



UNIVERSITÉ DE NANTES



IAE NANTES
ÉCONOMIE & MANAGEMENT

Variables Qualitatives 2 :
Etude sur les déterminants du respect des gestes barrières
et du maintien des habitudes sociales

Modèles Biprobit & Biprobit Récursif

Mounier-Tebas Teodoro, Aurouet Lucas

Master EKAP
IAE Nantes
01/10/2020

Résumé

L'analyse suivante consiste à étudier les déterminants qui incitent les étudiants de l'IAE de Nantes à respecter les gestes barrières depuis la fin du confinement tout en s'intéressant au phénomène de modification des habitudes sociales des étudiants durant cette même période. L'étude est construite à partir d'un sondage réalisé auprès des étudiants afin de récupérer l'information nécessaire à la construction des variables du modèle. La première partie consiste à expliquer le choix de nos deux variables à expliquer étant, le respect des gestes barrière et la modification des habitudes sociales. Mais aussi, à justifier et sélectionner l'ensemble des variables qui nous permettront d'expliquer ou du moins de mieux comprendre ces deux décisions. Le langage R a été utilisé pour la manipulation des données et Stata, pour la modélisation économétrique. Parmi les quatorze variables explicatives seulement huit d'entre elles ont été retenues pour notre modèle définitif. Dans la seconde partie, nous estimons quatre modèles Biprobit, dont deux après une sélection de variables par la méthodologie stepwise. Ensuite nous passons par l'estimation d'un modèle Biprobit récursif qui n'aboutira à aucune solution. Le modèle retenu est le modèle 'partiel'. Malgré une qualité de prédiction légèrement inférieure, la répartition des modalités correctement prédites est meilleure. Le modèle ne permettra pas de prédire la modalité (1,0) qui correspond aux étudiants qui respectent les gestes barrières mais n'ont pas vu leurs habitudes sociales changer. Enfin, nous nous assurons de la validité de notre modèle en mesurant sa représentativité et sa robustesse.

Mots clé : Biprobit, Biprobit récursif, Stepwise, Covid-19.

Table des matières

I	Introduction	4
II	Obtention de la base de données	8
II.1	Création des variables	8
II.1.1	Questionnaire	8
II.1.2	Variables expliquées	9
II.1.3	Variables explicatives	10
II.2	Analyse de la base de données	12
II.2.1	Manipulations	12
II.2.2	Détection des Outliers	12
II.2.3	Statistiques descriptives	13
II.2.4	Dépendance et corrélation entre les variables Khi-2	16
III	Méthodologie économétrique	17
III.1	Introduction	17
III.2	Estimations sans pré-sélection de variables explicatives	18
III.2.1	Hypothèses du modèle Biprobit	18
III.2.2	Estimation du Modèle ‘complet’	19
III.2.3	Estimation du Modèle ‘partiel’	20
III.2.4	Modèle retenu	21
III.3	Estimations avec pré-sélection de variables explicatives	23
III.3.1	Spécification du modèle	23
III.3.2	Algorithme Stepwise	24
III.3.3	Algorithme Stepwise sans Sexe	24
III.3.4	Choix du modèle définitif	25
III.4	Estimation d’un Biprobit récursif	28
III.4.1	Hypothèses d’un modèle récursif	28
III.4.2	Estimation du modèle 5 Biprobit récursif	28
IV	Présentation des résultats	29
IV.1	Introduction	29
IV.2	Interprétation des résultats	32
IV.3	Validité de notre modèle	35
IV.3.1	Représentativité	35
IV.3.2	Robustesse	36
V	Conclusions	37
VI	Annexes	38

I Introduction

À l'heure où nous réalisons notre étude, le coronavirus a déjà fait plus de 32.583 décès¹ en France, 529 dans les pays de la Loire, 185 en Loire-Atlantique. Le virus **SARS-CoV-2** initialement identifié en Chine en janvier 2020, donne la maladie du COVID-19, nom choisi par l'Organisation Mondiale de la Santé (OMS). Nous sommes, depuis le 11 mars 2020, face à une pandémie c'est-à-dire une épidémie au niveau mondial. Le COVID-19 est un virus dangereux car il est très contagieux et les symptômes se développent tardivement ce qui facilite une transmission très rapide.

Les **symptômes** provoqués par le virus sont ; fièvre, toux, maux de tête et courbatures ainsi qu'une fatigue inhabituelle. Une perte brutale de l'odorat ainsi qu'une disparition du goût peuvent aussi faire partie des symptômes. Bien entendu dans les cas les plus graves les difficultés respiratoires peuvent mener jusqu'au décès.

En ce qui concerne **la transmission**² de la maladie, quatre formes de transmission sont explicitées sur le site du gouvernement. La première étant : *"La projection de gouttelettes (comme les postillons) contaminées par une personne porteuse : en toussant, éternuant ou en cas de contacts étroits en l'absence de mesures de protection (distance physique, mesures barrières, port du masque). Les gouttelettes contaminées sont inhalées par la personne saine, et déclenchent la maladie"*. Le deuxième type de transmission peut-être *"par contact direct physique (poignées de main, accolade, bises...) entre une personne porteuse et une personne saine. Le virus est ensuite transmis à la personne saine quand elle porte ses mains à la bouche."* Nous avons la transmission *"par contact indirect, via des objets ou surfaces contaminées par une personne porteuse. Le virus est ensuite transmis à une personne saine qui manipule ces objets, quand elle porte ses mains à la bouche."* Et enfin celle-ci peut aussi se faire *"par l'air, essentiellement dans un espace confiné."*

Le masque s'est imposé comme *"une mesure de protection essentielle, notamment en milieu fermé ou lorsqu'il y a une forte densité de population (marché, rue fréquentée...)." En complément des gestes barrières (lavage des mains, distanciation physique), c'est le masque qui filtre les gouttelettes issues des éternuements, de la toux, voire d'une discussion prolongée."* Selon le Ministère des Solidarités et de la Santé. Le port du masque a donc été rendu obligatoire dans tous les établissements recevant du public (ERP) (ex : magasins de vente, centres commerciaux, établissements sportifs couverts, musées, ...). Bien entendu s'y ajoutent les gares routières et maritimes ainsi que les aéroports, les établissements scolaires et les transports en commun. Le non-respect de cette mesure peut être sanctionné, par une contravention de 4^e classe passible de 135 € d'amende.

1. <https://www.gouvernement.fr/info-coronavirus/carte-et-donnees> (consulté le 11/10/2020)

2. <https://www.gouvernement.fr/info-coronavirus/comprendre-la-covid-19> (consulté le 11/10/2020)

En ce qui concerne l'**éducation**, la rentrée scolaire a bien eu lieu en septembre 2020. C'était une urgence sociale et un impératif pédagogique selon le ministre de l'éducation nationale Jean-Michel Blanquer. L'organisation des classes et les règles de distanciation seront adaptées à l'évolution du virus, c'est ce qu'a précisé la circulaire de rentrée. Pour les universités, les salles de classe ou les amphithéâtres, ils ne pourront être remplis qu'à 50% de leur capacité au maximum, en zone d'alerte renforcée et maximale.

Au sujet de la **vie sociale**³ et notamment sportive le gouvernement rappelle que : *"le sport multiplie les occasions de contact, et donc les risques de diffusion de la Covid-19."* Mais pour les rassemblements familiaux dans les lieux privés, il n'y a pas d'obligation de porter le masque, même si cela est recommandé, afin de porter une plus grande attention aux plus fragiles, et notamment aux personnes âgées.

Le port du masque, ainsi que le lavage des mains au gel hydroalcoolique et la distanciation sociale, font partie des gestes barrières contre la propagation du virus. Gestes barrières qui ont mis du temps à intégrer le quotidien. Le masque est passé, d'une précieuse ressource, d'abord en pénurie, puis a été considéré comme inutile pour finir indispensable, aussi bien pour l'OMS que pour les gouvernements occidentaux. Le graphique ci-dessous représente la **chronologie**⁴ des revirements du gouvernement Français au sujet du port du masque obligatoire.



3. <https://www.gouvernement.fr/info-coronavirus> (consulté le 11/10/2020)

4. <https://www.ledauphine.com/sante/2020/08/03/chronologie> (consulté le 11/10/2020)

En France, deux enquêtes à propos du port du masque obligatoire ont été réalisées. L'une par **Harris Interactive**⁵ qui est une entreprise d'étude marketing et de sondage d'opinion. La première a été réalisée en août 2020 sur un échantillon de 1504 personnes, et avait pour objectif d'identifier comment les français percevaient le masque et dans quelle mesure et dans quelles conditions ils le portaient. Parmi les conclusions nous pouvons retenir que le port du masque est moins systématique chez les hommes, que 93 % des Français qui portent un masque ne l'utilisent pas dans de bonnes conditions et enfin que 51 % des Français indiquent faire confiance au gouvernement pour éviter un nouveau confinement.

La seconde étude a été réalisée par l'**Institut d'études opinion et marketing en france et à l'international (Ifop)**⁶. L'échantillon cette fois-ci était de 1009 personnes représentatives de la population française. On peut retenir de cette enquête que l'opinion en faveur de l'obligation du port du masque dans les espaces publics ouverts est une mesure moins soutenue par les jeunes, considérant comme jeune toutes personnes de moins de 35 ans. Ces jeunes sont également moins enclins à s'isoler en cas d'apparition des symptômes du COVID-19.

Le masque, la distanciation sociale et les gestes barrières, de manière générale, ont bouleversé nos relations sociales depuis la fin du confinement. Nous avons donc souhaité comprendre pourquoi certains respectaient les gestes barrières et d'autres non et si cela avait un impact sur les relations sociales. Étant étudiants, et ayant la possibilité de réaliser une enquête dont l'échantillon comprendrait principalement des étudiants de l'**Université de Nantes**, nous avons donc décidé de focaliser notre étude sur les étudiants. En effet, pour qu'une analyse économétrique ait une quelconque utilité il est important que l'échantillon à partir duquel les estimations sont réalisées soit représentatif de la population étudiée. Notre sujet consistera donc à étudier les facteurs influençant le fait de respecter les gestes barrières sans changer pour autant leurs habitudes en terme de relations sociales, chez les étudiants, depuis la fin du confinement.

Il est aisé de comprendre ce que signifie le respect des gestes barrières. Les gestes barrières consistent au respect du port du masque, de la distanciation sociale, ainsi que du lavage des mains avec du gel hydroalcoolique. Il est important de préciser que notre analyse ne s'intéresse pas au respect des gestes barrières dans les lieux où leur application est obligatoire sous peine de sanction financière. Notre réflexion s'intéresse au respect des gestes barrières dans la sphère privée des étudiants, hors famille, où il est recommandé de respecter ces gestes barrières mais où il est difficile de vérifier s'ils sont effectivement respectés, comme par exemple dans les soirées entre amis. En ce qui concerne le changement des habitudes en terme de relations sociales, nous souhaitons mesurer si le Covid-19 a eu un impact sur la fréquence des activités extra scolaire, c'est-à-dire des activités tel que le sport, les soirées entre amis, les engagements dans une association, la vie sociale étudiante d'une manière générale.

5. <https://harris-interactive.fr/opinion> (consulté le 14/10/2020)

6. <https://www.ifop.com/wp-content/uploads/2020/08/117539-Rapport-JDD-07.08.2020.pdf> (consulté le 14/10/2020)

La démarche que nous allons suivre consistera à répondre à la problématique suivante : **Quels sont les déterminants du respect des gestes barrières ainsi que de la modification des habitudes sociales chez les étudiants ? Y-a-t-il un lien entre les deux ?** Dans un premier temps nous procéderons à une justification du choix des variables que nous allons utiliser pour répondre à notre problématique. Puis dans un second temps nous démontrerons la méthodologie d'obtention et de nettoyage de la base ainsi qu'à une brève analyse des statistiques descriptives. Étant donné la multitude de méthodes d'analyse économétrique, nous avons décidé de procéder par deux méthodes distinctes de sélection des variables, puis de comparer les résultats. La première méthode consistera à sélectionner les variables seulement à partir de l'étude des dépendances entre elles, puis réaliser l'estimation par un Biprobit à partir de variables indépendantes et qui nous semblent pertinentes. La seconde méthode consistera à sélectionner les variables à l'aide de la méthode Stepwise, de manière automatique, puis de réaliser l'estimation via un Biprobit à partir de variables sélectionnées. Nous terminerons notre étude par la présentation des résultats du meilleur modèle sélectionné. Dans un objectif de validation de notre modèle nous avons fait le choix de conserver 20% de notre base de données afin de pouvoir tester la robustesse de notre modèle sur un échantillon test.

II Obtention de la base de données

II.1 Création des variables

II.1.1 Questionnaire

Nous avons choisi de réaliser une enquête pour obtenir nos données pour deux raisons principales. D'une part, le phénomène que l'on souhaite étudier est très récent ce qui signifie qu'il existe très peu de bases de données permettant de récupérer des informations sur le sujet. D'autre part, notre étude a pour cible les étudiants, c'est à dire un échantillon sur une population ciblée. Or les bases de données déjà existantes sont réalisées à l'échelle nationale. Cette méthode a, de plus, l'avantage d'obtenir une base de données contenant exactement l'information (en terme de variables dépendantes) que nous souhaitions pour la réalisation de notre étude.

Nos données ont donc été obtenues à l'aide d'un questionnaire envoyé à tous les étudiants de l'IAE. Ce questionnaire comportait 23 questions réparties en 4 sections de la manière suivante :

1. À propos du Covid-19 ?

Cette section comportait deux questions, l'une à propos de la gestion de la crise par le gouvernement actuel avec une réponse possible par une échelle linéaire allant de 1 à 10. Cette question consistait à récupérer la variable **Politique**. L'autre question consistait à classer les individus attentifs aux informations concernant le covid-19 provenant des médias et des réseaux sociaux. La variable **News** fut créée à partir de cette question.

2. Impact du Covid sur ta vie sociale ?

Cette section avait pour but de récupérer l'information concernant le respect des gestes barrières, ainsi que le changement des habitudes sociales des étudiants. Nous avons fait le choix de poser quatre questions pour chaque variable à expliquer afin de s'assurer des fréquences suffisantes si nous faisons face à une répartition inégale de la population dans les différentes modalités. Les questions à propos du port du masque, de la fréquence du lavage des mains au gel hydroalcoolique ou encore du respect de la distanciation sociale, en fait partie des questions pour la construction de la variable **Gbarriere**. En ce qui concerne la variable à expliquer **Rsocial** qui identifie les modifications dans les relations sociales regroupait des questions liées à la fréquentation des amis, à la pratique d'une activité sportive ou associative et enfin aux habitudes de soirée en général.

