

RÉPUBLIQUE DU SÉNÉGAL



Un peuple-Un but-Une foi

Ministère de L'Économie du Plan et de la Coopération

Agence Nationale de la Statistique et de la Démographie
(ANSD)



(ENSAE-Dakar)

TP

PROJET ECONOMETRIE DES VARIABLES QUALITATIVES

Rédigé par :

AMADOU Moussa

Élève Ingénieur Statisticien Économiste

Sous la supervision de :

Mamadou Abdoulaye DIALLO

Ingénieur Statisticien Economiste (ISE)

Chercheur au CRES

Enseignant à l'ENSAE

2023-2024

Table des matières

I	Modèle dichotomique simple	2
1	Construction d' une variable dichotomique qui permet d' identifier les personnes vaccinées contre la Covid-19	3
2	Tableau descriptif des différentes variables pertinentes.	3
3	Tableau synthétique des coefficients des modèles logit, probit et ols . . .	6
4	Estimer les Odds ratio et les effets marginaux	7
4.1	Tests de validation du modèle	9
4.2	Tirage aléatoire stratifié (milieu de résidence) parmi les non vaccinés	12
5	Comparaison dans un tableau les coefficients des deux modèles logit	13
6	Estimation séparée (ensemble, urbain, rural).	14
7	Estimation séparée de l'effet de la variable «litteracy» sur les indicateurs de santé infantile et Proposition des variables de contrôle	15
II	Modèle multinomial	20
8	Sous quelles conditions, il est possible de modéliser la variable type_contrat	21
9	Tableau synthétique des différentes variables juger pertinentes	21
10	Estimer et présenter le modèle logit multinomial en fixant comme modalité de référence « CDI ».	22
11	Estimer et présenter dans un tableau les risques relatifs.	23
12	Estimer et présenter dans un tableau les effets marginaux.	25
13	Effectuer le test IIA.	26
III	modèle Tobit et Heckman	28
14	Quel problème de modélisation soulève cette variable?	29
15	Tableau synthétique des variables pertinentes pour le problème posé : . .	29
16	Estimer un modèle linéaire sur l'ensemble de l'échantillon puis dans le sous échantillon des travailleurs	30
17	Estimer un modèle de sélection de Heckman avec les mêmes variables. . .	31

- 18 Est-ce que le fait d'être marié et le fait d'avoir un enfant sont-elles endogènes ? Justifier ? Quelles solutions préconisez-vous s'il y en a ? 32

Liste des tableaux

1	Statut de Vaccination	3
2	Description des variables	4
3	Croisement des variables	5
4	Comparaison des modèles logit probit ols	6
5	odds ratios-effets marginaux	7
6	Odds ratio et les effets marginaux avec modalité de référence	8
7	Qualité d'ajustement	9
8	Taux de prédiction à 50%	10
9	Taux de prédiction à 70%	11
10	Taux de prédiction à 80%	12
11	Comparaison des coefficients des modèles logit après rééchantillonnage	13
12	Estimation des modèles séparé	14
13	effet de la variable «literacy» sur les indicateurs de santé infantile	15
14	Tests de validation	16
15	Modèle avec variables de contrôle	17
16	Tests sur modèle avec variables de contrôle	17
17	lien entre les variables dépendantes	18
18	modèle probit multivarié	19
19	type_contrat	21
20	Variables pertinentes	21
21	modèle logit multinomial	22
22	risques relatifs	24
23	effets marginaux	25
24	Test IIA Wald	26
25	Tabulation sur le revenu	29
26	variables pertinentes	29
27	Estimer un modèle linéaire	30
28	Résumé du revenu des non travailleurs prédits	31
29	modèle de sélection de Heckman	31
30	Significativité globale	32
31	Revenu des non travailleurs prédits	32

Introduction

De nombreux comportements sont de nature qualitative. Par conséquent, il est fréquent que le chercheur, désireux d'expliquer ou de comprendre ces phénomènes, utilise l'analyse de régression multivariée, incluant ainsi des variables qualitatives. Ces variables peuvent être de deux types : dichotomiques d'une part, et polytomiques d'autre part. Les variables polytomiques représentent une généralisation des variables dichotomiques : c'est l'économétrie des variables qualitative. L'objet de ce travail est de mettre en pratique les notions apprises au cours afin d'être familiariser avec les aspects techniques et théorique. Le travail est scindé en trois parties à savoir : une première partie sera consacrée au modèle dichotomique simple, une deuxième partie sur le modèle multinomial et la dernière sur le tobit et Heckman.

Pour ce faire nous avons utilisé le logiciel stata pour le traitement et l'analyse des données. Ensuite les tableaux sont exportés sur le logiciel excel afin d'être mise en forme. Enfin nous avons utilisé latex pour la rédaction.

Première partie

Modèle dichotomique simple

1 Construction d' une variable dichotomique qui permet d' identifier les personnes vaccinées contre la Covid-19

Après avoir pris connaissance avec la base de données et les métadonnées (questionnaire) et en procédant à un résumé sommaire des variables de la base de données, il s'agit en fait de la base de données de afrobaromètre round9 sur le Niger. Au total il y a 1200 observations. La variable Q58A (la personne a reçu le vaccin contre la covid19) nous permettra de connaître les personnes vaccinées ou non. Elle a deux modalités à savoir non (avec le code 1) et oui (avec le code 0). Nous créons alors la variables dichotomique «Statut Vaccination» qui prends 1 si la personne est vaccinée et 0 sinon. Le tableau ci-dessous indique que 40,57% des personnes sont vaccinées.

TABLEAU 1 – Statut de Vaccination

	Freq	Percent	Cum
Non Vaccinée	718	59,83	59,83
Vaccinée	482	40,17	100
Total	1200	100	

Source : calcul de l' auteur données afrobaromètre Niger round9

2 Tableau descriptif des différentes variables pertinentes.

Plusieurs variables peuvent expliquer les raisons de vaccination. Il est nécessaire d'effectuer un croisement des variables que nous jugeons pertinentes et la variable dichotomique personne vaccinées. Dans un premier temps nous commençons par citer les variables pertinentes en se basant sur les travaux de recherche antérieurs ensuite nous présenterons un tableau descriptif des variables choisies et enfin nous présenterons un croisement entre les variables et notre variable dichotomique Statut Vaccination.

On retient essentiellement :

- > Facteurs sociodémographiques : L'âge, le sexe, le niveau d'éducation, lieu de résidence ;
- > Facteurs psychologiques : la confiance envers le gouvernement ;
- > Facteurs liés au système de santé : malade du covid

Le tableau ci-dessous présente les variables choisies. On retient de ce tableau que plus de la moitié des individus dans l'échantillon sont des hommes. On note également que ces individus sont majoritairement jeunes et ruraux. 49,17% n'ont une éducation formelle, moins de 1% sont atteints de la covid 19. Il faut souligner que 43% font confiance au gouvernement sur le fait que le vaccin de la covid est sûr.

