

STATISTIQUES ET ECONOMETRIE  
APPLIQUEES 2

**UNE ETUDE SUR LA CONSOMMATION  
DE BIÈRE, DE VIN ET DE SPIRITUEUX  
EN SUISSE**

Tea Danelutti

Juin 2003

Professeur Marius Brülhart  
Ecole des HEC - Université de Lausanne

# TABLES DES MATIERES

<b>1</b>	<b>INTRODUCTION</b> .....	<b>3</b>
<b>2</b>	<b>MODELE</b> .....	<b>3</b>
2.1	INTUITION.....	3
2.2	MODELE THEORIQUE.....	4
<b>3</b>	<b>BASE DE DONNEES</b> .....	<b>6</b>
3.1	DESCRIPTION DES VARIABLES.....	6
3.1.1	<i>VARIABLES SOCIO-DEMOGRAPHIQUES CLASSIQUES</i> .....	7
3.1.2	<i>VARIABLES SOCIO-DEMOGRAPHIQUES PARTICULIERES</i> .....	7
3.1.3	<i>VARIABLES CONCERNANT LA CONSOMMATION D'ALCOOL</i> .....	8
<b>4</b>	<b>ANALYSE ECONOMETRIQUE</b> .....	<b>9</b>
4.1	L'EQUATION DE SELECTION.....	9
4.2	L'EQUATION SUBSTANTIELLE.....	10
4.3	REMARQUE SUR LES SIGNES ATTENDUS.....	10
4.4	EXEMPLE DE LA BIERE.....	11
4.4.1	<i>PAR LA METHODE DU MAXIMUM DE VRAISEMBLANCE</i> .....	11
4.4.2	<i>PAR LA PROCEDURE EN DEUX ETAPES DE HECKMAN</i> .....	14
4.5	COMPARAISON DES TROIS MODELES.....	16
4.6	ESTIMATION SIMULTANEE DES TROIS MODELES.....	19
<b>5</b>	<b>CONCLUSION</b> .....	<b>23</b>
<b>6</b>	<b>BIBLIOGRAPHIE</b> .....	<b>24</b>
<b>7</b>	<b>ANNEXES</b> .....	<b>25</b>
7.1	ANNEXE 1.....	25
7.2	ANNEXE 2.....	26
7.3	ANNEXE 3.....	27
7.4	ANNEXE 4.....	33
7.5	ANNEXE 5.....	34
7.6	ANNEXE 6.....	36

# 1 INTRODUCTION

Au cours de cet hiver, et aujourd'hui encore, les panneaux publicitaires dénonçant la consommation abusive d'alcool se sont multipliés sur les murs de nos villes et dans les médias. Ces campagnes visent essentiellement à diminuer les ravages des boissons alcoolisées au sein de la population<sup>1</sup>. Beaucoup d'argent est ainsi dépensé dans un but d'information mais aussi de prévention, soulevant inévitablement de multiples interrogations quant à l'efficacité de telles mesures.

Dans le présent papier, nous étudierons la consommation de trois types d'alcools en Suisse : la bière, le vin et les spiritueux. Plus précisément, nous chercherons à déterminer si des réserves émises par un proche ou un médecin au sujet de la consommation d'alcool d'un individu ont un impact sur la consommation actuelle de ce dernier. En assimilant un tel effet aux campagnes de prévention dont il était question plus haut, nous pourrions exprimer un jugement sur l'efficacité de pareilles dépenses.

La première partie de cette étude traitera du type de modèle choisi et sera suivie d'une description des variables. L'analyse économétrique sera présentée dans une troisième partie. Elle laissera finalement place à une brève conclusion critique.

## 2 MODELE

### 2.1 INTUITION

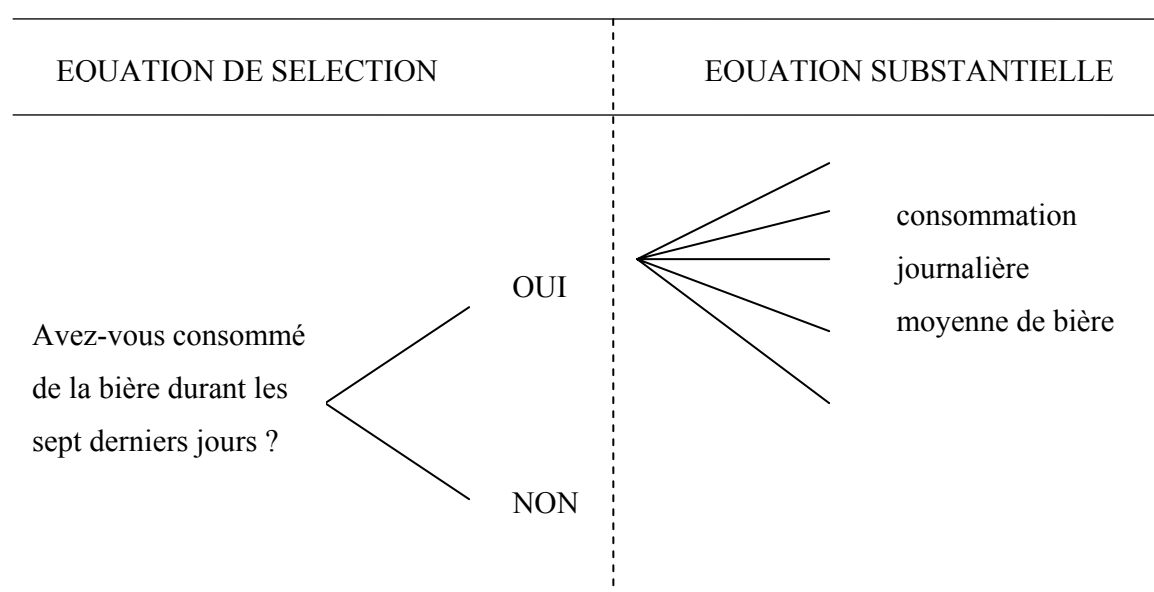
Le modèle économétrique envisagé pour cette étude relève du domaine des variables qualitatives, plus précisément des modèles de sélection. En effet, les variables que nous cherchons à expliquer sont les nombres de décilitres consommés pour la bière, le vin et les spiritueux. Or, il est certain que ces informations ne sont disponibles que pour les gens ayant

---

<sup>1</sup> Selon l'Institut suisse de prévention de l'alcoolisme et autres toxicomanies (2001), « l'abus d'alcool en Suisse occasionne chaque année des coûts sociaux d'un montant de trois milliards de francs ».

consommé l'un au moins de ces alcools durant la période considérée (dans notre cas, il s'agira des sept jours précédant l'enquête ; cf. Section 3).

Afin d'utiliser les informations disponibles pour les trois types d'alcool, nous allons procéder à l'analyse de trois modèles séparés, chacun ne différant des deux autres que par le type d'alcool étudié. Le modèle expliquant la quantité de bière consommée a été schématisé ci-dessous pour faciliter la compréhension.



Nous pouvons considérer ce modèle comme un modèle en deux étapes. Dans un premier temps, l'individu choisit de consommer de la bière ou non ; ensuite, le cas échéant, il décide de la quantité consommée. Cette approche est semblable à celle des « two-parts models » dont un aperçu peut être obtenu dans Manning (1997).

Mais voyons cela sur un plan plus théorique...

## 2.2 MODELE THEORIQUE

Dans notre échantillon, les données sont par définition tronquées. En effet, ce n'est que si une personne a une propension à consommer de la bière/du vin/des spiritueux supérieure à zéro, que nous pourrions étudier les facteurs l'ayant poussée à choisir telle quantité plutôt qu'une autre.

En utilisant la méthode développée par Heckman en 1979<sup>2</sup>, notre modèle peut se formaliser comme suit pour chaque individu  $i$  :

(1) Propension à consommer un certain type d'alcool : (équation de sélection)

$$\Rightarrow z_i^* = w_i\gamma + \mu_i \quad \text{on observe } z_i^* \text{ uniquement si l'individu } i \text{ consomme de cet alcool}$$

(2) Estimation de la quantité consommée : (équation substantielle<sup>3</sup>)

$$\Rightarrow y_i = x_i\beta + \varepsilon_i \quad \text{observable uniquement si } z_i^* > 0$$

avec  $w_i$  et  $x_i$  des variables socio-démographiques observables ;  $\mu_i$  suivant une loi normale  $N(0;1)$  et  $\varepsilon_i$  une loi normale  $N(0;\sigma_\varepsilon)$  ;  $\rho$  le coefficient de corrélation des termes d'erreur.

Ce genre de modèle est normalement estimé par la Méthode du Maximum de Vraisemblance (MMV). Cependant, comme la convergence s'avère parfois difficile, l'estimateur de Heckman, obtenu en deux étapes, lui est parfois préféré. L'équation de sélection est alors d'abord estimée par un modèle Probit ; ensuite, une régression par les Moindres Carrés Ordinaires (MCO) permet d'obtenir les coefficients de la deuxième équation. La méthode de Heckman permet ainsi une approximation des résultats trouvés par la MMV ; les estimateurs calculés seront légèrement biaisés.

Quelques mots encore sur le coefficient de corrélation ( $\rho$ )... Lorsque celui-ci est égal à 0, l'estimation des coefficients de l'équation substantielle par les MCO n'est pas biaisée<sup>4</sup>. Dans ce cas particulier où les termes d'erreur des deux équations ne sont pas corrélés entre eux, l'équation de sélection n'a plus de raison d'être car les deux décisions sont indépendantes.

Cette explication est certes courte et simplifiée, toutefois elle permet de présenter l'essentiel du fonctionnement de ce genre de modèle.

---

<sup>2</sup> Pour plus de détails, veuillez consulter Heckman (1979), Siegelman et Zeng (1999) ou encore Winship (1992).

<sup>3</sup> Traduction littérale de l'équation que l'on nomme « substantial equation » en anglais.

