Méthodes micro-économétriques
Estimation des dépenses de santé des particuliers en Australie

Pierre-Emmanuel Diot

Contents

1 Présentation des données

L'étude se base sur le jeu de données `SANTE.dta` et a pour double objectif d'analyser les facteurs qui influencent les dépenses de santé des particuliers australiens et de déterminer s'il existe des effets individuels liés à ces dépenses. En vue de répondre à ces deux problématiques, plusieurs méthodes d'estimation économétrique seront utilisées.

Table 1: Les 2 premiers individus de la base de données

ID	ANNEE	DEPSANTE	REV	AGE	ASSU
1	1	9	49	51	1
1	2	9	51	52	1
1	3	9	55	53	1
1	4	10	58	54	1
1	5	11	61	55	1
2	1	6	48	62	1
2	2	7	48	63	1
2	3	7	58	64	1
2	4	7	59	65	1
2	5	7	63	66	1

Table 2: Format des variables

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

Table 3: Statistiques descriptives

	Min	25%	Median	75%	Max	Mean	Sd_cor
DEPSANTE	0	3	5	7	13	4.86	2.84
REV	22	62	74	86	146	73.57	18.60
AGE	21	35	47	59	70	46.42	13.76

Balanced Panel: n = 200, T = 5, N = 1000

2 Revenu et dépenses de santé

Dans le cas d'un modèle niveau-niveau, β_2 représenterait la variation de la variable `DEPSANTE` en centaines de dollars si `REV` augmentait d'une unité de milliers de dollars, soit 1000 dollars, ce qui est difficilement interprétable.

La mise au logarithme du revenu permet d'obtenir la différence de niveau de la variable à expliquer `DEPSANTE` pour une certaine augmentation en poucentage de la variable explicative `REV`.

En notant y la variable associée aux dépenses de santé et x_1 la variable associée au revenu, on obtient:

$$\frac{\mathrm{d}y}{\mathrm{d}x_1} = \frac{\beta_2}{x_1} \iff \beta_2 = \frac{\mathrm{d}y}{\frac{\mathrm{d}x_1}{x_1}}$$

En divisant par 100 de chaque côté, on obtient alors:

$$\frac{\beta_2}{100} = \frac{\mathrm{d}y}{100\frac{\mathrm{d}x_1}{x_1}} = \frac{\mathrm{d}y}{\%\Delta x_1}$$

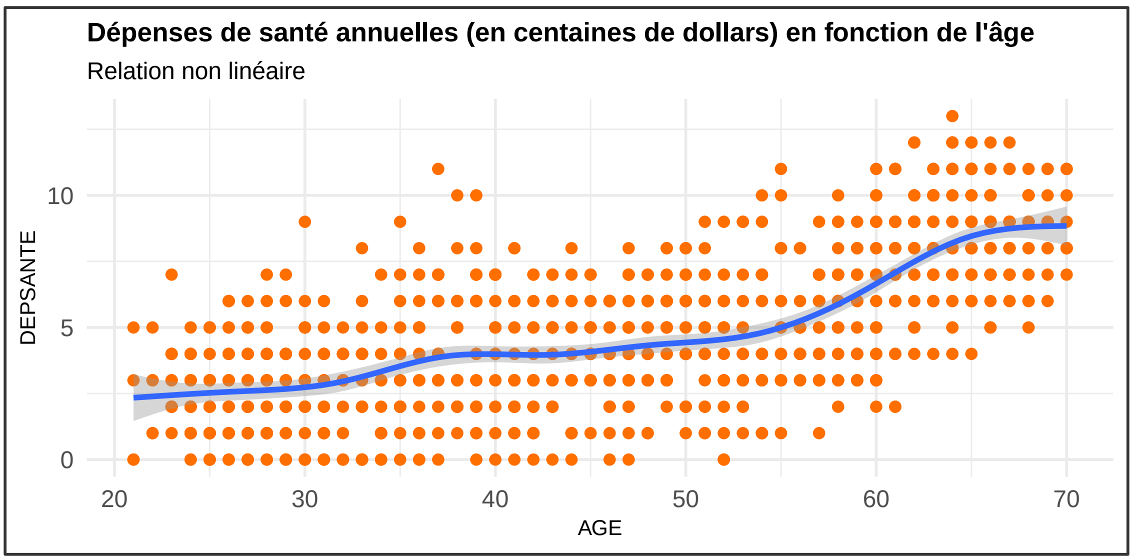
Il vient que lorsque x_1 augmente de 1% alors y varie de $\frac{\beta_2}{100}$ unités. Autrement dit, si le revenu annuel en milliers de dollars augmente de 1% alors les dépenses de santé annuelles varient de $\frac{\beta_2}{100}$ centaines de dollars.

