

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PARTICIPATION AU MARCHÉ DU TRAVAIL ET MALADIES CHRONIQUES :
UNE APPROCHE BAYÉSIENNE

THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN ÉCONOMIQUE

PAR
FRÉDÉRIC BROUSSAU

SEPTEMBRE 2014

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je désire prendre ces quelques lignes pour remercier toutes les personnes sans qui cette thèse n'aurait sans doute jamais vu le jour.

Merci à mon directeur Monsieur Pierre Ouellette et mon co-directeur Monsieur Phil Merrigan pour m'avoir donné de précieux conseils tout au long de ma thèse. Je tiens à les remercier pour tout le temps consacré à mon endroit. Merci à tous les professeurs et employés du département des sciences économiques de l'UQAM ainsi qu'aux employés du CIQSS

Merci à mes parents Raymonde et Roland pour m'avoir supporté jusqu'à la fin. Merci également aux amis et à ma famille qui m'ont permis de temps en temps de m'échapper un petit peu de mon difficile quotidien. Merci à ma copine Annick qui m'a supporté et aidé avec une patience exemplaire.

Et, merci à tous ceux qui de près et de loin m'ont soutenu et encouragé.

TABLE DES MATIÈRES

REMERCIEMENTS	ii
TABLE DES MATIÈRES	iii
LISTE DES FIGURES.....	vii
LISTE DES TABLEAUX.....	viii
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES	x
RÉSUMÉ	xii
INTRODUCTION	1
CHAPITRE I	
PROBIT MULTIVARIÉ.....	6
1.1 Introduction.....	6
1.2 Le modèle probit multivarié.....	6
1.3 Estimateur du maximum de vraisemblance simulé.....	9
1.4 Estimateur bayésien	11
1.4.1 Les densités <i>a priori</i>	12
1.4.2 Distribution conditionnelle <i>a posteriori</i> du modèle probit multivarié.....	16
CHAPITRE II	
L'ALGORITHME MCMC	20
2.1 Introduction.....	20
2.2 Échantillonnage des variables latentes.....	21
2.3 Échantillonnage des coefficients de régression.....	23
2.4 Comparaison des programmes bayésiens.....	25

CHAPITRE III	
PROBIT MULTIVARIÉ AVEC EFFETS INDIVIDUELS	28
3.1 Introduction	28
3.2 Modèle multivarié (SURE) avec individuels hiérarchiques.....	29
3.3 Modèle probit multivarié avec effets individuels hiérarchiques	32
3.4 Simulations du probit multivarié avec effets individuels hiérarchiques	37
CHAPITRE IV	
MARCHÉ DU TRAVAIL ET MALADIES CHRONIQUES	42
4.1 Introduction	42
4.1.1 La santé et ses facteurs de risque	45
4.1.2 Santé et participation au marché du travail	49
4.2 Les données	51
4.3 Modèle probit endogène avec structure réursive.....	58
4.4 Identification du modèle	61
4.5 Les résultats.....	63
4.5.1 L'arthrite	65
4.5.2 Les maux de dos.....	67
4.5.3 Les maladies pulmonaires obstructives chroniques	69
4.5.4 Les maladies cardio-vasculaires.....	71
4.5.5 Le diabète.....	74
4.5.6 Les coefficients de corrélation	76
4.5.7 Variances et covariances des effets individuels hiérarchiques	77
4.5.8 La participation au marché du travail	81
4.5.9 Les effets marginaux directs et indirects.....	87
4.5.10 Comparaison de modèle.....	91

4.5.11 Analyse de sensibilité.....	94
CONCLUSION.....	96
ANNEXE A	
PROBIT MULTIVARIÉ AVEC EFFETS INDIVIDUELS	100
ANNEXE B	
PROBIT UNIVARIÉ	104
APPENDICE C	
PROBIT UNIVARIÉ AVEC EFFETS INDIVIDUELS HIÉRARCHIQUES	107
APPENDICE D	
PROBIT MULTIVARIÉ.....	110
RÉFÉRENCES.....	120