<sup>4</sup> En effet, si  $\rho = 0$ , l'espérance conditionnelle des  $y$  observés est la suivante :  $E(y|z^* > 0) = X\beta$ .

### **3 BASE DE DONNEES**

Les données utilisées dans cette étude ont gentiment été mises à disposition par l'Institut d'Economie et de Management de la Santé (IEMS) et proviennent de l'Enquête Suisse sur la Santé 1997. Elles ont été récoltées par enquête téléphonique en 1997, sur demande de l'Office Fédéral de la Santé Publique. Il s'agit d'une source d'information très riche puisqu'elle fournit des renseignements pour 13'004 personnes de plus de 15 ans, domiciliées en Suisse.

Les thèmes abordés par l'enquête étant très nombreux (logement, travail, toxicomanies, pathologies...), le premier travail fut de sélectionner les variables relatives à la consommation d'alcool, ainsi que celles traitant des caractéristiques socio-démographiques des individus. Le nombre d'observations disponibles pour chacune d'elles n'étant cependant pas forcément le même, un deuxième nettoyage s'imposait. Les personnes pour lesquelles certaines réponses manquaient (sur le revenu ou l'état de santé, par exemple) furent retirées de l'échantillon<sup>5</sup>. Ainsi, suite à de multiples manipulations, seuls 7'947 individus furent conservés pour l'analyse.

#### **3.1 DESCRIPTION DES VARIABLES**

Face à l'ampleur de la base de données, le choix des variables pertinentes à inclure dans le modèle fut difficile. Les questions posées par les enquêteurs étant parfois très similaires, l'important était de déterminer quelle variable correspondait le plus à l'information recherchée. Douze d'entre elles furent finalement retenues<sup>6</sup>.

---

<sup>5</sup> Les données manquantes peuvent être dues à la combinaison de différents facteurs : le manque de temps pour répondre aux questions, la volonté de terminer rapidement, l'aspect tabou du sujet, la peur d'être jugé par l'interlocuteur...

<sup>6</sup> Par rapport à la base de données originale, certaines ont été transformées ou recodées pour les besoins de l'étude, en suivant les conseils de Hosmer et Lemeshow (1989).

Variable	Obs	Mean	Std. Dev.	Min	Max
biere	3085	2.708266	3.280162	.8571429	36
vin	5674	1.102838	1.053689	.2857143	10
spirits	1663	.0977064	.1010591	.0571429	1.2
age	7947	44.26576	16.00047	15	91
age2	7947	2215.441	1562.925	225	8281
revenu	7947	3900.303	1863.7	0	33333.31
femme	7947	.5018246	.5000281	0	1
mariage	7947	.547628	.4977577	0	1
maitrise	7947	.2777149	.4479002	0	1
sante	7947	.0240342	.1531651	0	1
avis	7947	.0357368	.1856448	0	1
avisfemme	7947	.01057	.1022723	0	1

### 3.1.1 VARIABLES SOCIO-DEMOGRAPHIQUES CLASSIQUES

➤ *age* et *age2*

*Age2* est tout simplement le carré de *age* pour chaque personne.

➤ *revenu*

Selon l'IEMS, cette variable correspond au revenu mensuel du ménage (en francs), « calculé sur la base des réponses données au questionnaire et corrigé selon divers critères de plausibilité ou compatibilité ». Il semblerait en effet que les réponses fournies étaient souvent lacunaires ou peu probables.

➤ *femme* et *mariage*

La variable dichotomique *femme* prend la valeur 1 lorsque l'individu est de sexe féminin (50.18% de l'échantillon) et 0 sinon. Pour *mariage*, 1 est attribué aux individus qui étaient mariés et vivaient en couple lors de l'enquête ; 0 aux personnes célibataires, veuves, divorcées ou séparées.

### 3.1.2 VARIABLES SOCIO-DEMOGRAPHIQUES PARTICULIERES

➤ *maitrise*

Cette variable dichotomique mesure le niveau de contrôle que pense avoir la personne interviewée sur sa propre vie : 1 si faible et 0 si moyen ou élevé.

➤ *sante*

*Sante* prend la valeur 0 si la personne estime que son état de santé est de moyen à très bon. Dans le cas contraire (de mal à très mauvais), le code attribué est 1.

➤ *avis*

Encore une variable muette... La question posée lors de l'enquête était la suivante : « Un de vos parents, un ami ou un médecin a-t-il émis des réserves au sujet de votre consommation d'alcool ou vous a conseillé d'en réduire la consommation ? » La personne avait le choix parmi les 3 réponses suivantes : (1) non, (2) oui, mais pas l'année dernière, (3) oui, l'année dernière. Les réponses positives ont été regroupées et le chiffre 1 leur a été attribué ; les individus n'ayant jamais reçu de pareille remarque sont représentés par un 0.

Cette variable semble a priori très intéressante pour notre analyse. En effet, si nous nous permettons de généraliser, nous pouvons la considérer comme une campagne de prévention contre l'abus d'alcool, mais ciblée sur un individu en particulier.

➤ *avisfemme*

*Avisfemme* a été créée en multipliant les variables *avis* et *femme*. Ainsi, les femmes ayant fait l'objet d'une remarque au sujet de leur consommation d'alcool sont désignées par 1, les autres par 0.

### **3.1.3 VARIABLES CONCERNANT LA CONSOMMATION D'ALCOOL**

➤ *biere, vin et spirits*<sup>7</sup>

Une variable représentant la quantité consommée par chaque individu a été construite pour chaque type d'alcool.

Me basant sur les études de Manning (1997) et Kenkel et Terza (2001), j'ai multiplié la fréquence de consommation durant les 7 jours précédant l'enquête par la quantité moyenne consommée à chaque occasion<sup>8</sup>. Les chiffres ainsi obtenus peuvent alors être interprétés comme la consommation journalière moyenne pour chaque type d'alcool<sup>9</sup>. Ces variables nouvellement créées sont considérées comme continues.

---

<sup>7</sup> La variable relative à la consommation des spiritueux a été appelée « spirits » par commodité ; cela permet de ne pas surcharger les diverses commandes dans Stata.

<sup>8</sup> Notons que pour chaque question, fréquence ou quantité, quatre ou cinq réponses étaient proposées par l'enquêteur. Avant tout calcul, il fallut donc retravailler les données : exprimer la fréquence en nombre de jours et la quantité en nombre de décilitres (cf. Annexe 1 pour un exemple).

<sup>9</sup> Cf. Annexe 2 pour un aperçu de la distribution de ces variables



## 4 ANALYSE ECONOMETRIQUE

Avant d'aborder l'analyse économétrique proprement dite, il faut choisir les variables entrant dans chacune des équations du modèle (sélection et substantielle). Pour ce faire, je me suis basée sur mon intuition, mais également sur l'étude de Kenkel et Terza (2001) ainsi que sur diverses régressions et estimations Probit exploratoires que je ne rapporterai pas dans ce papier.

### 4.1 L'EQUATION DE SELECTION

Les variables expliquant la propension des gens à consommer un certain type d'alcool sont les suivantes :

- *revenu* : l'alcool étant généralement cher, plus une personne gagne aisément sa vie et plus sa probabilité de consommer sera grande => POSITIF
- *maitrise* : moins les gens pensent maîtriser leur vie, plus ils seront tentés par les boissons alcoolisées => POSITIF
- *sante* : si une personne est en mauvaise santé elle boira probablement moins ; cependant, une sorte de désespoir peut s'installer et cette personne cherchera alors à noyer son chagrin dans l'alcool => AMBIGU
- *femme* : il existe une certaine pression sociale sur les femmes, qui les pousse à ne pas consommer d'alcool ; en tenant compte de cette « culture » étrange, nous pourrions penser que le fait d'être une femme a une influence négative sur la propension à consommer ce genre de boisson => NEGATIF
- *age* : par intuition, nous aurions tendance à croire que les personnes âgées ne boivent plus d'alcool par ce que leur état de santé ne le permet pas, mais cette idée reste très vague => AMBIGU

Les variables susmentionnées sont ainsi supposées avoir uniquement une influence sur le choix de consommer ou non. Cependant, si l'on estime que certaines d'entre elles ont

également un impact sur la quantité moyenne consommée, celles-ci apparaîtront dans les deux équations du modèle.

## 4.2 L'EQUATION SUBSTANTIELLE

Venons-en à l'estimation de la quantité consommée, dont voici les variables explicatives choisies :

- *age* : plus une personne est âgée, plus elle consommera => POSITIF
- *age2* : la consommation augmente avec l'âge, mais de moins en moins => NEGATIF
- *avis* : rappelons que cette dummy vaut 1 lorsque l'individu a reçu une remarque quant à sa consommation d'alcool ; nous pourrions donc nous attendre à ce que cet incident ait un impact négatif sur la quantité consommée => NEGATIF
- *avisfemme* : les femmes, considérées comme plus réceptives que les hommes vis-à-vis des conseils qui leur sont donnés, devraient réagir à une éventuelle remarque de la part de leur entourage en réduisant leur consommation d'alcool => NEGATIF
- *mariage* : vivre en couple ne va pas influencer sur la décision de consommer de l'alcool, mais plutôt modérer la consommation des époux puisque chacun se sent « contrôlé » par l'autre => NEGATIF

## 4.3 REMARQUE SUR LES SIGNES ATTENDUS

Les signes proposés ci-dessus se rapportent à la consommation de l'alcool en général. Néanmoins, des interprétations différentes sont envisageables lorsqu'il s'agit de la bière, du vin, ou encore des spiritueux. Cette remarque deviendra évidente lorsque nous discuterons des résultats obtenus pour les différents types d'alcool.

## 4.4 EXEMPLE DE LA BIÈRE

Pour ne pas surcharger l'analyse, seul le modèle pour la consommation de bière sera présenté en détail. La procédure suivie étant exactement la même pour les deux autres types d'alcool, je me contenterai de reporter les résultats en annexe.

