# Méthodes micro-économétriques
# Compte-rendu sur les dépenses de santé des Australiens

LEFAFTA Rémi, MANCER Djawed

## Table des matières

# 1 Introduction

## 1.1 Structure des données

Nous avons dans la base données 6 variables :

- `ID` : Identifiant de l'individu avec $i \in [1, 200]$.
- `ANNEE` : Année avec $t \in [1, 5]$.
- `DEPSANTE` : Dépenses en santé annuelles en centaines de dollars.
- `REV` : Revenu en milliers de dollars.
- `AGE` : Age en années.
- `ASSU` $= \begin{cases} 1 & \text{si l'individu a une assurance maladie privée au cours de l'année t} \\ 0 & \text{sinon.} \end{cases}$

Table 1 – Structure des données de notre panel

| col_name | col_index | col_class |
|---|---|---|
| ID | 1 | numeric |
| ANNEE | 2 | numeric |
| DEPSANTE | 3 | numeric |
| REV | 4 | numeric |
| AGE | 5 | numeric |
| ASSU | 6 | numeric |

Nous décidons de transformer la variable `ASSU` en facteur car elle n'a que deux valeurs possibles : 0 ou 1. On transforme notre variable $REV_{it}$ en logarithme, et on crée la variable $AGE^2_{it}$.

## 1.2 Sommaire

| **Variable** | **n** | **Min** | $\mathbf{q_1}$ | $\widetilde{\mathbf{x}}$ | $\bar{\mathbf{x}}$ | $\mathbf{q_3}$ | **Max** | **s** | **IQR** | **#NA** |
|---|---|---|---|---|---|---|---|---|---|---|
| DEPSANTE | 1000 | 0.0 | 3.0 | 5.0 | 4.9 | 7.0 | 13.0 | 2.8 | 4.0 | 0 |
| REV | 1000 | 22.0 | 62.0 | 74.0 | 73.6 | 86.0 | 146.0 | 18.6 | 24.0 | 0 |
| AGE | 1000 | 21.0 | 35.0 | 47.0 | 46.4 | 59.0 | 70.0 | 13.8 | 24.0 | 0 |
| lrev | 1000 | 3.1 | 4.1 | 4.3 | 4.3 | 4.5 | 5.0 | 0.3 | 0.3 | 0 |
| agesq | 1000 | 441.0 | 1225.0 | 2209.0 | 2344.1 | 3481.0 | 4900.0 | 1276.1 | 2256.0 | 0 |

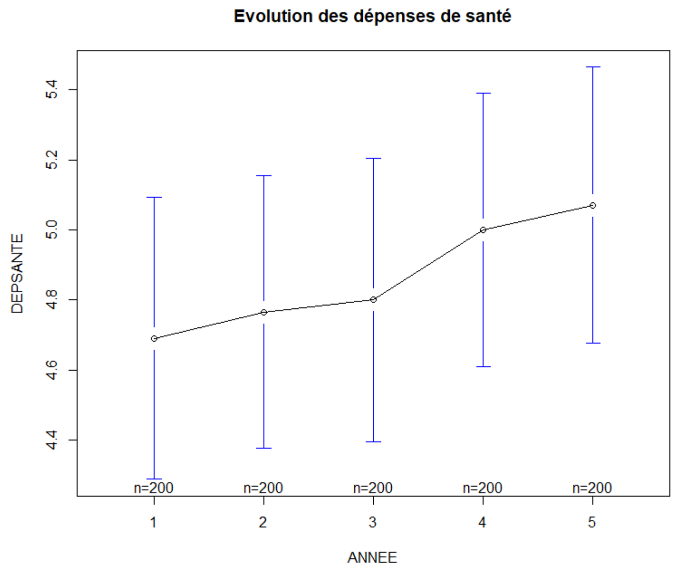
Table 2: Statistiques descriptives

Après vérification, il n'y a aucune donnée manquante.
On remarque que notre panel est très dispersé. En effet, l'étendue de l'âge et surtout celle du revenu est élevée. On note même qu'il y a certains individus qui n'ont rien dépensé en termes de santé dans une année.

