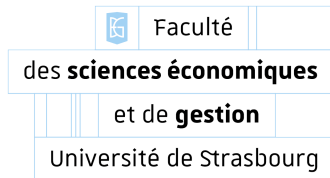


Quels sont les déterminants du prix des maisons a Ames dans l'Iowa?

-

Master Analyse et Politique Economique



Valentin Barthel

valentin.barthel@etu.unistra.fr

Georgiana Coroama

georgiana.coroama@etu.unistra.fr

Julien Schenk

julien.schenk@etu.unistra.fr

Introduction

House Prices, compiled by Dean De Cock - Advanced Regression Techniques.

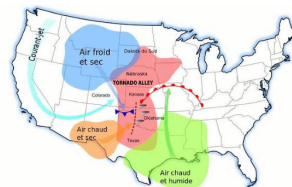
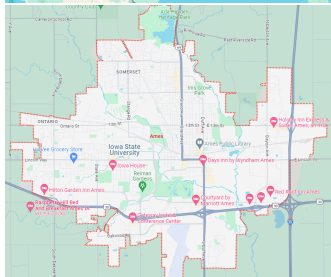
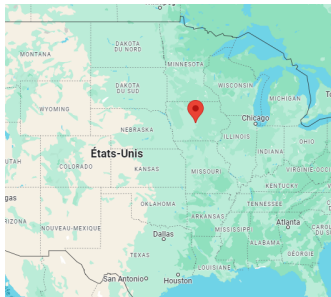
- **Nombre de variables:** 81
- **Nombre d'observations:** 1461
- **Variable cible:** Prix médian des maisons
- **Origine du jeu de données:** Ames, Iowa, USA
- **Année de collecte des données:** 2011
- **Types de données:**
 - Numérique: SalePrice, GrLivArea (surface habitable), LotArea (superficie du terrain), etc.
 - Catégorique: MSZoning (Localisation), BldgType (type du bâtiment), LotShape (forme du terrain), etc.

Table des Matières

- 1 Introduction au sujet
 - Présentation des données
- 2 Spécification du Modèle MCO
 - Présentation du modèle MCO
 - Présentation des résultats du modèle MCO
- 3 Spécification du Modèle IV
 - Présentation des différents quartiers
 - Endogénéité et choix de la variable instrumentale
 - Présentation du modèle IV
 - Présentation des résultats IV et comparaison avec le modèle MCO
- 4 Ajout des spécifications IV par Quartier et IV dummies pour les spécifications individuelles
 - Présentation des modèles et des résultats des spécifications IV par quartier et IV dummies
 - Limites et conclusion

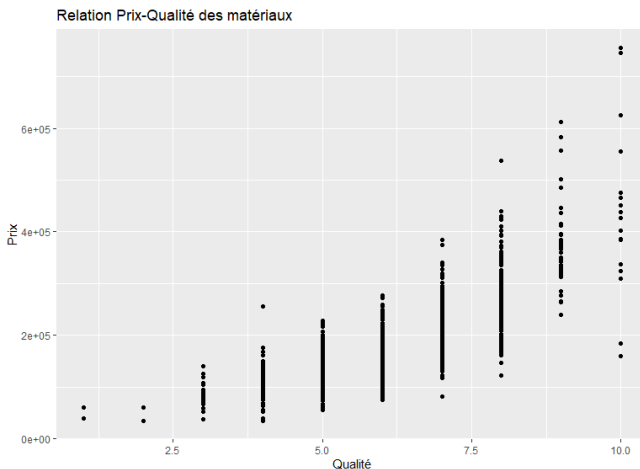
Présentation des données

Géographie et risques météorologiques

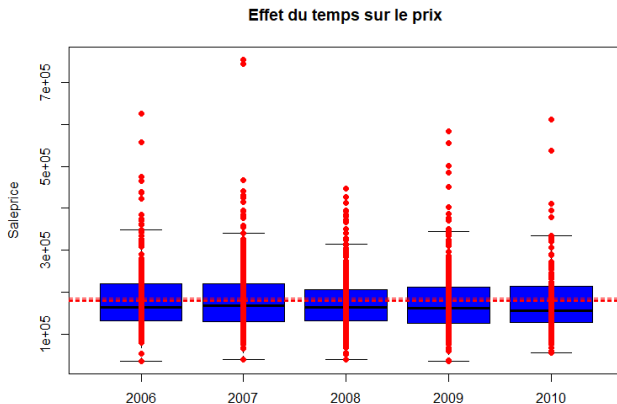


-> Importance de la qualité des matériaux (OverallQual)