Dans Stata, il existe deux commandes pour l'estimation des modèles de sélection : « heckman » et « heckprob ». La première est utilisée lorsque la variable dépendante  $y$  est une variable continue, alors que la seconde ne fonctionne qu'avec une variable  $y$  dichotomique. Dans notre cas, « heckman » est la commande la plus appropriée.

### 4.4.1 PAR LA METHODE DU MAXIMUM DE VRAISEMBLANCE

Lorsque la commande « heckman » est utilisée sans option particulière, l'estimation du modèle se fait par la MMV.

```
. heckman biere age age2 avis avisfemme mariage, select(revenu maitrise sante
  femme age)
```

```
Iteration 0:  log likelihood = -12562.706
Iteration 1:  log likelihood = -12515.855
Iteration 2:  log likelihood = -12513.634
Iteration 3:  log likelihood = -12513.626
Iteration 4:  log likelihood = -12513.626
```

```
Heckman selection model                Number of obs    =    7947
(regression model with sample selection) Censored obs     =    4862
                                         Uncensored obs   =    3085

                                         Wald chi2(5)     =    61.33
Log likelihood = -12513.63              Prob > chi2      =    0.0000
```

```
-----+-----
          |          Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
biere    |
   age   |   .0165953   .0232452     0.71   0.475   - .0289643   .062155
  age2   |  -.0001796   .0002409    -0.75   0.456   - .0006518   .0002926
   avis  |   1.895771   .278665     6.80   0.000   1.349598    2.441945
avisfemme| -2.249413   .6574349    -3.42   0.001  -3.537961   -.9608638
mariage |  -.423427   .1300103    -3.26   0.001  -.6782424  -1.1686116
   _cons |   3.455578   .4948141     6.98   0.000   2.48576    4.425396
-----+-----
select   |
revenu  |  4.03e-06   8.11e-06     0.50   0.619  - .0000119   .0000199
maitrise|  .0963863   .0335497     2.87   0.004   .03063     .1621426
sante   | -.0574484   .0996545    -0.58   0.564  - .2527676   .1378709
femme   | -1.15133    .030623    -37.60   0.000  -1.21135   -1.091309
   age   | -.0101968   .0009661   -10.55   0.000  - .0120902  -.0083033
   _cons |   .6468363   .0585451    11.05   0.000   .53209     .7615826
```

/athrho		-.3523075	.0340565	-10.34	0.000	-.4190571	-.2855579
/lnsigma		1.206053	.014598	82.62	0.000	1.177441	1.234664
rho		-.3384203	.0301561			-.3961358	-.2780412
sigma		3.340274	.0487614			3.246058	3.437225
lambda		-1.130417	.1099351			-1.345886	-.9149478
LR test of indep. eqns. (rho = 0):				chi2(1) =	86.24	Prob > chi2 = 0.0000	

La convergence est rapide puisqu'elle a lieu après 4 itérations seulement. Avec une statistique de Wald de 61.33, le modèle estimé paraît bien spécifié : l'hypothèse  $H_0$  que tous les coefficients sont égaux à zéro est aisément rejetée.

Nous constatons que l'échantillon total est composé de 7'947 individus, comme prévu. L'estimation de l'équation de sélection se fait à l'aide de la totalité des observations, que les individus aient consommé de la bière ou non. Pour la deuxième étape, Stata utilise les « uncensored observations », c'est-à-dire uniquement les observations des individus ayant choisi de consommer ce type d'alcool.

Si nous nous intéressons à la significativité des coefficients, nous constatons que pour le modèle de sélection les variables *revenu* et *sante* ne sont pas significatives. Quant à l'équation substantielle, elle ne semble pas du tout influencée par l'âge. Une fois les variables non significatives retirées du modèle, nous obtenons l'estimation suivante :

```
. heckman biere avis avisfemme mariage, select(maitrise femme age)

Iteration 0:  log likelihood = -12558.146
Iteration 1:  log likelihood = -12516.032
Iteration 2:  log likelihood = -12514.228
Iteration 3:  log likelihood = -12514.223
Iteration 4:  log likelihood = -12514.223

Heckman selection model                               Number of obs    =       7947
(regression model with sample selection)              Censored obs     =       4862
                                                       Uncensored obs   =       3085

Log likelihood = -12514.22                            Wald chi2(3)     =        60.89
                                                       Prob > chi2      =        0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
biere						
avis	1.901252	.2784869	6.83	0.000	1.355428	2.447076
avisfemme	-2.256953	.6557754	-3.44	0.001	-3.542249	-.9716571
mariage	-.3996621	.1167427	-3.42	0.001	-.6284737	-.1708506
_cons	3.783764	.1264288	29.93	0.000	3.535968	4.03156

```

-----+-----
select |
  maitrise | .0940739 .0333805 2.82 0.005 .0286494 .1594985
  femme | -1.153466 .0304171 -37.92 0.000 -1.213083 -1.09385
  age | -.0102304 .0009517 -10.75 0.000 -.0120957 -.0083651
  _cons | .6644044 .047758 13.91 0.000 .5708005 .7580084
-----+-----
  /athrho | -.3530199 .0335226 -10.53 0.000 -.418723 -.2873168
  /lnsigma | 1.206319 .0145228 83.06 0.000 1.177855 1.234783
-----+-----
  rho | -.339051 .029669 -.3958541 -.2796633
  sigma | 3.341164 .048523 3.247401 3.437634
  lambda | -1.132825 .1082066 -1.344906 -.9207439
-----+-----
LR test of indep. eqns. (rho = 0): chi2(1) = 89.98 Prob > chi2 = 0.0000
-----+-----

```

Les signes des coefficients estimés sont semblables à ceux attendus pour la plupart des variables. Selon les résultats obtenus, plus une personne avance en âge et plus la probabilité qu'elle consomme de la bière diminue<sup>10</sup>. Dans la même idée, plus un individu jugera faiblement le niveau de contrôle qu'il détient sur sa vie, plus il sera probable qu'il consomme de la bière (peut-être dans le but de se redonner confiance...). Cette même probabilité sera d'autant plus faible si la personne est une femme (signe négatif).

Concernant la quantité consommée de cette boisson, nous constatons que contrairement à nos attentes, l'estimateur de la variable *avis* affiche un signe positif. L'interprétation est la suivante : le fait que quelqu'un ait émis des réserves sur la consommation d'alcool d'un individu va provoquer une augmentation de la quantité consommée de bière par cette même personne. Un tel comportement paraît bien étrange... Par la suite, il sera intéressant de vérifier si le coefficient de cette variable est également positif dans les autres modèles.

Revenons maintenant à la table des résultats et attardons-nous quelques instants sur les autres estimations fournies par Stata. Relevons tout d'abord l'estimation *lambda* qui correspond à l'inverse du Mill's ratio. Quant à *rho*, il s'agit bien sûr du coefficient de corrélation des termes d'erreur des deux équations du modèle.

---

<sup>10</sup> Les coefficients de l'équation de sélection s'interprètent comme ceux d'un modèle Probit. Quant aux coefficients de l'équation substantielle, ils représentent le changement de la variable dépendante suite à une modification de la variable explicative, c'est-à-dire un effet marginal.

Lorsque le modèle est estimé par la MMV, le logiciel effectue automatiquement un test du  $\chi^2$  pour vérifier si  $\rho$  est significativement différent de 0 ( $H_0: \rho=0$ ). Le rejet de l'hypothèse nulle ( $pvalue < 0.05$  si seuil critique à 5%) signifie que l'équation substantielle n'est pas indépendante de l'équation de sélection; les deux décisions ne sont pas prises indépendamment l'une de l'autre. C'est pourquoi nous pouvons affirmer qu'une estimation du modèle par les MCO aurait fourni des estimateurs biaisés. Le modèle expliquant la consommation de bière correspond à cette situation puisque la  $pvalue$  est égale à 0.

#### 4.4.2 PAR LA PROCEDURE EN DEUX ETAPES DE HECKMAN

Les estimateurs de Heckman sont obtenus avec la commande « heckman », à laquelle est ajoutée l'option « twostep ». Les valeurs obtenues par les deux méthodes d'estimation sont similaires et s'interprètent pareillement, même si la procédure de Heckman reste une approximation de la MMV.

```
. heckman biere age age2 avis avisfemme mariage, select(revenu maitrise sante
  femme age) twostep
```

```
Heckman selection model -- two-step estimates      Number of obs      =      7947
(regression model with sample selection)          Censored obs       =      4862
                                                    Uncensored obs     =      3085

                                                    Wald chi2(6)       =      166.89
                                                    Prob > chi2        =      0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
biere						
age	.0214984	.0229421	0.94	0.349	-.0234673 .066464	
age2	-.0001704	.0002371	-0.72	0.472	-.0006352 .0002944	
avis	1.688597	.2783698	6.07	0.000	1.143002 2.234192	
avisfemme	-1.565469	.64504	-2.43	0.015	-2.829724 -.3012134	
mariage	-.4785596	.1285535	-3.72	0.000	-.7305199 -.2265993	
_cons	4.050915	.4997552	8.11	0.000	3.071412 5.030417	
-----						
select						
revenu	6.73e-06	8.20e-06	0.82	0.412	-9.35e-06 .0000228	
maitrise	.0906369	.0340737	2.66	0.008	.0238537 .1574201	
sante	-.0449891	.1010664	-0.45	0.656	-.2430756 .1530974	
femme	-1.120177	.0309904	-36.15	0.000	-1.180917 -1.059437	
age	-.010254	.0009651	-10.63	0.000	-.0121455 -.0083625	
_cons	.6299842	.058805	10.71	0.000	.5147286 .7452399	
-----						
mills						
lambda	-2.061243	.1816688	-11.35	0.000	-2.417307 -1.705179	
-----						
rho	-0.57806					
sigma	3.5658187					
lambda	-2.0612432	.1816688				
-----						