LISTE DES FIGURES

Figure		Page
4.1	Santé auto-déclarée moyenne - revenu (ESCC 2007).....	46
4.2	Santé auto-déclarée moyenne - scolarité (ESCC 2007)	46
4.3	Habitudes de vie selon le revenu - tabac (ESCC 2007)	48
4.4	Habitudes de vie selon le revenu – activité physique (ESCC 2007).....	48

LISTE DES TABLEAUX

Tableau		Page
1.1	Distribution conditionnelle <i>a posteriori</i> du modèle probit multivarié	19
2.1	Temps de traitement entre différents programmes.....	26
3.1	Distribution conditionnelle <i>a posteriori</i> du modèle SURE avec effets individuels hiérarchiques	32
3.2	Distribution conditionnelle <i>a posteriori</i> du modèle probit multivarié avec effets individuels aléatoires	33
3.3	Résultats des simulations, coefficients de régression	39
3.4	Résultats des simulations : corrélations du terme d'erreur	40
3.5	Résultats des simulations : moyenne des effets individuels hiérarchiques	40
3.6	Résultats des simulations : variances des effets individuels hiérarchiques	40
3.7	Résultats des simulations : covariances des effets individuels hiérarchiques	41
4.1	Facteurs de risque des maladies chroniques.....	43
4.2	Statistiques descriptives	54
4.3	Participation au marché du travail des individus selon qu'ils sont atteints ou non d'une maladie chronique	55
4.4	Comorbidités entre les maladies chroniques.....	57
4.5	Informations sur les modèles estimés	64

4.6a	Impact des facteurs de risque sur l'arthrite	66
4.6b	Impact des facteurs de risque sur les maux de dos.....	68
4.6c	Impact des facteurs de risque sur les maladies pulmonaires.....	70
4.6d	Impact des facteurs de risque sur les maladies cardio-vasculaires	72
4.6e	Impact des facteurs de risque sur le diabète.....	75
4.7	Corrélation des facteurs de risque exogènes non observés	76
4.8a	Variances des effets individuels hiérarchiques	78
4.8b	Covariances des effets individuels hiérarchiques	79
4.8c	Comparaison des corrélations des termes résiduels et des effets individuels hiérarchiques	80
4.9	Coefficients de la participation au marché du travail (facteurs exogènes) ...	82
4.10a	Coefficients des maladies chroniques sur la participation au marché du travail	84
4.10b	Probabilités conditionnelles des maladies chroniques sur la participation au marché du travail.....	85
4.11a	Coefficients de corrélation entre les maladies chroniques et la participation au marché du travail.....	86
4.11b	Variances et covariances des effets individuels hiérarchiques entre les maladies chroniques et la participation au marché du travail.....	87
4.12	Effets marginaux directs et indirects de la participation au marché du travail en points de pourcentage	88
4.13	Comparaisons de modèles.....	92
4.14	Analyse de sensibilité.....	95
B.1	Paramètres de la distribution <i>a posteriori</i> du probit univarié.....	104
C.1	Paramètres de la distribution <i>a posteriori</i> du probit univarié avec effets individuels hiérarchiques	107

D.1	Paramètres de la distribution <i>a posteriori</i> de la participation au marché du travail	110
D.2	Impact des facteurs de risque sur l'arthrite	112
D.3	Impact des facteurs de risque sur les maux de dos.....	113
D.4	Impact des facteurs de risque sur les maladies pulmonaires.....	114
D.5	Impact des facteurs de risque sur les maladies cardio-vasculaires.....	115
D.6	Impact des facteurs de risque sur le diabète.....	116
D.7	Corrélation des facteurs de risque exogènes non observés	117
D.8	Coefficients des maladies chroniques sur la participation au marché du travail	118
D.9	Coefficients de corrélation entre les maladies chroniques et la participation au marché du travail.....	119

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

ESCC	Enquête sur la santé dans les collectivités canadiennes
ENSP	Enquête nationale sur la santé de la population
DA	Data Augmentation
GHK	Méthode de Geweke-Hajivassiliou-Keane
HPDI	Highest posterior density interval
MCSE	Monte Carlo standard error
PXDA	Parameter expansion for data augmentation
SML	Simulated maximum likelihood
SURE	Seemingly unrelated regression equations