TABLEAU 2 – Description des variables

<i>Variables</i>	<i>Effectifs</i>	<i>Pourcentages %</i>
sexe		
Homme	602	50,17
Femme	598	49,83
milieu de résidence		
Urbain	216	18
Rural	984	82
classe d'âge		
18-25 ans	291	24,25
26-35 ans	283	23,58
36-45 ans	270	22,5
46-55 ans	169	14,08
56-65 ans	116	9,67
plus de 66 ans	71	5,92
niveau d'instruction		
Aucune éducation formelle	590	49,17
Primaire	405	33,75
Secondaire	167	13,92
Post secondaire	38	3,17
confiance au gouvernement		
Pas du tout	235	19,58
Juste un peu	213	17,75
Quelque peu	210	17,5
Beaucoup	516	43
Ne sait pas	26	2,17
tombé malade de la covid 19		
Oui	9	0,75
Non	1187	98,92
Ne sait pas	4	0,33

Source : calcul de l' auteur données afrobaromètre Niger round9

Le tableau ci-dessous présente un croisement entre la variable statut de vaccination et les variables choisies. Nous avons effectué un test de χ^2 . Il ressort que 43,7% des individus en milieu rural sont vaccinés et la corrélation entre le milieu de résidence et le statut de vaccination est significatif au seuil de 5%. De même on note une corrélation entre l'âge et le statut de vaccination. Par ailleurs 43,7% des individus n'ayant aucune éducation formelle ont été vaccinés et la corrélation est significative au seuil de 5% entre le niveau d'éducation avec le statut de vaccination. 59,9% des individus faisant beaucoup confiance au gouvernement sont vaccinés et la corrélation est significative entre la confiance au gouvernement et le statut de vaccination au seuil de 5%.

TABLEAU 3 – Croisement des variables

	<i>Statut de Vaccination</i>	
	<i>Non Vaccinée %</i>	<i>Vaccinée %</i>
sexe		
Homme	61.1	38.9
Femme	58.5	41.5
<i>Pearson $\chi^2(1) = 0.8446$ Pr = 0.358</i>		
milieu de résidence		
Urbain	75.9	24.1
Rural	56.3	43.7
<i>Pearson $\chi^2(1) = 28.3846$ Pr = 0.000</i>		
classe d'âge		
18-25 ans	68.0	32.0
26-35 ans	61.5	38.5
36-45 ans	54.8	45.2
46-55 ans	58.6	41.4
56-65 ans	50.0	50.0
plus de 66 ans	57.7	42.3
<i>Pearson $\chi^2(5) = 16.2140$ Pr = 0.006</i>		
niveau d'instruction		
Aucune éducation formelle	56.3	43.7
Primaire	59.0	41.0
Secondaire	68.9	31.1
Post secondaire	84.2	15.8
<i>Pearson $\chi^2(3) = 18.2894$ Pr = 0.000</i>		
confiance au gouvernement		
Pas du tout	87.2	12.8
Juste un peu	71.8	28.2
Quelque peu	62.4	37.6
Beaucoup	40.1	59.9
Ne sait pas	84.6	15.4
<i>Pearson $\chi^2(4) = 176.8522$ Pr = 0.000</i>		

Source : calcul de l'auteur données afrobaromètre Niger round9

3 Tableau synthétique des coefficients des modèles logit, probit et ols

TABLEAU 4 – Comparaison des modèles logit probit ols

	[1]	[2]	[3]
<i>VARIABLES</i>	<i>modele logit</i>	<i>modele probit</i>	<i>modele OLS</i>
sexe	0,133 [-0,121 - 0,386]	0,088 [-0,067 - 0,242]	0,028 [-0,027 - 0,084]
milieu	0,704*** [0,349 - 1,060]	0,432*** [0,220 - 0,645]	0,143*** [0,075 - 0,212]
classe d'age	0,117*** [0,032 - 0,202]	0,073*** [0,020 - 0,125]	0,026*** [0,008 - 0,045]
education	-0,062 [-0,228 - 0,104]	-0,042 [-0,143 - 0,059]	-0,011 [-0,046 - 0,024]
confiance gouvernement	0,349*** [0,220 - 0,477]	0,199*** [0,136 - 0,263]	0,075*** [0,051 - 0,098]
malade de covi19	-0,104 [-0,381 - 0,173]	-0,063 [-0,235 - 0,108]	-0,024 [-0,082 - 0,035]
Constant	-2,969*** [-4,157 - -1,781]	-1,793*** [-2,505 - -1,082]	-0,132 [-0,370 - 0,107]
Observations	1200	1200	1200
R-squared			0,084
Pseudo R-squared	0.0663	0.0657	
wald chi2	69,04	82,29	
Prob>chi2	0	0	
Robust ci ()			
***p<0,01,**p<0,05,*p<0,1			

Source : calcul de l' auteur données afrobaromètre Niger round9

Commentaire : Concernant la significativité globale du modèle : les modèles logit et probit sont globalement significatif car $\text{prob} > \chi^2$ est inférieure à 5%. De même le modèle ols est globalement significatif ($\text{prob} > F$ est inférieure à 5%). Concernant la significativité des coefficients : Pour les modèles logit, probit et ols tous les coefficients sont significatifs au seuil de 5% à l'exception des variables sexe, education et les malades du covid19. Concernant le R2 et le pseudo R2 : le modèle logit a un pseudo R2 0,0869 tant disque le modèle probit a un pseudo R2 0,066 et le modèle ols un R2 de 0,084. La spécification du modèle ols pour estimer un modèle dichotomique pose plusieurs problèmes à savoir le fait que la spécification est de nature sensible au codage. En se référant à la littérature on utilise généralement le modèle logit pour les phénomènes rare et probit pour les phénomènes normaux. Ainsi la pandémie de covid19 étant un phénomène rare par conséquent son vaccin l'est également. Au vu de tout ce qui précède on retient le modèle logit.

4 Estimer les Odds ratio et les effets marginaux

1- Odds ratio et effets marginaux

Nous procédons à l'interprétation des effets marginaux. Lorsque l'âge des individus augmentent la probabilité d'être vaccinée est de 0,028. De même plus la confiance envers le gouvernement augmente, la probabilité d'être vaccinée est 0,083. Lorsqu'on passe du milieu urbain au rural, la probabilité d'être vaccinée est 0,168.

TABLEAU 5 – odds ratios-effets marginaux

VARIABLES	Odds ratio	Effet marginal
Statut_Vaccination	.	
	(.)	
sexe	1.142	0.032
	(0.148)	(0.031)
milieu	2.023***	0.168***
	(0.367)	(0.043)
classe_age	1.124***	0.028***
	(0.049)	(0.010)
education	0.940	-0.015
	(0.080)	(0.020)
confiance_gouvernement	1.417***	0.083***
	(0.093)	(0.016)
malade_de_covi_19	0.901	-0.025
	(0.127)	(0.034)
Constant	0.051***	
	(0.031)	
Observations	1,200	1,200
wald chi2	69.04	69.04
Prob>chi2	0	0
Pseudo R-squared	0.0663	0.0663
Robust seeform ()		
***p<0.01, **p<0.05		

2- Odds ratio et effets marginaux en fixant la première modalité comme référence

Les résultats du tableau ci-dessous présente les odds ratios et les effets marginaux. Nous procédons à l'interprétation des odds ratio. Les individus vivant en milieu rural ont 2,06 fois plus de chance d'être vaccinées que ceux vivant en milieu urbain. De même les individus ayant un âge entre 56 et 65ans ont 2,06 fois plus de chance d'être vaccinées que ceux ayant un âge entre 18 et 25ans. En outre les individus faisant beaucoup confiance au gouvernement ont 9,05 fois plus de chance d'être vaccinées que ceux ne faisant aucune confiance au gouvernement.