Comme précédemment, un petit nettoyage s'impose... et voilà le modèle final !

```
. heckman biere avis avisfemme mariage, select(maitrise femme age) twostep
```

```
Heckman selection model -- two-step estimates      Number of obs      =      7947
(regression model with sample selection)          Censored obs       =      4862
                                                    Uncensored obs     =      3085

                                                    Wald chi2(3)       =      52.25
                                                    Prob > chi2        =      0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
biere						
avis	1.711558	.2778036	6.16	0.000	1.167073	2.256043
avisfemme	-1.660746	.6428939	-2.58	0.010	-2.920795	-.4006968
mariage	-.4027831	.1154699	-3.49	0.000	-.6290999	-.1764662
_cons	4.52698	.1749136	25.88	0.000	4.184156	4.869804
-----						
select						
maitrise	.0878416	.0339068	2.59	0.010	.0213855	.1542978
femme	-1.123437	.0307826	-36.50	0.000	-1.18377	-1.063104
age	-.0102545	.0009638	-10.64	0.000	-.0121436	-.0083654
_cons	.6576962	.0482897	13.62	0.000	.5630501	.7523423
-----						
mills						
lambda	-2.005841	.1750234	-11.46	0.000	-2.348881	-1.662802
-----						
rho	-0.56557					
sigma	3.5465885					
lambda	-2.0058412	.1750234				
-----						

D'une manière générale, les résultats sont très semblables : les variables éliminées sont les mêmes que pour l'estimation par la MMV et les signes des coefficients demeurent également inchangés. Les valeurs, elles, sont très similaires.

Une remarque peut être ajoutée quant à l'indépendance des 2 équations. En effet, aucun test n'est disponible ici pour vérifier si  $\rho$  est significativement différent de 0 ou non, comme avec la MMV. Toutefois, notons qu'un coefficient a été estimé pour l'inverse du Mill's ratio ( $\lambda$ ). Ce sera désormais la significativité de cet estimateur, et non plus de  $\rho$ , qui nous permettra de déterminer si un modèle de sélection permet d'obtenir des estimateurs plus efficaces qu'avec une simple estimation par les MCO.

## 4.5 COMPARAISON DES TROIS MODELES

Une présentation synthétique des coefficients estimés pour chaque type d'alcool est possible sur Stata8 grâce à la commande « estimates ». Celle-ci existait déjà dans l'ancienne version, mais ses fonctionnalités ont été améliorées.

Voici la procédure utilisée :

- (1) Estimation de chacun des trois modèles dans sa forme extensive et mise en mémoire immédiate de l'output :

```
. heckman biere age age2 avis avisfemme mariage, select(revenu maitrise
sante femme age) score(b*)11
. est store biere, title(consommation de bière)
. heckman vin age age2 avis avisfemme mariage, select(revenu maitrise
sante femme age) score(v*)
. est store vin, title(consommation de vin)
. heckman spirits age age2 avis avisfemme mariage, select(revenu maitrise
sante femme age) score(s*)
. est store spirits, title(consommation de spiritueux)
```

Liste des modèles estimés en mémoire :

```
. est dir
-----
model      | command  depvar      npar  title
-----+-----
      biere | heckman   biere       14  consommation de bière
      vin   | heckman   vin         14  consommation de vin
spirits    | heckman   spirits     14  consommation de spirirueux
-----
```

- (2) Demande de restitution des résultats dans un seul tableau (cf. Annexe 5) :

```
. est table biere vin spirits, star(0.1 0.05 0.01) stats(N N_cens rho p_c)
```

Nous pouvons procéder de la même manière pour les modèles finaux, c'est-à-dire ceux résultant de l'élimination des variables non significatives. Voici le tableau obtenu dans ce cas (après réarrangement des lignes pour une lecture facilitée) :

```
. est table biere1 vin1 spirits1, star(0.1 0.05 0.01) stats(N N_cens rho p_c)
```

---

<sup>11</sup> Cette option n'est pas vraiment nécessaire à cette étape ; toutefois, il s'agit là d'une anticipation pour pouvoir utiliser la commande « suest » par la suite (cf. 4.6).



Variable	biere1	vin1	spirits1
-----			
biere			
avis	1.9012522***		
avisfemme	-2.2569533***		
mariage	-.39966212***		
_cons	3.7837641***		
-----			
vin			
age		.03630546***	
age2		-.00028853***	
avis		.82007727***	
avisfemme		-.27154993*	
_cons		.15674143	
-----			
spirits			
age			-.0034383***
age2			.00003354***
avis			.0618149***
_cons			.20900142***
-----			
select			
maitrise	.09407391***		.13484722***
femme	-1.1534664***	-.06038813*	-.48175485***
age	-.01023038***	.01469016***	
revenu		.00009583***	.00003324***
sante		-.34320813***	
_cons	.66440442***	-.39551042***	-.76198642***
-----			
athrho			
_cons	-.35301987***	-.1531266***	-.27994186***
-----			
lnsigma			
_cons	1.2063193***	.0365541***	-2.2801498***
-----			
Statistics			
N	7947	7947	7947
N_cens	4862	2273	6284
rho	-.33905099	-.1519409	-.27285127
p_c	2.402e-21	.01501692	.00014227

legend: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01

Pour tous les modèles, supposer la décision de participation dépendante de la décision de consommation paraît tout à fait justifié. En effet, les *pvalues* du test de significativité de *rho* (*p\_c* dans le tableau ci-dessus) sont toujours inférieures à 5%.

Avec l'âge, les consommations de vin et de spiritueux évoluent en sens inverse. Alors que la première augmente mais de moins en moins, la consommation de spiritueux diminue de plus en plus. Pour les alcools forts, nous nous apercevons que les signes des coefficients de *age* et *age2* sont donc contraires à ceux attendus.

*Avisfemme* se comporte comme prévu pour le vin et la bière : une femme ayant reçu une remarque négative au sujet de sa consommation consommera moins qu'un homme célibataire

n'ayant reçu aucune remarque. Toutefois, il semblerait que l'impact sur la consommation du troisième type d'alcool ne soit pas significatif. Pour *mariage* aussi l'effet est conforme à nos attentes, mais pour la bière uniquement : une personne mariée consommera une quantité de bière significativement moins importante qu'un célibataire.

La variable *avis*, dans laquelle nous avons mis beaucoup d'espoir, s'avère significative dans tous les modèles, cependant son signe est positif<sup>12</sup> alors que nous nous attendions au signe contraire. Peut-être ce problème découle-t-il de la formulation imprécise de la question relative à cette variable (cf. 3.1.2) : les réserves émises par l'entourage ne peuvent en effet pas être datées précisément. D'autre part, il est aussi possible qu'une personne ayant reçu une telle remarque abuse réellement de l'alcool, ce qui expliquerait le signe positif du coefficient. Quelles que soient leurs caractéristiques socio-démographiques, ces individus consommeraient donc une quantité minimale d'alcool supérieure au reste de l'échantillon. Cette variable ne serait donc non pas exogène, mais endogène.

Dans l'équation substantielle, la constante estimée est toujours positive, mais non significative pour le vin. Elle peut s'interpréter comme la quantité minimale consommée pour le type d'alcool étudié. Pour la bière et les spiritueux, sa valeur est respectivement de 3.8 dl et 0.2 dl environ, ce qui correspond à peu près à la consommation d'un verre à chaque fois.

Voyons maintenant ce que nous suggère l'équation de sélection...

Comme prévu, le fait d'être une femme diminue la probabilité de consommer l'un des trois types d'alcool analysés. Pour la bière, cette probabilité est de 0.18<sup>13</sup> en moyenne pour les femmes alors que ce chiffre s'élève à 0.59 pour les hommes.

La propension à consommer de la bière ou des spiritueux paraît plus importante pour les gens qui détiennent un faible niveau de contrôle sur leur vie (signe positif des coefficients). Une mauvaise santé, quant à elle, aura seulement un impact négatif sur la probabilité de consommer du vin.

---

<sup>12</sup> Le fait que l'entourage ait émis des réserves au sujet de la consommation d'alcool semble augmenter la quantité consommée par l'individu.

<sup>13</sup> Option « psel » de predict, après le modèle restreint ; puis « sum [varname] if femme==1 ».

L'effet de l'âge est significativement négatif pour la bière, alors qu'il est positif pour le vin. Ainsi, un jeune aura une propension plus grande à consommer de la bière que du vin, ce qui correspond en général à la réalité. La richesse, de son côté, a un effet significativement positif sur la probabilité de consommer des spiritueux ainsi que du vin. Ce phénomène pourrait s'expliquer par la cherté relative de ces boissons par rapport à la bière.

## 4.6 ESTIMATION SIMULTANEE DES TROIS MODELES

Même si les modèles estimés ne considèrent qu'un type d'alcool à chaque fois, ils découlent tous d'un thème plus général qui est la consommation d'alcool. Il semble alors justifié d'imaginer qu'ils puissent être liés et donc que les termes d'erreur puissent être corrélés entre eux. La consommation de bière, par exemple, pourrait être partiellement expliquée par la consommation de spiritueux et vice-versa... Il existerait alors une corrélation entre ces deux modèles, nécessitant un ajustement des écarts-types.

Pour parer à une telle éventualité, nous pouvons estimer les trois modèles simultanément avec la commande « `suest` »<sup>14</sup>, qui signifie « *seemingly unrelated estimation* ». La valeur des coefficients estimés sera identique à celle trouvée par des estimations séparées ; seuls les écart-types, et par conséquent les *t* de Student, seront corrigés. Le logiciel essaie d'améliorer l'efficacité des coefficients estimés en tenant compte de possibles corrélation entre les consommations de bière, de vin et de spiritueux. Il calcule une matrice d'information pour les trois modèles simultanément et génère ainsi de nouveaux écarts-types.

Avant d'utiliser « `suest` », les trois modèles doivent être estimés séparément et conservés en mémoire. Reprenons l'exemple de la bière<sup>15</sup> :

```
. heckman biere age age2 avis avisfemme mariage, select(revenu maitrise
  sante femme age) score(b*)
. est store biere, title(consommation de bière)
```

---

<sup>14</sup> Cette commande n'est disponible que sur Stata8.

<sup>15</sup> Notez que l'option « `score([varname])` » doit être ajoutée lors de l'estimation afin de pouvoir utiliser plus tard la commande « `suest` ».

Les modèles choisis sont les modèles initiaux où toutes les variables explicatives apparaissent dans les équations.