3. À propos de ta Famille ?

Cette section avait pour but de regrouper les questions dont l'information récupérée concernait la famille de l'individu. Les trois questions étaient au sujet de la fréquentation des grands parents (**GParents**), si un membre de la famille était à considérer comme personne à risque (**PRisque**) ou encore si l'un des parents appartenait au corps médical, (**ParentMed**).

4. À propos de toi ?

La dernière section avait pour objectif de récupérer l'information générale sur le répondant comme ; son genre pour la variable **Sexe**, son âge **Age**, son niveau d'étude avec le choix entre Licence, Master ou Doctorat pour la variable **Etude**. Ensuite nous avons souhaité connaître la situation financière de l'individu à l'aide de proxys tels que le **Logement**, le fait que l'étudiant soit **Boursier** ou non ou encore s'il a un **Emploi** en dehors de ses études. Pour finir nous avons posé des questions sur la **Sante** de l'étudiant et sur son **Stress** vis-à-vis du Covid-19.

II.1.2 Variables expliquées

1. Y1 : Le respect des gestes barrières

La variable Y_1 , le respect des gestes barrières prend une valeur de 1 si l'individu respecte les gestes barrières et de 0 sinon. Cette variable a été construite d'une manière particulière que nous détaillerons ici. Dans notre questionnaire nous avons utilisé un total de 4 questions pour déterminer si un individu respecte les gestes barrières. La première est une question à choix multiples (3 choix) dans laquelle l'individu doit indiquer les mesures qu'il ou elle respecte. Les choix sont : distance de sécurité, lavage des mains, port du masque et aucune. Nous avons construit notre variable de telle sorte qu'un individu doive respecter les 3 consignes pour être considéré comme respectant les gestes barrières. Un individu qui ne respecte qu'une ou deux mesures sera classé comme ne respectant pas les gestes barrières. Dans un souci de représentativité et n'ayant pas encore obtenu les réponses, nous avons pris la décision d'ajouter des questions supplémentaires au cas où les effectifs dans chaque catégorie seraient trop hétérogènes. Ces questions nous permettent d'avoir une idée plus claire de la composition de notre échantillon en terme du respect des mesures sanitaires et ainsi d'ajuster au besoin nos critères afin d'obtenir des effectifs suffisants à l'utilisation d'un modèle Biprobit. Les questions supplémentaires concernent le respect des 3 consignes citées plus haut. Nous avons demandé aux répondants à quelle fréquence se lavaient-ils les mains (sur une échelle de 1 à 10, par jour), s'ils respectaient les distances de sécurité en tout temps, seulement quand c'était obligatoire ou bien jamais. Et enfin s'ils portaient le masque en tout temps, avec des amis ou jamais. A la réception des réponses, nous avons observé que notre découpage était suffisant pour obtenir des effectifs relativement homogènes. nous avons donc décidé de garder cette stratégie pour la construction de Y_1 , à savoir, $Y_1 = 1$ si l'individu respecte les 3 mesures sanitaires, sinon 0.

2. **Y2 : Changement des habitudes en terme de relations sociales**

Pour ce qui concerne Y_2 , le bouleversement des habitudes sociales, nous avons suivi la même logique, avec une variable binaire qui prend en compte si l'individu a choisi de changer ses habitudes en termes de relations sociales depuis la fin du confinement ($Y_2 = 1$, 0 sinon). Puis nous avons ajouté des questions supplémentaires pour nous assurer de pouvoir avoir des effectifs suffisants dans chaque catégorie. Les questions supplémentaires sont au nombre de 3 et nous renseignent sur la pratique d'une activité sportive ou associative, sur les habitudes de soirées et sur la fréquence à laquelle un individu choisit de voir ses amis. Pour chacune d'entre elles l'individu devait choisir s'il estimait que le Covid-19 avait eu un impact significatif ou bien s'il avait conservé ses habitudes. Tout comme notre Y_1 nous avons constaté, à la réception des réponses, que les questions supplémentaires n'étaient pas nécessaires puisque la première nous donnait déjà une répartition très correcte (41% de non contre 59% de oui). Nous avons donc recodé cette variable en 1 si l'individu considère que le Covid-19 a significativement changé ses habitudes de socialisation et en 0 sinon.

II.1.3 Variables explicatives

1. Les Caractéristiques individuelles de l'individu :

- X1 : Sexe = 1 si l'étudiant est une femme, 0 sinon. Notre intuition est de considérer que les hommes sont moins susceptibles de respecter les gestes barrières et de changer leurs habitudes sociales.
- X2 : Âge, âge de l'étudiant. Cette variable nous permet de vérifier si l'âge a un impact positif sur nos variables expliquées.
- X3 : Logement = 1 si l'étudiant habite chez les parents pendant la semaine, 0 sinon. Nous pensons que si l'étudiant vit seul ou en collocation, et voit rarement ses parents, il anticipera moins la possibilité de transmettre le virus à ses proches et serait donc moins susceptible de respecter les gestes barrières ou de changer ses habitudes par méfiance.
- X4 : Boursier = 1 si l'étudiant est boursier, 0 sinon. Nous nous servons de cette variable pour déterminer si le revenu a un impact sur nos variables expliquées (le statut de boursier étant lui aussi un proxy, bien qu'imparfait, du niveau de revenu).
- X5 : Etude = 1 si l'étudiant est en Licence, 0 si en Master. Nous pensons que le niveau d'étude et le temps passé à l'université peut agir positivement sur le respect des gestes barrières.
- X6 : Santé = 1 si l'étudiant se considère à risque (problème respiratoire, obésité, etc...), 0 sinon. Un individu avec des problèmes de santé est plus susceptible

de développer des complications au contact du virus et est supposément plus attentif au respect des gestes barrières.

- X7 : Stress = 1 si l'étudiant est généralement stressé, 0 sinon. un niveau de stress plus élevé peut conduire à une attention accrue aux mesures d'hygiène.
- X8 : Covid = 1 si l'étudiant a eu le covid, 0 sinon. Le taux de ré-infection ⁷ du Covid-19 étant relativement faible, un individu l'ayant déjà contracté sera probablement moins soucieux de respecter les mesures sanitaires.

2. Dimension familiale de l'individu :

- X9 : GParents = 1 si l'étudiant voit ses grands parents au moins une fois par mois, 0 sinon. Nous savons que le Covid-19 peut entraîner des complications fréquentes chez les personnes âgées. Cette variable reprend la même logique que 'Logement'.
- X10 : PRisque = 1 si l'étudiant a un membre de sa famille à risque (hors GP), 0 sinon.
- X11 : ParentMed = 1 si au moins l'un des parents de l'étudiant travaille dans le domaine médical. Avoir un proche en contact du virus peut potentiellement entraîner une plus grande précaution et un respect plus strict des gestes barrières et des consignes en terme d'hygiène.

3. Dimension sociale de l'individu :

- X12 : News = 1 si l'étudiant suit la propagation du Covid-19 dans les médias. Tout comme le Stress, le suivi actif des informations peut entraîner une méfiance accrue et être positivement corrélé avec la probabilité de changer ses habitudes sociales et de respecter les gestes barrières.
- X13 : Politique = une échelle de 1 à 10 sur laquelle l'étudiant indique sa satisfaction vis-à-vis de la gestion de la crise par le gouvernement actuel (0 mauvaise, 10 parfaite). Nous pensons que cette variable peut être déterminante aux vues des récentes polémiques et déboires autour du port du masque et de la fermeture réglementée de certains établissements.
- X14 : Emploi = 1 si l'étudiant occupe un emploi salarié en dehors des cours, 0 sinon. L'inclusion de cette variable est justifiée par notre expérience personnelle. Nous pensons qu'être salarié peut désensibiliser l'individu et le pousser à être moins attentif aux mesures d'hygiène.

7. personnes testées positives au virus après l'avoir contracté une première fois auparavant

II.2 Analyse de la base de données

II.2.1 Manipulations

Nous avons récolté 555 réponses de la part des étudiants de l'université de Nantes. Ces données étant brutes, il a fallu, bien entendu, traiter les données afin d'obtenir une base permettant la réalisation d'estimations économétriques. Nos deux variables à expliquer ont été renommées **Gbarriere** et **Rsocial**. Les manipulations ont notamment consisté à retirer toutes les valeurs manquantes de la base. Nous avons donc dû retirer les 44 observations qui correspondaient aux étudiants qui n'avaient pas souhaité répondre à la question : "*Vois-tu tes grands-parents régulièrement ? (au moins 1 fois par mois)*". Nous pouvons rajouter à ces 44 observations retirées, 67 valeurs manquantes entre les variables Santé, Boursier, Logement et Emploi. A la fin de la manipulation de la base de données, nous avons donc retiré 86 observations correspondant à 20% de la base initiale. Nous avons décidé de ne pas imposer de réponse à chaque question, afin de respecter la liberté du répondant ne désirant pas répondre aux questions pouvant être intrusives.

II.2.2 Détection des Outliers

Pour terminer le nettoyage de la base il est nécessaire de lui retirer les valeurs atypiques. Nos variables étant majoritairement qualitatives, la détection des outliers va donc se faire sur l'unique variable quantitative continue de notre base, à savoir la variable **Age**. A partir du boxplot suivant de la variable Age nous pouvons observer qu'il y a 13 points atypiques potentiels. Nous allons confirmer leurs atypicités avec un test de Rosner.

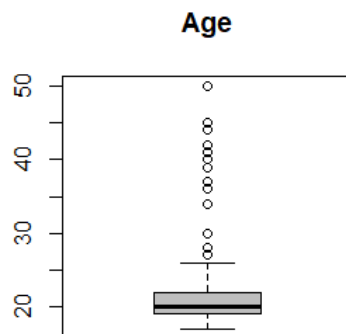


Figure 1 – Boxplot de la variable Age

Le test de Rosner a pour hypothèse nulle, l'absence de valeur(s) atypique(s) dans la série. Nous avons supposé 13 valeurs atypiques à partir du boxplot ci-dessus et les résultats du test de Rosner confirment ces points comme étant atypiques. En ordonnant ces valeurs nous pouvons constater qu'il s'agit des individus 212, 220, 115, 424, 425, 181, 317, 278, 169, 376, 369, 362 et 40. En effet, l'âge des individus retirés s'étendait de 37 ans à 50 ans. Afin de réaliser une estimation correcte, il est important de retirer ces individus atypiques de part leur âge plus avancé par rapport à la moyenne. Les résultats du test sont présentés

en annexe. Après avoir retiré ces 13 individus de notre base de données nous terminons avec une base de 431 observations et 16 variables.

Le reste des variables étant exclusivement qualitatives, il n’y a pas de détection d’outliers à proprement parler mais nous devons regarder les effectifs dans chaque catégorie pour chaque variable.

II.2.3 Statistiques descriptives

Afin d’avoir une vision plus nette de notre base de donnée, il est important de réaliser les statistiques descriptives sur nos variables résumées dans l’annexe 2. Nos variables étant majoritairement qualitatives, nous allons procéder à une observation de la répartition des individus dans chaque modalité de chacune des variables. En ce qui concerne les variables à expliquer, **Gbarriere** et **Rsocial** nous devons vérifier que chaque modalité soit équitablement répartie pour pouvoir réaliser une estimation ou chaque modalité est représentée. Dans la figure suivante, ainsi que celle présente dans l’annexe 3, nous représentons cette répartition.

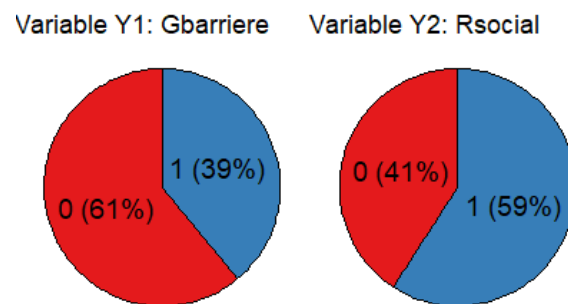


Figure 2 – pie chart des variables à expliquer

On peut constater qu’environ 39% des étudiants de l’IAE de Nantes respectent les gestes barrières en dehors des lieux où il y a une obligation de les respecter. En ce qui concerne les relations sociales, la majorité des étudiants ont changé leurs habitudes de soirée, d’engagement associatif, leurs habitudes sportives et amicales, depuis la fin du confinement jusqu’à aujourd’hui. Dans l’ensemble, la répartition de nos deux variables à expliquer est correcte.

Nous pouvons réaliser la même démarche avec les variables explicatives cette fois-ci. Dans un premier temps nous allons observer la répartition des variables représentatives des caractéristiques individuelles des étudiants.

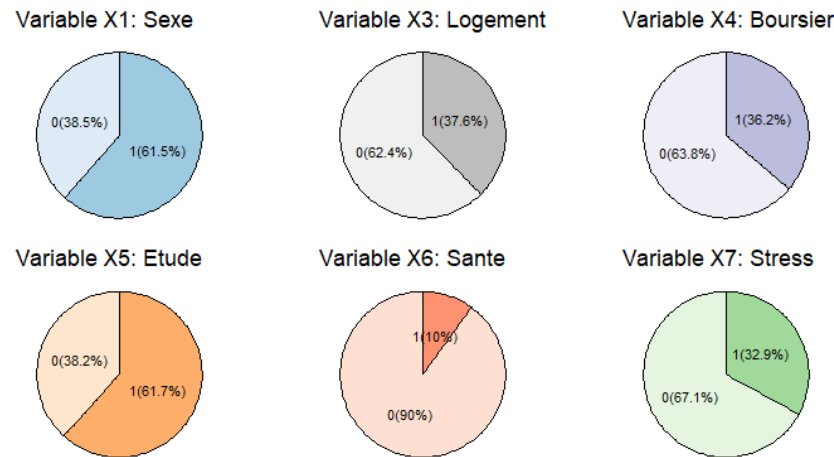


Figure 3 – pie chart des variables explicatives

Pour ces variables, nous pouvons constater que la variable Santé est l'unique variable très injustement répartie avec une majorité d'étudiants n'ayant aucun problème de santé de manière générale, ils correspondent à 90% de notre échantillon. Pour le reste des variables, on peut constater que parmi les répondants au questionnaire, les 2/3 sont des femmes, sont en Licence, n'habitent pas chez leur parents durant la semaine, ne sont pas boursiers et ne sont pas stressés de manière générale. Observons maintenant la répartition des effectifs pour les variables récupérant l'information sur les dimensions familiale et sociale des étudiants.

