

Valentin Barthel

22 avril 2024

Table des matières

1 Question 1 : Modèles Probit et Logit expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation	2
2 Question 2 : Modèles Probit et Logit à effets aléatoires expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation	5
3 Question 3 : Modèles Probit et Logit à effets fixes expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation	9

1 Question 1 : Modèles Probit et Logit expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation

```
-----Question 1-----*/  
  
/*Chargement de la base de données depuis l'espace de stockage*/  
libname maLib '/home/u63824485/sasuser.v94';  
  
data merged_panelex3;  
    set maLib.merged_panelex3;  
run;  
  
/*Modélisation Probit et Logit*/  
/*Probit*/  
proc logistic data=merged_panelex3;  
    class sexe;  
    model actifp(event='1') = lrd0 sexeFemme age primaire secondaire  
deuxieme_cycle professionnel_court professionnel_long / link=probit;  
    title "Modèle probit expliquant l'activité principale par le revenu d'inactivité,  
l'âge et le degré de formation";  
run;  
  
/*Logit*/  
proc logistic data=merged_panelex3;  
    class sexe;  
    model actifp(event='1') = lrd0 sexeFemme age primaire secondaire  
deuxieme_cycle professionnel_court professionnel_long / link=logit;  
    title "Modèle logit expliquant l'activité principale par le revenu d'inactivité,  
l'âge et le degré de formation";  
run;
```

Figure 1. Résultats - Modèle probit expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation

Analyse des valeurs estimées du maximum de vraisemblance					
Paramètre	DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > khi-2
Intercept	1	10.9108	0.3793	827.3423	<.0001
Ird0	1	-0.9416	0.0359	686.1464	<.0001
sexefemme	1	0.2816	0.0326	74.6890	<.0001
agea	1	0.00281	0.00163	2.9884	0.0839
primaire	1	-0.7842	0.0695	127.2594	<.0001
secondaire	1	-0.6163	0.0666	85.6584	<.0001
deuxieme_cycle	1	-0.0595	0.0670	0.7881	0.3747
professionnel_court	1	-0.4326	0.0623	48.1833	<.0001
professionnel_long	1	-0.1208	0.0813	2.2084	0.1373

Figure 2. Résultats - Modèle logit expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation

Analyse des valeurs estimées du maximum de vraisemblance					
Paramètre	DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > khi-2
Intercept	1	20.0510	0.6965	828.6693	<.0001
Ird0	1	-1.7278	0.0654	698.2775	<.0001
sexefemme	1	0.5349	0.0583	84.1625	<.0001
agea	1	0.00196	0.00291	0.4552	0.4999
primaire	1	-1.4120	0.1304	117.1772	<.0001
secondaire	1	-1.1249	0.1265	79.0743	<.0001
deuxieme_cycle	1	-0.1355	0.1312	1.0666	0.3017
professionnel_court	1	-0.8417	0.1206	48.7316	<.0001
professionnel_long	1	-0.2753	0.1556	3.1323	0.0768

L'analyse des valeurs estimées pour nos modèles Probit et Logit à l'ajout des variables d'éducation et de l'âge nous présente des coefficients statistiquement non significatifs ou peu significatifs. En effet, dans le cadre du modèle Probit l'âge est statistiquement significatif au seuil 10% pour expliquer l'activité principale, les parcours de professionnel long et deuxième cycle ne se révèleront pas significatifs tandis que les autres coefficients le seront. Cependant, dans le modèle Logit, l'âge ne sera plus statistiquement significatif pour expliquer l'activité principale et le parcours professionnel long le deviendra au seuil 10%

Dans nos modèles Probit et Logit, le signe positif du coefficient de "sexefemme" nous montre une probabilité accrue pour une femme d'avoir une activité principale par rapport à un homme. Par ailleurs, si l'on se réfère au signe du coefficient de "lrd0", nous observons également un effet désincitatif du revenu d'inactivité envers l'activité principale.

Figure 3. Effets marginaux de la hausse d'une revenu d'inactivité sur l'activité principale

Calculons les effets marginaux de lrd0

L'objectif étant de voir un effet désincitatif marginal du revenu d'inactivité sur l'activité principale.