Le choix de la qualité



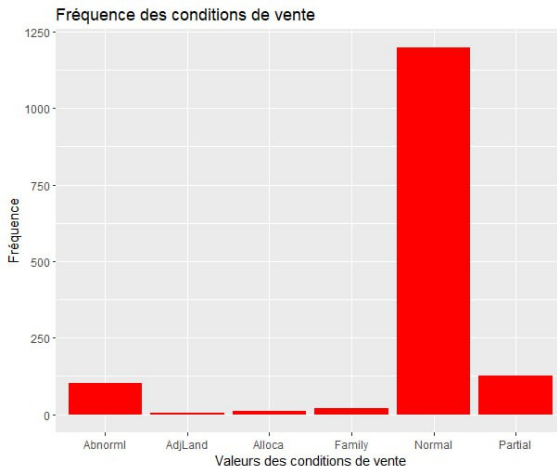
Effet du temps sur le prix



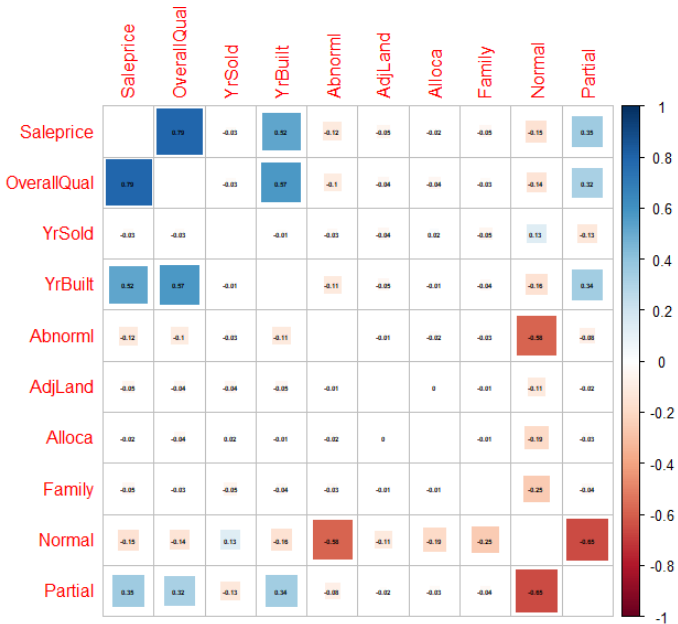
Présentation des différentes conditions de vente

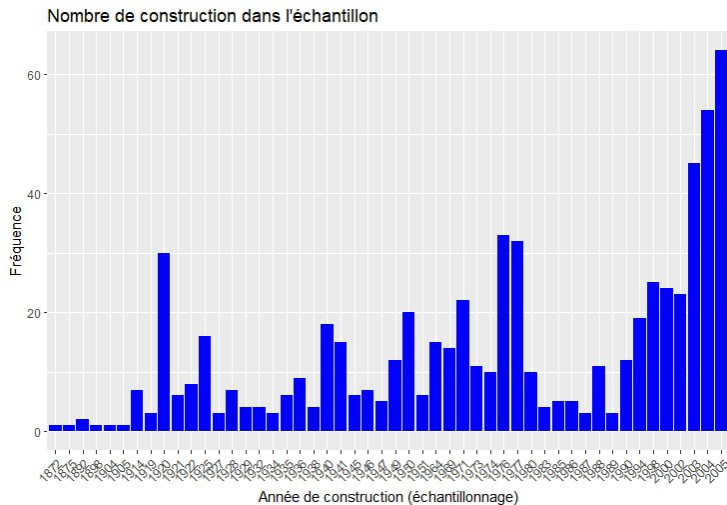
Numérisation de la variable nominale par encodage à chaud

```
condi <- -model.matrix(~ SaleCondi-1, data = nonnumericHomeprice)
```



Présentation des données

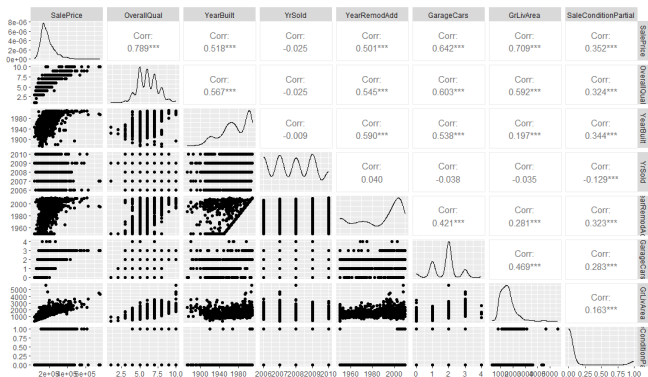




Présentation du modèle MCO

Modèle MCO

$$\begin{aligned}
 \text{Prix}_i = & \beta_0 + \beta_1 \text{Qualite}_i + \beta_2 \text{Renovation}_i + \beta_3 \text{Garage}_i \\
 & + \beta_4 \text{Surface}_i + \beta_4 \text{Partielle}_i + \epsilon_i
 \end{aligned}
 \quad (1)$$