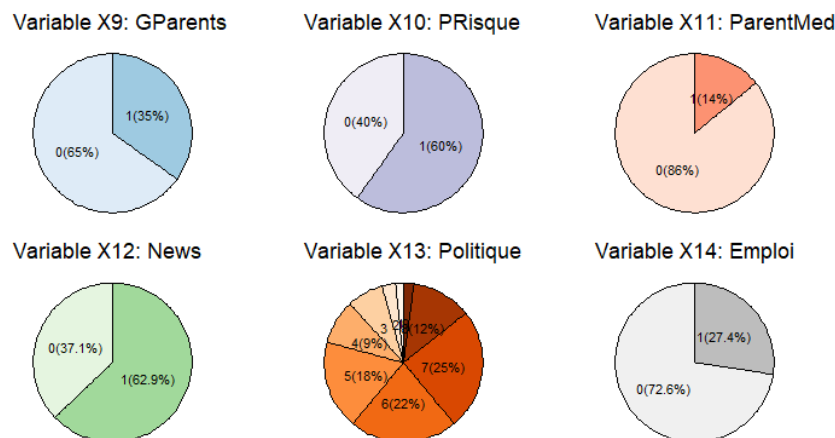


Figure 4 – pie chart des variables explicatives

D'après le diagramme ci-dessus, cette fois-ci c'est la variable ParentMed qui est mal répartie avec seulement 14% des étudiants ayant un parent travaillant dans le corps médical. Cela était à prévoir mais nous voulions l'inclure pour voir si elle pouvait être significative dans la modalité ($Y1 = 1$, $Y2 = 1$) c'est à dire respecter les gestes barrières et avoir changé ses habitudes en terme de relations sociales. De façon surprenante, deux-tiers des étudiants de notre échantillon ont un parent à risque. Seulement 35% des étudiants visitent leurs grands parents au moins une fois par mois. Parmi les étudiants, environ 63% suivent activement les nouvelles liées au Covid et la majorité des étudiants considère que la gestion de la crise sanitaire par le gouvernement actuel a été plutôt bien gérée dans l'ensemble avec une note supérieure à 5 sur 10 pour les deux-tiers des étudiants.

La variable Covid prend la valeur de 1 si l'individu a déjà été testé positif au Covid-19 et 0 sinon. Le diagramme suivant montre une trop mauvaise répartition qui la rend inutile pour l'estimation. Nous la retirerons donc de nos variables explicatives lors de l'estimation des modèles.

Variable X8: Covid

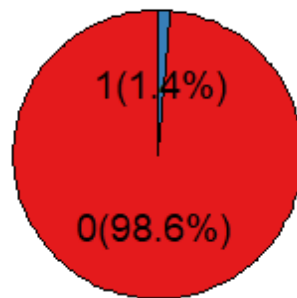


Figure 5 – pie chart de la variable Covid

Dans l'ensemble, nos variables sont correctement réparties (annexe 4), avec toujours un léger biais pour une modalité, $\frac{2}{3} - \frac{1}{3}$. Les variables concernées par une répartition déséquilibrée sont les variables Santé et ParentMed. Enfin, Comme nous l'avons vu précédemment, la variable Covid, est très mal distribuée avec avec seulement 1% des étudiants testés positifs dans notre échantillon. Avant de prendre la décision de retirer les variables explicatives mal réparties nous pouvons observer les dépendances entre-elles et retirer, parmi les variables dépendantes, celles qui ont une mauvaise répartition.

II.2.4 Dépendance et corrélation entre les variables Khi-2

Pour vérifier l'indépendance de nos variables qualitatives nous utilisons le test d'indépendance du Khi^2 . Le tableau ci-dessous reprend les p-values associées aux tests sur chaque paire de variables.

	GParents	PRisque	ParentsMed	Sexe	Etude	Logement	Boursier	Sante	Stress	Covid	News	Emploi
GParents	1	0.090	0.608	0.321	0.017	0.241	0.488	0.181	0.441	0.939	0.947	0.372
PRisque	-	1	0.326	0.091	0.171	0.313	0.131	0.162	0.021	0.732	0.442	0.889
ParentsMed	-	-	1	0.888	0.701	0.160	0.241	0.377	0.537	0.000	0.337	0.598
Sexe	-	-	-	1	0.124	0.120	0.823	0.885	0.000	0.154	0.249	0.571
Etude	-	-	-	-	1	0.000	0.276	0.879	0.069	0.052	0.238	0.794
Logement	-	-	-	-	-	1	0.170	0.700	0.895	0.287	0.518	0.300
Boursier	-	-	-	-	-	-	1	0.884	0.766	0.316	0.308	0.083
Sante	-	-	-	-	-	-	-	1	0.190	0.001	0.990	0.245
Stress	-	-	-	-	-	-	-	-	1	0.084	0.004	0.177
Covid	-	-	-	-	-	-	-	-	-	1	0.511	0.742
News	-	-	-	-	-	-	-	-	-	-	1	0.965
Emploi	-	-	-	-	-	-	-	-	-	-	-	1

Nous pouvons constater que certaines variables apparaissent dépendantes les unes des autres (p-values inférieures à 0.1 donc rejet de l'hypothèse H_0 d'indépendance). Il est important de retirer les variables dépendantes entre elles car elles pourraient biaiser nos estimations et rendre notre modèle statistiquement invalide. Étant donné que les variables sont dépendantes 2 à 2, un processus de sélection des variables doit d'ores et déjà se faire. Les variables **Stress** et **Covid** sont deux variables que nous n'avons pas le choix de retirer si nous souhaitons conserver quelques variables explicatives car elles sont dépendantes (cf annexe 5) avec cinq autres variables explicatives qui sont **PRisque**, **ParentsMed**, **Sexe**, **Etude** et enfin **Sante**. En ce qui concerne les autres dépendances nous avons fait le choix de retirer la variable **GParents** dépendante avec **PRisque** et **Etude** afin de pouvoir les conserver. De plus, il a fallu retirer la variable **Logement** dépendante avec **Etude**. Enfin, avec l'intention de conserver la variable **Boursier** dans nos estimation nous avons dû retirer la variable **Emploi**. Il est important de rappeler que ce choix des variables explicatives à retirer ne s'appuie sur aucun test statistique, mais bien sûr une décision basée sur la volonté de conserver les variables qui nous semblaient les plus pertinentes.

III Méthodologie économétrique

III.1 Introduction

L'étude que nous réalisons consiste à déterminer quels sont les facteurs expliquant le fait qu'une personne respecte tous les gestes barrières et ait changé ses habitudes en terme de relations sociales depuis la fin du confinement. Cela consiste à expliquer deux phénomènes distincts ne pouvant prendre que deux modalités : respecter ou non les gestes barrières, changement ou maintien des relations sociales. Pour expliquer ces deux phénomènes nous avons choisi un ensemble de variables explicatives. Si nous avions supposé que ces deux phénomènes sont indépendants l'un de l'autre, ou que nous avions fait le choix de comprendre ces phénomènes séparément l'estimation aurait pu se faire via deux Probit séparés. Dans notre étude nous supposons que ces deux phénomènes sont liés, c'est-à-dire qu'un étudiant qui respecte les gestes barrières aura aussi changé ses habitudes en terme de relations sociales et inversement. Nous sommes face à quatre cas de figure, dont les effectifs pour chaque modalités sont en annexe 6 :

- $(Y_1 = 1; Y_2 = 1)$: C'est à dire que l'étudiant respecte les gestes barrières et change ses habitudes sociales.
- $(Y_1 = 0; Y_2 = 0)$: L'étudiant ne respecte pas les gestes barrières et ne change pas ses habitudes sociales.
- $(Y_1 = 1; Y_2 = 0)$: L'étudiant respecte les gestes barrières mais ne change pas ses habitudes sociales.
- $(Y_1 = 0; Y_2 = 1)$: L'étudiant ne respecte pas les gestes barrières mais change ses habitudes sociales.

Le modèle Biprobit est donc le modèle qui va nous permettre, à la fois de connaître les facteurs explicatifs pour chaque décision, ainsi que de savoir si les deux décisions sont bien corrélées entre elles via le **test de Wald**. test qui consiste à mesurer la significativité du paramètre ρ qui correspond à la corrélation entre les résidus des deux équations ϵ_{i1} et ϵ_{i2} . En d'autres termes, s'il n'y a pas de liens entre les termes d'erreurs l'estimation devra se réaliser via deux Probit séparés sinon, l'estimation via un Biprobit est justifiée.

Cette section va se dérouler de la manière suivante, Dans un premier temps nous réaliserons l'estimation de deux modèles Biprobit avec un certain nombre de variables que nous pensons pertinentes. Dans un second temps nous procéderons à la méthodologie **Stepwise** afin de pré-sélectionner les variables les plus pertinentes (de manière automatique) à inclure pour l'estimation du Biprobit. Pour terminer nous comparerons tous les modèles Biprobit estimés dans chaque partie afin de sélectionner le meilleur modèle à partir duquel nous pourrions interpréter les résultats et conclure. L'emploi de l'algorithme stepwise est quant à lui justifié par notre nombre conséquent de variables explicatives potentielles. La sélection algorithmique de variables statistiquement significatives nous permet d'éliminer de manière relativement ⁸ efficace.

8. D'autres méthodes telles que la régression LASSO est empiriquement plus efficace que le stepwise

III.2 Estimations sans pré-sélection de variables explicatives

III.2.1 Hypothèses du modèle Biprobit

➤ Modèle Biprobit dans sa forme théorique :

$$\begin{cases} y_{i1}^* = x'_{i1}\beta_1 + \epsilon_{i1}, y_{i1} = 1 \text{ si } y_{i1}^* > 0, 0 \text{ sinon} \\ y_{i2}^* = x'_{i2}\beta_2 + \epsilon_{i2}, y_{i2} = 1 \text{ si } y_{i2}^* > 0, 0 \text{ sinon} \end{cases} \quad (1)$$

➤ Hypothèses :

- L'espérance des termes d'erreurs est nulle $E[\epsilon_{i1}] = E[\epsilon_{i2}] = 0$
- Homoscedasticité des erreurs (variance constante) $Var(\epsilon_{i1}) = Var(\epsilon_{i2}) = 1$
- Covariance entre les termes d'erreurs $Cov(\epsilon_{i1}, \epsilon_{i2}) = \rho$
- Les termes d'erreurs sont i.i.d, normaux

➤ Sous forme Matricielle :

$$\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right] \quad (2)$$

la matrice $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ qui correspond à la matrice de variance covariance des résidus suppose une homoscedasticité des termes d'erreur car la première diagonale est formée par des 1. Dans notre étude, nous allons nous prémunir d'une possible hétéroscédasticité des erreurs en estimant les modèles de manière **Robust**. Ce qui vient modifier les hypothèses sur les résidus de la manière suivante :

$$\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{i1} & \rho \\ \rho & \sigma_{i2} \end{pmatrix} \right] \quad (3)$$

On constate qu'à l'inverse d'un modèle homoscedastique, ici la première diagonale de la matrice de variance covariance est constituée de la variance de chaque terme d'erreur et elles ne sont pas toutes égales. L'estimation des paramètres β se fera par maximum de vraisemblance.

III.2.2 Estimation du Modèle ‘complet’

1. Spécification du Modèle ‘complet’

Nous commençons nos estimations par un modèle regroupant toutes les variables explicatives pour les deux équations car nous supposons dans un premier temps que toutes les variables peuvent potentiellement expliquer à la fois le respect des gestes barrières mais aussi les changements des relations sociales. La spécification de notre premier modèle est la suivante :

$$\begin{aligned} \mathbf{Gbarriere} &= \beta_0 + \beta_1 \mathbf{PRisque} + \beta_2 \mathbf{ParentMed} + \beta_3 \mathbf{Age} + \beta_4 \mathbf{Sexe} + \beta_5 \mathbf{Boursier} \\ &\quad + \beta_6 \mathbf{Etude} + \beta_7 \mathbf{Sante} + \beta_8 \mathbf{News} + \beta_9 \mathbf{Politique} + \epsilon \\ \mathbf{Rsocial} &= \beta_0 + \beta_1 \mathbf{PRisque} + \beta_2 \mathbf{ParentMed} + \beta_3 \mathbf{Age} + \beta_4 \mathbf{Sexe} + \beta_5 \mathbf{Boursier} \\ &\quad + \beta_6 \mathbf{Etude} + \beta_7 \mathbf{Sante} + \beta_8 \mathbf{News} + \beta_9 \mathbf{Politique} + \epsilon \end{aligned} \quad (4)$$

2. Estimation

Après estimation du modèle complet de façon robust⁹(c.f. annexe 7), nous obtenons un $\rho = 0.39$, significatif à un seuil de risque de 1%. Il est donc raisonnable d’estimer un modèle Biprobit car il existe une corrélation significative entre les résidus des deux équations. Le tableau suivant reprend les résultats des effets marginaux des variables explicatives sur les quatre probabilités conjointes ainsi que leur significativité respective. Nous ne rentrerons pas dans l’interprétation des coefficients dans cette section, mais seulement une fois que le meilleur modèle aura été retenu. Nous pouvons tout de même constater que la variable **ParentMed** n’est significative pour aucune modalité.

Variables	Effets Marginaux			
	$P(Y_1 = 1; Y_2 = 1)$	$P(Y_1 = 0; Y_2 = 0)$	$P(Y_1 = 1; Y_2 = 0)$	$P(Y_1 = 0; Y_2 = 1)$
PRisque	.06	-.096**	-.031	.058
ParentMed	.02	-.004	.019	-.036
Age	.048***	-.050***	-.004	.005
Sexe	.151***	-.124***	.027	-.054
Boursier	.031	.022	.064**	-.117***
Etude	.084	-.111*	-.030	.058
Sante	.150*	-.147**	-.033	.030
News	.161***	-.145***	.020	-.036
Politique	.039***	-.043***	-.006	.010

Table 1 – Niveaux de significativité : * = 0.1, ** = 0.05, *** = 0.01

9. On estime de manière robust afin de ne pas se préoccuper de l’hypothèse d’homocédasticité des erreurs vue dans la partie des hypothèses d’un Biprobit.