RÉSUMÉ

L'objet de cette thèse est de quantifier la relation entre les déterminants de la santé, la santé et la participation au marché du travail. Nous estimons l'impact de cinq maladies chroniques sur la participation au marché du travail à l'aide d'un probit multivarié avec structure réursive endogène et avec effets individuels hiérarchiques. Ce modèle a été estimé avec deux échantillons d'hommes et de femmes de 30 à 64 ans formés à partir des données de l'enquête nationale sur la santé des populations de Statistique Canada. Pour éviter les problèmes numériques liés aux méthodes classiques d'estimation, les méthodes bayésiennes ont été utilisées et les distributions conditionnelles *a posteriori* du probit multivarié avec effets individuels hiérarchiques ont été dérivées. De plus, le nombre d'équations, de paramètres et d'observations étant important, un algorithme très efficace du point de vue numérique a été développé à partir des dernières avancées de la littérature.

Les résultats montrent que les facteurs tels que l'obésité et le tabagisme affectent les maladies chroniques sur le long terme et indirectement la participation au marché du travail. Par contre, une fois ces facteurs contrôlés, l'éducation a peu d'impact sur les maladies chroniques. Également, les résultats montrent que les effets directs et indirects de l'obésité sur la participation au marché du travail sont plus importants que ceux du tabagisme. Finalement, les résultats obtenus peuvent être utilisés pour chiffrer les retombées d'éventuelles interventions gouvernementales en santé publique visant l'obésité et le tabagisme.

Mots clés : méthode bayésienne, probit multivarié, effets individuels, maladies chroniques, participation au marché du travail, obésité, tabagisme, éducation.

INTRODUCTION

Les modèles à choix discrets multivariés supposent que plusieurs réponses d'un individu sont corrélées entre elles. Ce type de modèle permet d'étudier, par exemple, des modèles à choix discrets longitudinaux, des essais cliniques randomisés ou des modèles à choix discrets simultanés. Le modèle probit multivarié introduit par Ashford et Snowden (1970) suppose que la réponse multivariée observée est simplement une représentation discrète de variables continues sous-jacentes non observées appelées variables latentes. Ce modèle suppose également que, étant donné les variables explicatives, les variables latentes suivent une distribution normale multivariée.

Avant l'apparition des méthodes de simulation, les modèles probits multivariés étaient très difficiles à estimer par maximum de vraisemblance. Cette difficulté était due à la forme particulière de la fonction de vraisemblance qui se caractérise par une intégrale multidimensionnelle. Si le nombre de dimensions est très faible, il est possible d'approximer la fonction de vraisemblance par la méthode des quadratures (Butler et Moffitt, 1982). Par contre, si ce nombre est élevé, la simulation de la fonction de vraisemblance est la seule façon d'obtenir les paramètres qui caractérisent le modèle étudié. Lerman et Manski (1981) ont proposé l'estimateur du maximum de vraisemblance simulé (SML) pour estimer le modèle probit multivarié, mais la méthode qu'ils ont proposée pour simuler la fonction de vraisemblance n'était pas assez précise pour satisfaire les chercheurs. Au début des années 90, de nouvelles méthodes de simulation plus satisfaisantes ont vu le jour comme la méthode GHK proposée par Geweke (1991), Hajivassiliou et McFadden (1998) et Keane (1990). Cette méthode

permet également d'estimer le modèle probit multivarié à partir de la méthode des moments simulés (McFadden, 1989). Les bayésiens ont, de leur côté, proposé leur propre méthode pour estimer le probit multivarié (McCulloch et Rossi, 1994).

Théoriquement, si le nombre de données est suffisant pour éviter le problème d'*overfitting*, il n'y a pas de limite au nombre d'équations que peut contenir le modèle probit multivarié ou multinomial. Néanmoins, l'augmentation du nombre d'équations et de paramètres à estimer peut entraîner des problèmes numériques importants lorsque le modèle probit est estimé par maximum de vraisemblance simulé. Plus le nombre d'équations augmente, plus la maximisation de l'intégrale multidimensionnelle est difficile à effectuer. Si certaines valeurs de départ sont éloignées des vraies valeurs, la convergence de l'estimateur peut être lente, voire impossible.