TABLEAU 6 – Odds ratio et les effets marginaux avec modalité de référence

<i>Statut _ Vaccination</i>	<i>Références</i>	<i>Odd.ratio</i>	<i>effets.marginaux</i>
sexe	Homme		
Femme		1.131 (0.155)	0.025 (0.027)
milieu	Urbain		
Rural		2.063*** (0.401)	0.141*** (0.036)
classe_age	18-25ans		
26-35 ans		1.384 (0.273)	0.064 (0.039)
36-45 ans		1.598** (0.321)	0.093** (0.040)
46-55 ans		1.292 (0.297)	0.050 (0.045)
56-65 ans		2.061*** (0.530)	0.146*** (0.052)
plus de 66 ans		1.570 (0.470)	0.090 (0.060)
education	Aucune Education formelle		
Primaire		0.994 (0.147)	-0.001 (0.029)
Secondaire		1.056 (0.232)	0.011 (0.044)
Post secondaire		0.806 (0.345)	-0.042 (0.083)
confiance_gouvernement	Pas du tout		
Juste un peu		2.353*** (0.601)	0.136*** (0.039)
Quelque peu		3.660*** (0.919)	0.230*** (0.041)
Beaucoup		9.051*** (2.031)	0.445*** (0.033)
Ne sait pas		1.012 (0.573)	0.001 (0.069)
malade_de_covi_19	oui		
Non		2.540 (2.156)	0.170 (0.134)
Ne sait pas		0.702 (1.054)	-0.052 (0.214)
Constant		0.024*** (0.022)	
Observations		1,200	1,200
Robust seeform ()			
***p<0.01, **p<0.05			

4.1 Tests de validation du modèle

1- Qualité d'ajustement du modèle

Le tableau 8 présente le test d'adéquation de Hosmer Lemeshow. On a une pvalue qui est supérieure au seuil de 5% donc le modèle est bien spécifié.

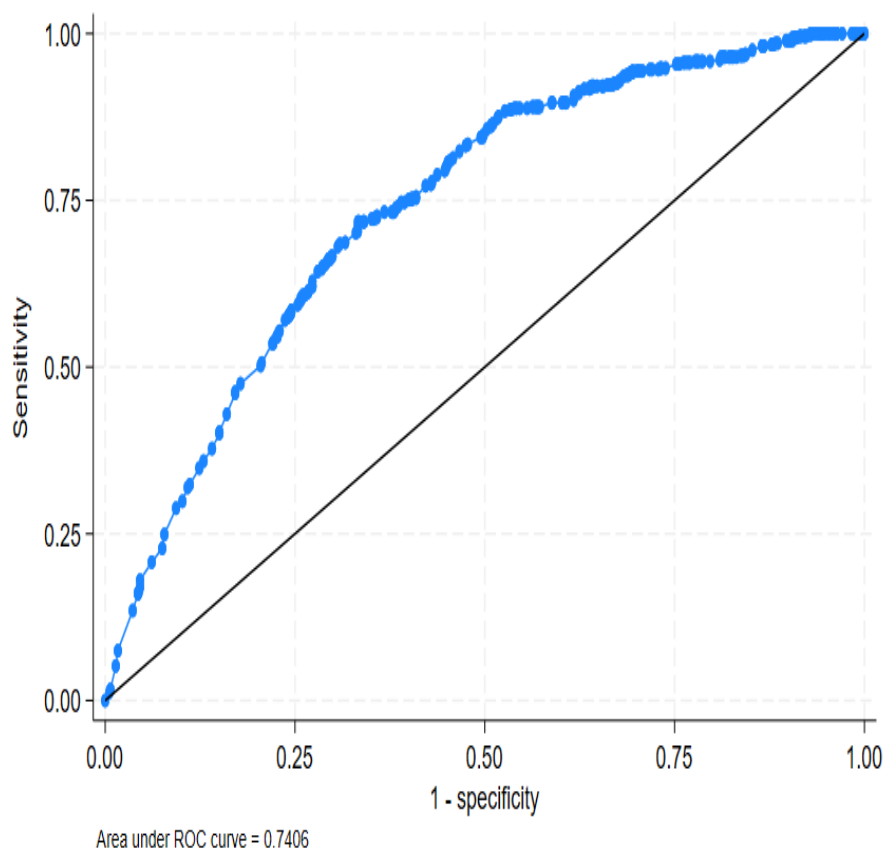
TABLEAU 7 – Qualité d'ajustement

<i>Variable : Statut_Vaccination</i>	
<i>Number of observations</i>	1200
<i>Number of covariate patterns</i>	257
<i>Pearson chi2(240)</i>	232,62
<i>Prob > chi2</i>	0,6216

Source :calcul de l'auteur données afrobaromètre Niger round9

2- Qualité discriminatoire

Le graphique ci-dessous présente la courbe de ROC. Graphiquement, plus la courbe s'écarte de la première bissectrice, meilleure est la discrimination et donc meilleur est le modèle. Ici nous avons une faible discrimination (Area under ROC curve=0.7406)



Source :calcul de l'auteur données afrobaromètre Niger round9

4- Pouvoir Prédictif du modèle

Taux de prédiction à 50%

Dans le tableau ci-dessous, le nombre de bonnes prédictions se trouve sur la diagonale(279+544), soit un taux de 68,58%.

TABLEAU 8 – Taux de prédiction à 50%

	True ——— Classified D D	Total
	279 174	453
	203 544	747
Total	482 718	1200
Classified + if predicted $\Pr(D) \geq .5$		
True D defined as Statut_Vaccination!= 0		
Sensitivity $\Pr(+ D)$	57.88%	
Specificity $\Pr(- D)$	75.77%	
Positive predictive value $\Pr(D +)$	61.59%	
Negative predictive value $\Pr(D -)$	72.82%	
False + rate for true D $\Pr(+ D)$	24.23%	
False - rate for true D $\Pr(- D)$	42.12%	
False + rate for classified + $\Pr(D +)$	38.41%	
False - rate for classified - $\Pr(D -)$	27.18%	
Correctly classified	68.58%	

Source : calcul de l'auteur données afrobaromètre Niger round9

Taux de prédiction à 70%

De même le tableau ci-dessous, le nombre de bonnes prédictions se trouve également sur la diagonale(25+708), soit un taux de 61,08%.