3. Qualité de prévision

Modèle	Modalités	Observé	Prédit
Biprobit ‘complet’	$Y_1 = 1; Y_2 = 1$	110	67 (60.90%)
	$Y_1 = 0; Y_2 = 0$	103	65 (63.11%)
	$Y_1 = 0; Y_2 = 1$	101	28 (27.72%)
	$Y_1 = 1; Y_2 = 0$	31	0 (0%)

Ce modèle prévoit correctement 46.38% (160/345) de notre échantillon (in-sample). 60.90% des individus sont correctement classés dans la première probabilité jointe.

III.2.3 Estimation du Modèle ‘partiel’

1. Spécification du Modèle ‘partiel’

Ce modèle contient un sous ensemble de variables qui nous semblent être les plus pertinentes pour notre analyse(c.f. annexe 8). Nous avons également décidé d’omettre la variable ParentMed pour cause de non-significativité (c.f. Table ci-dessus).

$$\begin{aligned}
 \mathbf{Gbarriere} &= \beta_0 + \beta_1 Age + \beta_2 Sexe + \beta_3 Boursier + \beta_4 News + \beta_5 Politique + \epsilon \\
 \mathbf{Rsocial} &= \beta_0 + \beta_1 PRisque + \beta_2 Age + \beta_3 Sexe + \beta_4 Etude + \beta_5 Sante + \beta_6 News \\
 &\quad + \beta_7 Politique + \epsilon
 \end{aligned}
 \tag{5}$$

2. Estimation

Nous avons estimé le modèle partiel de manière robuste pour les mêmes raisons que le modèle complet. De même, le test de Wald continue à être significatif avec cette spécification, l’emploi d’un Biprobit est donc justifié.

Variables	Effets Marginaux			
	$P(Y_1 = 1; Y_2 = 1)$	$P(Y_1 = 0; Y_2 = 0)$	$P(Y_1 = 1; Y_2 = 0)$	$P(Y_1 = 0; Y_2 = 1)$
PRisque	.044**	-.082**	-.044**	.082**
Age	.043***	-.046***	-.005	.009
Sexe	.148***	-.124***	.025	-.049
Boursier	.073**	-.036*	.036*	-.073**
Etude	.047*	-.087*	-.047*	.087*
Santé	.048*	-.104*	-.048*	.104*
News	.159***	-.142***	.021	-.038
Politique	.038***	-.043***	-.007	.012

Table 2 – Niveaux de significativité : * = 0.1, ** = 0.05, *** = 0.01

3. Qualité de prévision

Modèle	Modalités	Observé	Prédit
BiProbit ‘partiel’	$Y_1 = 1; Y_2 = 1$	110	62 (56.36%)
	$Y_1 = 0; Y_2 = 0$	103	61 (59.22%)
	$Y_1 = 0; Y_2 = 1$	101	32 (31.68%)
	$Y_1 = 1; Y_2 = 0$	31	0 (0%)

Ce modèle prévoit correctement 44.93% (155/345) de notre échantillon (in-sample). 56.36% des individus sont correctement classés dans la première probabilité jointe.

III.2.4 Modèle retenu

Afin de choisir un modèle final nous utiliserons l’AIC et le BIC pour discriminer les modèles et trouver le plus performant, nous nous baserons également sur les qualités de prédiction. Cette information complète les critères d’informations qui ont le désavantage de ne pas tenir compte de ce paramètre.

1. Comparaison AIC/BIC

Modèle	AIC	BIC
BiProbit ‘complet’	873.46	954.17
BiProbit ‘partiel’	866.69	924.34

Comme nous pouvons l’observer, les critères d’informations suggèrent un seul et même modèle. Les AIC¹⁰ des deux modèles sont 873.46 pour le modèle ‘complet’ contre 866.69 pour le modèle ‘partiel’ ce qui revient à sélectionner le modèle ‘partiel’. Pour confirmer notre choix, nous pouvons aussi regarder le BIC¹¹, celui-ci passe de 954.17 dans le modèle ‘complet’ à 924.34 dans le modèle ‘partiel’. Ce qui revient à dire que le modèle ‘partiel’ est meilleur (en terme d’ajustement rapporté à la complexité) que le modèle ‘complet’.

10. $AIC = 2 \log(\mathcal{L}) - 2p$ avec \mathcal{L} le Log de Vraisemblance du modèle et p le nombre de paramètres.

11. Le BIC se calcule de manière analogue mais appliquera une pénalité plus forte au nombre de paramètres. $AIC > BIC$ dès lors que $\log(N) > 2$ avec N la taille de l’échantillon

2. Comparaison des prévisions

Modèle	Modalités	Observé	Prédit
BiProbit ‘complet’	$Y_1 = 1; Y_2 = 1$	110	67 (60.90%)
	$Y_1 = 0; Y_2 = 0$	103	65 (63.11%)
	$Y_1 = 0; Y_2 = 1$	101	28 (27.72%)
	$Y_1 = 1; Y_2 = 0$	31	0 (0%)
BiProbit ‘partiel’	$Y_1 = 1; Y_2 = 1$	110	62 (56.36%)
	$Y_1 = 0; Y_2 = 0$	103	61 (59.22%)
	$Y_1 = 0; Y_2 = 1$	101	32 (31.68%)
	$Y_1 = 1; Y_2 = 0$	31	0 (0%)

On constate que les deux modèles sont encore une fois très proches. La qualité de prédiction globale est meilleure dans le modèle ‘complet’ (44.93% contre 46.38%) avec 5 individus supplémentaires correctement classifiés par ce même modèle. Le modèle complet prédit mieux les catégories ($Y_1 = 1; Y_2 = 1$) et ($Y_1 = 0; Y_2 = 0$) avec 60.90% et 63.11% respectivement comparé au modèle partiel qui ne prévoit que 56.36% et 59.22% pour ces deux modalités. En ce qui concerne la modalité ($Y_1 = 1; Y_2 = 0$), les deux modèles sont identiques avec 0 individus prédits sur les 31 observés. La grande différence entre les deux modèles s’opère sur la modalité ($Y_1 = 0; Y_2 = 1$) où le modèle ‘partiel’ prédit mieux que le modèle ‘complet’ avec 32 individus prédits sur 101 observés contre seulement 28 pour le modèle ‘complet’.

3. Conclusion

D’une manière globale, les deux modèles possèdent la même capacité explicative, ceci est dû au faible pouvoir explicatif de notre modèle en général. Malgré tout, notre objectif est de conserver un seul modèle c’est pour cela que nous les avons comparé à l’aide des critères AIC et BIC ainsi que sur leurs qualités de prévisions. Le modèle BiProbit ‘partiel’ ressort favori avec un AIC et un BIC inférieur à celui du BiProbit ‘complet’. Dans la section précédente nous avons mis en évidence la moins bonne qualité de prévision du modèle ‘partiel’ pour les modalités ($Y_1 = 1; Y_2 = 1$) et ($Y_1 = 0; Y_2 = 0$) mais à contrario, il est meilleur pour prédire les étudiants qui ne respectent pas les gestes barrières mais qui changent tout de même leurs habitudes sociales. Dans cette optique, nous choisissons donc le modèle ‘partiel’, qui est très légèrement moins efficace en terme de prédictions que le modèle ‘complet’. Mais ce modèle a l’avantage de mieux expliquer la modalité où le modèle ‘complet’ est le moins performant.

III.3 Estimations avec pré-sélection de variables explicatives

III.3.1 Spécification du modèle

Dans cette seconde partie, il nous a semblé important de nous concentrer sur une spécification plus rigoureuse de notre modèle. La spécification du modèle est déterminante à la fois pour la qualité de prédiction du modèle mais aussi sur son interprétabilité. Ces deux critères peuvent parfois être conflictuels, un modèle contenant de nombreuses variables explicatives aura une qualité de prédiction élevée. Mais un tel modèle, avec des variables choisies uniquement pour leur participation relative à la qualité de prédiction, peut perdre son sens économique et social et devenir incroyablement difficile à interpréter d'un point de vue théorique. A l'inverse, un modèle trop parcimonieux peut plus facilement s'interpréter, mais sera moins bon en terme de prédictions. Le sur-ajustement des données est également un problème récurrent qui entraîne souvent un ajustement in-sample très favorable mais qui se détériore rapidement dès lors que l'on tente des prédictions out-of-sample. Trouver un modèle intelligible, performant et stable au travers des échantillons est une tâche fortement déterminée par la spécification du modèle en elle-même.

Dans cette section, nous avons fait le choix de passer par une série de procédures destinées à sélectionner les meilleures variables au sens des critères que nous avons cité plus haut. Pour ce processus de sélection de variables, nous allons utiliser les algorithmes Stepwise (backward et forward). Le principal inconvénient de cette méthode est qu'elle s'applique uniquement à des modèles probits, et non à des biprobit. Nous allons donc devoir considérer le modèle biprobit comme deux modèles probit séparés pour la sélection des variables, puis nous estimerons le modèle biprobit à partir de ces variables sélectionnées. Cet écart méthodologique est justifié plus loin par la similitude des coefficients et des résultats aux tests d'hypothèses des modèles Biprobites et des deux probits séparés.

Notre intérêt se porte principalement sur le backward stepwise, notre idée étant de conserver un maximum de variables explicatives car elles sont toutes, à priori, intéressantes. Nous utilisons également la procédure forward, pour confirmer les résultats. A l'inverse, si backward et forward donnent des résultats différents, cela constitue une information en soi que nous pourrions utiliser. La classification stepwise backward part d'un modèle comprenant toutes les variables fournies et retire une à une chacune des variables en fonction de leur significativité et d'un test de Wald¹² qui compare les modèles avec et sans la variable en question. Le modèle obtenu au bout d'un certain nombre d'itérations est un modèle potentiellement intéressant, qui est supposé avoir une bonne qualité de prédiction (bonne, relativement aux autres modèles) et une parcimonie suffisante pour être stable et faire sens.

12. Le test de Wald est parfois remplacé par un test de ratio de vraisemblance ou par la comparaison stricte de l'AIC et du BIC, ici nous gardons le critère par défaut à savoir le test de Wald

III.3.2 Algorithme Stepwise

Une fois les variables dépendantes retirées, nous lançons les estimations et obtenons deux modèles probits contenant les variables les plus intéressantes¹³. Les procédures backward et forward donnent les mêmes résultats (c.f. Annexe 9). A savoir ; les deux modèles retenus sont :

$$\mathbf{Gbarriere} = \beta_0 + \beta_1 Age + \beta_2 Sexe + \beta_3 News + \beta_4 Politique + \epsilon \quad (6)$$

$$\mathbf{Rsocial} = \beta_0 + \beta_1 Age + \beta_2 PRisque + \beta_3 Politique + \beta_4 News + \beta_5 Sante + \beta_6 Etude + \beta_7 Sexe + \epsilon \quad (7)$$

Pour chaque modèle, la variable Sexe est gardée. Comme nous avons pu le constater précédemment cette variable est légèrement dépendante de la variable PRisque. Nous avons donc décidé de forcer l'abandon de cette variable et de lancer des stepwise sans la prendre en compte. Le but étant d'observer si nous obtenons les mêmes modèles une fois la variable exclue. Nous devons ensuite discriminer entre les modèles obtenus pour déterminer si la variable Sexe possède un quelconque intérêt et mérite d'être gardée.

III.3.3 Algorithme Stepwise sans Sexe

Avec cette configuration, les deux méthodes (forward et backward) nous donnent également les mêmes résultats (c.f. Annexe 10). Les deux modèles retenus sont :

$$\mathbf{Gbarriere} = \beta_0 + \beta_1 Age + \beta_2 News + \beta_4 Politique + \epsilon \quad (8)$$

$$\mathbf{Rsocial} = \beta_0 + \beta_1 Age + \beta_2 PRisque + \beta_3 Politique + \beta_4 News + \beta_5 Sante + \beta_6 Etude + \epsilon \quad (9)$$

Ces modèles sont identiques aux précédents (à l'exception de la variable Sexe qui n'apparaît évidemment plus). Aucune variable supplémentaire n'a été ajoutée ou retirée. Nous avons donc deux modèles différents l'un contenant un ensemble de variables significatives dont Sexe, l'autre contenant le même ensemble de variables explicatives à l'exception de Sexe. Nous devons désormais discriminer l'un de ces deux modèles. Pour ce faire, nous nous baserons sur les critères d'information communément utilisés ainsi que sur la qualité de prédiction de chaque modèle.

Modèle	AIC	BIC
GBarriere (contenant Sexe)	442.12	461.33
GBarriere (Sexe retirée)	450.09	465.47
RSocial (contenant Sexe)	441.11	471.85
RSocial (Sexe retirée)	442.35	469.26

13. qui maximisent le Log de Vraisemblance

Les résultats obtenus sont logiques et en accord avec notre méthodologie. Dans le cas de Gbarrière, les deux critères nous suggèrent d'utiliser la variable Sexe. Cette variable possède un pouvoir explicatif sur Gbarrière, l'inclure revient donc à augmenter le Log de Vraisemblance, donc à faire augmenter l'AIC (resp. le BIC). L'inclusion d'une variable supplémentaire entraîne parallèlement, via la pénalité, une baisse de l'AIC (resp. BIC), le tableau ci-dessus nous indique donc que l'augmentation du Log de Vraisemblance contrebalance plus que suffisamment la pénalité appliquée à l'ajout d'une variable. En revanche pour Rsocial, le BIC montre le résultat contraire, l'augmentation de la qualité d'ajustement n'est pas suffisante pour justifier l'ajout d'une variable au modèle. L'AIC nous indique qu'il serait préférable d'ajouter Sexe au modèle. La différence entre les deux critères est due au fait que le BIC pénalise plus fortement la complexité du modèle (c.f. note de bas de page). Sur ces informations, nous pensons qu'il serait judicieux de conserver Sexe dans notre modèle, le BIC étant le seul critère à suggérer son retrait.