→ Soit un homme de 25 ans de parcours "troisième cycle" ayant un revenu d'inactivité de 10

Probabil:

$$\text{actifp} = 10,9108 + 2,5 \times 0,002831 - 10 \times 0,3416$$

$$\text{actifp} = 1,56505$$

$$P = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\text{actifp}^2}{2}\right)$$

$$\Leftrightarrow P = 0,1172$$

L'effet marginal : $P_{lrd0} \times P = -0,9416 \times 0,1172$
 $= -0,1103$

Logit:

$$\text{actifp} = 20,0510 + 2,5 \times 0,00196 - 10 \times 1,7278$$

$$\text{actifp} = 2,822$$

$$P = \frac{1}{1 + \exp(\text{actifp})}$$

$$P = 0,0561$$

L'effet marginal : $P_{lrd0} \times P(1-P) = -1,7278 \times 0,0561(1-0,0561)$
 $= -0,091$

D'après nos modèles, augmenter le revenu d'inactivité de 1% d'un homme de 25 ans ayant un niveau de "troisième cycle" aura pour effet désincitatif de réduire de près de 10% les chances de réalisation d'une activité principale (respectivement 11% et 9,1% pour les modèles Probit et Logit).

2 Question 2 : Modèles Probit et Logit à effets aléatoires expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation

```
/*-----Question 2-----*/
/*Modèles à effets aléatoires*/
/*Modèle probit à effets aléatoires*/
proc glimmix data=merged_panelex3;
  class mident mois;
  model actifp(event='1') = lrd0 sexeFemme age primaire secondaire
    deuxieme_cycle professionnel_court professionnel_long / dist=binary link=probit solution;
  random int/ subject=mident;
  title "Modèle probit à effets aléatoires expliquant l'activité principale par le revenu d'inactivité,
    l'âge et le degré de formation";
run;

/*Modèle logit à effets aléatoires*/
proc glimmix data=merged_panelex3;
  class mident mois;
  model actifp(event='1') = lrd0 sexeFemme age primaire secondaire
    deuxieme_cycle professionnel_court professionnel_long / dist=binary link=logit solution;
  random int / subject = mident;
  title "Modèle logit à effets aléatoires expliquant l'activité principale par le revenu d'inactivité,
    l'âge et le degré de formation";
run;
```

Figure 4. Résultats - Modèle probit à effets aléatoires expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation

Critère de convergence (PCONV=1.11022E-8) respecté.

Tests d'ajustement	
-2 Log Pseudo-vraisemblance rés.	49413.40
Khi-2 généralisé	2013.11
Khi-2 génér. / DDL	0.19

Valeur estimée du paramètre de covariance			
Param. de cov.	Sujet	Estimation	Erreur type
Intercept	mident	2.9200	0.1739

Solutions pour effets fixes					
Effet	Estimation	Erreur type	DDL	Valeur du test t	Pr > t
Intercept	13.5157	1.5131	872	8.93	<.0001
lrd0	-1.1273	0.1447	9667	-7.79	<.0001
sexeFemme	0.3141	0.1476	9667	2.13	0.0334
agea	0.008006	0.007304	9667	1.10	0.2730
primaire	-1.3426	0.3159	9667	-4.25	<.0001
secondaire	-1.0145	0.2982	9667	-3.40	0.0007
deuxieme_cycle	-0.1547	0.2892	9667	-0.53	0.5928
professionnel_court	-0.7175	0.2740	9667	-2.62	0.0088
professionnel_long	-0.2857	0.3544	9667	-0.81	0.4202

Tests des effets fixes - Type III				
Effet	DDL num.	DDL den.	Valeur F	Pr > F
lrd0	1	9667	60.67	<.0001
sexeFemme	1	9667	4.53	0.0334
agea	1	9667	1.20	0.2730
primaire	1	9667	18.06	<.0001
secondaire	1	9667	11.57	0.0007
deuxieme_cycle	1	9667	0.29	0.5928

m/SASStudio/sasexec/submissions/ad77a34a-afc8-49af-9dcf-769b59f01362/results

Résultats : Exercice4.sas

Tests des effets fixes - Type III

Effet	DDL num.	DDL den.	Valeur F	Pr > F
professionnel_court	1	9667	6.86	0.0088
professionnel_long	1	9667	0.65	0.4202

Figure 5. Résultats - Modèle logit à effets aléatoires expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation

Critère de convergence (PCONV=1.11022E-8) respecté.