Présentation des résultats du modèle MCO

Résultats du modèle MCO

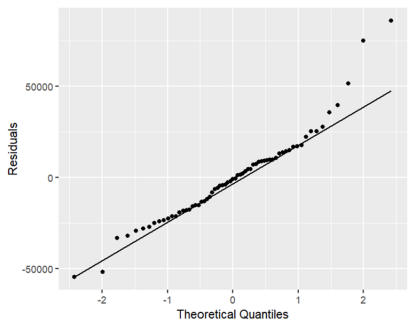
<i>Dependent variable:</i>	
Prix	
Qualite	23,251.970*** (1,161.211)
Renovation	309.042*** (62.063)
Garage	18,707.670*** (1,806.200)
Surface	52.866*** (2.507)
Partielle	24,974.730*** (4,005.672)
Constant	-689,481.700*** (120,498.400)
Observations	1,449
R ²	0.751
Adjusted R ²	0.750
Residual Std. Error	39,592.920 (df = 1443)
F Statistic	870.234*** (df = 5; 1443)

Note: *p<0.1; **p<0.05; ***p<0.01

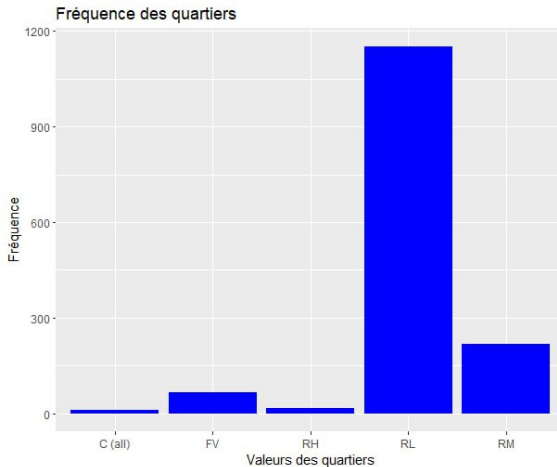
Notes:

Breusch-Pagan P-Value: 5.799408e-62

Durbin-Watson P-Value: 0.4497174



Présentation des différents quartiers



Présentation des différents quartiers



Endogénéité et choix de IV

- Biais de variable omise: la propriété a été incluse dans la SFHA (special flood hazard area) \Rightarrow si oui, réglementations spécifiques imposent des modifications structurelles pour minimiser les risques d'inondation \Rightarrow *Surface habitable au-dessus du sol* **endogène**
- **influence (-)** sur la surface **habitable** au-dessus du sol (exigence de construction de colonnes de soutien à l'intérieur de la maison, élévation de la structure)
- **influence (-)** sur le prix de vente par la réduction de l'attractivité et par la perception du niveau de sécurité de la zone

Endogénéité et choix de IV

- Choix de IV: **Localisation des half-baths**
- Variable binaire: 1, s'il y a au moins une half-bath située au-dessus du sol, 0 sinon
- Intuition: Pièce complémentaire aux salles de bain standard, plutôt spécifique aux maisons de grande taille (positivement corrélée à la surface habitable au-dessus du sol, $r(\text{LocalisationHB}, \text{Surface}) \approx 0.5$).
- Pas d'impact direct de la localisation sur le prix de vente autrement que par l'effet sur la surface habitable au-dessus du sol.

Spécification IV (2LS)

$$\begin{cases} Surface_i = \beta_{0_{FS}} + \beta_{1_{FS}} Localisation_HB_i + \beta_{3_{FS}} x'_i + \nu_i \\ \begin{aligned} Prix_i = & \beta_{0_{IV}} + \beta_{1_{2LS}} Qualite_i + \beta_{2_{2LS}} Renovation_i + \beta_{3_{2LS}} Garage_i \\ & + \beta_{4_{2LS}} Surface_i + \beta_{5_{2LS}} Partielle_i + \epsilon_i \end{aligned} \end{cases} \quad (2)$$

Présentation des résultats IV et comparaison avec le modèle MCO

MCO vs IV

Figure: Régression MCO et IV

	<i>Dependent variable:</i>	
	Prix	
	<i>OLS</i>	<i>instrumental variable</i>
	MCO	IV
	(1)	(2)
Qualite	23,251.970*** (1,161.211)	26,848.510*** (1,935.236)
Renovation	309.042*** (62.063)	274.587*** (64.817)
Garage	18,707.670*** (1,806.200)	21,116.350*** (2,104.683)
Surface	52.866*** (2.507)	35.112*** (7.985)
Partielle	24,974.730*** (4,005.672)	23,693.350*** (4,111.110)
Constant	-689,481.700*** (120,498.400)	-620,285.800*** (126,072.300)
Observations	1,449	1,449
R ²	0.751	0.742
Adjusted R ²	0.750	0.741
Residual Std. Error (df = 1443)	39,592.920	40,274.920
F Statistic	870.234*** (df = 5; 1443)	
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		
Wu-Hausman p.value	: 0.033 *	
Weak instrument p.value	: <0.001 ***	