3 Age et dépenses de santé

L'ajout de la variable `AGE` au carré dans le modèle permet de prendre en compte la non linéarité entre `AGE` et `DEPSANTE`. En effet, on peut voir sur le graphique suivant que les dépenses de santé annuelles augmentent très légèrement jusqu'à l'âge de 55 ans. A partir de 55 ans, les dépenses de santé augmentent exponentiellement jusqu'à l'âge de 65 ans puis semblent atteindre un plateau. Ainsi `DEPSANTE` et `AGE` ne semble pas évoluer proportionnellement, i.e. la relation n'est pas linéaire entre ces deux variables. On estimera l'effet marginal de l'âge sur les dépenses de santé comme suit:

$$\frac{\partial y}{\partial x_2} = \beta_3 + 2\beta_4 x_2$$

où y représente `DEPSANTE` et x_2 est associé à `AGE`.



4 Estimations de différents modèles

Dans chaque modèle, les coefficients associés aux variables explicatives `LREV`, `AGE`, `AGE`^{2} et `ASSU` sont communs à chaque individu. En d'autres termes, on a:

$$\forall i \in [\![1;200]\!] \ ; \ \forall t \in [\![1;5]\!] \ ; \ \forall k \in [\![2;5]\!] \ ; \ \beta_{k,it} = \beta_k$$

On utilisera la fonction `plm` du package `plm` pour estimer les différentes modèles.

4.1 Modèle pooled

Dans le modèle pooled, on considère que les paramètres β_{1i} sont tous égaux à β_1. On estime le modèle pooled suivant par la méthode MCO:

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

Le tableau suivant montre que chaque paramètre du modèle est significatif, bien que β_2 ne soit significatif qu'à 10%.

On remarque que si le revenu en milliers de dollars augmente de 1%, alors les dépenses de santé augmentent de 0.00392 centaines de dollars, soit un peu moins de 0.40 dollars. Dit autrement si le revenu en milliers de dollars double, ie augmente de 100%, alors le modèle pooled prévoit que les dépenses de santé augmenteront d'un peu moins de 40 dollars.

L'effet marginal de l'âge sur les dépenses de santé est égale à $0.008 \times AGE - 0.208$. On constate que cet effet marginal augmente avec l'âge, ce qui confirme les résultats trouvés graphiquement à la question 2.

Enfin, l'estimation de β_5 indique que la possession d'une assurance maladie privée entraine une augmentation des dépenses de santé de 1.517 centaines de dollars, soit 151.70 dollars.

Table 4: Estimation du modèle pooled

	Dependent variable:
	DEPSANTE
LREV	0.392^{*}
	(0.220)
AGE	-0.208^{***}
	(0.033)
I(AGEˆ2)	0.004^{***}
	(0.0004)
ASSU	1.517^{***}
	(0.123)
Constant	3.439^{***}
	(1.221)
Observations	1,000
R^2	0.550
Adjusted R^2	0.548
F Statistic	303.859^{***} (df = 4; 995)
Note:	*p<0.1; **p<0.05; ***p<0.01

4.2 Modèle à effets fixes individuels

Cette fois le modèle à estimer s'écrit comme suit :

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

où β_{1i} représente un effet fixe spécifique à l'individu i en matière de dépenses de santé.

On peut estimer ce modèle à l'aide de l'opérateur Within. Cet opérateur lorqu'il est appliqué à un vecteur d'observations pour un individu i calcule pour chaque observation son écart par rapport à la moyenne individuelle. Le modèle devient alors:

$$\boxed{\begin{gathered}DEPSANTE_{it} - DEPSANTE_{i.} = \beta_2\Big(ln(REV_{it}) - ln(REV)_{i.}\Big) + \beta_3\Big(AGE_{it} - AGE_{i.}\Big) + \\ \beta_4\Big(AGE^2_{it} - AGE^2_{i.}\Big) + \beta_5\Big(ASSU_{it} - ASSU_{i.}\Big) + \varepsilon_{it} - \varepsilon_{i.}\ \forall i \in [\![1;200]\!]\ ;\ \forall t \in [\![1;5]\!]\end{gathered}}$$

où la notation "$_{i.}$" désigne la moyenne temporelle de chaque variable, telle que:

$$\forall i \in [\![1;200]\!]\ ;\ x_{i.} = \frac{1}{5}\sum_{t=1}^{5} x_{it}$$

Il est pertinent de relever que l'effet individuel β_{1i} a disparu du modèle après le centrage des variables puisque:

$$\beta_{1i.} = \frac{1}{5}\sum_{t=1}^{5}\beta_{1i} = \beta_{1i}$$

On peut donc utiliser le modèle `within` du package `plm`.