TABLEAU 9 – Taux de prédiction à 70%

True ——— Classified D D		Total
25	10	35
457	708	1165
Total	482	718
Total		1200
Classified + if predicted $\Pr(D) \geq .7$		
True D defined as Statut_Vaccination!= 0		
Sensitivity $\Pr(+ D)$		5.19%
Specificity $\Pr(- D)$		98.61%
Positive predictive value $\Pr(D +)$		71.43%
Negative predictive value $\Pr(D -)$		60.77%
False + rate for true D $\Pr(+ D)$		1.39%
False - rate for true D $\Pr(- D)$		94.81%
False + rate for classified + $\Pr(D +)$		28.57%
False - rate for classified - $\Pr(D -)$		39.23%
Correctly classified		61.08%

Source : calcul de l'auteur données afrobaromètre Niger round9

Taux de prédiction à 80%

Enfin dans le tableau ci-dessous, le nombre de bonnes prédictions se trouve également sur la diagonale(0+718), soit un taux de 59,83%.

TABLEAU 10 – Taux de prédiction à 80%

	True ——— Classified D	D	Total
	0	0	0
	482	718	1200
Total	482	718	1200
Classified + if predicted $\Pr(D) \geq .8$			
True D defined as Statut_Vaccination!= 0			
	Sensitivity $\Pr(+ D)$		0.00%
	Specificity $\Pr(- D)$		100.00%
	Positive predictive value $\Pr(D +)$.%
	Negative predictive value $\Pr(D -)$		59.83%
	False + rate for true $D \Pr(+ D)$		0.00%
	False - rate for true $D \Pr(- D)$		100.00%
	False + rate for classified + $\Pr(D +)$.%
	False - rate for classified - $\Pr(D -)$		40.17%
	Correctly classified		59.83%

Source : calcul de l'auteur données afrobaromètre Niger round9

En résumé on retient que plus le seuil fixé est élevé moins la prédiction est bonne. Pour avoir une bonne classification il va falloir prendre un seuil faible

4.2 Tirage aléatoire stratifié (milieu de résidence) parmi les non vaccinés

La taille de l'échantillon est de 1200. Les personnes vaccinées représentent 40,17% de l'échantillon(482) : on a donc un pourcentage supérieur à 40%. Ainsi en se basant sur le rapport «Connaissances, attitudes et pratiques du personnel de santé face à la Covid-19 à Maradi, Niger en 2021» on retient un taux raisonnable de 45% des personnes vaccinées dans le nouvel echantillon. Determinons alors la taille de l'echantillon : soit n la taille du nouvel echantillon on a : $(482/n)*100=45\% \rightarrow n=(482*100)/45=1077,1111$.

La taille du nouvel échantillon est donc sensiblement égal à 1071. De même la taille des personnes non vaccinées à tirer est : $1077-482 = 589$. On aura donc à tirer 589 personnes parmi les non vaccinées. Procédure du tirage aléatoire stratifié par milieu. Nous avons 22,84% de personnes non vaccinées en milieu urbain et 77,16% en milieu rural. Ainsi en milieu urbain on aura à tirer $589 \times 22,84/100 = 134$ et en milieu rural on aura $589 \times 77,16/100 = 455$. Les estimateurs obtenus sont sans biais.

5 Comparaison dans un tableau les coefficients des deux modèles logit

Le tableau ci-dessous présente une comparaison entre les coefficients du modèle de base estimé sur 1200 observations et le modèle estimé après rééchantillonnage sur 1071 observations. Les deux modèles sont globalement significatifs. Après rééchantillonnage, on a un pseudo R² élevé. On note également que pour les deux modèles les variables milieu, âge, confiance au gouvernement sont significatifs au seuil de 1%.

TABLEAU 11 – Comparaison des coefficients des modèles logit après rééchantillonnage

	(modèle de base)	(modèle après rééchantillonnage)
<i>VARIABLES</i>	<i>coefficients</i>	<i>coefficients</i>
sexe	0.133 (-0.121 - 0.386)	0.092 (-0.170 - 0.354)
milieu	0.704*** (0.349 - 1.060)	0.697*** (0.330 - 1.064)
classe_age	0.117*** (0.032 - 0.202)	0.124*** (0.035 - 0.213)
education	-0.062 (-0.228 - 0.104)	-0.064 (-0.234 - 0.107)
confiance_gouvernement	0.349*** (0.220 - 0.477)	0.372*** (0.228 - 0.516)
malade_de_covi_19	-0.104 (-0.381 - 0.173)	0.060 (-0.311 - 0.432)
Constant	-2.969*** (-4.157 - -1.781)	-3.114*** (-4.410 - -1.818)
Observations	1,200	1,071
wald chi2	69.04	63.50
Prob>chi2	0	0
Pseudo R-squared	0.0663	0.0706
Robust ci ()		
***p<0.01, **p<0.05		

Source : calcul de l'auteur données afrobaromètre Niger round9

6 Estimation séparée (ensemble, urbain, rural).

Le tableau ci-dessous présente les odds ratio du modèle d'ensemble, en milieu urbain et enfin en milieu rural. Lorsqu'on passe du modèle d'ensemble au modèle séparées en milieu urbain on a 185 observations et en milieu rural 885. Cependant la variable classe d'âge n'est pas significatif lorsqu'on passe en milieu urbain. Ainsi lorsque l'âge des individus en milieu rural augmente, les individus ont 1,132 fois chance d'être vaccinées.

TABLEAU 12 – Estimation des modèles séparé

	<i>(ensemble)</i>	<i>(urbain)</i>	<i>(rural)</i>
VARIABLES	Odds ratio	Odds ratio	Odds ratio
Statut_Vaccination	.	.	.
	(.)	(.)	(.)
sexe	1.097 (0.147)	1.153 (0.430)	1.087 (0.156)
milieu	2.007*** (0.376)		
classe_age	1.132*** (0.052)	1.181 (0.131)	1.132** (0.057)
education	0.938 (0.082)	0.987 (0.182)	0.959 (0.098)
confiance_gouvernement	1.451*** (0.107)	2.131*** (0.456)	1.371*** (0.102)
malade_de_covi_19	1.062 (0.201)		1.054 (0.197)
o.malade_de_covi_19		-	
Constant	0.044*** (0.029)	0.024*** (0.029)	0.212*** (0.117)
Observations	1,071	185	885
Robust seeform ()			
***p<0.01, **p<0.05			

Source : calcul de l'auteur données afrobaromètre Niger round9

En reprenant l'estimation du modèle et en prenant les variables explicatives une à une comme variable expliquée, on remarque qu'aucune variable n'est susceptible d'être endogène au modèle.

