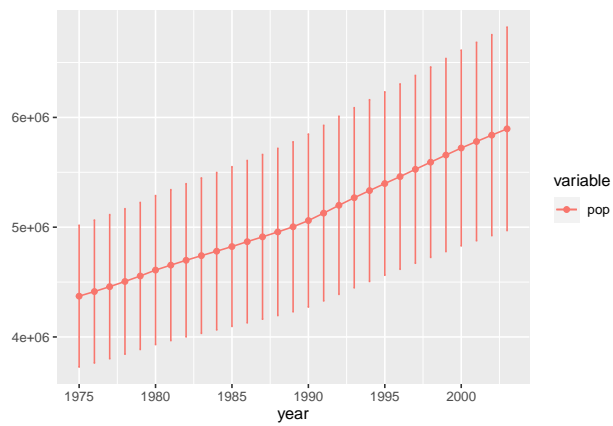


Figure 5: Evolution de la population en fonction du temps

## 6 Annexe : Code

L'intégralité du code utilisé est disponible sur ma page [GitHub](#)

## References

Reichert, A. K. (1990). The impact of interest rates, income, and employment upon regional housing prices. *The Journal of Real Estate Finance and Economics*, 3(4):373–391. Cited on page 1.