D'après le tableau suivant, les coefficients estimés β_k pour $k = 2, 3, 4$ ne sont pas significatifs. Seul l'effet de l'assurance maladie privée sur les dépenses de santé est significatif au risque 1%.

Table 5: Estimation du modèle à effets fixes individuels

	Dependent variable:
	DEPSANTE
LREV	−0.105
	(0.754)
AGE	0.065
	(0.094)
I(AGEˆ2)	0.0003
	(0.001)
ASSU	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

Grâce à la fonction `fixef` du package `plm`, on peut obtenir les effets individuels pour les 200 individus de la base de données. On trouve que l'effet individuel moyen est $\beta_1 \approx 0.93$. On peut donc réécrire le modèle comme suit :

$$\boxed{\begin{array}{c} DEPSANTE_{it} = \beta_1 + \gamma_i + \beta_2 ln(REV_{it}) + \beta_3 AGE_{it} + \beta_4 AGE_{it}^2 + \beta_5 ASSU_{it} + \varepsilon_{it} \\ \forall i \in [\![1;200]\!] \ ; \ \forall t \in [\![1;5]\!] \end{array}}$$

Ce modèle conserve une constante commune β_1 et 200 constantes individuelles γ_i correspondant à la différence $\beta_{1i} - \beta_1$.

Table 6: Estimation des 10 premiers effets fixes individuels

	Estimate	Std. Error	t-value	Pr(>|t|)
1	4.360	2.286	1.907	0.057
2	0.452	2.195	0.206	0.837
3	-0.485	2.236	-0.217	0.828
4	0.222	2.392	0.093	0.926
5	0.563	2.564	0.220	0.826
6	3.562	2.228	1.599	0.110
7	0.212	2.427	0.088	0.930
8	2.774	2.233	1.242	0.214
9	4.884	2.207	2.213	0.027
10	-0.040	2.425	-0.017	0.987

4.3 Modèle à effets aléatoires

Cette fois on souhaite estimer le modèle à effets aléatoires dans lequel l'effet spécifique β_{1i} n'est plus considéré comme un effet fixe mais comme une variable aléatoire. Le modèle s'écrit alors :

$$\boxed{\begin{array}{c} DEPSANTE_{it} = \beta_0 + \beta_2 ln(REV_{it}) + \beta_3 AGE_{it} + \beta_4 AGE_{it}^2 + \beta_5 ASSU_{it} + v_{it} \\ \forall i \in [\![1;200]\!] \ ; \ \forall t \in [\![1;5]\!] \end{array}}$$

où v_{it} est le terme d'erreur composée du terme d'erreur idiosyncratique et de l'effet individuel tel que $v_{it} = \varepsilon_{it} + \beta_{1i}$.

Comme l'introduction d'un effet aléatoire dans le modèle conduit à une autocorrélation temporelle des perturbations v_{it}, on réduit le poids de chaque perturbation par l'introduction de l'opérateur θ. Le modèle peut se réécrire comme suit et sera estimé avec la méthodes des Moindres Carrés Quasi Généralisés (MCQG):

$$\boxed{\begin{array}{c} DEPSANTE_{it} - \theta DEPSANTE_{i.} = \beta_0(1-\theta) + \Big(ln(REV_{it}) - \theta ln(REV)_{i.}\Big)\beta_2 + \Big(AGE_{it} - \theta AGE_{i.)}\Big)\beta_3 \\ \Big(AGE_{it}^2 - \theta AGE_{i.}^2\Big)\beta_4 + \Big(ASSU_{it} - \theta ASSU_{i.}\Big)\beta_5 + (1-\theta)\beta_{1i} + (1-\theta)\varepsilon_i \\ \forall i \in [\![1;200]\!] \ ; \ \forall t \in [\![1;5]\!] \end{array}}$$

θ est calculé avec la formule suivante :

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

où le terme $\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_{\beta_1}^2}$ représente le poids des variations intra-individuelles dans les variations totales.