7 Estimation séparée de l'effet de la variable «literacy» sur les indicateurs de santé infantile et Proposition des variables de contrôle

Le tableau ci-dessous présente l'effet de la variable literacy sur les indicateurs de santé infantile. On peut remarquer que les modèles portant l'effet de la variable literacy sur les indicateurs stunted_growth et Underweight sont globalement significatifs. Il faut noter que la variable literacy a un effet significatif au seuil de 1% sur les indicateurs stunted_growth(retard de croissance) et Underweight (insuffisance pondérale). Ainsi les femmes alphabétisées ont moins de chance d'avoir un des enfants qui ont un retard de croissance par rapport aux femmes non alphabétisation. De même les femmes alphabétisées ont moins de chance d'avoir un des enfants qui ont une insuffisance pondérale par rapport aux femmes non alphabétisation.

TABLEAU 13 – effet de la variable «literacy» sur les indicateurs de santé infantile

		(stunted_growth)	(emaciation)	(Underweight)
VARIABLES	référence	coef	coef	coef
literacy	no			
yes		-0.429*** (0.086)	-0.212 (0.115)	-0.485*** (0.095)
Constant		-1.282*** (0.040)	-2.226*** (0.055)	-1.517*** (0.043)
Observations		5,045	5,045	5,045
wald chi2		26.15	3.496	27.91
Prob>chi2		3.15e-07	0.0615	1.27e-07
Pseudo R-squared		0.00517	0.00112	0.00620
Standard errors ()				
***p<0.01,**p<0.05				

Source : calcul de l'auteur données afrobaromètre Niger round9

7-1 Tests de validation

Le tableau ci-dessous présente les tests de validation du modèle du tableau 13. Il ressort du tableau que seul le modèle effet de la variable literacy sur l'indicateur emaciation n'est pas significatif. Par ailleurs les pseudo de McFadden R^2 pour les modèles effets de la variable literacy sur les indicateurs stunted_growth(retard de croissance) et Underweight (insuffisance pondérale) sont respectivement 0,005 et 0,006, ils sont donc loin de 0. On peut dire que ces deux modèles sont relativement bon.

TABLEAU 14 – Tests de validation

	<i>stunted_growth</i>	<i>emaciation</i>	<i>Underweight</i>
Log-likelihood			
Model	-2514.223	-1558.876	-2236.727
Intercept-only	-2527.300	-1560.624	-2250.680
Chi-square			
Deviance(df=5043)	5028.447	3117.753	4473.454
LR(df=1)	26.154	3.496	27.907
p-value	0.000	0.062	0.000
R2			
McFadden	0.005	0.001	0.006
McFadden(adjusted)	0.004	-0.000	0.005
McKelvey & Zavoina	0.011	0.003	0.014
Cox-Snell/ML	0.005	0.001	0.006
Cragg-Uhler/Nagelkerke	0.008	0.002	0.009
Efron	0.005	0.001	0.005
Tjur's D	0.005	0.001	0.005
Count	0.800	0.907	0.836
Count(adjusted)	0.000	0.000	0.000
IC			
AIC	5032.447	3121.753	4477.454
AIC divided by N	0.998	0.619	0.888
BIC(df=2)	5045.499	3134.805	4490.506
Variance of			
e	3.290	3.290	3.290
y-star	3.325	3.299	3.335

7-2 Variables de contrôle pour affiner les résultats

Pour le choix des variables de controle nous avons choisi v701 que nous avons renommer en nivinst(niveau d'instruction) et residential_zone(milieu de residence). Ce choix de ces variables nous permettra de voir la structure suivant le niveau d'instruction et le milieu de résidence. Ainsi les femmes ayant un niveau d'éducation élevé ont moins de chance d'avoir des enfants ayant un retard de croissance (stunted_growth) par rapport à celles qui n'ont aucune éducation. De même les femmes rurales ont plus de chance d'avoir des enfants ayant un retard de croissance (stunted_growth) et une insuffisance pondérale (Underweight) par rapport aux femmes vivant en milieu urbain.

TABLEAU 15 – Modèle avec variables de contrôle

		stuntedgrowth	emaciation	Underweight
VARIABLES	Références	coef	coef	coef
literacy	no			
yes		-0.229** (0.096)	-0.138 (0.130)	-0.281*** (0.105)
nivinst	no education			
primary		0.007 (0.124)	-0.092 (0.174)	-0.056 (0.137)
secondary		-0.203 (0.148)	-0.057 (0.191)	-0.244 (0.163)
higher		-0.775** (0.301)	-0.111 (0.310)	-0.560 (0.303)
don't know		-0.107 (0.168)	-0.270 (0.246)	-0.094 (0.181)
residential_zone	urbain			
rural		0.518*** (0.094)	0.030 (0.120)	0.352*** (0.100)
Constant		-1.684*** (0.092)	-2.247*** (0.116)	-1.796*** (0.097)
Observations		4,860	4,860	4,860
wald chi2		74.99	3.926	45.74
Prob>chi2		0	0.687	3.33e-08
Pseudo R-squared		0.0154	0.00132	0.0106
Standard errors ()				
***p<0.01,**p<0.05				

7-3 Tests d'hypothèses

Le tableau n°16 présente les tests sur les modèles avec les variables de contrôle. Seuls les modèles portant sur les indicateurs retard de croissance (stunted_growth) et insuffisance pondérale (Underweight) sont significatifs. Le R^2 de McFadden de ces deux modèles significatifs est respectivement 0,015 et 0,011 donc ils sont assez loin de 0, on peut donc dire que ces deux modèles sont relativement bon.

TABLEAU 16 – Tests sur modèle avec variables de contrôle

	stunted_growth	emaciation	Underweight
Log-likelihood			
Model	-2399.994	-1490.466	-2136.008
Intercept-only	-2437.491	-1492.430	-2158.879
Chi-square			
Deviance(df=4853)	4799.988	2980.933	4272.016

LR(df=6)	74.993	3.926	45.742
p-value	0.000	0.687	0.000
R2			
McFadden	0.015	0.001	0.011
McFadden(adjusted)	0.013	-0.003	0.007
McKelvey & Zavoina	0.036	0.003	0.025
Cox-Snell/ML	0.015	0.001	0.009
Cragg-Uhler/Nagelkerke	0.024	0.002	0.016
Efron	0.014	0.001	0.009
Tjur's D	0.014	0.001	0.009
Count	0.799	0.908	0.837
Count(adjusted)	0.000	0.000	0.000
IC			
AIC	4813.988	2994.933	4286.016
AIC divided by N	0.991	0.616	0.882
BIC(df=7)	4859.410	3040.354	4331.437
Variance of			
e	3.290	3.290	3.290
y-star	3.412	3.300	3.374

7-4 Existe -t-il un lien entre les variables dépendantes ?

Le tableau ci-dessous présente un croisement entre les variables dépendantes. Le test de χ^2 a été réalisé afin de voir s'il y a un lien entre ces variables. Au vu des résultats dans le tableau les p_values étant toutes inférieure à 5% on en déduit que les variables dépendantes sont liées au seuil fixé.