```
. suest biere vin spirits
```

```
Simultaneous results for biere, vin, spirits
```

```
Obs = 7947
```

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
biere_biere						
age	.0165953	.0235953	0.70	0.482	-.0296506	.0628413
age2	-.0001796	.0002503	-0.72	0.473	-.0006702	.0003109
avis	1.895771	.4161186	4.56	0.000	1.080194	2.711349
avisfemme	-2.249413	.5649306	-3.98	0.000	-3.356656	-1.142169
mariage	-.423427	.1356277	-3.12	0.002	-.6892525	-.1576015
_cons	3.455578	.4796417	7.20	0.000	2.515498	4.395659
-----						
biere_select						
revenu	4.03e-06	8.06e-06	0.50	0.617	-.0000118	.0000198
maitrise	.0963863	.0339256	2.84	0.004	.0298933	.1628793
sante	-.0574484	.0995467	-0.58	0.564	-.2525563	.1376596
femme	-1.15133	.029881	-38.53	0.000	-1.209895	-1.092764
age	-.0101968	.0009718	-10.49	0.000	-.0121015	-.008292
_cons	.6468363	.0588977	10.98	0.000	.531399	.7622736
-----						
biere_atrho						
_cons	-.3523075	.0193778	-18.18	0.000	-.3902873	-.3143276
-----						
biere_lnsi~a						
_cons	1.206053	.0488213	24.70	0.000	1.110365	1.301741
-----						
vin_vin						
age	.0345126	.0052609	6.56	0.000	.0242014	.0448238
age2	-.0002719	.000053	-5.13	0.000	-.0003758	-.000168
avis	.8228508	.1245383	6.61	0.000	.5787602	1.066942
avisfemme	-.2696774	.1981188	-1.36	0.173	-.6579831	.1186283
mariage	.0288942	.0292942	0.99	0.324	-.0285213	.0863098
_cons	.1836879	.1147564	1.60	0.109	-.0412305	.4086063
-----						
vin_select						
revenu	.0000958	9.44e-06	10.15	0.000	.0000773	.0001143
maitrise	-.0124139	.0337567	-0.37	0.713	-.0785758	.053748
sante	-.339961	.0955019	-3.56	0.000	-.5271412	-.1527808
femme	-.0600682	.036416	-1.65	0.099	-.1314423	.0113058
age	.0146891	.0010196	14.41	0.000	.0126907	.0166875
_cons	-.3920111	.0623473	-6.29	0.000	-.5142095	-.2698127
-----						
vin_atrho						
_cons	-.1542831	.0296636	-5.20	0.000	-.2124227	-.0961435
-----						
vin_lnsigma						
_cons	.0365748	.0242071	1.51	0.131	-.0108703	.0840198
-----						
spirits_sp~s						
age	-.0030805	.0011626	-2.65	0.008	-.0053592	-.0008019
age2	.0000302	.0000113	2.68	0.007	8.09e-06	.0000524
avis	.0676836	.0219481	3.08	0.002	.0246661	.1107011
avisfemme	-.0412541	.0296745	-1.39	0.164	-.0994151	.0169069
mariage	-.0071647	.00506	-1.42	0.157	-.0170821	.0027528
_cons	.2026834	.0278677	7.27	0.000	.1480637	.2573032

```

-----+-----
spirits_se~t |
  revenu | .0000326 8.31e-06 3.92 0.000 .0000163 .0000489
  maitrise | .1381403 .0356042 3.88 0.000 .0683574 .2079232
  sante | -.1641061 .119663 -1.37 0.170 -.3986412 .070429
  femme | -.4800775 .0324922 -14.78 0.000 -.5437611 -.4163938
  age | .0006028 .0010187 0.59 0.554 -.0013939 .0025995
  _cons | -.7841544 .0613121 -12.79 0.000 -.9043239 -.6639848
-----+-----
spirits_at~o |
  _cons | -.2679771 .0427458 -6.27 0.000 -.3517574 -.1841968
-----+-----
spirits_ln~a |
  _cons | -2.283835 .077872 -29.33 0.000 -2.436461 -2.131209
-----+-----

```

Les coefficients estimés sont identiques à ceux des modèles séparés ; néanmoins, la valeur des écarts-types a été quelque peu modifiée. Ce changement reste relativement faible et ne provoque pas de grandes surprises au niveau de la significativité des variables, exception faite pour le modèle VIN. La variable *avisfemme* était significative à un seuil de 10% auparavant (*pvalue* = 0.086), mais ne l'est plus du tout maintenant ; et *femme* avait une *pvalue* de 0.71, alors qu'elle frôle dorénavant le seuil critique des 10% (*pvalue* = 0.099).

Grâce à l'estimation simultanée, nous pouvons désormais procéder à quelques tests d'hypothèses sur les coefficients des modèles BIERE, VIN et SPIRITS.

Vérifions d'abord si le modèle pour la consommation de bière est identique à celui pour la consommation de vin. A priori, nous penserions que cette hypothèse sera rejetée puisque les variables statistiquement significatives ne sont pas toujours les mêmes dans chaque cas.

```

. test [biere_biere]=vin_vin]

( 1) [biere_biere]age - [vin_vin]age = 0
( 2) [biere_biere]age2 - [vin_vin]age2 = 0
( 3) [biere_biere]avis - [vin_vin]avis = 0
( 4) [biere_biere]avisfemme - [vin_vin]avisfemme = 0
( 5) [biere_biere]mariage - [vin_vin]mariage = 0

      chi2( 5) =    40.42
      Prob > chi2 =    0.0000

```

A la vue de ce résultat, il est facile de conclure que les coefficients estimés pour le modèle BIERE ne peuvent pas être considérés égaux à ceux du modèle VIN. Ce qui confirme notre pressentiment. Si nous remplaçons le modèle VIN par SPIRITS, l'hypothèse nulle est à nouveau rejetée. Même chose pour une combinaison de VIN et SPIRITS (cf. Annexe 5).

Des tests similaires peuvent être effectués pour l'équation de sélection. D'après notre précédente analyse, il semblerait que le modèle de sélection soit différent pour chaque type d'alcool. Cela est-il confirmé ? Oui, car à chaque fois  $H_0$  est rejetée.

Parallèlement à ces hypothèses que je qualifierai de « globales », nous pouvons ensuite tester l'égalité des coefficients estimés pour l'équation substantielle. Voyons par exemple si les modèles BIERE et VIN ont des coefficients identiques pour la variable *avisfemme* :

```
. test ([biere_biere]avisfemme=[vin_vin]avisfemme)
( 1)  [biere_biere]avisfemme - [vin_vin]avisfemme = 0
      chi2( 1) =    10.81
      Prob > chi2 =    0.0010
```

La *pvalue* nous indique clairement que cette hypothèse doit être rejetée. Toutefois, rappelons qu'il existe deux hypothèses sous-jacentes dans ce genre de test : (1) signe du coefficient identique et (2) valeur absolue identique. Si  $H_0$  est rejetée, nous ne savons pas exactement quelle en est la raison. Ici, *avisfemme* apparaît avec le même signe dans chacun des deux modèles, mais rien de plus ne peut être affirmé.

De la même manière, d'autres hypothèses ont été testées, à la fois sur les équations de sélection et sur les équations substantielles. L'hypothèse d'égalité des coefficients de *maitrise* pour la bière et les spiritueux n'a par exemple pas pu être rejetée au seuil de 5% :