Pour l'estimation de ce modèle, on utilise le modèle `random` et la méthode d'estimation `walhus` du package `plm` qui permet de trouver $\hat{\beta}_{MCQG}$ le vecteur des coefficients estimés du modèle.

Tous les coefficients estimés sont significatifs à l'exception de $\hat{\beta}_2$ comme l'indique le tableau suivant.

Table 7: Estimation du modèle à effets aléatoires

	Dependent variable:
	DEPSANTE
LREV	-0.149
	(0.292)
AGE	-0.090^{*}
	(0.051)
I(AGEˆ2)	0.002^{***}
	(0.001)
ASSU	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

Le résumé de la régression indique que $\hat{\theta} \approx 0.73$. Il est donc plus proche de 1 que de 0, ce qui signifie que l'estimateur MCQG est plus proche de l'estimateur Within que de l'estimateur du modèle pooled.

5 Significativité des constantes individuelles

Tester l'égalité des termes constants en comparant les résultats du modèle pooled à ceux du modèles à effets fixes individuels revient à appliquer le test 3 de la procédure d'Hsiao. Ici, on fait l'hypothèse que les coefficients associés aux variables explicatives du modèle sont identiques pour les 200 individus.

$$\begin{cases} H_0: & \beta_{1i} = \beta_1 \quad \forall i \in [\![1;200]\!] \\ H_1: & \exists i \neq j \mid \beta_{1i} \neq \beta_{1j} \quad (i,j) \in [\![1;200]\!]^2 \end{cases}$$

La statistique de Fisher associée au test 3 de la procédure d'Hsiao est définie comme suit :

$$F = \frac{\big(SCR_P - SCR_W\big)/(N-1)}{SCR_W/\big(N(T-1)-K\big)} \sim F\Big((N-1), N(T-1)-K\Big)$$

avec $N = 200$ le nombre d'individus, $T = 5$ le nombre de périodes, $K = 4$ le nombre de variables explicatives des deux modèles, SCR_P la somme des carrés des résidus du modèle contraint sous H_0 (modèle pooled) et SCR_W somme des carrés des résidus du modèle contraint sous H_1 (modèle à effets fixes individuels).

Règle de décision: On trouve $F \approx 13.18 > F_{5\%}(199, 796) \approx 1.20$, on rejette donc l'hypothèse nulle d'égalité des termes constants. Au risque $\alpha = 5\%$, le modèle à effets fixes individuels est donc significativement meilleur que le modèle pooled pour représenter le processus qui génère les données.

6 Effet du revenu sur les dépenses de santé selon le modèle estimé

Tout d'abord, notons que β_2 n'est significatif que pour le modèle pooled. Ensuite, le tableau suivant présente la variation en niveau des dépenses de santé en centaines de dollars pour une augmentation du revenu de 1%. On remarque que les modèles à effets fixes individuels et à effets aléatoires estiment une diminution des dépenses de santé lorsque le revenu augmente de 1%. Supposons que le revenu double, ie augmente de 100%, alors les modèles à effets fixes individuels et à effets aléatoires prévoient respectivement une diminution du niveau des dépenses de santé de 11 dollars et 15 dollars. Le modèle pooled prévoit quant à lui une augmentation des dépenses de santé d'environ 39 dollars. La prise en compte des effets individuels entraine un changement de l'influence du revenu sur les dépenses de santé, bien que ce changement ne soit pas siginificatif d'un point de vue statistique.

Table 8: Effet d'une augmentation de 1 % du revenu sur les dépenses de santé pour les 3 modèles

	pooled	within	random
Effet marginal de REV	0.0039	-0.0011	-0.0015

7 Comparaison des modèles à effets fixes et à effets aléatoires

Le modèle à effets aléatoires semble plus adapté pour décrire le processus qui génère les données dans la mesure où tous ses coefficients sont significatifs à l'exception de β_2 le coefficient associé à `LREV`.