TABLEAU 17 – lien entre les variables dépendantes

stunted_growth		Underweight		
		0	1	Total
0	3,749	285	4,034	
1	469	542	1,011	
Total	4,218	827	5,045	
Pearson chi2(1) = 1.3e+03 Pr = 0.000				
stunted_growth		emaciation		
		0	1	Total
0	3,709	325	4,034	
1	867	144	1,011	
Total	4,576	469	5,045	
Pearson chi2(1) = 36.6963 Pr = 0.000				

Underweight	emaciation		
	0	1	Total
0	4,064	154	4,218
1	512	315	827
Total	4,576	469	5,045
Pearson chi2(1) = 972.5274 Pr = 0.000			

Pour prendre en compte le lien entre les variables dépendantes, il convient d'utiliser un modèle multivarié qui permettrait de prendre en compte ce lien, un modèle probit multivarié serait un bon candidat.

7-5 Modèle qui prend en compte l'interdépendance potentielle.

Le modèle est globalement significatif. La variable literacy est significatif au seuil de 5% sur toutes les variables dépendantes stunted_growth, Underweight et emaciation.

TABLEAU 18 – modèle probit multivarié

	<i>(stunted_growth)</i>	<i>(Underweight)</i>	<i>(emaciation)</i>
VARIABLES	coef	coef	coef
literacy	-0.237*** (0.047)	-0.267*** (0.050)	-0.122** (0.059)
Constant	-0.780*** (0.023)	-0.920*** (0.024)	-1.282*** (0.029)
Observations	5,045	5,045	5,045
wald chi2	36.54	36.54	36.54
Prob>chi2	5.77e-08	5.77e-08	5.77e-08
Robust standard errors ()			
***p<0.01,**p<0.05			

Deuxième partie

Modèle multinomial

8 Sous quelles conditions, il est possible de modéliser la variable type_contrat

Le tableau ci-dessous nous présente la variable type_contrat. Il ressort de ce tableau que cette variable est catégorielle avec 4 modalités. Nous sommes dans le cadre d'un modèle multinomial non ordonné. Ainsi ce modèle repose sur les hypothèses suivantes :

- > l'Indépendance des Alternatives Non Pertinentes (IANP ou IIA en anglais pour Independence of Irrelevant Alternative). La condition est que le rapport de probabilités entre deux modalités est indépendant des autres modalités de la variable d'intérêt.
- > les co-variables sont les mêmes quelle que soit la modalité de y considéré (x dépend de i mais pas de j)

TABLEAU 19 – type_contrat

E9_Type contrat	Freq.	Percent	Cum.	
	CDI	717	17.78	17.78
	CDD	343	8.51	26.29
	Contrat prestation service	420	10.42	36.71
	Sans contrat	2,552	63.29	100.00
	Total	4,032	100.00	

Source : calcul de l'auteur données afrobaromètre Niger round9

9 Tableau synthétique des différentes variables juger pertinentes

Le tableau ci-dessous présente une description des variables jugées indépendantes. Nous avons retenus essentiellement le sexe, l'âge, la situation matrimoniale, le niveau d'instruction, la maladie, l'incidence de pauvreté et enfin le nombre d'enfant.

TABLEAU 20 – Variables pertinentes

Variable	Storage	Display	Value	
name	type	format	label	Variable label
b2	double	%10.0g	b2	B2_Sexe
b3	double	%10.0g	b3	B3_Age
b4	double	%10.0g	b4	B4_Situation matrimoniale
nivinst	double	%10.0g	nivinst	Niveau d'instruction
malade	double	%10.0g	malade	
p0	double	%10.0g	p0	Incidence de la pauvreté
nbr_enfant	float	%9.0g		

10 Estimer et présenter le modèle logit multinomial en fixant comme modalité de référence « CDI ».

Le tableau ci-dessous présente l'estimation du modèle logistique multinomial en gardant comme référence la modalité «CDI» de la variable type de contrat. Il ressort de ce tableau que les individus de sexe féminin ont plus de chance d'être sans contrat que d'avoir un CDI par rapport aux individus de sexe masculin. De même au fur et à mesure que l'âge augmente les individus ont moins de chance d'avoir un contrat de type CDD que d'avoir CDI. Il en est de même pour les contrats de type prestation de service et sans contrat. Par ailleurs les 1^{ère} épouse ont plus de chance d'avoir une prestation de service qu'un contrat CDI par rapport au marié monogame. Les individus ayant un niveau d'instruction primaire ont de chance d'avoir un CDD que d'avoir un CDI par rapport à ceux n'ayant aucune instruction. Cependant les individus ayant un niveau d'instruction supérieur ont moins de chance d'être sans contrat que d'avoir un CDI par rapport aux individus sans niveau d'instruction. Cependant les individus non malade ont moins de chance d'avoir une prestation de service que d'avoir un CDI par rapport aux malades. Les pauvres ont plus de chance d'avoir un CDD qu'un CDI par rapport aux non pauvre.

TABLEAU 21 – modèle logit multinomial

		(2)	(3)	(4)
VARIABLES	Réf	CDD	Prestation.services	Sans.contrat
sexe	Masculin			
Feminin		0.280 (0.163)	-0.126 (0.171)	0.517*** (0.137)
Age		-0.021*** (0.008)	-0.033*** (0.007)	-0.028*** (0.006)
situation matrimoniale	Marié.monog			
1ereipouse		0.347 (0.286)	1.090*** (0.265)	0.454 (0.237)
2ipouses/2hmeipouse		-0.151 (0.254)	-0.293 (0.272)	-0.150 (0.204)
3ipouses/3hmeipouse		-1.234 (0.780)	-0.163 (0.525)	-0.550 (0.427)
4ipousesou+/4hmeipouseou+		15.053 (1,692.283)	-0.816 (2,792.505)	13.801 (1,692.283)
Celibataire		0.728*** (0.194)	0.785*** (0.203)	1.252*** (0.167)
Veuf(ve)		-0.500 (0.624)	0.326 (0.511)	0.267 (0.410)

Divorci(e)		0.336 (0.416)	0.497 (0.433)	0.301 (0.351)
Concubinage/Union libre		14.942 (825.229)	0.097 (1,558.167)	15.475 (825.229)
Niveau d'instruction	Sans.instru			
Primaire		-0.651** (0.255)	-1.050*** (0.226)	-1.147*** (0.197)
Moyen		-1.066*** (0.234)	-2.286*** (0.226)	-2.991*** (0.188)
Secondaire		-1.740*** (0.235)	-3.459*** (0.257)	-4.650*** (0.209)
Supirieur		-1.950*** (0.254)	-4.134*** (0.336)	-5.882*** (0.297)
malade	Oui			
Non		-0.136 (0.156)	-0.486*** (0.156)	-0.484*** (0.129)
nbr_enfant		0.027 (0.046)	0.040 (0.045)	0.092** (0.038)
Incidence de la pauvreté	non.pauvre			
pauvre		0.696*** (0.161)	0.707*** (0.161)	0.458*** (0.137)
Observations		3981	3981	3981
Pseudo R2		0.236	0.236	0.236
Standard errors()				
***p<0.01,**p<0.05				

11 Estimer et présenter dans un tableau les risques relatifs.

Le tableau ci-dessous présente les risques relatifs. Il ressort que les individus de sexe féminin ont 1,324 fois plus de risque d'avoir un CDD qu'un CDI que ceux de sexe masculin. De même plus l'âge augmente les individus ont 0,979 (respectivement 0,968 et 0,972) fois moins de risque d'avoir un CDD (respectivement une prestation de service et sans contrat) que d'avoir un CDI. Les célibataires ont 3,496 fois plus de risque d'être sans contrat que d'avoir un CDI que les marié monogame. Les individus ayant un niveau supérieur ont 0,00279 fois moins de risque d'être sans contrat que d'avoir un CDI par rapport à ceux n'ayant aucun niveau d'instruction. Ceux ayant un niveau primaire ont 0,522 fois moins de risque d'avoir un CDD qu'un CDI par rapport à ceux n'ayant aucune instruction. Par ailleurs les individus pauvres ont 2,005 fois plus de risque d'être CDD que CDI par rapport aux non pauvres. Plus le nombre d'enfant augmente les individus ont 1,096 fois plus de risque d'être sans contrat.