A partir des variables sélectionnées par la méthodologie stepwise nous avons donc estimé deux modèles BiProbit l'un avec la variable Sexe et l'autre sans (c.f. Annexe 11 et 12). Nous avons mis les résultats en Annexe. Le tableau suivant récapitule les qualités de prédiction de chaque BiProbit estimé.

Modèle	Modalités	Observé	Prédit
BiProbit Avec Sexe	$Y_1 = 1; Y_2 = 1$	110	65 (59.09%)
	$Y_1 = 0; Y_2 = 0$	103	61 (60.19%)
	$Y_1 = 0; Y_2 = 1$	101	28 (27.72%)
	$Y_1 = 1; Y_2 = 0$	31	0 (0%)
BiProbit Sans Sexe	$Y_1 = 1; Y_2 = 1$	110	61 (55.45%)
	$Y_1 = 0; Y_2 = 0$	103	57 (55.34%)
	$Y_1 = 0; Y_2 = 1$	101	28 (27.72%)
	$Y_1 = 1; Y_2 = 0$	31	0 (0%)

On peut observer dans le tableau ci-dessus le résumé des prédictions (in-sample) des deux modèles estimés avec la méthodologie stepwise. Les prédictions semblent être meilleures lorsque Sexe est conservée. les critères d'informations ne suggèrent pas drastiquement de retirer cette variable et la dépendance avec la variable PRisque n'est significative qu'à 10%. En terme de qualité prédictive le modèle BiProbit avec Sexe est donc meilleur.

III.3.4 Choix du modèle définitif

Nous avons estimé différents modèles, il est désormais nécessaire de faire un choix définitif quant au modèle que nous allons interpréter et à partir duquel nous ferons nos prédictions out-of-sample. nous avons constaté que les modèles ne différaient que très peu en terme d'ajustement et de prédictions. De plus, les modèles choisis par Stepwise sont quasiment identiques au modèle 'partiel' estimé plus haut. nous ne nous attendons donc pas à une différence significative entre les modèles, en revanche, nous choisirons tout de même

le meilleur, à l'aide des mêmes critères que nous avons déjà utilisés (critères d'informations et qualités de prédictions).

Modèle	AIC	BIC
BiProbit 'complet'	873.46	954.17
Biprobit 'partiel'	866.69	924.34
Biprobit Avec Sexe	868.50	922.31
BiProbit Sans Sexe	875.39	921.52

Le BIC semble suggérer le modèle stepwise sans Sexe, et l'AIC suggère le modèle 'partiel' où les variables ont été choisies à priori. On peut savoir que cette différence est due au fait que le premier modèle comprend moins de paramètres et aura donc forcément, à log de vraisemblance égale, un BIC inférieur. Cependant, l'AIC nous indique que le modèle 'partiel' est mieux ajusté toutes choses égales par ailleurs. nous supposons donc que le modèle stepwise sans sexe n'est pas objectivement meilleur, même si son BIC est le plus faible, mais bénéficie simplement de son faible nombre de paramètres.

Modèle	Modalités	Observé	Prédit
BiProbit 'complet'	$Y_1 = 1; Y_2 = 1$	110	67 (60.90%)
	$Y_1 = 0; Y_2 = 0$	103	65 (63.11%)
	$Y_1 = 0; Y_2 = 1$	101	28 (27.72%)
	$Y_1 = 1; Y_2 = 0$	31	0 (0%)
BiProbit 'partiel'	$Y_1 = 1; Y_2 = 1$	110	62 (56.36%)
	$Y_1 = 0; Y_2 = 0$	103	61 (59.22%)
	$Y_1 = 0; Y_2 = 1$	101	32 (31.68%)
	$Y_1 = 1; Y_2 = 0$	31	0 (0%)
BiProbit Avec Sexe	$Y_1 = 1; Y_2 = 1$	110	65 (59.09%)
	$Y_1 = 0; Y_2 = 0$	103	62 (60.19%)
	$Y_1 = 0; Y_2 = 1$	101	28 (27.72%)
	$Y_1 = 1; Y_2 = 0$	31	0 (0%)
BiProbit Sans Sexe	$Y_1 = 1; Y_2 = 1$	110	61 (55.45%)
	$Y_1 = 0; Y_2 = 0$	103	57 (55.34%)
	$Y_1 = 0; Y_2 = 1$	101	28 (27.72%)
	$Y_1 = 1; Y_2 = 0$	31	0 (0%)

En terme de prédiction on peut constater que chaque modèle prédit mieux une modalité particulière mais il n'y a pas de modèle qui se démarque de l'ensemble des prédictions. Sur la base des critères AIC/BIC, nous retiendrons le Biprobit 'partiel'. ce choix est directement suggéré par l'AIC. De plus, notre décision est motivée par plusieurs critères ; comme nous l'avons dit plus haut, nous trouvons pertinent de perdre quelques points de précision dans les catégories $P(Y_1 = 1, Y_2 = 1)$ et $P(Y_1 = 0, Y_2 = 0)$ au profit

de la troisième catégorie $P(Y_1 = 0, Y_2 = 1)$. Nos prédictions sont ainsi légèrement moins bonnes mais permettent de mieux mettre en évidence les différents phénomènes en jeu. Au fil des différentes estimations, nous avons compris que la variable Boursier est en grande partie responsable de la qualité de prédiction dans cette catégorie. Nous choisissons donc de l'inclure dans le modèle, au mépris des procédures stepwise qui retireraient systématiquement cette variable. La Section III.3, qui reprend les procédures stepwise, nous indique que les qualités de prévisions et les critères d'informations sont meilleurs lorsque Sexe est conservée¹⁴. Notre modèle final doit donc conserver Sexe et Boursier pour les raisons citées plus haut. En partant du stepwise avec Sexe, et en rajoutant les deux variables citées plus haut, nous retombons sur le modèle 'partiel' ce qui justifie d'autant plus sa sélection en tant que modèle définitif. Du point de vue de la qualité de prédiction in-sample, le modèle partiel n'est pas strictement meilleur (44.93% de classifications correctes contre 46.38% pour le modèle 'complet', soit 5 individus de plus correctement classifiés dans le modèle 'complet'). Mais nous gardons à l'esprit que ces prédictions sont effectuées sur le même échantillon que nous avons utilisé pour estimer les modèles. Un biais de sur-apprentissage est possible et certaines variables sont de toute façon, inexploitable et doivent forcément être retirées, raison pour laquelle nous excluons le modèle 'complet' malgré sa qualité de prédiction supérieure.

Notre choix final se porte donc sur le modèle 'partiel', dont nous nous servirons pour étudier les effets marginaux sur les probabilités jointes et les probabilités conditionnelles. Ce modèle nous servira également à produire des prédictions out-of-sample sur les 20% des données conservées à cet effet.

14. Le BIC est inférieur mais cela se justifie aisément, cf Section III.3

III.4 Estimation d'un Biprobit récursif

III.4.1 Hypothèses d'un modèle récursif

Les hypothèses sur les erreurs d'un modèle Biprobit récursif ou "*Recursive Bivariate Probit Model*" sont les mêmes qu'un Biprobit "normal" c'est à dire une espérance nul, l'homocedasticité et la variance covariance non nul des résidus. La seule hypothèse que nous réalisons c'est de supposer que l'une des deux variables à expliquer est explicatives de l'autre. ce qui revient à dire :

$$\begin{cases} y_{i1}^* = y_{i2}^* \beta_1 + x'_{i1} \beta_2 + \epsilon_{i1}, & y_{i1} = 1 \text{ si } y_{i1}^* > 0, 0 \text{ sinon} \\ y_{i2}^* = x'_{i2} \beta_2 + \epsilon_{i2}, & y_{i2} = 1 \text{ si } y_{i2}^* > 0, 0 \text{ sinon} \end{cases} \quad (10)$$

Dans cette spécification la seule chose qui a changé c'est que Y_2 intervient dans la première équation d'où la récursivité.

III.4.2 Estimation du modèle 5 Biprobit récursif

1. Spécification du modèle Biprobit récursif

Nous avons donc supposé pour cette estimation que la variable **Rsocial** pouvait expliquer la variable **Gbarriere** c'est pour cela que nous la retrouvons en tant que variable explicative dans la première équation. Cela revient à supposer que le fait de ne pas changer ses relations sociales est à l'origine du non respect des gestes barrières.

$$\begin{aligned} \mathbf{Gbarriere} &= \beta_0 + \beta_1 \mathbf{Rsocial} + \beta_2 \mathbf{Age} + \beta_3 \mathbf{Sexe} + \beta_4 \mathbf{Boursier} + \beta_5 \mathbf{News} \\ &\quad + \beta_6 \mathbf{Politique} + \epsilon \\ \mathbf{Rsocial} &= \beta_0 + \beta_1 \mathbf{PRisque} + \beta_2 \mathbf{Age} + \beta_3 \mathbf{Sexe} + \beta_4 \mathbf{Etude} + \beta_5 \mathbf{Sante} + \beta_6 \mathbf{News} \\ &\quad + \beta_6 \mathbf{Politique} + \epsilon \end{aligned} \quad (11)$$

2. Estimation

Nous avons estimé le modèle biprobit récursif de façon robuste comme pour les précédents modèles. Cependant le modèle n'a pas abouti sur une solution car "*non concave*". Ce qui rend impossible l'estimation du modèle biprobit récursif (c.f. Annexe13). Nous retiendrons donc seulement le modèle partiel pour les interprétations.

IV Présentation des résultats

IV.1 Introduction

Suite au processus de sélection de variables et plus largement, de modèle, nous optons pour le modèle suivant :

$$\left\{ \begin{array}{l} \mathbf{Gbarriere} = \beta_0 + \beta_1 Age + \beta_2 Sexe + \beta_3 Boursier + \beta_4 News + \beta_5 Politique + \epsilon \\ \mathbf{Rsocial} = \beta_0 + \beta_1 PRisque + \beta_2 Age + \beta_3 Sexe + \beta_4 Etude + \beta_5 Sante + \beta_6 News \\ \quad \quad \quad + \beta_7 Politique + \epsilon \end{array} \right. \quad (12)$$

Le respect des gestes barrières et la modification des habitudes sociales seront donc conjointement expliqués par un biprobit non récursif. Nous présenterons ici les résultats du modèle en termes d'ajustement et de prédiction out-of-sample. Comme expliqué plus haut, nous avons conservé 20% (90 observations environ) de l'échantillon pour mesurer la qualité de prédiction du modèle. Le tableau ci-dessous reprend les effectifs observés et prédits pour chacune des catégories.

Modèle	Modalités	Observé	Prédit
biprobit partiel	$Y_1 = 1; Y_2 = 1$	19	13 (68.42%)
-	$Y_1 = 0; Y_2 = 0$	34	19 (55.88%)
-	$Y_1 = 0; Y_2 = 1$	25	4 (16%)
-	$Y_1 = 1; Y_2 = 0$	8	0 (0%)

D'une manière générale, notre modèle a correctement assigné 41.86% de notre échantillon. Sur 19 individus, 68% ont été correctement assignés à la catégorie (1, 1) c'est à dire respecter les gestes barrières et avoir changé ses habitudes sociales. Sur 34 individus, 55.88% ont été correctement assignés à la catégorie (0, 0) c'est à dire ne pas respecter les gestes barrières et ne pas avoir changé ses habitudes sociales. Sur 25, 16% ont été correctement assignés à (0, 1) à savoir ; ne pas respecter les gestes barrières en ayant changé ses habitudes sociales. Comme pendant la phase d'entraînement, le modèle ne parvient pas à capter les déterminants de la catégorie (1, 0). Les graphiques suivants illustrent ce phénomène.