## 1.3 Vérification si panel cylindré ou non

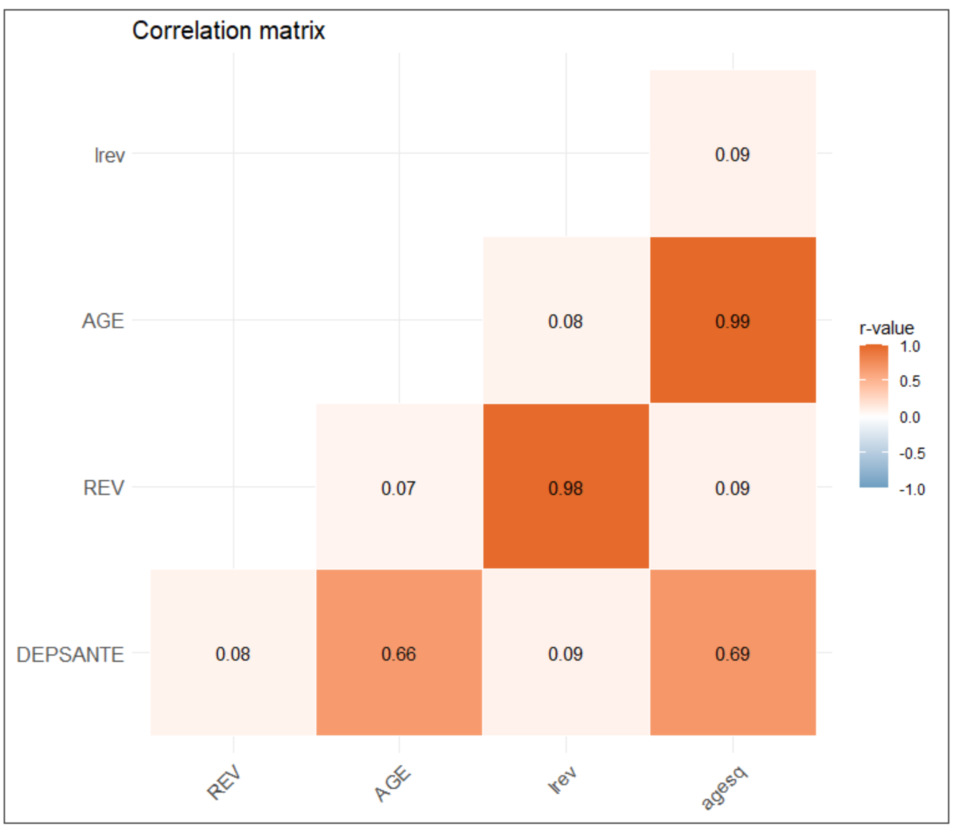
Balanced Panel : n = 200, T = 5, N = 1000.
On conclue que le panel est bel et bien cylindré, on peut donc commencer à l'étudier.

## 1.4 Evolution des dépenses de santé




On remarque une hause des dépenses de santé au fil du temps.

## 1.5 Matrice des corrélations




De la matrice des corrélations, on remarque que toutes les variables exercent un impact positif sur les dépenses de santé. La variable la plus corrélée aux dépenses de santé la plus forte est l'âge au carré (et de fait l'âge également). On note cependant que le revenu et les dépenses de santé ont une corrélation assez proche de 0.

## 2 Le revenu au logarithme

Lorsque l'on a une variable explicative en logarithme, on peut directement étudier le changement en unité de dépense de santé suite à une variation en pourcentage du revenu. Démontrons cela : Notre modèle s'écrit de la manière suivante :

$$DEPSANTE_{it} = \beta_{1i} + \beta_2 ln(REV_{it}) + \beta_3 AGE_{it} + \beta_4 AGE_{it}^2 + \beta_5 ASSU_{it} + \epsilon_{it} \quad \forall \quad i \in [1, 200],\ t \in [1, 5]$$

Si l'on dérive $DEPSANTE_{it}$ par rapport à $REV_{it}$ on obtient :

$$\frac{\partial DEPSANTE_{it}}{\partial REV_{it}} = \frac{\beta_2}{REV_{it}} \leftrightarrow \beta_2 = \frac{\partial DEPSANTEit}{\frac{\partial REV_{it}}{REV_{it}}}$$