```
. test ([biere_select]maitrise=[spirits_select]maitrise)
( 1)  [biere_select]maitrise - [spirits_select]maitrise = 0
      chi2( 1) =    0.80
      Prob > chi2 =    0.3709
```

Ainsi, *age* et *age2* auraient le même effet sur la bière et le vin ; alors que les consommations de vin et de spiritueux seraient pareillement influencées par *avisfemme* et *mariage*. Par ailleurs, il n'est pas possible de rejeter l'hypothèse qu'une mauvaise santé affecte similairement la probabilité de consommer l'un des trois types d'alcool en question. D'autres interprétations seraient bien sûr possibles, cependant je préfère ne pas surcharger inutilement l'analyse qui me semble déjà très étendue et vous invite à consulter l'Annexe 6 pour un récapitulatif des résultats des différents tests.

## 5 CONCLUSION

Après une brève présentation théorique du modèle de sélection, diverses estimations ont été effectuées avec la commande « heckman », par la MMV et par la procédure en deux étapes de Heckman.

Trois modèles différents ont été estimés : un par type d'alcool (bière, vin et spiritueux). Nous avons constaté que la combinaison de variables statistiquement significatives était différente pour chaque modèle, même si les signes de leur coefficient se conformaient généralement à nos attentes. L'exception la plus marquante fut la variable *avis*. Nous pensions a priori que si quelqu'un avait déjà reçu une remarque négative quant à sa consommation d'alcool, cet incident aurait un impact négatif sur sa consommation actuelle. Or, il s'avère que pour chaque type d'alcool le coefficient est significativement positif. Peut-être est-ce dû à l'imprécision de la question posée lors de l'enquête ... Une codification différente de cette variable aurait peut-être été nécessaire.

Cependant, un tel phénomène pourrait tout simplement s'expliquer par le fait qu'une partie de l'échantillon abuse effectivement de bière, vin ou spiritueux. Peut-être qu'en observant l'évolution de la quantité consommée par ces individus sur un certain laps de temps, nous nous apercevions que le nombre de décilitres bus a toutefois diminué depuis l'incident. Un simple modèle de sélection ne réussit pas à expliquer cela. La variable *avis* ne nous permet donc finalement pas d'émettre un jugement sur l'efficacité des dépenses consacrées aux différentes campagnes d'information et de prévention contre l'alcool. La question reste ouverte...

Dans la dernière partie, une estimation simultanée des trois modèles a été effectuée afin d'obtenir les estimateurs les plus efficaces (à variance minimale). Une corrélation entre les trois modèles était en effet suspectée. Etant donné la faible ampleur de la modification des écarts-types et donc des *pvalues*, nous pourrions imaginer que des corrélations existent effectivement mais qu'elles semblent être peu importantes. Il aurait été intéressant d'obtenir une estimation des corrélations des trois modèles entre eux, cependant celle-ci aurait nécessité un travail plus approfondi encore... Je pense toutefois qu'il faut savoir s'arrêter un jour et ne pas vouloir résoudre tous les mystères de l'économétrie en même temps.

## 6 BIBLIOGRAPHIE ET REFERENCES

Heckman, James (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, 47: 153-162

Hosmer, David W., and Stanley Lemeshow (1989), "Applied Logistic Regression", Wiley Interscience Publication.

Institut suisse de prévention de l'alcoolisme et autres toxicomanies (2001), "Plan National d'Action Alcool 2000", version actualisée  
(disponible à l'adresse : [http://www.sfa-isp.ch/Politique/francais/akt\\_plan/naap\\_f.pdf](http://www.sfa-isp.ch/Politique/francais/akt_plan/naap_f.pdf))

Kenkel, Donald S., and Joseph V. Terza (2001), "The Effect of Physician Advice on Alcohol Consumption: Count Regression with an Endogenous Treatment Effect", *Journal of Applied Econometrics*, 16: 165-184.

Manning, W. G. (1997), "Alternative Econometric Models of Alcohol Demand" in Bryant, Kendall J., Michael T. Windle et Stephen G. West (1997), "The Science of Prevention: Methodological advances from alcohol and substance abuse research" (pp.101-121), Washington, DC: American Psychological Association.

Sigelman, Lee and Langche Zeng (1999), "Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models", *Political Analysis*, 8:2.

Stata Press (2003), "Stata Base Reference Manual".

Winship, Christopher, and Robert D. Mare (1992), "Models for Sample Selection Bias", *Annual Review of Sociology*, 18: 327-350.



# 7 ANNEXES

## 7.1 ANNEXE 1

### CREATION DES VARIABLES BIÈRE, VIN ET SPIRITS

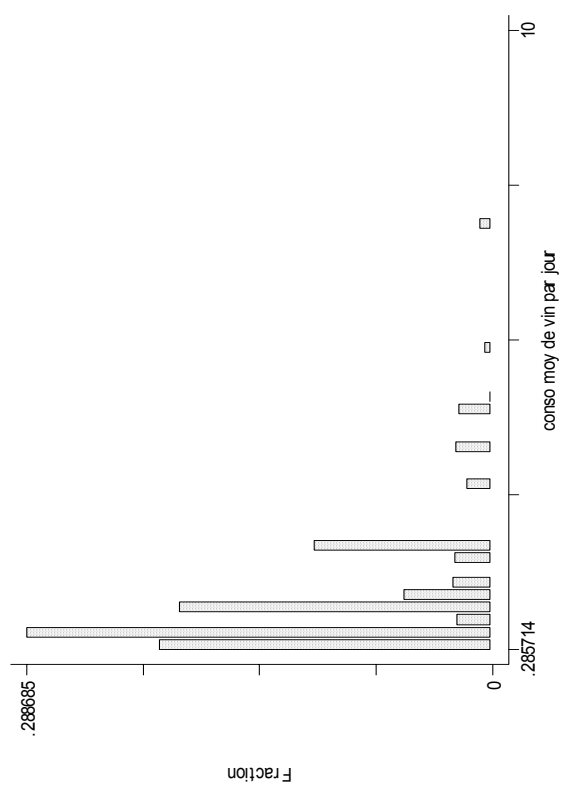
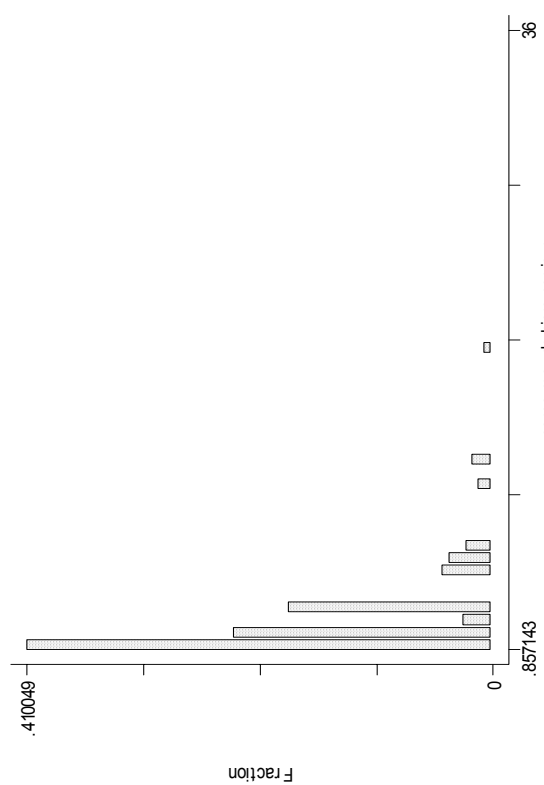
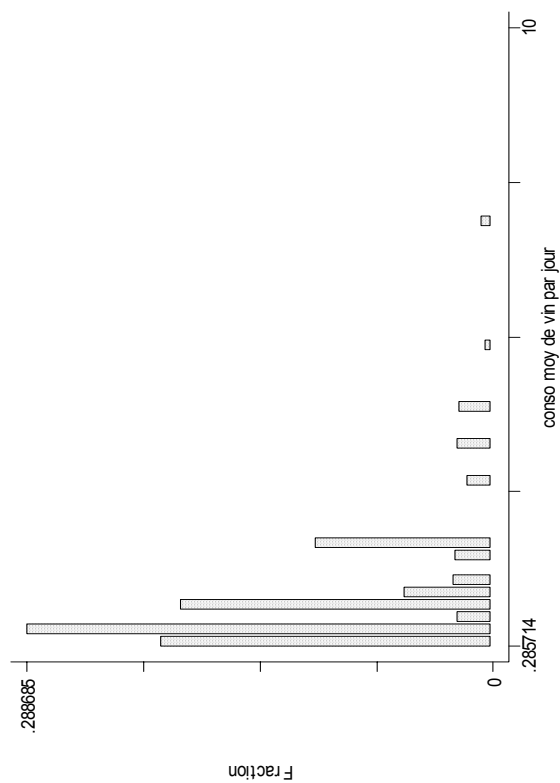
Pour chaque type d'alcool, les questions suivantes avaient été posées : (exemple de la bière)

- « Avez-vous bu de la bière au cours des derniers 7 jours ? »
  - (1) oui
  - (2) non
  
- « Combien de fois ? »                      Transformation de la variable
  - (1) 3 fois par jour ou plus                      => 3 jours
  - (2) 2 fois par jour                                      => 2 jours
  - (3) 1 fois par jour                                      => 1 jour
  - (4) presque tous les jours                      => 6/7 jour
  - (5) 3 ou 4 fois cette semaine                      => 4/7 jour
  - (6) 1 ou 2 fois cette semaine                      => 2/7 jour
  
- « En quelle quantité chaque fois ? »
  - (1) 5 verres / petites bouteilles de 3 dl ou 3 bouteilles de 6 dl ou plus                      => 18 dl
  - (2) 3-4 verres / petites bouteilles de 3 dl ou 2 bouteilles de 6 dl                      => 12 dl
  - (3) 2 verres / petites bouteilles de 3 dl ou 1 bouteille de 6 dl                      => 6 dl
  - (4) 1 verre / petite bouteille de 3 dl                      => 3 dl

Une fois ces transformations effectuées et afin d'obtenir la quantité journalière de bière consommée par chaque personne, il suffit de multiplier la fréquence journalière par la quantité consommée à chaque occasion. Le processus reste le même pour le vin et les spiritueux, seule l'échelle de quantité change (de 1 à 5 dl pour le vin et de 0.2 à 1.2 dl pour les spiritueux).

## 7.2 ANNEXE 2

### DISTRIBUTION DE BIÈRE, VIN ET SPIRITS



## 7.3 ANNEXE 3

### ESTIMATION DU MODELE INITIAL (MMV), POUR LES 3 TYPES D'ALCOOL

> BIERE

```
. heckman biere age age2 avis avisfemme mariage, select(revenu maitrise sante
femme age)
```

```
Iteration 0: log likelihood = -12562.706
Iteration 1: log likelihood = -12515.855
Iteration 2: log likelihood = -12513.634
Iteration 3: log likelihood = -12513.626
Iteration 4: log likelihood = -12513.626
```

```
Heckman selection model           Number of obs       =       7947
(regression model with sample selection)  Censored obs       =       4862
                                           Uncensored obs     =       3085
```

```
Log likelihood = -12513.63          Wald chi2(5)       =       61.33
                                           Prob > chi2       =       0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
biere						
age	.0165953	.0232452	0.71	0.475	-.0289643	.062155
age2	-.0001796	.0002409	-0.75	0.456	-.0006518	.0002926
avis	1.895771	.278665	6.80	0.000	1.349598	2.441945
avisfemme	-2.249413	.6574349	-3.42	0.001	-3.537961	-.9608638
mariage	-.423427	.1300103	-3.26	0.001	-.6782424	-.1686116
_cons	3.455578	.4948141	6.98	0.000	2.48576	4.425396
-----						
select						
revenu	4.03e-06	8.11e-06	0.50	0.619	-.0000119	.0000199
maitrise	.0963863	.0335497	2.87	0.004	.03063	.1621426
sante	-.0574484	.0996545	-0.58	0.564	-.2527676	.1378709
femme	-1.15133	.030623	-37.60	0.000	-1.21135	-1.091309
age	-.0101968	.0009661	-10.55	0.000	-.0120902	-.0083033
_cons	.6468363	.0585451	11.05	0.000	.53209	.7615826
-----						
/athrho	-.3523075	.0340565	-10.34	0.000	-.4190571	-.2855579
/lnsigma	1.206053	.014598	82.62	0.000	1.177441	1.234664
-----						
rho	-.3384203	.0301561			-.3961358	-.2780412
sigma	3.340274	.0487614			3.246058	3.437225
lambda	-1.130417	.1099351			-1.345886	-.9149478
-----						
LR test of indep. eqns. (rho = 0):	chi2(1) =	86.24	Prob > chi2 =	0.0000		
-----						