TABLEAU 22 – risques relatifs

		Risk relatif	Risk relatif	Risk relatif
VARIABLES	Références	CDD	prest.servic	Sanscontrat
sexe	Masculin			
Feminin		1,324*	0,881	1,677***
		-0,22	-0,154	-0,234
Age		0,979***	0,968***	0,972***
		-0,00688	-0,00694	-0,00564
situation matrimoniale	Mariémono			
1ere ipouse		1,415	2,974***	1,574**
		-0,402	-0,753	-0,355
2 ipouses/2hme ipouse		0,86	0,746	0,861
		-0,217	-0,206	-0,181
3 ipouses/3hme ipouse		0,291	0,85	0,577
		-0,229	-0,452	-0,251
4ipousesou+/4hmeipouseou+		3,446e+06***	0,442***	985,699***
		-3,65E+06	-0,0891	(590,245)
Cilibataire		2,072***	2,191***	3,496***
		-0,42	-0,447	-0,608
Veuf(ve)		0,607	1,386	1,306
		-0,359	-0,722	-0,498
Divorci(e)		1,4	1,644	1,351
		-0,565	-0,698	-0,467
Concubinage/Union libre		3,086e+06***	1,101	5,257e+06***
		-2,66E+06	-0,335	-2,10E+06
Niveau d'instruction	Sans instru			
Primaire		0,522**	0,350***	0,318***
		-0,134	-0,0807	-0,0641
Moyen		0,344***	0,102***	0,0502***
		-0,0818	-0,0238	-0,00982
Secondaire		0,176***	0,0315***	0,00956***
		-0,0422	-0,00837	-0,00209
Supirieur		0,142***	0,0160***	0,00279***
		-0,0363	-0,00548	-0,000847
malade	Oui			
Non		0,873	0,615***	0,616***
		-0,135	-0,0952	-0,0778
Incidence de la pauvreté	nonpauvre			
pauvre		2,005***	2,027***	1,581***
		-0,323	-0,318	-0,211
nbr_enfant		1,027	1,04	1,096**
		-0,0485	-0,047	-0,0426
Constant		2,458**	11,73***	54,16***
		-1,048	-4,854	-19,5
Observations		3,981	3,981	3,981

Robuststandarderrorsinparen

*** p<0,01, ** p<0,05, * p<0,1

12 Estimer et présenter dans un tableau les effets marginaux.

Le tableau ci-dessous présente les effets marginaux. Il ressort que les individus de sexe féminin ont 0,03 fois moins de chance d'avoir un CDI que ceux de sexe masculin. Plus l'âge augmente les individus ont 0,002 fois plus de chance d'être en CDI. Les pauvres ont 0,037 fois moins de chance d'avoir un CDI que les non pauvres. Concernant le niveau d'instruction les individus ayant un niveau moyen ont 0,378 fois plus de chance d'être en CDI que ceux n'ayant aucune instruction. Cependant les célibataires ont 0,077 fois moins de chance d'être en CDI par rapport aux marié monogame. Les individus vivant en concubinage ou union libre ont 0,087 fois moins de chance d'avoir un CDI que les marié monogame.

TABLEAU 23 – effets marginaux

		(CDI)
VARIABLES	Réf	effet marginaux
sexe	Masculin	
Féminin		-0,0300***
		-0,00952
situation matrimoniale	Mariémono	
1ere ipouse		-0,0350***
		-0,0112
2 ipouses/2hme ipouse		0,0137
		-0,0172
3 ipouses/3hme ipouse		0,0514
		-0,0487
4ipousesou+/4hmeipouseou+		-0,0849***
		-0,00639
Celibataire		-0,0773***
		-0,0103
Veuf(ve)		-0,0152
		-0,0244
Divorci(e)		-0,0224
		-0,0192
Concubinage/Union libre		-0,0874***
		-0,00656

Niveau d'instruction	Sans instru	
Primaire		0,104*** -0,0222
Moyen		0,378*** -0,0354
Secondaire		0,642*** -0,0327
Supirieur		0,738*** -0,0304
malade	Oui	
Non		0,0324*** -0,00844
Age		0,00216*** -0,00042
	nbr_enfant	-0,00597** -0,0029
Incidence de la pauvreté	nonpauvre	
pauvre		-0,0378*** -0,00908
Observations		3,981
Robust standard errors in parentheses		
*** p<0,01, ** p<0,05, * p<0,1		

13 Effectuer le test IIA.

L'hypothèse IIA n'est donc pas vérifiée. Ainsi, on peut conclure que la probabilité qu'un individu prenne un caractère n'est pas indépendante d'un autre. Elle a tendance à varier et donc être peu fiable. Ainsi, le modèle multinomial n'est donc pas commode pour cette étude.

TABLEAU 24 – Test IIA Wald

	chi2	df	P>chi2
2.b2	34.038	3	0.000
b3	24.889	3	0.000
2.b4	19.923	3	0.000
3.b4	1.222	3	0.748
4.b4	3.719	3	0.293
5.b4	1.182	3	0.757
6.b4	58.751	3	0.000
7.b4	2.280	3	0.516

8.b4	1.431	3	0.698
9.b4	0.428	3	0.934
2.nivinst	39.207	3	0.000
3.nivinst	318.199	3	0.000
4.nivinst	542.117	3	0.000
5.nivinst	441.254	3	0.000
2.malade	16.990	3	0.001
nbr_enfant	8.757	3	0.033
1.p0	25.882	3	0.000

Troisième partie

modèle Tobit et Heckman

14 Quel problème de modélisation soulève cette variable ?

TABLEAU 25 – Tabulation sur le revenu

montant mensuelle de votre rémunération ?	Freq.	Percent	Cum.
0	615	100.00	100.00
Total	615	100.00	

D'après le tableau ci-dessus le problème de modélisation soulevé par la variable « revenu » est lié à la présence de données censurées. En effet, les données sur le revenu ne sont observées que pour les individus qui travaillent, les autres (non-travailleurs) ont un revenu non observé. C'est un problème de sélection qui nécessite une modélisation appropriée, comme le modèle de Heckman.