Table 9: Comparaison des modèles within et random

	Dependent variable:	
	DEPSANTE	
	(1)	(2)
LREV	−0.105	−0.149
	(0.754)	(0.292)
AGE	0.065	−0.090*
	(0.094)	(0.051)
I(AGE^2)	0.0003	0.002***
	(0.001)	(0.001)
ASSU	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

Testons maintenant si l'on peut réellement appliquer le modèle à effets aléatoires. On doit vérifier si l'effet aléatoire β_{1i} n'est pas corrélé aux vecteurs des variables explicatives X_i. Cela revient à tester si $\mathbb{E}\big(\beta_{1i}|X_i\big) = 0$. Pour ce faire on a recours au test d'Hausman dont les hypothèses sont les suivantes :

$$\begin{cases} H_0 : & \mathbb{E}\big(\beta_{1i}|X_i\big) = 0 \\ H_1 : & \mathbb{E}\big(\beta_{1i}|X_i\big) \neq 0 \end{cases}$$

Après avoir procédé au test on trouve $p_c \approx 0.0023 < 0.05$, ce qui nous amène à rejet l'hypothèse nulle de non corrélation entre les effets aléatoires et les variables explicatives du modèle. Le modèle à effets aléatoires est donc inconsistent. On retient donc le modèle à effets fixes individuels.

8 Analyse du modèle à effets aléatoires

Dans cette question, on utilisera les résultats du modèle à effets aléatoires.

8.1 Effet de l'âge sur les dépenses de santé

L'effet marginal de l'âge sur les dépenses de santé est défini comme suit :

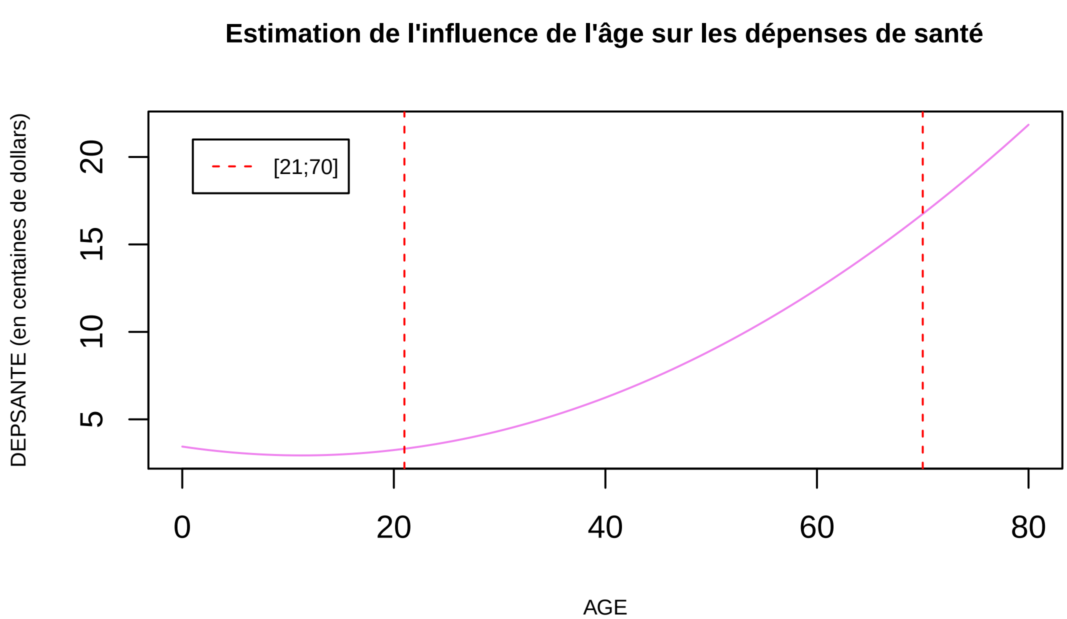
$$\frac{\partial DEPSANTE}{\partial AGE} = \hat{\beta}_3 + 2\hat{\beta}_4 AGE = 0.004 \times AGE - 0.090$$

L'effet marginal de l'âge sur les dépenses de santé croit proportionnellement avec l'âge de l'individu considéré puisque :

$$\frac{\partial^2 DEPSANTE}{\partial AGE^2} > 0$$

L'effet de l'âge sur les dépenses de santé est donc une fonction convexe définie comme suit :