> VIN

```
. heckman vin age age2 avis avisfemme mariage, select(revenu maitrise sante  
femme age)
```

```
Iteration 0: log likelihood = -12925.924  
Iteration 1: log likelihood = -12808.691  
Iteration 2: log likelihood = -12797.414  
Iteration 3: log likelihood = -12797.062  
Iteration 4: log likelihood = -12797.062
```

```
Heckman selection model          Number of obs    =    7947  
(regression model with sample selection)  Censored obs      =    2273  
                                           Uncensored obs    =    5674
```

```
Log likelihood = -12797.06          Wald chi2(5)      =    207.77  
                                           Prob > chi2       =    0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
vin						
age	.0345126	.0057277	6.03	0.000	.0232865	.0457388
age2	-.0002719	.0000566	-4.80	0.000	-.0003829	-.0001609
avis	.8228508	.0823353	9.99	0.000	.6614767	.984225
avisfemme	-.2696774	.1568851	-1.72	0.086	-.5771665	.0378117
mariage	.0288942	.0295281	0.98	0.328	-.0289798	.0867682
_cons	.1836879	.135432	1.36	0.175	-.081754	.4491298
-----						
select						
revenu	.0000958	9.04e-06	10.60	0.000	.0000781	.0001135
maitrise	-.0124139	.0338227	-0.37	0.714	-.0787051	.0538773
sante	-.339961	.0941788	-3.61	0.000	-.5245481	-.1553739
femme	-.0600682	.0332337	-1.81	0.071	-.1252052	.0050687
age	.0146891	.0009687	15.16	0.000	.0127905	.0165877
_cons	-.3920111	.0603096	-6.50	0.000	-.5102157	-.2738066
-----						
/athrho	-.1542831	.04768	-3.24	0.001	-.2477342	-.0608319
/lnsigma	.0365748	.0100694	3.63	0.000	.0168391	.0563105
-----						
rho	-.1530705	.0465629			-.2427876	-.060757
sigma	1.037252	.0104445			1.016982	1.057926
lambda	-.1587726	.0489002			-.2546152	-.0629301
-----						
LR test of indep. eqns. (rho = 0):	chi2(1) =	6.08	Prob > chi2 =	0.0137		
-----						

> SPIRITUEUX

```
. heckman spirits age age2 avis avisfemme mariage, select(revenu maitrise
sante femme age)
```

```
Iteration 0: log likelihood = -2490.6904
Iteration 1: log likelihood = -2467.7091
Iteration 2: log likelihood = -2464.9077
Iteration 3: log likelihood = -2464.8519
Iteration 4: log likelihood = -2464.8518
```

```
Heckman selection model          Number of obs      =      7947
(regression model with sample selection)  Censored obs      =      6284
                                           Uncensored obs    =      1663
```

```
Log likelihood = -2464.852          Wald chi2(5)      =      55.98
                                           Prob > chi2      =      0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
spirits						
age	-.0030805	.0009349	-3.30	0.001	-.0049129	-.0012482
age2	.0000302	9.45e-06	3.20	0.001	.0000117	.0000487
avis	.0676836	.0114342	5.92	0.000	.0452729	.0900943
avisfemme	-.0412541	.0280368	-1.47	0.141	-.0962052	.013697
mariage	-.0071647	.0053237	-1.35	0.178	-.0175988	.0032695
_cons	.2026834	.022248	9.11	0.000	.1590781	.2462887
-----						
select						
revenu	.0000326	8.46e-06	3.86	0.000	.000016	.0000492
maitrise	.1381403	.0354787	3.89	0.000	.0686033	.2076773
sante	-.1641061	.1118687	-1.47	0.142	-.3833648	.0551526
femme	-.4800775	.0326975	-14.68	0.000	-.5441633	-.4159916
age	.0006028	.0010065	0.60	0.549	-.0013699	.0025755
_cons	-.7841544	.0615327	-12.74	0.000	-.9047562	-.6635525
-----						
/athrho	-.2679771	.0660093	-4.06	0.000	-.397353	-.1386011
/lnsigma	-2.283835	.0220013	-103.80	0.000	-2.326957	-2.240713
-----						
rho	-.2617415	.0614871			-.3776818	-.1377204
sigma	.1018927	.0022418			.0975923	.1063826
lambda	-.0266695	.0066437			-.0396909	-.0136482
-----						
LR test of indep. eqns. (rho = 0):	chi2(1) =	12.79	Prob > chi2 =	0.0003		
-----						

## ESTIMATION DU MODELE RESTREINT (MMV), POUR LES 3 TYPES D'ALCOOL

> BIERE

```
. heckman biere avis avisfemme mariage, select(maitrise femme age)
```

```
Iteration 0: log likelihood = -12558.146
Iteration 1: log likelihood = -12516.032
Iteration 2: log likelihood = -12514.228
Iteration 3: log likelihood = -12514.223
Iteration 4: log likelihood = -12514.223
```

```
Heckman selection model          Number of obs    =      7947
(regression model with sample selection)  Censored obs    =      4862
                                          Uncensored obs  =      3085

Log likelihood = -12514.22          Wald chi2(3)     =       60.89
                                          Prob > chi2     =       0.0000
```

```
-----+-----
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
biere						
avis	1.901252	.2784869	6.83	0.000	1.355428	2.447076
avisfemme	-2.256953	.6557754	-3.44	0.001	-3.542249	-.9716571
mariage	-.3996621	.1167427	-3.42	0.001	-.6284737	-.1708506
_cons	3.783764	.1264288	29.93	0.000	3.535968	4.03156
-----+-----						
select						
maitrise	.0940739	.0333805	2.82	0.005	.0286494	.1594985
femme	-1.153466	.0304171	-37.92	0.000	-1.213083	-1.09385
age	-.0102304	.0009517	-10.75	0.000	-.0120957	-.0083651
_cons	.6644044	.047758	13.91	0.000	.5708005	.7580084
-----+-----						
/athrho	-.3530199	.0335226	-10.53	0.000	-.418723	-.2873168
/lnsigma	1.206319	.0145228	83.06	0.000	1.177855	1.234783
-----+-----						
rho	-.339051	.029669			-.3958541	-.2796633
sigma	3.341164	.048523			3.247401	3.437634
lambda	-1.132825	.1082066			-1.344906	-.9207439
-----+-----						
LR test of indep. eqns. (rho = 0):	chi2(1) =	89.98	Prob > chi2 =	0.0000		
-----+-----						

> VIN

```
. heckman vin age age2 avis avisfemme, select(revenu sante femme age)
```

```
Iteration 0: log likelihood = -12907.661
Iteration 1: log likelihood = -12807.2
Iteration 2: log likelihood = -12797.89
Iteration 3: log likelihood = -12797.611
Iteration 4: log likelihood = -12797.611
```

```
Heckman selection model          Number of obs    =    7947
(regression model with sample selection)  Censored obs    =    2273
                                           Uncensored obs  =    5674
```

```
Log likelihood = -12797.61          Wald chi2(4)     =    206.50
                                           Prob > chi2     =    0.0000
```

```
-----+-----
            |          Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
vin
   age |   .0363055   .0054224     6.70  0.000   .0256777   .0469333
  age2 |  -.0002885   .0000054    -5.34  0.000  -.0003944  -.0001827
   avis |   .8200773   .0822964     9.96  0.000   .6587793   .9813753
avisfemme | -.2715499   .1568845    -1.73  0.083  -.5790378   .035938
   _cons |   .1567414   .1325312     1.18  0.237  -.1030149   .4164978
-----+-----
select
  revenu |   .0000958   9.02e-06    10.62  0.000   .0000781   .0001135
   sante |  -.3432081   .0938726    -3.66  0.000  -.527195   -.1592213
   femme |  -.0603881   .0332655    -1.82  0.069  -.1255874   .0048111
   age |   .0146902   .0009686    15.17  0.000   .0127918   .0165886
   _cons |  -.3955104   .0593504    -6.66  0.000  -.5118352  -.2791857
-----+-----
  /athrho |  -.1531266   .0479108    -3.20  0.001  -.24703   -.0592232
  /lnsigma |   .0365541   .0100651     3.63  0.000   .0168269   .0562813
-----+-----
   rho |  -.1519409   .0468047          - .2421248  -.0591541
  sigma |   1.03723   .0104398          1.016969  1.057895
 lambda |  -.1575977   .0491434          -.2539169  -.0612785
-----+-----
LR test of indep. eqns. (rho = 0):   chi2(1) =    5.91   Prob > chi2 = 0.0150
-----+-----
```

> SPIRITUEUX

```
. heckman spirits age age2 avis, select(revenu maitrise femme)
```

```
Iteration 0: log likelihood = -2498.6756
Iteration 1: log likelihood = -2471.5864
Iteration 2: log likelihood = -2468.1427
Iteration 3: log likelihood = -2468.0647
Iteration 4: log likelihood = -2468.0646
```

```
Heckman selection model          Number of obs    =    7947
(regression model with sample selection)  Censored obs    =    6284
                                          Uncensored obs  =    1663