Spécification IV par quartier et IV Dummies

$$\begin{cases} Surface_i = \beta_{0_{FS}} + \beta_{1_{FS}} Localisation_HB_i + \beta_{3_{FS}} x'_i + \nu_i \\ \begin{aligned} Prix_i = & \beta_{0_{Q_{2LS}}} + \beta_{1_{Q_{2LS}}} Qualite_i + \beta_{2_{Q_{2LS}}} Renovation_i \\ & + \beta_{3_{Q_{2LS}}} Garage_i + \beta_{4_{Q_{2LS}}} Surface_i + \beta_{5_{Q_{2LS}}} Partielle_i + \epsilon_i \end{aligned} \end{cases} \quad (3)$$

$$\begin{cases} Surface_i = \beta_{0_{FS}} + \beta_{1_{FS}} Localisation_HB_i + \beta_{3_{FS}} x'_i + \epsilon_i \\ \begin{aligned} Prix_i = & \beta_{0_{2LS}} + \beta_{1_{2LS}} Qualite_i + \beta_{2_{2LS}} Renovation_i + \\ & + \beta_{3_{2LS}} Garage_i + \beta_{4_{2LS}} Surface_i + \beta_{5_{2LS}} Partielle_i \\ & + \beta_{6_{2LS}} RM_i + \beta_{7_{2LS}} RH_i + \beta_{8_{2LS}} FV_i + \epsilon_i \end{aligned} \end{cases} \quad (4)$$

Présentation des modèles et des résultats des spécifications IV par quartier et IV dummies

Régressions IV par quartier et IV dummies

Figure: Régressions IV par quartier et IV dummies

	Dependent variable:				
	RL (1)	RM (2)	Prix RH (3)	FV (4)	IV Dummies (5)
Quartier_RM					-24,083.280*** (2,998.523)
Quartier_RH					-13,450.360 (10,068.440)
Quartier_FV					-21,951.540*** (5,319.694)
Qualite	28,531.500*** (2,027.411)	17,445.180** (7,001.501)	5,632.566 (12,137.470)	-48,533.860* (26,944.250)	26,649.500*** (1,841.143)
Renovation	200.790*** (77.109)	329.398** (151.618)	680.605** (300.612)	-1,134.056 (3,995.338)	279.291*** (63.116)
Garage	19,423.970*** (2,397.402)	17,860.480*** (6,021.986)	10,029.750 (10,630.940)	66,678.440 (50,109.600)	18,991.150*** (2,036.491)
Surface	36.636*** (7.720)	21.906 (37.952)	57.426 (53.664)	-98.756 (169.929)	34.998*** (7.457)
Partielle	22,399.540*** (4,711.474)	58,356.330*** (14,782.780)		23,108.860 (21,302.580)	25,393.190*** (4,021.494)
Constant	-478,713.100*** (149,734.900)	-674,586.000** (303,246.900)	-1,338,125.000** (501,562.500)	2,144,272.000 (7,389,695.000)	-619,860.500*** (122,826.200)
Observations	1,150	218	16	65	1,419
R ²	0.745	0.658	0.656	-0.276	0.756
Adjusted R ²	0.744	0.650	0.531	-0.384	0.754
Residual Std. Error	40,888.140 (df = 1144)	28,701.760 (df = 212)	24,462.970 (df = 11)	61,607.630 (df = 59)	39,245.010 (df = 1440)

Note:

*p<0.1; **p<0.05; ***p<0.01

Wu-Hausman p.value IV_RL	:	0.054	Wu-Hausman p.value IV_RM	:	0.652
Weak instrument p.value IV_RL	:	<0.001 ***	Weak instrument p.value IV_RM	:	0.043 *
Wu-Hausman p.value IV_RH	:	0.252	Wu-Hausman p.value IV_FV	:	0.007 **
Weak instrument p.value IV_RH	:	0.171	Weak instrument p.value IV_FV	:	0.200
Wu-Hausman p.value dummies	:	0.030 *			
Weak instrument p.value dummies	:	<0.001 ***			

Limites et conclusions

- Les variables sont très corrélées entre elles.
- Spécification adaptée pour l'ensemble des quartiers, mais pas pour chaque quartier pris individuellement.
- Potentielle violation de la condition d'exclusion, $r(\text{LocalisationHB}, \text{Prix de vente}) \approx 0.25$).
- Potentielle présence d'autres problèmes d'endogénéité non identifiés.