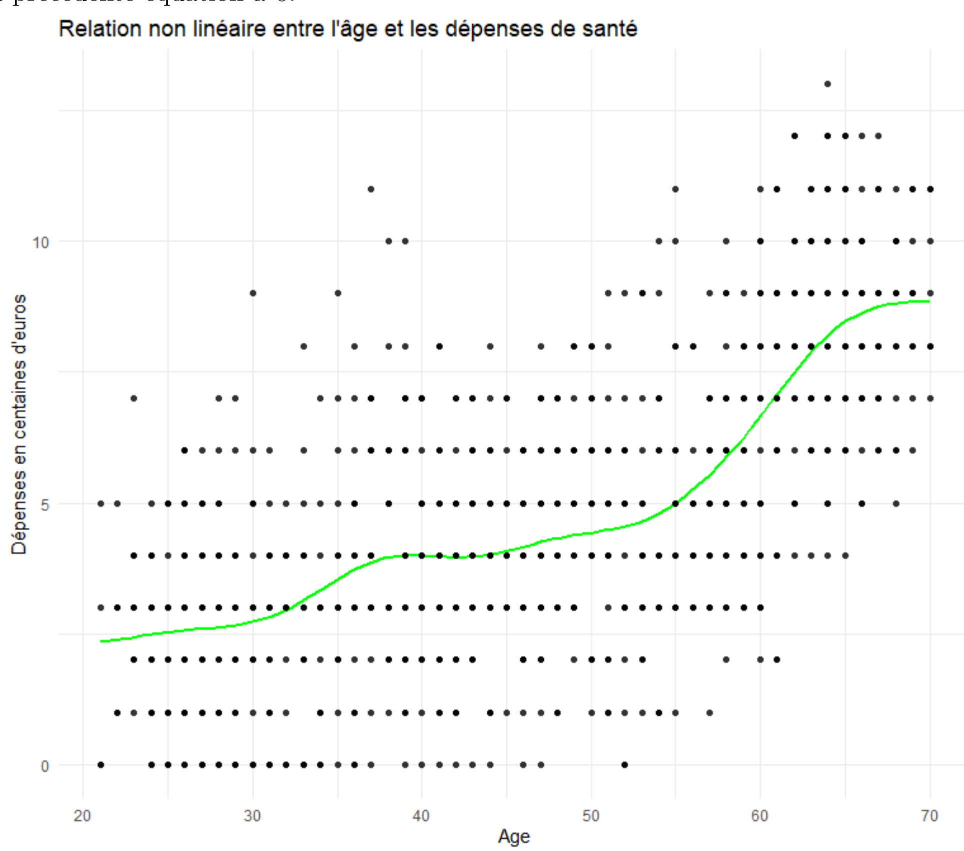
On finit par diviser par 100 et l'on obtient enfin : $\frac{\beta_2}{100} = \frac{\partial DEPSANTE_{it}}{\%\Delta REV_{it}}$

## 3 Utilité de l'ajout d'une variable au carré

On sait que l'âge n'exerce pas forcément un impact linéaire sur la variable à expliquer. Si l'on calcule

$$\frac{\partial DEPSANTE_{it}}{\partial AGE_i t} = \beta_3 + 2\beta_4 AGE_{it}$$

On voit que l'effet marginal de l'âge sur les dépenses de santé est linéaire, ce qui signifie que l'effet de l'âge sur les dépenses de santé est non linéaire. Cela montre que l'âge permet d'avoir un impact non linéaire sur les dépenses de santé. En effet, si par exemple $\beta_3$ est positif et $\beta_4$ négatif, on comprend que l'âge exerce un effet positif sur les dépenses de santé mais que l'effet positif s'estompera et sera de moins en moins fort au fur et à mesure que l'âge croît. Pour déterminer quand est-ce que l'impact de l'âge devient non linéaire, il suffit d'égaliser la précédente équation à 0.




On voit visuellement, que il y une relation non-linéaire entre la variable `AGE` et `DEPSANTE`.

# 4 Estimations

## 4.1 MCO

$$DEPSANTE_{it} = \beta_1 + \beta_2 ln(REV_{it}) + \beta_3 AGE_{it} + \beta_4 AGE_{it}^2 + \beta_5 ASSU_{it} + \epsilon_{it}$$
$$\forall i \in [1, 200],\ t \in [1, 5]$$

Table 3 – Estimation du modèle pooled

| | *Variable expliquée :* |
|---|---|
| | DEPSANTE |
| lrev | $0.392^{*}$ |
| | (0.220) |
| AGE | $-0.208^{***}$ |
| | (0.033) |
| agesq | $0.004^{***}$ |
| | (0.0004) |
| ASSU1 | $1.517^{***}$ |
| | (0.123) |
| Constant | $3.439^{***}$ |
| | (1.221) |
| Observations | 1,000 |
| $R^2$ | 0.550 |
| Adjusted $R^2$ | 0.548 |
| Residual Std. Error | 1.910 (df = 995) |
| F Statistic | $303.859^{***}$ (df = 4 ; 995) |
| *Note :* | $^{*}$p<0.1 ; $^{**}$p<0.05 ; $^{***}$p<0.01 |

Pour le modèle estimé par la méthode des MCO, nous avons un $R^2$ de 0.55. Tous nos coefficients sont significatif à 1% sauf celui associé ($\beta_2$) à `lrev` qui est significatif à seulement 10%. Par exemple, une personne ayant une assurance privé, paye en moyenne 151.7$ de plus qu'une personne non assurée.