                                          Wald chi2(3)    =    51.58
Log likelihood = -2468.065         Prob > chi2     =    0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
spirits						
age	-.0034383	.0008744	-3.93	0.000	-.0051521	-.0017245
age2	.0000335	8.96e-06	3.74	0.000	.000016	.0000511
avis	.0618149	.010521	5.88	0.000	.0411942	.0824356
_cons	.2090014	.0216497	9.65	0.000	.1665688	.251434
-----						
select						
revenu	.0000332	8.43e-06	3.94	0.000	.0000167	.0000498
maitrise	.1348472	.0353591	3.81	0.000	.0655446	.2041499
femme	-.4817548	.0326222	-14.77	0.000	-.5456931	-.4178166
_cons	-.7619864	.0427388	-17.83	0.000	-.845753	-.6782198
-----						
/athrho	-.2799419	.0648254	-4.32	0.000	-.4069972	-.1528865
/lnsigma	-2.28015	.0222065	-102.68	0.000	-2.323674	-2.236626
-----						
rho	-.2728513	.0599993			-.3859201	-.1517063
sigma	.1022689	.002271			.0979132	.1068183
lambda	-.0279042	.0065426			-.0407274	-.015081
-----						
LR test of indep. eqns. (rho = 0):	chi2(1) =	14.47	Prob > chi2 =	0.0001		
-----						



## 7.4 ANNEXE 4

TABLEAU RECAPITULATIF DE L'ESTIMATION INITIALE DE CHAQUE MODELE

. est table biere vin spirits, star(0.1 0.05 0.01) stats(N N\_cens rho p\_c)

Variable	biere	vin	spirits
-----			
biere			
age	.01659535		
age2	-.00017961		
avis	1.8957713***		
avisfemme	-2.2494126***		
mariage	-.42342703***		
_cons	3.4555782***		
-----			
vin			
age		.03451264***	
age2		-.00027192***	
avis		.82285084***	
avisfemme		-.2696774*	
mariage		.02889423	
_cons		.18368788	
-----			
spirits			
age			-.00308055***
age2			.00003022***
avis			.06768359***
avisfemme			-.0412541
mariage			-.00716467
_cons			.20268343***
-----			
select			
revenu	4.031e-06	.00009577***	.00003261***
maitrise	.09638629***	-.0124139	.1381403***
sante	-.05744835	-.339961***	-.16410613
femme	-1.1513295***	-.06006822*	-.48007745***
age	-.01019675***	.01468908***	.00060282
_cons	.64683631***	-.39201111***	-.78415436***
-----			
athrho			
_cons	-.35230746***	-.15428307***	-.26797707***
-----			
lnsigma			
_cons	1.2060529***	.03657476***	-2.2838351***
-----			
Statistics			
N	7947	7947	7947
N_cens	4862	2273	6284
rho	-.33842033	-.15307047	-.26174149
p_c	1.597e-20	.01368109	.00034931
-----			

legend: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01

Les lignes du tableau ont été partiellement réarrangées pour faciliter la lecture (déplacement de « select », « athrho » et « lnsigma » vers le bas du tableau).

## 7.5 ANNEXE 5

### TESTS D'EGALITE DE TOUS LES COEFFICIENTS : EQUATION SUBSTANTIELLE

```
. test [biere_biere=vin_vin]

( 1) [biere_biere]age - [vin_vin]age = 0
( 2) [biere_biere]age2 - [vin_vin]age2 = 0
( 3) [biere_biere]avis - [vin_vin]avis = 0
( 4) [biere_biere]avisfemme - [vin_vin]avisfemme = 0
( 5) [biere_biere]mariage - [vin_vin]mariage = 0

      chi2( 5) =    40.42
      Prob > chi2 =    0.0000

. test [biere_biere=spirits_spirits]

( 1) [biere_biere]age - [spirits_spirits]age = 0
( 2) [biere_biere]age2 - [spirits_spirits]age2 = 0
( 3) [biere_biere]avis - [spirits_spirits]avis = 0
( 4) [biere_biere]avisfemme - [spirits_spirits]avisfemme = 0
( 5) [biere_biere]mariage - [spirits_spirits]mariage = 0

      chi2( 5) =    32.69
      Prob > chi2 =    0.0000

. test [vin_vin=spirits_spirits]

( 1) [vin_vin]age - [spirits_spirits]age = 0
( 2) [vin_vin]age2 - [spirits_spirits]age2 = 0
( 3) [vin_vin]avis - [spirits_spirits]avis = 0
( 4) [vin_vin]avisfemme - [spirits_spirits]avisfemme = 0
( 5) [vin_vin]mariage - [spirits_spirits]mariage = 0

      chi2( 5) =   172.99
      Prob > chi2 =    0.0000

. test [biere_biere=vin_vin=spirits_spirits]

( 1) [biere_biere]age - [vin_vin]age = 0
( 2) [biere_biere]age2 - [vin_vin]age2 = 0
( 3) [biere_biere]avis - [vin_vin]avis = 0
( 4) [biere_biere]avisfemme - [vin_vin]avisfemme = 0
( 5) [biere_biere]mariage - [vin_vin]mariage = 0
( 6) [biere_biere]age - [spirits_spirits]age = 0
( 7) [biere_biere]age2 - [spirits_spirits]age2 = 0
( 8) [biere_biere]avis - [spirits_spirits]avis = 0
( 9) [biere_biere]avisfemme - [spirits_spirits]avisfemme = 0
(10) [biere_biere]mariage - [spirits_spirits]mariage = 0

      chi2( 10) =   209.50
      Prob > chi2 =    0.0000
```

## TESTS D'EGALITE DE TOUS LES COEFFICIENTS : EQUATION DE SELECTION

```
. test [biere_select=vin_select]

( 1) [biere_select]revenu - [vin_select]revenu = 0
( 2) [biere_select]maitrise - [vin_select]maitrise = 0
( 3) [biere_select]sante - [vin_select]sante = 0
( 4) [biere_select]femme - [vin_select]femme = 0
( 5) [biere_select]age - [vin_select]age = 0

      chi2( 5) = 911.97
      Prob > chi2 = 0.0000

. test [biere_select=spirits_select]

( 1) [biere_select]revenu - [spirits_select]revenu = 0
( 2) [biere_select]maitrise - [spirits_select]maitrise = 0
( 3) [biere_select]sante - [spirits_select]sante = 0
( 4) [biere_select]femme - [spirits_select]femme = 0
( 5) [biere_select]age - [spirits_select]age = 0

      chi2( 5) = 312.64
      Prob > chi2 = 0.0000

. test [vin_select=spirits_select]

( 1) [vin_select]revenu - [spirits_select]revenu = 0
( 2) [vin_select]maitrise - [spirits_select]maitrise = 0
( 3) [vin_select]sante - [spirits_select]sante = 0
( 4) [vin_select]femme - [spirits_select]femme = 0
( 5) [vin_select]age - [spirits_select]age = 0

      chi2( 5) = 222.24
      Prob > chi2 = 0.0000

. test [biere_select=vin_select=spirits_select]

( 1) [biere_select]revenu - [vin_select]revenu = 0
( 2) [biere_select]maitrise - [vin_select]maitrise = 0
( 3) [biere_select]sante - [vin_select]sante = 0
( 4) [biere_select]femme - [vin_select]femme = 0
( 5) [biere_select]age - [vin_select]age = 0
( 6) [biere_select]revenu - [spirits_select]revenu = 0
( 7) [biere_select]maitrise - [spirits_select]maitrise = 0
( 8) [biere_select]sante - [spirits_select]sante = 0
( 9) [biere_select]femme - [spirits_select]femme = 0
(10) [biere_select]age - [spirits_select]age = 0

      chi2( 10) = 944.13
      Prob > chi2 = 0.0000
```

## 7.6 ANNEXE 6

### RECAPITULATIF DES TESTS D'HYPOTHESE SUR LES COEFFICIENTS

BIERE	age	H <sub>0</sub> non rejetée
VIN	age2	H <sub>0</sub> non rejetée
	avis	H <sub>0</sub> rejetée
	avisfemme	H <sub>0</sub> rejetée
	mariage	H <sub>0</sub> rejetée
	revenu	H <sub>0</sub> rejetée
	maitrise	H <sub>0</sub> rejetée
	sante	H <sub>0</sub> rejetée
	femme	H <sub>0</sub> rejetée
	age	H <sub>0</sub> rejetée

BIERE	age	H <sub>0</sub> non rejetée
SPIRITS	age2	H <sub>0</sub> non rejetée
	avis	H <sub>0</sub> rejetée
	avisfemme	H <sub>0</sub> rejetée
	mariage	H <sub>0</sub> rejetée
	revenu	H <sub>0</sub> rejetée
	maitrise	H <sub>0</sub> non rejetée
	sante	H <sub>0</sub> non rejetée
	femme	H <sub>0</sub> rejetée
	age	H <sub>0</sub> rejetée

VIN	age	H <sub>0</sub> rejetée
SPIRITS	age2	H <sub>0</sub> rejetée
	avis	H <sub>0</sub> rejetée
	avisfemme	H <sub>0</sub> non rejetée
	mariage	H <sub>0</sub> non rejetée
	revenu	H <sub>0</sub> rejetée
	maitrise	H <sub>0</sub> rejetée
	sante	H <sub>0</sub> non rejetée
	femme	H <sub>0</sub> rejetée
	age	H <sub>0</sub> rejetée

BIERE	age	H <sub>0</sub> rejetée
VIN	age2	H <sub>0</sub> rejetée
SPIRITS	avis	H <sub>0</sub> rejetée
	avisfemme	H <sub>0</sub> rejetée
	mariage	H <sub>0</sub> rejetée
	revenu	H <sub>0</sub> rejetée
	maitrise	H <sub>0</sub> rejetée
	sante	H <sub>0</sub> non rejetée
	femme	H <sub>0</sub> rejetée
	age	H <sub>0</sub> rejetée

Le seuil critique considéré est de 5%.

Pour l'équation de sélection, une hypothèse nulle non rejetée signifie que le signe du coefficient de la variable considérée est le même pour les modèles en questions.