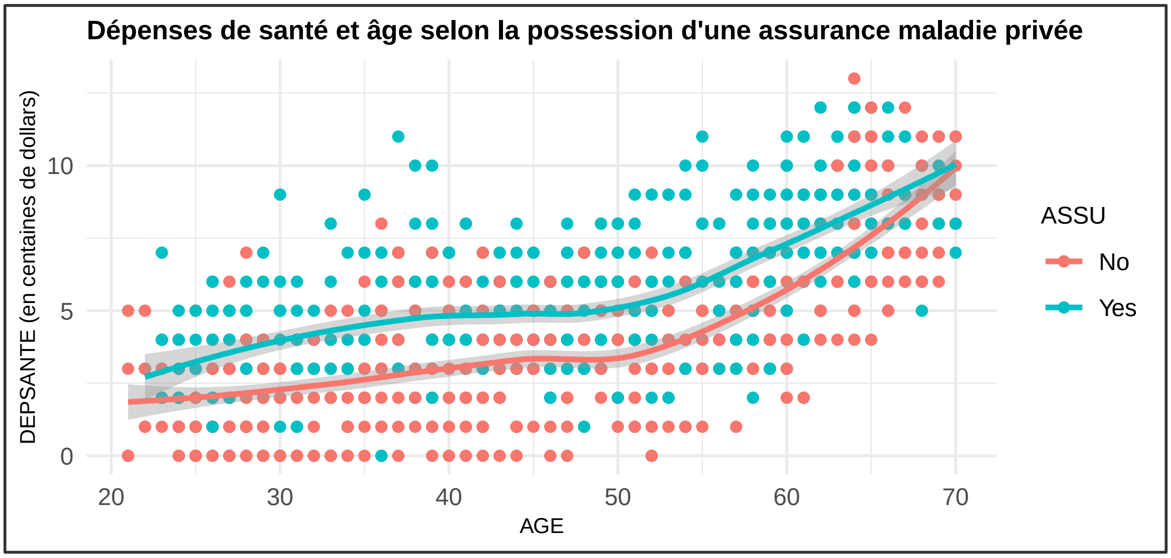
$$DEPSANTE = 0.004AGE^2 - 0.09AGE + k, \;\; k \in \mathbb{R}$$



Le graphique précédent représente la relation entre `DEPSANTE` et `AGE`. Le paramètre k a été estimé par la valeur de la constante $\hat{\beta}_0$ du modèle à effets aléatoires. On obtient bien une courbe convexe avec un minimum des dépenses de santé atteint à l'age de 11 ans et 3 mois. Les lignes en pointillés représentent l'étendue de l'âge des individus de la bases de données.

8.2 Effet de l'assurance maladie privée sur les dépenses de santé

L'estimation du modèle à effets aléatoires donne $\hat{\beta}_5 = 1.362$. Or, on sait que β_5 différencie un individu présentant une assurance maladie privée d'un individu qui n'en possède pas en termes de dépenses annuelles de santé. Ainsi, la possession d'une assurance maladie privée entraine une augmentation estimée des dépenses annuelles de santé de 1.362 centaines de dollars, soit 136.20 dollars. Notons cependant que l'écart de dépenses de santé entre ceux qui ont une assurance privée et ceux qui n'en ont pas tend à diminuer avec l'âge.



9 Limites du modèle à effets fixes individuels

Si l'on suppose que personne n'a changé de statut d'assurance on a alors $ASSU_{it} = 0$ ou 1, $\forall t \in [\![1;5]\!]$. Si l'on calcule la moyenne temporelle de cette variable pour l'individu i, on obtient :

$$ASSU_{i.} = \frac{1}{5}\sum_{t=1}^{5} ASSU_{it} = ASSU_{it}$$

En centrant les variables du modèles pour appliquer la méthode d'estimation Within, on a alors :

$$ASSU_{it} - ASSUi. = 0$$

Pour tester empiriquement ce résultat on décide d'estimer un nouveau modèle avec la méthode d'estimation Within. Pour ce faire on crée une nouvelle variable `ASSUcst` telle que $ASSUcst_{it} = ASSU_{i1}$, i.e. 0 ou 1, avec $t \in [\![1;5]\!]$.

Table 10: Aperçu de la nouvelle variable ASSUcst (individu 4)

ID	ANNEE	ASSU	ASSUcst
4	1	1	1
4	2	1	1
4	3	1	1
4	4	0	1
4	5	0	1

En appliquant la méthode Within avec la variable `ASSUcst`, i.e. la variable `ASSU` constante dans la dimension temporelle pour l'individu i, on remarque que β_5 n'a pas pu être estimé. L'estimateur à effets fixes devient donc inefficace dans la mesure où il ne peut pas déterminer l'influence de l'assurance maladie privée sur les dépenses de santé annuelles. Notons aussi que la significativité globale du modèle est faible contrairement aux modèles estimés précédemment, ce qui témoigne de l'inefficacité de l'estimateur Within.

Table 11: Estimation Within sans changement de statut d'assurance maladie privée

	Dependent variable:
	DEPSANTE
LREV	0.267
	(0.817)
AGE	0.047
	(0.102)
I(AGEˆ2)	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