## 4.2 Modèle à effets fixes

On utilise un modèle à effets fixes lorsque l'on cherche à analyser l'impact des variables qui changent au cours du temps. Dans un modèle à effets fixes individuels, on considère que les paramètres $\beta_k$ avec $k = 1, 2, 3, 4$ sont homogènes pour tous les individus $i \in [1, 200]$ et où l'hétérogénéité se voit modélisée grâce à une constante individuelle. On a ainsi, pour l'individu $i$, le modèle suivant :

$$DEPSANTE_{it} = \beta_i + x_{it}'\beta + \epsilon_{it}$$

.

On réécrit le modèle comme suit :

$$DEPSANTE_{it} = \beta_{1i} + \beta_2 ln(REV_{it}) + \beta_3 AGE_{it} + \beta_4 AGE_{it}^2 +$$

$$\beta_5 ASSU_{it} + \epsilon_{it}$$
$$\forall i \in [1, 200],\ t \in [1, 5]$$

#### 4.2.1 Within

L'estimation Within est une estimation où les observations sont centrées sur la moyenne individuelle. Lorsque les observations sont centrées, on estime par la méthode des MCO.

$$DEPSANTE_{it} - \overline{DEPSANTE_i} = (\beta_{1i} - \overline{\beta_i}) + \beta_2(ln(REV_{it}) - ln(\overline{REV_i})) + \beta_3(AGE_{it} - \overline{AGE_i}) +$$
$$\beta_4(AGE_{it}^2 - \overline{AGE_i^2}) + \beta_5(ASSU_{it} - \overline{ASSU_i}) + \epsilon_{it} - \epsilon_i$$
$$\forall i \in [1, 200],\ t \in [1, 5]$$

Table 4 – Estimation du modele within

| | *Variable expliquée :* |
|---|---|
| | DEPSANTE |
| lrev | −0.105 |
| | (0.754) |
| AGE | 0.065 |
| | (0.094) |
| agesq | 0.0003 |
| | (0.001) |
| ASSU1 | 1.351*** |
| | (0.114) |
| Observations | 1,000 |
| $R^2$ | 0.166 |
| Adjusted $R^2$ | −0.046 |
| F Statistic | 39.678*** (df = 4 ; 796) |
| *Note :* | *p<0.1 ; **p<0.05 ; ***p<0.01 |

Le coefficient associé à la variable indicatrice correspond à l'estimation de la constante pour l'individu $i$. Le coefficient estimé de la variable `lrev` indique que si le revenu augmente de 1% alors les dépenses de santé diminue de 0.105$.
Pour l'interprétation du coefficient associé à la variable `age` il faut prendre en compte aussi la variable `agesq`. Les deux s'interprétent ensemble, en effet $\frac{\partial DEPSANTE}{\partial AGE} = 0.065 + 2 * 0.0003 AGE_{it}$. Si l'individu $i$ a une assurance alors il dépense en moyenne 135$ en plus que une personne non assuré. Avec la méthode Within nous avons un $R^2 = 0.16$, seul le coefficient associé à la variable `ASSU` est significatif à 1%. Enfin, nous aurions pu, grâce à la fonction `fixef` du package `plm`, obtenir les effets fixes individuels pour chacun des individus. On obtient un effet individuel moyen de $\approx 0.93$.

### 4.3 Modèle à effets aléatoires par MCQG

Le modèle qu'on estime est :

$$DEPSANTE_{it} = \beta_1 + \beta_2 ln(REV_{it}) + \beta_3 AGE_{it} + \beta_4 AGE_{it}^2 + \beta_5 ASSU_{it} + \epsilon_{it} + \mu_i$$

$$\forall i \in [1, 200],\ t \in [1, 5]$$

Ce modèle est à erreurs composées, c'est à dire que que nous avons deux termes d'erreurs $\epsilon_{it}$ qui est identique à celui des MCO. Mais aussi le terme d'erreur $\mu_i$ qui est une variable aléatoire. $\epsilon_{it}$ correspond à la variance intra-individuelle alors que $\mu_i$ correspond lui à la variance inter. Le modèle à effets aléatoires a un avantage sur le modèle à effets fixes : celui de pouvoir intégrer les variables invariantes au cours du temps. Ce modèle a également l'incorporation du terme $\theta$. Ce terme se calcule comme suit :

$$\theta = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_{\beta_1}^2}}$$

Comment l'interpéter ? Le terme qui est mis en racine carrée n'est ni plus ni moins que la part des variations intra-individuelles parmi les variations totales. Autrement dit, $\approx 27\%$ de la moyenne individuelle est retirée de chaque variable pour l'estimation par MCQG. Ainsi, $\widehat{\theta}$ valant $\approx 73\%$, on en conclue que l'on se rapproche davantage d'un modèle à effets individuels que d'un modèle pooled.