Tests d'ajustement					
-2 Log Pseudo-vraisemblance rés.					66056.69
Khi-2 généralisé					2006.70
Khi-2 génér. / DDL					0.19
Valeur estimée du paramètre de covariance					
Param. de cov.	Sujet	Estimation	Erreur type		
Intercept	mident	10.7799	0.6912		
Solutions pour effets fixes					
Effet	Estimation	Erreur type	DDL	Valeur du test t	Pr > t
Intercept	25.2535	2.9860	872	8.46	<.0001
lrd0	-2.1035	0.2855	9667	-7.37	<.0001
sexeFemme	0.6110	0.2981	9667	2.05	0.0404
agea	0.01737	0.01473	9667	1.18	0.2382
primaire	-2.6673	0.6478	9667	-4.12	<.0001
secondaire	-2.0297	0.6138	9667	-3.31	0.0009
deuxieme_cycle	-0.3431	0.8004	9667	-0.57	0.5677
professionnel_court	-1.4637	0.5674	9667	-2.58	0.0099
professionnel_long	-0.6358	0.7291	9667	-0.87	0.3832
Tests des effets fixes - Type III					
Effet	DDL num.	DDL den.	Valeur F	Pr > F	
lrd0	1	9667	54.30	<.0001	
sexeFemme	1	9667	4.20	0.0404	
agea	1	9667	1.39	0.2382	
primaire	1	9667	16.95	<.0001	

m/SASStudio/sasexec/submissions/ad77a34a-afc8-49af-9dcf-769b59f01362/results

Résultats : Exercice4.sas

Tests des effets fixes - Type III				
Effet	DDL num.	DDL den.	Valeur F	Pr > F
secondaire	1	9667	10.93	0.0009
deuxieme_cycle	1	9667	0.33	0.5677
professionnel_court	1	9667	6.65	0.0099
professionnel_long	1	9667	0.76	0.3832

Premièrement, les résultats pour les modèles Probit et Logit à effets aléatoires indiquent que le critère de convergence est respecté, c'est-à-dire que les paramètres estimés sont fiables, sous les hypothèses de ces modèles (spécification correcte des modèles et sans biais, indépendance des observations, distribution de l'erreur, non-colinéarité, homoscédasticité et taille de l'échantillon). Ainsi, nous avons bien trouvé un maximum de la log-vraisemblance.

Nous observons que désormais, pour les modèles Probit et Logit à effets aléatoires, le coefficient associé au sexe de l'individu n'est plus que statistiquement significatif au seuil de 5% et que l'âge et les formations ne sont pas statistiquement pas significatifs pour expliquer l'activité principale. Par ailleurs, nous pouvons lire dans nos tableaux que le signe des coefficients pour nos modèles à effets aléatoires sont identiques à ceux des modèles étudiés dans la question précédente. Dans ces modèles, l'effet désincitatif persistera.

Les valeurs élevées du paramètre de covariances pour l'intercept (respectivement 2,92 et 10,78 pour les modèles Probit et Logit) suggèrent qu'il y a beaucoup de variabilité dans les données qui n'est pas expliquée par le revenu d'inactivité, le sexe, l'âge ou encore la formation. De ce fait, ces valeurs indiquent une forte hétérogénéité entre les individus observés.

3 Question 3 : Modèles Probit et Logit à effets fixes expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation

```
-----Question 3-----*/  
/*Modèles à effets fixes*/  
/*Modèle probit à effet fixes*/  
proc genmod data=merged_panlex3;  
    class mident mois;  
    model actifp(event='1') = mident lrd0 sexeFemme age primaire secondaire  
        deuxieme_cycle professionnel_court professionnel_long / dist=binomial link=probit;  
    title "Modèle probit à effets fixes expliquant l'activité principale par le revenu d'inactivité,  
        l'âge et le degré de formation";  
run;  
  
/*Modèle logit à effets fixes*/  
proc genmod data=merged_panlex3;  
    class mident mois;  
    model actifp(event='1') = mident lrd0 sexeFemme age primaire secondaire  
        deuxieme_cycle professionnel_court professionnel_long / dist=binomial link=logit;  
    title "Modèle logit à effets fixes expliquant l'activité principale par le revenu d'inactivité,  
        l'âge et le degré de formation";  
run;
```

Figure 6. Résultats - Modèle probit à effets fixes expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation

WARNING: Negative of Hessian not positive definite.

Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald à95%	Khi-2 de Wald	Pr > khi-2
Intercept		1	-16.8298	4.6515	-25.9466 -7.7131	13.09	0.0003
lrd0		1	-0.2197	0.3448	-0.8956 0.4561	0.41	0.5240
sexefemme		1	-2.2790	12894.07	-25274.2 25269.63	0.00	0.9999
agea		1	0.6010	0.1364	0.3337 0.8684	19.41	<.0001
primaire		1	-4.8949	31621.39	-61981.7 61971.89	0.00	0.9999
secondaire		1	-15.5900	12894.07	-25287.5 25256.32	0.00	0.9990
deuxieme_cycle		1	6.8349	0.6002	5.6585 8.0112	129.69	<.0001
professionnel_court		1	1.5860	0.8784	-0.1356 3.3076	3.26	0.0710
professionnel_long		1	-0.6318	12894.07	-25272.5 25271.28	0.00	1.0000
mident	1100080001	1	-27.4540	28862.38	-56596.7 56541.77	0.00	0.9992
mident	1100220001	1	-1.6753	12894.07	-25273.6 25270.23	0.00	0.9999
mident	1100239001	1	-2.6806	31566.91	-61872.7 61867.33	0.00	0.9999

Figure 7. Résultats - Modèle logit à effets fixes expliquant l'activité principale par le revenu d'inactivité, l'âge et le degré de formation

WARNING: Negative of Hessian not positive definite.

Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald à95%	Khi-2 de Wald	Pr > khi-2
Intercept		1	-29.5475	8.1047	-45.4325 -13.6625	13.29	0.0003
lrd0		1	-0.3037	0.5997	-1.4792 0.8717	0.26	0.6125
sexefemme		1	8.3839	25098.04	-49182.9 49199.63	0.00	0.9997
agea		1	1.0309	0.2343	0.5717 1.4901	19.36	<.0001
primaire		1	-8.4208	61485.88	-120519 120501.7	0.00	0.9999
secondaire		1	-39.1857	25098.04	-49230.4 49152.06	0.00	0.9988
deuxieme_cycle		1	11.6714	0.9922	9.7266 13.6161	138.36	<.0001
professionnel_court		1	2.6561	1.4803	-0.2453 5.5575	3.22	0.0728
professionnel_long		1	11.1781	25098.04	-49180.1 49202.42	0.00	0.9996
mident	1100080001	1	-59.5142	56148.43	-110108 109989.4	0.00	0.9992
mident	1100220001	1	-15.1690	25098.04	-49206.4 49176.08	0.00	0.9995
mident	1100239001	1	-4.5735	61466.80	-120477 120468.1	0.00	0.9999

Les tables des résultats étant trop grandes en raison du nombre de constantes, j'ai choisi de reporter les résultats principaux. Il y a une constante par individu, indiquant une valeur différente pour chacun des individus, ce qui est caractéristique des modèles à effets fixes, contrairement aux modèles à effets aléatoires.

Contrairement aux modèles à effets aléatoires de la question précédente, les résultats des modèles à effets fixes présentent chacun une matrice hessienne négative non définie positivement. Cela suggère que le point trouvé par SAS est un point selle (non optimum) ou un minimum local, entraînant des coefficients estimés non fiables, d'où la non/faible significativité statistique de la plupart des coefficients estimés pour expliquer l'activité principale. Ainsi, contrairement aux modèles à effets aléatoires, ces modèles à effets fixes présentent des coefficients associés au niveau d'étude, au sexe et au revenu d'inactivité qui ne sont pas statistiquement significatifs pour expliquer l'activité principale, seuls les coefficients associés à l'âge et au niveau de deuxième cycle restent statistiquement significatifs au seuil 1%.

Nous remarquons que le signe associé au coefficient du niveau de deuxième cycle est désormais positif dans nos modèles à effets fixes, montrant une augmentation de chance de la réalisation d'activités principales pour les individus ayant ce niveau, contrairement aux modèles étudiés dans les questions précédentes. Concernant les coefficients qui ne sont pas statistiquement significatifs, nous pouvons lire des coefficients positifs pour certains niveaux d'études qui ne l'étaient pas dans nos modèles précédents. Bien que le coefficient associé au revenu d'inactivité ne soit pas statistiquement significatif pour expliquer l'activité principale dans nos modèles à effets fixes, ce dernier sera à nouveau de signe négatif, et donc mènera à nouveau à un effet désincitatif du revenu d'inactivité envers la probabilité de réalisation de l'activité principale.

Les modèles à effets aléatoires supposent qu'il y a des différences entre les individus qui ne sont pas expliquées par le modèle tandis que dans le modèle à effets fixes, nous aurons une intercept correspondant à chacun des individus observés, absorbant la variabilité entre individus. Dans notre cas, si l'on suppose des effets fixes, le revenu d'inactivité, l'âge, le sexe et la formation ne suffiront pas à expliquer l'activité principale avec la différence entre les individus, ce qui est cohérent avec nos valeurs estimées du paramètres de covariance qui nous indiquait une forte hétérogénéité entre nos individus. En d'autres mots, les effets fixes absorbent la variabilité intra-individu et ne permettent pas d'expliquer l'activité principale.