Pour éviter les problèmes numériques associés à l'estimateur SML, nous avons opté pour l'estimateur bayésien. Étant donné que l'estimateur bayésien ne nécessite pas la maximisation de la fonction de vraisemblance, l'augmentation du nombre d'équations d'un probit multivarié ne pose pas de problème numérique. Néanmoins, les nombres d'équations, de paramètres et d'observations peuvent devenir rapidement un obstacle si l'algorithme utilisé n'est pas conçu pour estimer ce type de modèle. Par exemple, la fonction *rmvpGibbs*¹ dans *R* programmé par Peter Rossi (Rossi, Allenby et McCulloch, 2005) peut prendre jusqu'à 24 heures pour estimer un probit multivarié à huit équations, 40 coefficients de régression et 50 000 observations. Jusqu'à maintenant, la vitesse des ordinateurs permettait aux chercheurs de ne pas trop s'attarder sur l'efficacité computationnelle des algorithmes qu'ils utilisaient, mais avec l'avènement

¹ La fonction *rmvpGibbs* fait partie du package *bayesm*.

du *Big Data*, il n'est pas rare maintenant de voir des banques de données qui accumulent des téraoctets de données. Dans cette optique, il est nécessaire d'adopter des stratégies qui permettent une estimation plus efficiente des modèles utilisés.

Déjà, plusieurs méthodes ont été proposées pour augmenter l'efficacité numérique de l'estimateur bayésien des modèles discrets multivariés. Par exemple, Liu et Wu (1999) et Talhouk, Doucet et Murphy (2012) pour le modèle probit multivarié et Dyk et Meng (2001) pour le modèle probit multinomial ont développé des techniques qui permettent de réduire l'autocorrélation entre les tirages de la chaîne de Markov augmentant ainsi le vitesse de convergence de l'estimateur bayésien. Parallèlement à ces développements, d'autres stratégies ont été développées pour augmenter la vitesse computationnelle des algorithmes utilisés (Chopin, 2011).

Même si ces méthodes permettent d'améliorer l'efficacité computationnelle de l'estimateur bayésien, aucun algorithme n'est disponible sur le marché mettant à profit tous ces développements de façon simultanée. Pour répondre à cette lacune, nous proposons notre propre algorithme incorporant l'ensemble des développements récents proposés dans la littérature. En plus de ces développements, nous discuterons d'une stratégie supplémentaire permettant d'augmenter de façon drastique la vitesse d'échantillonnage de l'estimateur bayésien.

Depuis quelques années, la multiplication des enquêtes longitudinales produites par Statistiques Canada permet aux chercheurs d'incorporer des effets fixes ou aléatoires à leurs modèles et de contrôler certains facteurs non observés chez les individus qui sont constants dans le temps. Pour profiter de cette possibilité, nous proposons de généraliser le modèle probit multivarié pour y inclure des effets individuels et de dériver les distributions conditionnelles *a posteriori* permettant d'échantillonner la dis-

tribution *a posteriori* de ce type de modèle. De plus, pour profiter pleinement des méthodes décrites plus haut, nous généralisons la méthode *PXDA* de Talhouk, Doucet et Murphy (2012) pour y inclure des effets individuels hiérarchiques (effets aléatoires). La performance de ce modèle est testée au moyen d'une étude Monte Carlo.

Les maladies chroniques sont la principale cause de décès et d'incapacités à travers le monde. Selon l'Organisation Mondiale de la Santé (OMS), les maladies chroniques sont responsables de 60 % des décès à travers le monde, même si dans la plupart des cas, ces décès étaient évitables.² Les maladies chroniques ne sont pas uniquement responsables de nombreux décès, elles sont également responsables d'une perte importante de bien-être liée à la souffrance physique et à la détresse psychologique. Les coûts directs constitués des dépenses de santé, tels que le salaire des médecins, les frais d'hospitalisation et les médicaments, représentaient au Canada plus de 10 % du PIB en 2006, c'est-à-dire 4 867 \$ par habitant. De ces dépenses, 15,2 % étaient consacrées aux maladies cardiovasculaires, 9,4 % aux maladies pulmonaires et 10,1 % aux cancers.³ Les coûts indirects sont également colossaux. La baisse de la participation au marché du travail, l'absentéisme, la perte de productivité, les retraites hâtives et les décès prématurés sont tous des coûts supplémentaires engendrés par les maladies chroniques. Si rien n'est fait, la progression des taux de morbidité associé à chacune des maladies chroniques risque de faire exploser les coûts directs et indirects dans un proche avenir.