Table 5 – Estimation du modèle à effets aléatoires

| | *Variable expliquée :* |
|---|---|
| | DEPSANTE |
| lrev | $-0.149$ |
| | (0.292) |
| AGE | $-0.090^{*}$ |
| | (0.051) |
| agesq | $0.002^{***}$ |
| | (0.001) |
| ASSU1 | $1.362^{***}$ |
| | (0.107) |
| Constant | $3.443^{**}$ |
| | (1.642) |
| Observations | 1,000 |
| $R^2$ | 0.308 |
| Adjusted $R^2$ | 0.305 |
| F Statistic | $441.870^{***}$ |
| *Note :* | $^{*}$p<0.1 ; $^{**}$p<0.05 ; $^{***}$p<0.01 |

Les coefficients associés aux variabless `ASSU` et `agesq` sont significatifs à 1%.

# 5 Test d'homogénéité des constantes

Afin de répondre à cette question, on se sert du troisième test de la procédure de Hsiao. Les hypothèses de test sont :

$$H_0^1 : \alpha_i = \alpha \quad \forall i \in [1, N]$$
$$H_1^1 : \exists\ (i, j) \in [1, N]\ /\ \alpha_i \neq \alpha_j$$

On suppose que les $\beta_i$ sont homogènes. On a des contraintes uniquement sur les constantes.

La statistique de test de Fisher s'écrit :

$$F = \frac{SCR_{pooled} - SCR_W/[N-1]}{SCR_W/[N(T-1)-K]} \sim F[(N-1), N((T-1)-K]$$

Si F est supérieur au seuil théorique de la distribution de Fisher avec $F(199, 796)$ et 5% de risque, on rejette $H_0^1$ alors les constantes sont individuelles. Si on conserve $H_0^1$ alors on a une constante commune à tous nos individus.

On a $N-1 = 199$ contraintes et $N(T-1)-K = 796$ degrès de liberté.

On récupère les sommes des carrées des résidus du modèle pooled et within.

$$F = \frac{3630.26 - 845.43}{845.43}\frac{796}{199} \approx 13.18 > F^{0.05}(199, 796) = 1.19$$

On rejette $H_0^1$, on a donc des constantes individuelles.

# 6 L'effet du revenu sur les dépenses de santé

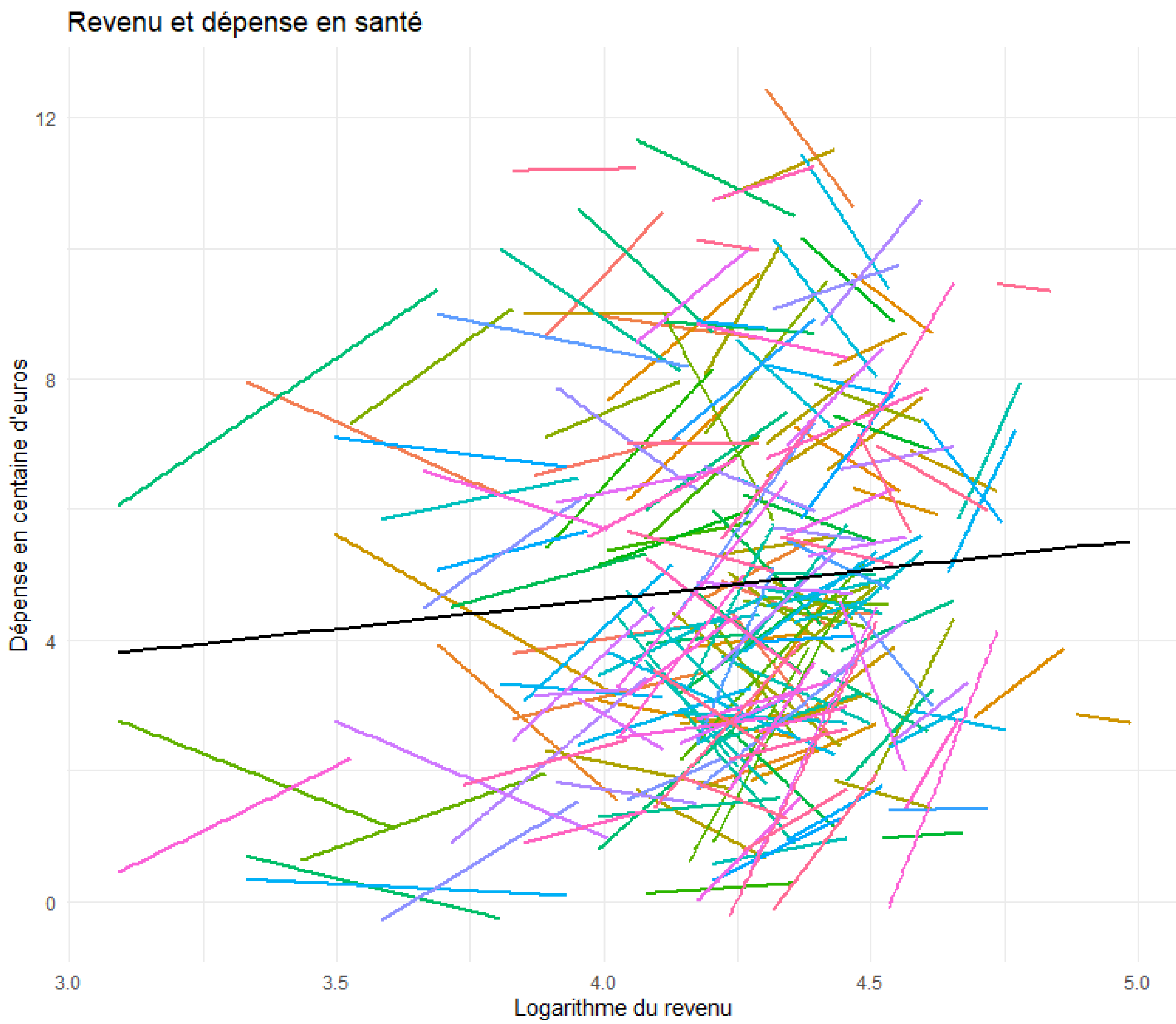
Table 6 – Comparaison des modèles MCO, Within et Random