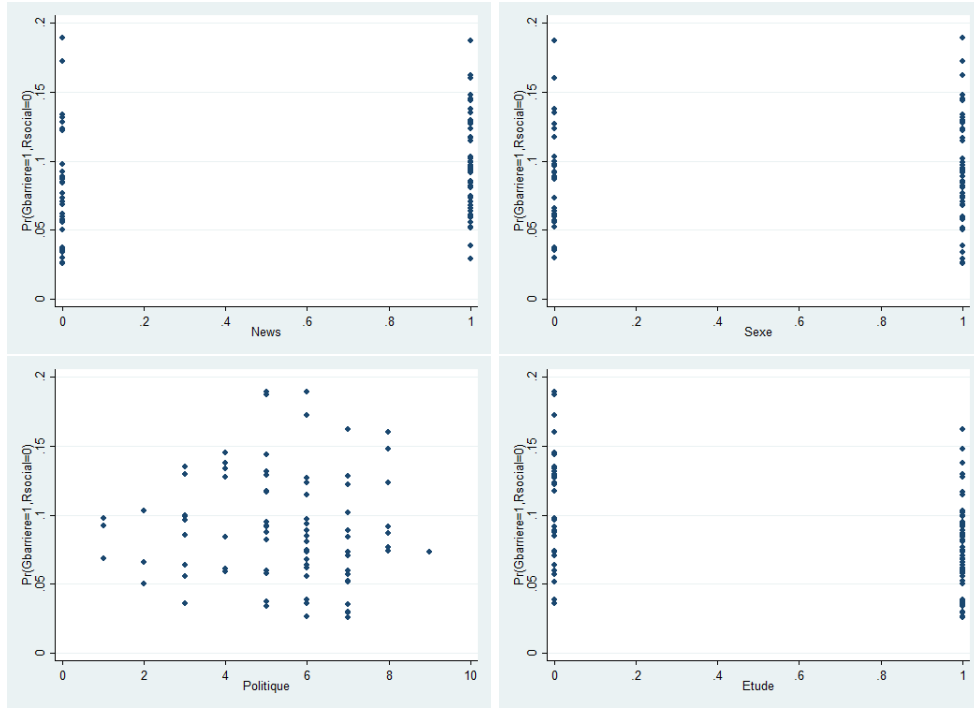


Figure 6 – Évolution de $P(Y_1 = 1, Y_2 = 0 | X = x)$

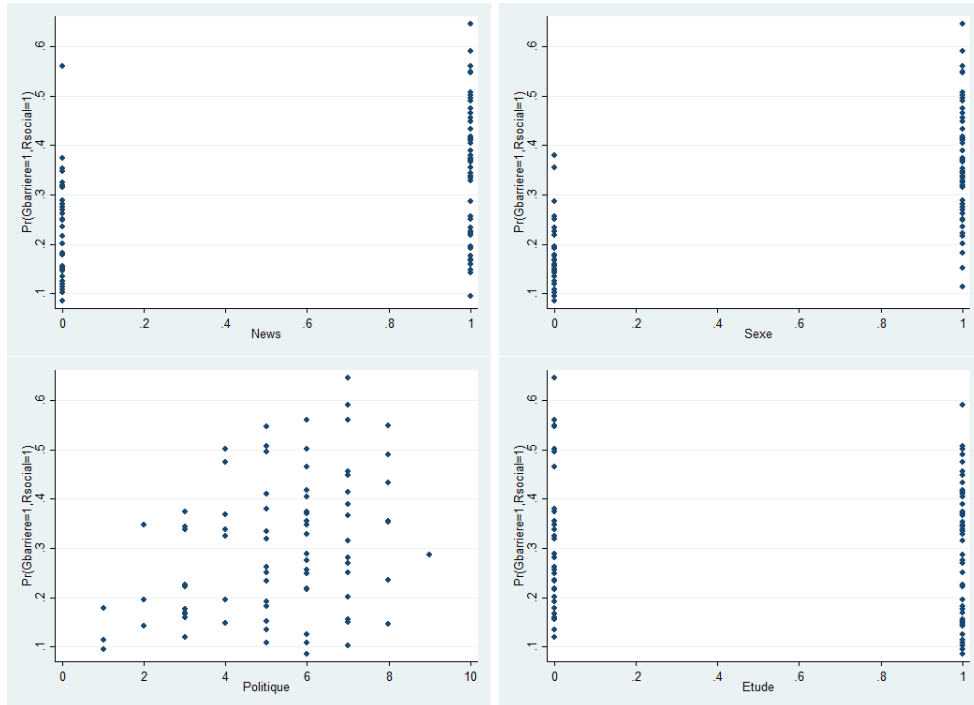


Figure 7 – Évolution de $P(Y_1 = 1, Y_2 = 1 | X = x)$

Comme on peut le constater ; la probabilité $P(Y_1 = 1, Y_2 = 0)$ ne change quasiment pas quelque soit la valeur de la variable explicative. Pour les variables binaires (Sexe, News, Stress) $P(Y_1 = 1, Y_2 = 0)$ est identique, que la variable prenne une valeur de 0 ou 1. Autrement dit, $P(Y_1 = 1, Y_2 = 0)$ est indépendante de Sexe, News et Stress. Pour la variable Politique, quantitative discrète, le nuage de point semble n'indiquer aucune corrélation notable. On constate avec ces graphiques que les variables explicatives que nous avons choisies ne permettent pas d'expliquer $P(Y_1 = 1, Y_2 = 0)$. En revanche, ces relations sont plus claires et identifiables dès lors que l'on s'intéresse à $P(Y_1 = 1, Y_2 = 1)$. Les distributions de $P(Y_1 = 1, Y_2 = 1)$ sont différentes selon la valeur prise par nos variables. On constate par exemple que l'avis d'un individu sur le gouvernement actuel a un impact positif sur la probabilité de respecter les gestes barrières et de changer ses habitudes sociales. Cette relation, bien que très hétérogène, est plus facilement identifiable que dans le cas $(Y_1 = 1, Y_2 = 0)$. La même observation se vérifie au niveau des variables qualitatives binaires. $P(Y_1 = 1, Y_2 = 1)$ varie selon que ces variables prennent la valeur 0 ou 1. Ce que signifie que cette probabilité est conditionnelle aux valeurs de nos variables qualitatives ; $P(Y_1 = 1, Y_2 = 1|X = 0) \neq P(Y_1 = 1, Y_2 = 1|X = 1)$. Contrairement au cas $(Y_1 = 0, Y_2 = 1)$, nos variables sont capables, dans une certaine mesure, d'expliquer les décisions d'un individu en terme de respect des gestes barrières et de changement des habitudes sociales.

IV.2 Interprétation des résultats

Avant de présenter les résultats, nous avons fait le choix de ré-estimer le modèle partiel choisi à partir de la base entière et non seulement la base d'entraînement comme nous l'avons fait dans la partie précédente. Le tableau ci-dessous représente donc les effets marginaux pour les 4 probabilités conjointes calculés pour un étudiant moyen de l'échantillon.

Variables	Effets Marginaux			
	$P(Y_1 = 1; Y_2 = 1)$	$P(Y_1 = 0; Y_2 = 0)$	$P(Y_1 = 1; Y_2 = 0)$	$P(Y_1 = 0; Y_2 = 1)$
PRisque	.047***	-.093**	-.048***	.093***
Age	.038***	-.043***	-.004	.008
Sexe	.146***	-.142***	.019	-.023
Boursier	.059*	-.032*	.032*	-.059*
Etude	.039*	-.076*	-.039*	.076*
Santé	.026	-.053	-.025	.053
News	.153***	-.166***	.005	.008
Politique	.042***	-.048***	-.005	.011

Table 3 – Effets Marginaux sur probabilités jointes, niveaux de significativité : * = 0.1, ** = 0.05, *** = 0.01

Sur notre échantillon entier, nous observons les relations suivantes :

Avoir un **parent à risque** augmente de 0.047 la probabilité de respecter les gestes barrières et de modifier (réduire) ses fréquentations, et de 0.093 la probabilité que l'étudiant ne respecte pas les gestes barrières mais voit ses habitudes sociales changer. Cette même variable diminue de 0.093 la probabilité de ne pas respecter les gestes barrières et de ne pas modifier ses habitudes sociales ainsi que la probabilité de respecter les gestes barrières sans pour autant modifier ses fréquentations. L'effet de la variable dans les quatre modalités est significatif au seuil de 1%. Ce que l'on peut constater c'est que le signe des coefficients change lorsque Y2 change, c'est à dire lorsque l'on passe d'un changement des habitudes sociales à pas de changement dans les habitudes sociales depuis la fin du confinement. En d'autres termes, avoir un parent à risque modifie significativement les habitudes sociales, peu importe le respect des gestes barrières.

Quand l'**âge** augmente d'un an, la probabilité de respecter les gestes barrières et de modifier ses habitudes sociales augmente d'une manière significative de 0.038. A l'inverse, cela diminue de manière significative la probabilité de ne pas respecter les mesures sanitaires et de maintenir ses relations sociales. Cela pourrait s'expliquer par la plus grande maturité des étudiants plus âgés, c'est un argument, bien entendu, purement basé sur notre intuition.

Nous pouvons rappeler que pour une variable de type qualitatif binaire, l'effet marginal mesure l'impact sur la probabilité étudiée d'un changement de 0 à 1 de cette

variable. Nous avons donc pour la variable **Sexe** des effets marginaux récupérant l'impact sur la probabilité d'un changement d'un homme codé 0 à une femme codé 1 dans notre base. Une femme a donc plus de chance de respecter les gestes barrières et de changer ses habitudes sociales qu'un homme (la probabilité jointe de ces deux événements augmentent de 0.146 quand l'individu est une femme), à contrario, la probabilité de ne pas respecter les gestes barrières et de ne pas modifier ses relations diminue de 0.142 si l'individu est une femme. Ces deux modalités étant significatives au seuil de risque de 1%. Au regard de ce résultat, nous sommes surpris de constater que ce sont les femmes, d'une manière générale, qui respectent le plus les gestes barrières et modifient leurs habitudes sociales conjointement.

Il est intéressant d'observer qu'un étudiant **boursier** a une probabilité positive et significative de respecter les gestes barrières, peu importe s'il change d'habitudes sociales ou non, par rapport à un étudiant non boursier. (la probabilité augmente de 0.059). Les coefficients s'inversent de manière significative aussi pour les modalités ($Y_1 = 0, Y_2 = 0$) et ($Y_1 = 0, Y_2 = 1$).

Tout comme la variable Boursier, la variable **Etude** est significative pour expliquer les quatre modalités du modèle biprobit. Contrairement à ce que nous avons présupposé lors de l'intégration de cette variable explicative à notre modèle, un étudiant en licence a étonnamment plus de chance de respecter les gestes barrières tout en modifiant ses habitudes qu'un étudiant en master avec une probabilité augmentant de 0.039, significative à un seuil de 10%). Nous pourrions l'expliquer par le fait que les étudiants en licence ont souffert de changements plus importants en ce qui concerne l'enseignement notamment, ce qui est moins le cas des étudiants en master.

Pour terminer, les variables **News** et **Politiques** ont un effet significatif pour les deux premières modalités uniquement. Un individu qui s'intéresse activement aux nouvelles qui touchent au Covid-19 augmente la probabilité de respecter les gestes barrières et de modifier ses habitudes de 0.153 par rapport à un individu qui ne le fait pas et diminue parallèlement la probabilité de ne pas respecter les gestes barrières ni de changer ses habitudes sociales de 0.166. La satisfaction du répondant quant à la gestion de la crise par le gouvernement actuel est positivement corrélée avec la probabilité de respecter les gestes et de changer ses habitudes. Quand la satisfaction augmente d'un point (sur notre échelle de 1 à 10) la probabilité de respecter les deux mesures simultanément augmente de 0.042. La probabilité de ne respecter aucune des deux diminue de 0.048 pour chaque point supplémentaire.

En règle général, on peut constater que peu de variables sont significatives sur les quatre modalités, notamment sur les modalités ($Y_1 = 1, Y_2 = 0$) et ($Y_1 = 0, Y_2 = 1$) qui sont les deux modalités pour lesquelles le modèle affiche une mauvaise qualité de prédiction. Nous pouvions donc nous attendre à avoir peu de variables significatives dans ces deux catégories. La variable Santé n'est jamais significative, ceci peut être dû au fait que son ajout est déterminé par la méthodologie stepwise, conduite sur deux Probit séparés, et non pas un Biprobit.

Maintenant que nous avons interprété les effets marginaux sur les modalités conjointes il peut être intéressant de calculer les effets marginaux sur les probabilités marginales du modèle. C'est ce que résume le tableau ci-dessous.

Variables	Effets Marginaux	
	$P(Y_1 = 1)$	$P(Y_2 = 1)$
PRisque	0	.141***
Age	.034***	.046***
Sexe	.165***	.124***
Boursier	.091*	0
Etude	0	.116*
Sante	0	.079
News	.158***	.161***
Politique	.037***	.054***

Table 4 – Effets Marginaux sur probabilités marginales, niveaux de significativité : * = 0.1, ** = 0.05, *** = 0.01

Au seuil de risque de 1%, le fait d'être plus âgé, d'être une femme, de suivre les informations liées au covid et enfin de considérer la politique du gouvernement satisfaisante dans la gestion de la crise, a un impact positif sur la probabilité de respecter les gestes barrières indépendamment des habitudes sociales. Avec un seuil de 10%, être Boursier aussi impacte positivement ce respect des gestes barrières indépendamment de Y2. Les effets marginaux des probabilités marginales des variables PRisque, Etude et Santé, ne sont pas significatives.

Sur le même principe, les variables PRisque, Age, Sexe, News et politique ont un impact positif sur la probabilité de changer ses habitudes en terme de relations sociales indépendamment du respect des gestes barrières, au seuil de risque de 1% et 10% pour Etude.

Variables	Effets Marginaux	
	$P(Y_1 = 1 Y_2 = 1)$	$P(Y_2 = 1 Y_1 = 1)$
PRisque	-.032**	.126***
Age	.026**	.033***
Sexe	.154***	.069
Boursier	.098*	-.021*
Etude	-.027	.103*
Santé	-.017	.067
News	.138***	.105**
Politique	.028*	.038***

Table 5 – Effets Marginaux sur probabilités conditionnelles, niveaux de significativité : * = 0.1, ** = 0.05, *** = 0.01

Nous pouvons terminer les interprétations avec les effets marginaux en abordant les probabilités conditionnelles cette fois. Sachant que l'étudiant respecte les gestes barrières, avoir un parent à risque augmente sa probabilité de modifier ses habitudes sociales de 0.126, avec un seuil de risque de 1%. De la même façon, sachant que l'étudiant a modifié ses habitudes sociales, qu'il suit les informations, il augmentera significativement sa probabilité de respecter les gestes barrières de 0.138.

IV.3 Validité de notre modèle

IV.3.1 Représentativité

Il est important de vérifier la représentativité de notre échantillon lorsque nous souhaitons étudier un phénomène sur une population donnée. Dans notre cas nous voulions étudier le respect des gestes barrières ainsi que le changement en terme de relations sociales parmi les étudiants de l'IAE de Nantes. Nous avons donc vérifié si la répartition des étudiants de l'échantillon en fonction de leur niveau d'étude ainsi que de leur genre correspondait à la répartition générale des étudiants de l'IAE de Nantes pour l'année 2020. A partir des données transmises par l'Université, nous avons fait le choix de sommer les **Diplômes Universitaires** (DU) aux Licenciés afin de faciliter le comptage et les interprétations.

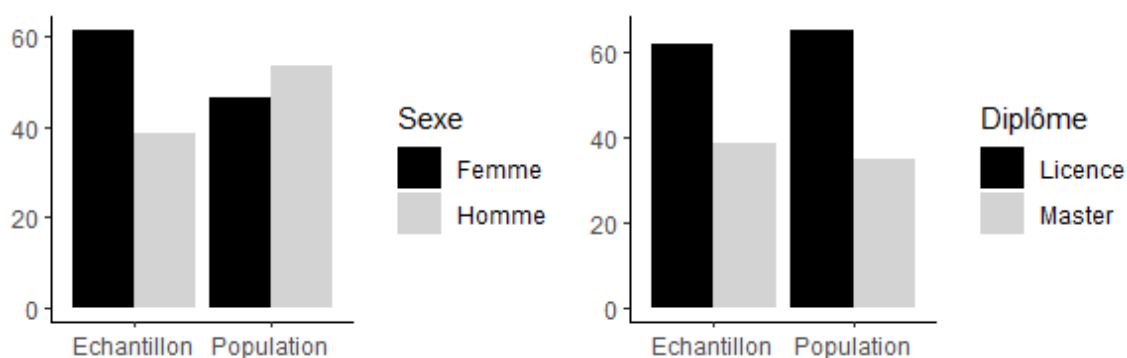


Figure 8 – Répartition des effectifs entre l'échantillon et la population étudiée

Les deux diagrammes à bar ci-dessus nous permettent de vérifier la représentativité de notre échantillon en terme de répartition homme/femme ainsi que la proportion entre les étudiants en Licences et ceux en master. Nous pouvons constater que notre échantillon est bien représentatif de la population en ce qui concerne le niveau d'étude mais malheureusement ce n'est pas le cas pour le genre. En effet ce sont majoritairement des femmes qui ont pris le temps de répondre à notre questionnaire malgré une proportion plus importante de la gente masculine au sein de l'IAE selon les données de 2020.