15 Tableau synthétique des variables pertinentes pour le problème posé :

TABLEAU 26 – variables pertinentes

variable	Storage	Display	Value
name	type	format	label Variable label
age	double	%12.0g	Q1. Quel âge avez-vous ?
sexe	double	%12.0g	Q2 Q2. Sexe
revenu	double	%12.0g	Q7a_1. Quel était le montant mensuelle (en FCFA) de votre rémunération ?
niv_educ	double	%26.0g	lab_niv_educ RECODE of Q4 (Q4. Quel est le niveau scolaire le plus élevé que vous ayez fréquenté)
nbr_enfant	double	%12.0g	Q11. Combien d'enfants avez-vous ?
stat_matrim	double	%12.0g	lab_sit_mat RECODE of q9 (Q9. Quel est votre statut matrimonial actuel?)
labor_part	float	%15.0g	labor partie travail
health_problem	float	%9.0g	
nbr_year_educ	float	%9.0g	Nombre d'années d'éducation

Source : calcul de l'auteur données afrobaromètre Niger round9

Ces variables permettent d'étudier les déterminants du revenu tout en tenant compte du fait que le revenu n'est observé que pour les travailleurs.

16 Estimer un modèle linéaire sur l'ensemble de l'échantillon puis dans le sous échantillon des travailleurs

Lorsqu'on réalise le modèle ols sur l'ensemble et sur l'échantillon des travailleurs on peut remarquer qu'on a le même nombre de coefficients significatifs. Cependant sur l'ensemble de l'échantillon la constante n'est pas significative par contre elle est significative dans l'échantillon des travailleurs

TABLEAU 27 – Estimer un modèle linéaire

	(1)	(2)
VARIABLES	ensemble_echantillon	echantillon_travailleurs
age	0.0192*** (0.00398)	0.0317*** (0.00715)
sexe	-0.207*** (0.0379)	-0.417*** (0.0690)
nbr_enfant	-0.0515*** (0.0165)	-0.0696** (0.0293)
milieu	0.0903** (0.0369)	0.197*** (0.0660)
niv_educ	0.0644 (0.0773)	0.144 (0.141)
stat_matrim	-0.0556 (0.0372)	-0.0397 (0.0648)
labor_part	10.83*** (0.0430)	
health_problem	-0.108*** (0.0356)	-0.205*** (0.0655)
year_exp	0.0300*** (0.00756)	0.0297*** (0.0107)
nbr_year_educ	-0.00395 (0.0185)	-0.0118 (0.0335)
region	-0.0156*** (0.00587)	-0.0245** (0.0105)
o.labor_part		-
Constant	-0.0883 (0.150)	10.53*** (0.267)
Observations	1,300	685
R-squared	0.988	0.199
Standard errors ()		
***p<0.01,**p<0.05,*p<0.1		

Source : calcul de l'auteur données afrobaromètre Niger round9

16-1 Prédire les revenus pour les non travailleurs

Le tableau ci-dessous présente un résumé du revenu prédit des non travailleurs. Le revenu moyen des non travailleurs prédit est 49759,87. On peut remarquer que les valeurs prédites du revenu des non travailleurs s'écartent de la vraie valeur du revenu des travailleurs. Le modèle ols surestime le revenu prédit

TABLEAU 28 – Résumé du revenu des non travailleurs prédits

Variable	Obs	Mean	Std. dev.	Min	Max
revenu_pred	615	49759.87	19902.02	18823.53	150077.7

17 Estimer un modèle de sélection de Heckman avec les mêmes variables.

La pvalue étant supérieur à 5% , le ratio de mill n'est pas significatif, il n'y a donc pas un biais de sélection dans notre modèle.

TABLEAU 29 – modèle de sélection de Heckman

	Coefficient	Std. err.	z	P>z	[95% conf. interval]
ln_wage					
sexe	-.4030369	.0689682	-5.84	0.000	-.5382122 -.2678617
age	.0352602	.0070733	4.99	0.000	.0213969 .0491236
milieu	.2320832	.0650023	3.57	0.000	.1046809 .3594854
year_exp	.0254055	.0106395	2.39	0.017	.0045524 .0462586
health_problem	-.2266683	.0649511	-3.49	0.000	-.3539701 -.0993664
region	-.0291245	.0104076	-2.80	0.005	-.049523 -.008726
labor_part	0	(omitted)			
nbr_enfant	-.076542	.0278517	-2.75	0.006	-.1311302 -.0219537
_cons	10.08394	.4057751	24.85	0.000	9.288635 10.87924
travailleur					
stat_matrim	-.1265746	.0648591	-1.95	0.051	-.2536961 .0005469
niv_educ	-.4348255	.1576819	-2.76	0.006	-.7438764 -.1257746
nbr_year_educ	.097702	.0378158	2.58	0.010	.0235843 .1718197
_cons	.2385621	.1139658	2.09	0.036	.0151933 .461931
/mills					
lambda	.5523262	.4510679	1.22	0.221	-.3317507 1.436403
rho	0.60315				
sigma	.91574026				

17-1 Sgnificativité globale du modèle

Le modèle est réalisé sur 1300 observations, elle est globalement significative.

TABLEAU 30 – Significativité globale

<i>Number of obs</i>	=	<i>1300</i>
<i>Selected</i>	=	<i>685</i>
<i>Nonselected</i>	=	<i>615</i>
<i>Wald chi2(7)</i>	=	<i>144.84</i>
<i>Prob > chi2</i>	=	<i>0.0000</i>

17-2 Prédire les revenus pour les non travailleurs

Le tableau ci-dessous présente un résumé du revenu prédit des non travailleurs dans le cas d'un modèle de Heckman. Le revenu moyen des non travailleurs prédit est 32147,06. Ainsi cette prédiction ne s'écarte pas trop du revenu des travailleurs. On peut donc dire que le modèle de selection de heckman prédit mieux le revenu des non travailleurs.

TABLEAU 31 – Revenu des non travailleurs prédits

Variable	Obs	Mean	Std. dev.	Min	Max
salairé_pred	615	32147.06	11854.8	10896.95	84969.94

17-3 Comparer les résultats des prédictions avec ceux

En comparant les deux prédictions on peut tout simplement dire que le modèle de selection de heckman prédit mieux le revenu des non travailleurs et cela s'explique facilement car lorsque la variable dépendante est censurée le modèle adapté est celui de heckman.

18 Est-ce que le fait d'être marié et le fait d'avoir un enfant sont-elles endogènes ? Justifier ? Quelles solutions préconisez-vous s'il y en a ?

Le fait d'être marié et d'avoir un enfant peuvent être endogènes, c'est-à-dire que ces variables peuvent être corrélées avec des facteurs non observés qui influencent également le revenu. Pour traiter ce problème, on pourrait envisager d'utiliser des variables instrumentales qui seraient corrélées avec le statut marital et le nombre d'enfants, mais pas avec le revenu.