| | *Variable expliquée :* | | |
|---|---|---|---|
| | DEPSANTE | | |
| | MCO | Within | Random |
| lrev | 0.392* | −0.105 | −0.149 |
| | (0.220) | (0.754) | (0.292) |
| AGE | −0.208*** | 0.065 | −0.090* |
| | (0.033) | (0.094) | (0.051) |
| agesq | 0.004*** | 0.0003 | 0.002*** |
| | (0.0004) | (0.001) | (0.001) |
| ASSU1 | 1.517*** | 1.351*** | 1.362*** |
| | (0.123) | (0.114) | (0.107) |
| Constant | 3.439*** | | 3.443** |
| | (1.221) | | (1.642) |
| Observations | 1,000 | 1,000 | 1,000 |
| $R^2$ | 0.550 | 0.166 | 0.308 |
| Adjusted $R^2$ | 0.548 | −0.046 | 0.305 |
| Residual Std. Error | 1.910 (df = 995) | | |
| F Statistic | 303.859*** (df = 4 ; 995) | 39.678*** (df = 4 ; 796) | 441.870*** |
| *Note :* | | | *p<0.1 ; **p<0.05 ; ***p<0.01 |

Le seul modèle où le revenu a un effet positif est le modèle pooled. De plus, seulement pour le modèle pooled le coefficient associé à la variable `lrev` est significatif à 10% et entraine une augmentation des dépenses de 0.39$.

En revanche les modèles Within et Random, indiquent que les dépenses diminuent respectivement de 0.11$ et 0.15$. La prise en compte des effets individuels engendre un effet marginal négatif du revenu.

### 6.0.1 Compléments microéconomiques sur l'effet du revenu sur les dépénses de santé de chaque individu




Grâce à cette représentation graphique, on peut voir que certains individus considèrent les dépenses de santé comme un bien inférieur. En effet, lorsque leur revenu augmente il baisse leur dépense de santé. A l'inverse, d'autres individus, considèrent la santé comme un bien normal, plus leur revenu est élevé plus ils dépensent en santé.

# 7 Comparaison des modèles Within et Random

Afin de déterminer le bon choix de modèle entre un à effets fixes et un à effets aléatoires, on performe un test d'Hausman dans lequel l'hypothèse nulle correspond à un modèle à effets aléatoires.