IV.3.2 Robustesse

Utiliser un jeu de données pour estimer le modèle et un second pour la prédiction permet de s'affranchir de plusieurs problèmes. Si le modèle est testé sur les mêmes données utilisées pour l'estimer, la qualité de prédiction peut être influencée par le fait que le modèle est très spécifiquement adapté à ces données. Utiliser un jeu de données qui n'a pas été utilisé pour les prédictions, permet de passer outre un éventuel sur-apprentissage ou biais dû aux fluctuations d'échantillonnage. Cela permet de vérifier si le modèle est stable, qu'il explique correctement nos variables, peu importe l'échantillon de variables explicatives pris en compte. Ainsi nous nous assurons que le modèle n'est pas biaisé, ni par le sur-apprentissage, ni pas les fluctuations d'échantillonnage. Il n'y a pas de règles décisives sur la taille des échantillons, d'entraînement et de test. Nous avons choisi 80%-20%, avec 20% pour le test, en accord avec le consensus général. Le but final étant de garder suffisamment de données pour avoir des prédictions en nombre suffisant. Ici, nous conservons 86 observations pour le test, qui nous semblent pleinement suffisantes. Les qualités de prédictions sont similaires, que nous soyons sur le jeu d'entraînement ou de test, ce qui nous indique que le modèle est stable. D'autres méthodes, comme la cross-validation par exemple (qui repose cependant sur le même principe) auraient pu être utilisées pour être encore plus rigoureux.

V Conclusions

Nous avons construit plusieurs modèles Biprobit, à l'aide de différentes méthodes. Nous avons sélectionné ensuite le meilleur de ces modèles pour prédire et expliquer le respect des gestes barrières et la modification des habitudes sociales. Deux des modèles ont été construits avec les variables que nous pensions importantes : un modèle 'complet' qui contient toutes les variables utilisables et un modèle 'partiel' qui contient un sous-ensemble de ces variables choisies selon leur significativité et notre intuition sur leur importance explicative. Nous avons pris ensuite une approche plus rigoureuse pour la sélection de variables, en l'occurrence, nous avons utilisé les algorithmes stepwise (backward et forward). Nous avons choisi le modèle 'partiel' comme étant le meilleur de tous. Malgré une qualité de prédiction légèrement inférieure (44.93% contre 46.68% pour le modèle 'complet'), les prédictions justes, étant mieux réparties, ont donc permis de mieux identifier les phénomènes en jeu. De plus les critères d'informations pointaient également dans la même direction. Nous avons utilisé, par conséquent, le modèle 'partiel' pour estimer les effets marginaux de chaque variables sur les probabilités jointes, marginales et conditionnelles. Nous découvrons qu'avoir un parent à risque, le sexe, l'âge, la satisfaction vis-à-vis de la politique du gouvernement et la fréquence de consultation des médias ont eu un impact positif sur la probabilité de respecter les gestes barrières et de modifier ses habitudes sociales. La critique majeure de notre modèle s'articule autour de la mauvaise qualité de prédictions, notamment sur la modalité ($Y_1 = 1, Y_2 = 0$). Cette modalité n'est jamais prédite par notre modèle alors qu'elle est belle et bien présente dans nos échantillons ¹⁵.

15. Ce problème est présent dans les deux échantillons, d'entraînement et de test

VI Annexes

Annexe 1 : Résultat Test de Rosner

```

$all.stats
      i Mean.i      SD.i value Obs.Num      R.i+1 lambda.i+1 outlier
1    0 21.11486 4.283187    50     212 6.743841    3.831464    TRUE
2    1 21.04966 4.061436    45     220 5.897012    3.830859    TRUE
3    2 20.99548 3.902436    44     115 5.894915    3.830252    TRUE
4    3 20.94331 3.749419    42     424 5.615987    3.829643    TRUE
5    4 20.89545 3.616336    41     181 5.559369    3.829033    TRUE
6    5 20.84966 3.490390    41     425 5.773091    3.828422    TRUE
7    6 20.80365 3.358487    40     278 5.715772    3.827809    TRUE
8    7 20.75973 3.233913    40     317 5.949534    3.827194    TRUE
9    8 20.71560 3.103107    39     169 5.892289    3.826578    TRUE
10   9 20.67356 2.979827    37      40 5.478988    3.825960    TRUE
11  10 20.63594 2.877996    37     362 5.685920    3.825340    TRUE
12  11 20.59815 2.771414    37     369 5.918223    3.824719    TRUE
13  12 20.56019 2.659499    37     376 6.181546    3.824096    TRUE

```



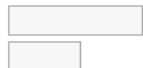


Annexe 2 : Statistiques descriptives à partir du package summarytools

Data Frame Summary

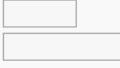
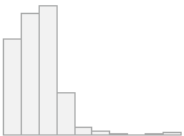

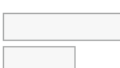


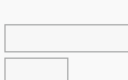

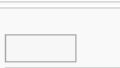
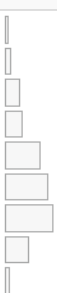
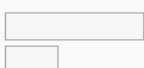
df_clean_ind

Dimensions: 431 x 16

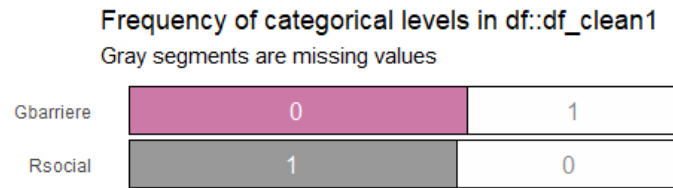
Duplicates: 15

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	Gbarriere [numeric]	Min : 0 Mean : 0.4 Max : 1	0: 263 (61.0%) 1: 168 (39.0%)		431 (100%)	0 (0%)
2	Rsocial [numeric]	Min : 0 Mean : 0.6 Max : 1	0: 176 (40.8%) 1: 255 (59.2%)		431 (100%)	0 (0%)
3	GParents [numeric]	Min : 0 Mean : 0.3 Max : 1	0: 281 (65.2%) 1: 150 (34.8%)		431 (100%)	0 (0%)
4	PRisque [numeric]	Min : 0 Mean : 0.6 Max : 1	0: 173 (40.1%) 1: 258 (59.9%)		431 (100%)	0 (0%)
5	ParentMed [numeric]	Min : 0 Mean : 0.1 Max : 1	0: 370 (85.9%) 1: 61 (14.1%)		431 (100%)	0 (0%)

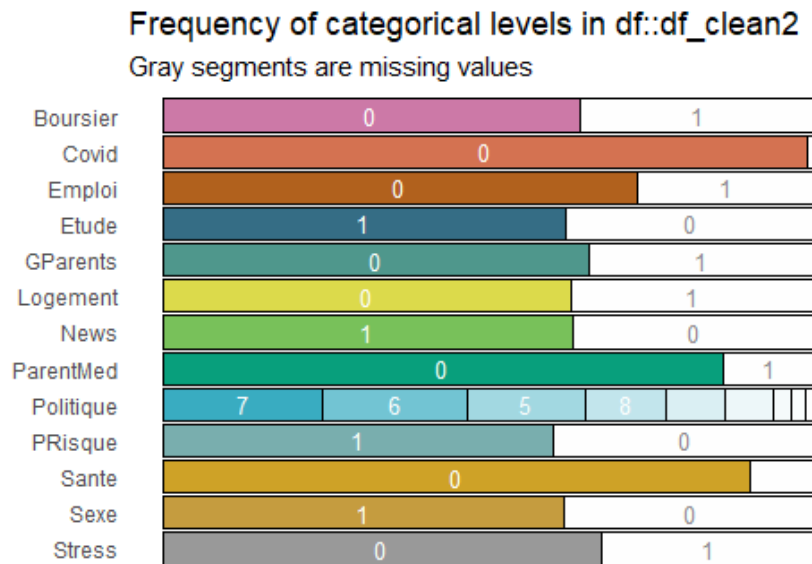
Annexe 2 : Suite

6	Sexe [numeric]	Min : 0 Mean : 0.6 Max : 1	0: 166 (38.5%) 1: 265 (61.5%)		431 (100%)	0 (0%)
7	Age [numeric]	Mean (sd) : 20.5 (2.5) min < med < max: 17 < 20 < 36 IQR (CV) : 3 (0.1)	15 distinct values		431 (100%)	0 (0%)
8	Etude [numeric]	Min : 0 Mean : 0.6 Max : 1	0: 165 (38.3%) 1: 266 (61.7%)		431 (100%)	0 (0%)
9	Logement [numeric]	Min : 0 Mean : 0.4 Max : 1	0: 269 (62.4%) 1: 162 (37.6%)		431 (100%)	0 (0%)
10	Boursier [numeric]	Min : 0 Mean : 0.4 Max : 1	0: 275 (63.8%) 1: 156 (36.2%)		431 (100%)	0 (0%)
11	Sante [numeric]	Min : 0 Mean : 0.1 Max : 1	0: 388 (90.0%) 1: 43 (10.0%)		431 (100%)	0 (0%)
12	Stress [numeric]	Min : 0 Mean : 0.3 Max : 1	0: 289 (67.0%) 1: 142 (33.0%)		431 (100%)	0 (0%)
13	Covid [numeric]	Min : 0 Mean : 0 Max : 1	0: 425 (98.6%) 1: 6 (1.4%)		431 (100%)	0 (0%)
14	News [numeric]	Min : 0 Mean : 0.6 Max : 1	0: 160 (37.1%) 1: 271 (62.9%)		431 (100%)	0 (0%)
15	Politique [numeric]	Mean (sd) : 5.8 (1.7) min < med < max: 1 < 6 < 9 IQR (CV) : 2 (0.3)	1: 7 (1.6%) 2: 12 (2.8%) 3: 32 (7.4%) 4: 39 (9.0%) 5: 78 (18.1%) 6: 95 (22.0%) 7: 106 (24.6%) 8: 53 (12.3%) 9: 9 (2.1%)		431 (100%)	0 (0%)
16	Emploi [numeric]	Min : 0 Mean : 0.3 Max : 1	0: 313 (72.6%) 1: 118 (27.4%)		431 (100%)	0 (0%)

Annexe 3 : Fréquence des variables à expliquées



Annexe 4 : Fréquence des variables explicatives



Annexe 5 : Dépendances des variables Stress et Covid

. tabulate ParentMed Covid, chi2

ParentMed	Covid		Total
	0	1	
0	368	2	370
1	57	4	61
Total	425	6	431

Pearson chi2(1) = 13.8104 Pr = 0.000

. tabulate Sexe Stress, chi2

Sexe	Stress		Total
	0	1	
0	136	30	166
1	153	112	265
Total	289	142	431

Pearson chi2(1) = 27.0386 Pr = 0.000

Annexe 6 : Effectif pour chacune des modalités

. tab Gbarriere Rsocial

Gbarriere	Rsocial		Total
	0	1	
0	137	126	263
1	39	129	168
Total	176	255	431

Annexe 7 : Estimation du modèle complet sur la base train

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
Gbarriere						
PRisque	.096478	.1456611	0.66	0.508	-.1890125	.3819685
ParentMed	.1003008	.2112964	0.47	0.635	-.3138327	.5144342
Age	.1149524	.0452183	2.54	0.011	.0263262	.2035786
Sexe	.4729148	.1482682	3.19	0.001	.1823143	.7635152
Boursier	.2455934	.1468225	1.67	0.094	-.0421734	.5333603
Etude	.1382317	.1967595	0.70	0.482	-.2474099	.5238733
Sante	.2950908	.253343	1.16	0.244	-.2014525	.791634
News	.4811238	.1480259	3.25	0.001	.1909983	.7712492
Politique	.0859255	.0424868	2.02	0.043	.0026529	.1691982
_cons	-3.979468	1.091757	-3.65	0.000	-6.119272	-1.839664
Rsocial						
PRisque	.3327774	.1447474	2.30	0.022	.0490777	.6164772
Boursier	-.2248822	.1482506	-1.52	0.129	-.515448	.0656836
ParentMed	-.0398263	.2121662	-0.19	0.851	-.4556645	.3760118
Age	.141741	.0405548	3.50	0.000	.0622551	.2212269
Sexe	.2536415	.1456321	1.74	0.082	-.0317923	.5390752
Etude	.3702183	.1993747	1.86	0.063	-.020549	.7609856
Sante	.5247826	.277921	1.89	0.059	-.0199326	1.069498
News	.3280752	.1479758	2.22	0.027	.0380479	.6181024
Politique	.1297654	.0444457	2.92	0.004	.0426534	.2168774
_cons	-4.102139	1.032034	-3.97	0.000	-6.124889	-2.079389
/athrho	.4121031	.0975645	4.22	0.000	.2208802	.603326
rho	.390257	.0827054			.2173569	.5394121
Wald test of rho=0: chi2(1) = 17.8414 Prob > chi2 = 0.0000						

Annexe 8 : Estimation du modèle partiel sur la base train

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Gbarriere						
Age	.0964526	.0326373	2.96	0.003	.0324847	.1604206
Sexe	.4584479	.146255	3.13	0.002	.1717933	.7451025
Boursier	.2805866	.1424854	1.97	0.049	.0013203	.5598529
News	.4799953	.1482615	3.24	0.001	.1894081	.7705825
Politique	.0784782	.0422774	1.86	0.063	-.0043841	.1613405
_cons	-3.376238	.7578786	-4.45	0.000	-4.861653	-1.890823
Rsocial						
PRisque	.3290161	.1388408	2.37	0.018	.0568931	.6011391
Age	.135987	.0390572	3.48	0.000	.0594363	.2125377
Sexe	.2586224	.1455675	1.78	0.076	-.0266847	.5439295
Etude	.35023	.192743	1.82	0.069	-.0275394	.7279993
Sante	.4332937	.2785217	1.56	0.120	-.1125988	.9791862
News	.3160435	.147239	2.15	0.032	.0274603	.6046267
Politique	.1320513	.0441042	2.99	0.003	.0456087	.218494
_cons	-4.059223	.9875323	-4.11	0.000	-5.994751	-2.123696
/athrho	.4105322	.0971908	4.22	0.000	.2200417	.6010226
rho	.3889244	.0824895			.2165578	.5377768
Wald test of rho=0:						
			chi2(1) =	17.842	Prob > chi2 = 0.0000	

Annexe 9 : Sélection de variables méthodologie Stepwise avec Sexe

```

> stepwise probit Gbarriere GParents PRisque ParentMed Age Sexe Etude Bourasier Logement
> Sante Stress New Politique pr(0.1)
> begin with full model
p = 0.5907 >= 0.1000 removing PRisque
p = 0.4828 >= 0.1000 removing Logement
p = 0.4655 >= 0.1000 removing ParentMed
p = 0.4115 >= 0.1000 removing GParents
p = 0.3814 >= 0.1000 removing Etude
p = 0.2443 >= 0.1000 removing Sante

Probit regression                                Number of obs      =       345
                                                LR chi2(6)         =      42.56
                                                Prob > chi2         =     0.0000
                                                Pseudo R2          =     0.0912

Log likelihood = -212.06997

```

Gbarriere	Coeff.	Std. Err.	z	P> z	[95% Conf. Interval]	
Bourasier	2.475742	1.485245	1.67	0.096	-0.435285	.538677
Politique	0.840228	0.435866	1.93	0.054	-.0014054	.159451
Stress	.3707255	.1559322	2.38	0.017	.0651043	.6763475
Age	0.952224	.0282271	3.28	0.001	.3711982	.1478465
Sexe	.3873933	.1514533	2.56	0.011	.0905503	.6842364
Etude	.4411704	.1516751	2.91	0.004	.1438927	.738448
_cons	-3.370088	.6939111	-4.86	0.000	-4.730128	-2.010047

```

> stepAe probit Gbarriere PRISque ParentMed Age Sexe Etude Boursier Logement
> > Santa Stress News Polittique_pe(0.1)
begin with empty model
p = 0.0004 < 0.1000 adding Stress
p = 0.0026 < 0.1000 adding Age
p = 0.0068 < 0.1000 adding News
p = 0.0082 < 0.1000 adding Sexe
p = 0.0743 < 0.1000 adding Polittique
p = 0.0955 < 0.1000 adding Boursier

Probit regression
Number of obs      =      345
LR chi2(4)         =      42.56
Prob > chi2        =      0.0000
Pseudo R2          =      0.0912

Log likelihood = -212.06997

```

Gbarriere	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Stress	.3707259	.1559322	2.38	0.017	.0651043 .6763475
Age	.0952224	.0282271	3.28	0.001	.0371982 .1478465
News	.4411704	.1516751	2.91	0.004	.1438927 .738448
Sexe	.3873933	.1514533	2.56	0.011	.0905503 .6842364
Polittique	.0840228	.0435866	1.93	0.054	-.0114054 .169451
Boursier	.2475742	.1465245	1.67	0.096	-.0435265 .5365777
_cons	-3.370088	.6939111	-4.86	0.000	-4.730128 -2.010047

```

> stepwise probit Rsocial GParents Priisque ParentMed Age Sexe Etude Boursier Logement
> Sante Stress News Politique, pr(0.1)
begin with full model
p = 0.9491 >= 0.1000 removing ParentMed
p = 0.5749 >= 0.1000 removing Logement
p = 0.4521 >= 0.1000 removing GParents
p = 0.2427 >= 0.1000 removing Sexe
p = 0.1656 >= 0.1000 removing Boursier
p = 0.1000 >= 0.1000 removing News

Probit regression

Number of obs      =      345
LR chi2(4)         =      36.28
Prob > chi2        =      0.0000
Pseudo R2          =      0.0830

Log likelihood = -211.93103

```

Rsocial	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Stress	4.952234	1.58568	3.12	0.002	1.844358 . 806011
Priisque	3.211483	1.448411	2.22	0.027	0.37265 . 6050316
Politique	1.422538	0.433569	3.28	0.001	0.5572759 . 2.272318
Age	1.339284	0.056376	3.07	0.002	0.0504805 . 2.239764
Sante	5.278323	2.718157	1.94	0.052	-.0049167 . 1.060581
Etude	4.235854	2.027076	2.09	0.037	0.026289 . 8.205858
_cons	-4.042397	1.091081	-3.71	0.000	-6.181477 -1.904517

```

. stepwise probit rsocial cParents Frisique ParentMed Age Sexe Etude Boursier Logement
> > Santa Stress New Politique, pse(0.1)
                                begin with empty model.

p = 0.0011 < 0.1000 adding Stress
p = 0.0135 < 0.1000 adding Politique
p = 0.0111 < 0.1000 adding Age
p = 0.0185 < 0.1000 adding Frisique
p = 0.0376 < 0.1000 adding Etude
p = 0.0522 < 0.1000 adding Santa

Probit regression                                Number of obs   =       345
                                                LR chi2(6)       =       38.28
                                                Prob > chi2      =       0.0000
                                                Pseudo R2       =       0.0830

Log likelihood = -211.33103

```

rsocial	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Stress	.4952234	.158568	3.12	0.002	.1844358 .806011
Politique	.1422538	.0433569	3.28	0.001	.0572759 .2272318
Age	.1399284	.0456376	3.07	0.002	.0504005 .2293764
Frisique	.3211483	.1448411	2.22	0.027	.037265 .605316
Etude	.4235984	.2027076	2.09	0.037	-.0262589 .820858
Santa	.5278523	.2718157	1.94	0.052	-.0049172 .1.060581
_cons	-4.042997	1.091082	-3.71	0.000	-.6.181477 -1.904517

Annexe 10 : Sélection de variables méthodologie Stepwise sans Sexe

```
. stepwise probit Gbarriere GParents PRisque ParentMed Age Etude Boursier Logem
> ent Sante News Politique, pr(0.1)
begin with full model
```

p = 0.7629 >= 0.1000 removing Etude
p = 0.7364 >= 0.1000 removing GParents
p = 0.5544 >= 0.1000 removing ParentMed
p = 0.4931 >= 0.1000 removing Logement
p = 0.4471 >= 0.1000 removing PRisque
p = 0.2545 >= 0.1000 removing Sante
p = 0.1443 >= 0.1000 removing Boursier

Probit regression Number of obs = 345
 LR chi2(3) = 24.61
 Prob > chi2 = 0.0000
Log likelihood = -221.04836 Pseudo R2 = 0.0527

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
News	.4607724	.1468018	3.14	0.002	.1730461 .7484987
Age	.0991808	.0282747	3.51	0.000	.0437635 .1545982
Politique	.0773138	.0423864	1.82	0.068	-.0057621 .1603897
_cons	-3.023733	.6803721	-4.44	0.000	-4.357238 -1.690228

```
. stepwise probit Gbarriere GParents PRisque ParentMed Age Etude Boursier Logem
> ent Sante News Politique, pe(0.1)
begin with empty model
```

p = 0.0011 < 0.1000 adding Age
p = 0.0017 < 0.1000 adding News
p = 0.0681 < 0.1000 adding Politique

Probit regression Number of obs = 345
 LR chi2(3) = 24.61
 Prob > chi2 = 0.0000
Log likelihood = -221.04836 Pseudo R2 = 0.0527

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Age	.0991808	.0282747	3.51	0.000	.0437635 .1545982
News	.4607724	.1468018	3.14	0.002	.1730461 .7484987
Politique	.0773138	.0423864	1.82	0.068	-.0057621 .1603897
_cons	-3.023733	.6803721	-4.44	0.000	-4.357238 -1.690228

```
. stepwise probit Rsocial GParents PRisque ParentMed Age Etude Boursier Logemen
> t Sante News Politique, pr(0.1)
begin with full model
```

p = 0.9548 >= 0.1000 removing ParentMed
p = 0.6387 >= 0.1000 removing GParents
p = 0.4994 >= 0.1000 removing Logement
p = 0.1139 >= 0.1000 removing Boursier

Probit regression Number of obs = 345
 LR chi2(6) = 32.58
 Prob > chi2 = 0.0000
Log likelihood = -214.17792 Pseudo R2 = 0.0707

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
News	.3021387	.1459138	2.07	0.038	.016153 .5881244
PRisque	.3592988	.1437102	2.50	0.012	.0776319 .6409656
Politique	.1366779	.0431505	3.17	0.002	.0521044 .2212514
Age	.1418414	.0459895	3.08	0.002	.0517037 .2319791
Etude	.3556419	.201554	1.76	0.078	-.0393967 .7506806
Sante	.5122666	.2692334	1.90	0.057	-.0154212 1.039954
_cons	-4.065617	1.100614	-3.69	0.000	-6.222781 -1.908453

```
. * forward probit(2)
stepwise probit Rsocial GParents PRisque ParentMed Age Etude Boursier Logemen
> t Sante News Politique, pe(0.1)
begin with empty model
```

p = 0.0131 < 0.1000 adding Age
p = 0.0076 < 0.1000 adding Politique
p = 0.0082 < 0.1000 adding PRisque
p = 0.0331 < 0.1000 adding News
p = 0.0576 < 0.1000 adding Sante
p = 0.0776 < 0.1000 adding Etude

Probit regression Number of obs = 345
 LR chi2(6) = 32.58
 Prob > chi2 = 0.0000
Log likelihood = -214.17792 Pseudo R2 = 0.0707

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Age	.1418414	.0459895	3.08	0.002	.0517037 .2319791
Politique	.1366779	.0431505	3.17	0.002	.0521044 .2212514
PRisque	.3592988	.1437102	2.50	0.012	.0776319 .6409656
News	.3021387	.1459138	2.07	0.038	.016153 .5881244
Sante	.5122666	.2692334	1.90	0.057	-.0154212 1.039954
Etude	.3556419	.201554	1.76	0.078	-.0393967 .7506806
_cons	-4.065617	1.100614	-3.69	0.000	-6.222781 -1.908453

Annexe 11 : Estimation du modèle Biprobit avec Sexe

	Robust					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Gbarriere						
Age	.0969327	.0339617	2.85	0.004	.0303689	.1634965
Sexe	.4528205	.1459938	3.10	0.002	.1666779	.7389632
News	.4923969	.1478817	3.33	0.001	.2025541	.7822398
Politique	.0701637	.0424129	1.65	0.098	-.012964	.1532915
_cons	-3.239368	.7832092	-4.14	0.000	-4.77443	-1.704306
Rsocial						
Age	.1363347	.0388232	3.51	0.000	.0602426	.2124268
PRisque	.3363633	.139077	2.42	0.016	.0637774	.6089493
Sante	.4336219	.2797492	1.55	0.121	-.1146765	.9819203
Etude	.3565982	.1925725	1.85	0.064	-.020837	.7340333
Politique	.1322424	.0441405	3.00	0.003	.0457287	.2187562
News	.315923	.1471496	2.15	0.032	.0275152	.6043309
Sexe	.2607149	.1455273	1.79	0.073	-.0245134	.5459431
_cons	-4.077103	.9823265	-4.15	0.000	-6.002427	-2.151778
/athrho	.3890195	.0963654	4.04	0.000	.2001468	.5778922
rho	.3705146	.0831362			.1975164	.5211317

Wald test of rho=0: chi2(1) = 16.2967 Prob > chi2 = 0.0001

Annexe 12 : Estimation du modèle Biprobit sans Sexe

	Robust				[95% Conf. Interval]	
	Coef.	Std. Err.	z	P> z		
Gbarriere						
Age	.0985944	.0323654	3.05	0.002	.0351595	.1620294
News	.4567353	.146241	3.12	0.002	.1701082	.7433623
Politique	.0741309	.0419003	1.77	0.077	-.0079923	.156254
_cons	-2.989398	.7447681	-4.01	0.000	-4.449117	-1.529679
Rsocial						
Age	.1329486	.0375151	3.54	0.000	.0594204	.2064768
PRisque	.3393654	.1383069	2.45	0.014	.068289	.6104419
News	.2985456	.1460228	2.04	0.041	.0123463	.584745
Politique	.1345559	.0436871	3.08	0.002	.0489308	.220181
Etude	.3279174	.1888856	1.74	0.083	-.0422915	.6981264
Sante	.4249703	.275935	1.54	0.124	-.1158525	.965793
_cons	-3.83602	.9365317	-4.10	0.000	-5.671589	-2.000452
/athrho	.4106942	.0958507	4.28	0.000	.2228304	.5985581
rho	.389062	.0813418			.2192141	.5360227

Wald test of rho=0: chi2(1) = 18.3589 Prob > chi2 = 0.0000

Annexe 13 : Estimation du modèle Biprobit Récursif

```
. biprobit ( Gbarriere = Rsocial PRisque Age Sexe Boursier News Politique
> ) (Rsocial = PRisque Age Sexe Boursier News Politique), robust

Fitting comparison equation 1:

Iteration 0:  log pseudolikelihood = -233.3512
Iteration 1:  log pseudolikelihood = -204.53975
Iteration 2:  log pseudolikelihood = -204.49018
Iteration 3:  log pseudolikelihood = -204.49018

Fitting comparison equation 2:

Iteration 0:  log pseudolikelihood = -230.47022
Iteration 1:  log pseudolikelihood = -215.35795
Iteration 2:  log pseudolikelihood = -215.27419
Iteration 3:  log pseudolikelihood = -215.27417

Comparison:  log pseudolikelihood = -419.76436

Fitting full model:

Iteration 0:  log pseudolikelihood = -419.76436 (not concave)
Iteration 1:  log pseudolikelihood = -419.76317 (not concave)
Iteration 2:  log pseudolikelihood = -419.76158 (not concave)
Iteration 3:  log pseudolikelihood = -419.76033 (not concave)
Iteration 4:  log pseudolikelihood = -419.75924 (not concave)
Iteration 5:  log pseudolikelihood = -419.75834 (not concave)
Iteration 6:  log pseudolikelihood = -419.75754 (not concave)
Iteration 7:  log pseudolikelihood = -419.75675 (not concave)
Iteration 8:  log pseudolikelihood = -419.75598 (not concave)
```