Table 7 –

| | *Variable expliquée :* | |
|---|---|---|
| | DEPSANTE | |
| | Within | Random |
| lrev | $-0.105$ | $-0.149$ |
| | (0.754) | (0.292) |
| AGE | 0.065 | $-0.090^{*}$ |
| | (0.094) | (0.051) |
| agesq | 0.0003 | $0.002^{***}$ |
| | (0.001) | (0.001) |
| ASSU1 | $1.351^{***}$ | $1.362^{***}$ |
| | (0.114) | (0.107) |
| Constant | | $3.443^{**}$ |
| | | (1.642) |
| Observations | 1,000 | 1,000 |
| $R^2$ | 0.166 | 0.308 |
| Adjusted $R^2$ | $-0.046$ | 0.305 |
| F Statistic | $39.678^{***}$ (df = 4 ; 796) | $441.870^{***}$ |
| *Note :* | $^{*}$p<0.1 ; $^{**}$p<0.05 ; $^{***}$p<0.01 | |

Pour le modèle Within, le seul coefficient significatif est celui associé à la variable `ASSU`. A l'inverse, pour le modèle random, tous les coefficients sont significatifs sauf $\beta_2$ associé à la variable `lrev`. De plus, on remarque que l'effet marginal exercé par l'âge est opposé, l'un étant positif (Within) et l'autre négatif (random). Le coefficient associé à la variable `agesq` est positif, au fur et à mesure qu'un individu vieillit, il dépense de plus en plus en santé. Un $R^2$ de 30% pour le modèle random est très bon pour des données microéconométrique. On ne peut comparer cependant les deux $R^2$ car nous ne disposons pas des mêmes classes de modèle ni d'estimateur.

$$\begin{cases} H_0: & \mathbb{E}[\beta_{1i}|X_i] = 0 \\ H_1: & \mathbb{E}[\beta_{1i}|X_i] \neq 0. \end{cases}$$

Si l'on conserve $H_0$ alors on préfère garder le modèle à effet aléatoire et donc l'estimateur des $\widehat{\beta}_{MCQG}$ qui d'après les hypothèses de Gauss-Markov est BLUE. Sous $H_1$, le modèle à effets fixes individuelles est préferé car l'estimateur $\widehat{\beta}_W$ est non biaisé.

La statistque de test :

$$H = (\widehat{\beta}_{MCQG} - \widehat{\beta}_W)'[Var(\widehat{\beta}_{MCQG} - \widehat{\beta}_W)]^{-1} - (\widehat{\beta}_{MCQG} - \widehat{\beta}_W) \sim \chi^2(K)$$

Si la statistique de test est supérieure à la valeur critique du $\chi^2_{0.95}$ alors on rejette $H_0$.

La réalisation de la statistique de test nous donne $H = 16.608$.

On a de plus $H^{0.05}(4) = 9.49$

On a donc $H > H^{0.05}(4)$ ce qui implique le rejet de $H_0$, on préfère conserver le modèle Within.

# 8 Le modèle Random

## 8.1 L'effet marginal de l'âge sur les dépenses de santé

$$\frac{\partial DEPSANTE_{it}}{\partial AGE_{it}} = -0.09 + 2 * 0.002 AGE_{it}$$

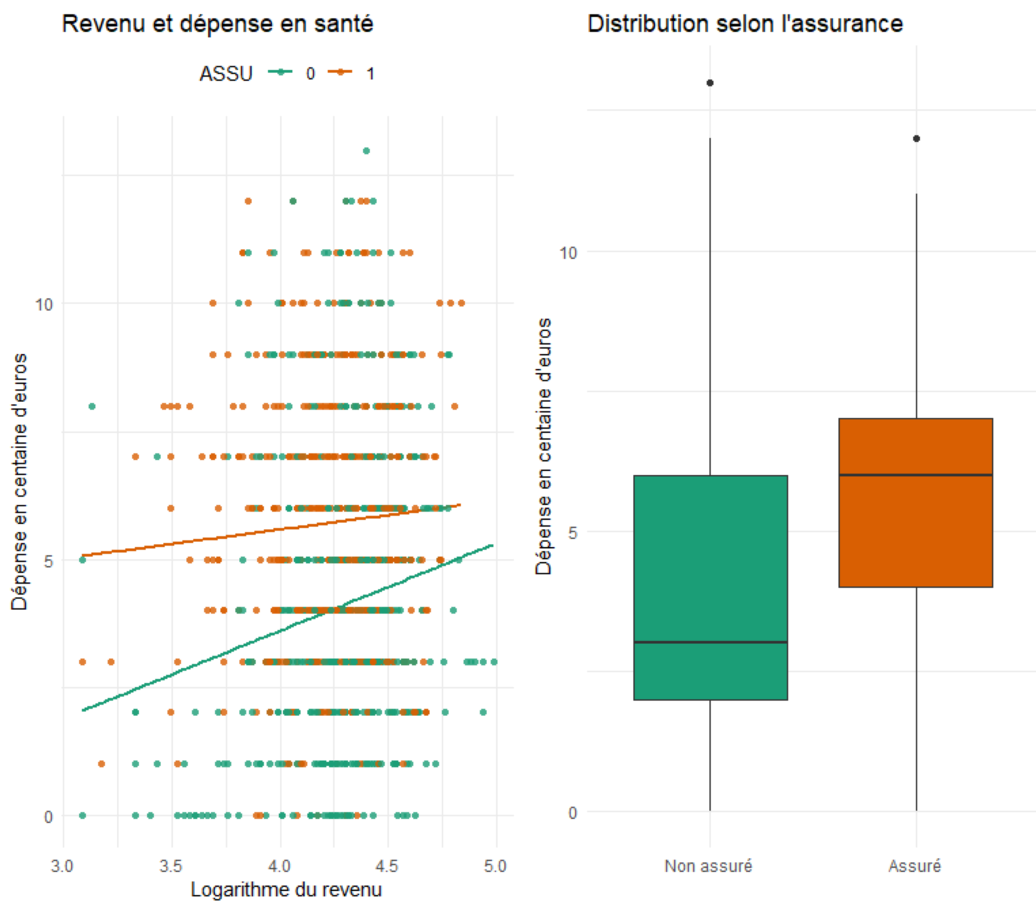
La dérivée seconde est :

$$\frac{\partial^2 DEPSANTE_{it}}{\partial AGE_{it}^2} = 0.004 > 0$$

Cela confirme ce qu'on a dit précédemment, les dépenses de santé augmentent plus on prend de l'âge.

## 8.2 L'effet de l'assurance sur les dépenses de santé

Sachant que le coefficient associé à l'assurance est positif, ceux qui ont souscrit à une assurance privé dépensent plus en santé.




On voit graphiquement très bien le résultat énoncé. Donc ceux qui ont une assurance santé dépensent en moyenne 136$ de plus que ceux qui ne sont pas assurés. Donc, sur le premier graphique, on observe que deux individus ayant des revenus faibles, ont un grand écart en terme de dépense de santé en fonction de s'ils sont assurés ou non. Cet écart s'estompe lorsque les revenus augmentent.
Sur le second graphique, on voit que l'écart inter-quartile est plus grand pour les non assurés que les assurés. La médiane des assurés est plus élévés que le troisième quartile des non assurés. A noter cependant que l'individu qui dépense le plus en santé est non assuré.

# 9 L'inefficacité du modèle Within

Le modèle Within consistant à soustraire à chaque variable sa moyenne, si l'on a : $ASSU_{it} = 0 \ ou \ 1 \ \forall t \in [1;5]$ alors on aura :

$$ASSU_{it} - \overline{ASSU_{it}} = ASSU_{it} - ASSU_i = 0$$

D'où la disparition de la variable `ÀSSU` de notre estimation. Effectivement, la moyenne temporelle de la variable `ÀSSU` pour un individu $i$ est : $ASSU_it$.

Si `ASSU` est invariant alors il n'y a pas de variabilité intra dans la variable `ASSU` pour chaque individu et donc les valeurs de la variable `ASSU` transformée Within seront toujours égales à 0. On en revient à un modèle avec pour variables explicatives : `age`, `age^2` et `ln(rev)`.
En outre, on sait qu'un modèle à effets fixes n'est pas du tout efficace pour des données dont la variance intra des individus au fil du temps est faible. Il souffre également en terme d'efficacité si les variables évoluent très peu au cours du temps. Ainsi, si ASSU est invariant, alors le modèle Within est inefficace pour estimer l'impact de l'assurance maladie privée sur les dépenses de santé. En effet, les variables invaraiantes sont parfaitement colinéaires avec l'individu. Enfin, le modèle à effets fixes est étudié afin d'étudier les causes de changement d'un individu.

Table 8 – Estimation du modèle within avec Assurance invariante

| | *Dependent variable :* |
|---|---|
| | DEPSANTE |
| lrev | 0.267 |
| | (0.817) |
| AGE | 0.047 |
| | (0.102) |
| agesq | 0.0004 |
| | (0.001) |
| Observations | 1,000 |
| $R^2$ | 0.020 |
| Adjusted $R^2$ | −0.229 |
| F Statistic | $5.386^{***}$ (df = 3 ; 797) |
| *Note :* | $^{*}$p<0.1 ; $^{**}$p<0.05 ; $^{***}$p<0.01 |

On remarque que aucun des coefficients n'est significatif.

# 10 Conclusion

En guise de conclusion, suite à nos différentes estimations, nous avons vu qu'en fonction du type de modèle choisi, les variables impactaient différement les dépenses de santé.
L'asssurance joue, peu importe le modèle, un rôle positif sur les dépenses de santé. A l'inverse, l'effet du revenu doit être davantage nuancé.
Il aurait peut-être été intéréssant d'effectuer une ACP afin de regrouper nos individus et nos variables.