² http://www.who.int/topics/chronic_disease/fr/

³ Les soins de santé au Canada 2008, institut Canadien d'information sur la santé, http://secure.cihi.ca/cihiweb/products/HCIC_2008_f.pdf

Pour contribuer à ce débat, nous proposons d'utiliser un modèle probit multivarié avec effets individuels et structure réursive pour estimer l'effet des maladies chroniques sur la participation au marché du travail. À partir de cinq équations de santé représentant l'arthrite, les maux de dos, le diabète, les maladies cardiaques et les maladies pulmonaires, nous estimons une équation de participation au marché du travail avec ces cinq maladies comme variables endogènes. Ce modèle nous permettra d'estimer l'impact des maladies chroniques sur la participation au marché du travail en tenant compte de l'endogénéité des maladies chroniques dans l'équation de travail. De plus, en incorporant des effets individuels dans le modèle, nous pourrions tenir compte des facteurs de risque non observables spécifiques aux individus. À partir de deux échantillons, d'hommes et de femmes de 30 à 64 ans, construits à partir de l'Enquête nationale sur la santé des populations (ENSP), nous pourrions mesurer l'impact des maladies chroniques sur la participation au marché du travail et mesurer les effets directs et indirects, via les maladies chroniques, de l'âge, de l'éducation, de l'obésité et du tabagisme sur la participation au marché du travail.

La structure de la thèse est la suivante. Le premier chapitre introduit le probit multivarié ainsi que la notation utilisée tout au long de la thèse. En particulier, on discutera du problème d'identification du probit multivarié et de sa modélisation. Le deuxième chapitre discute de l'algorithme utilisé dans cette thèse et des stratégies adoptées permettant de réduire le temps de traitement numérique. La troisième partie développe un modèle probit multivarié avec effets individuels hiérarchiques (effets aléatoires) et la quatrième partie porte sur l'application de ce modèle sur la participation au marché du travail et les maladies chroniques.

CHAPITRE I

PROBIT MULTIVARIÉ

1.1 Introduction

Ce chapitre discute du modèle probit multivarié ainsi que deux stratégies possibles pour estimer ce modèle : l'estimateur du maximum de vraisemblance simulé et l'estimateur bayésien. Il sert à introduire les différents concepts et notations qui serviront tout au long de la thèse et se veut une revue de littérature sur la question, en particulier sur l'approche bayésienne de ce modèle.

1.2 Le modèle probit multivarié

Le modèle probit multivarié est un modèle à m équations où chaque variable dépendante représente un choix ou un état discret. Ce type de modèle permet d'étudier par exemple la possession de m biens durables de façon simultanée ou la possession d'un bien durable sur m périodes différentes. On suppose que les valeurs des variables dépendantes sont la représentation discrète d'une variable continue sous-jacente non observée appelée variable latente. Dans le cas d'un bien durable, la variable latente représenterait l'utilité supplémentaire procurée par la possession d'un bien durable quelconque. Soit $y_i = (y_{i1}, y_{i2}, \dots, y_{im})$ et $w_i = (w_{i1}, w_{i2}, \dots, w_{im})$ les vecteurs de

dimensions m des réponses discrètes observées et des variables latentes non observées d'un individu $i = 1, \dots, n$. Soit X_i la matrice $m \times k$ des variables explicatives et β le vecteur $k \times 1$ des coefficients de régression. Le modèle probit multivarié peut s'écrire de la façon suivante :

$$w_i = X_i \beta + \varepsilon_i, \quad i = 1, \dots, n$$

$$y_{ij} = \begin{cases} 1 & \text{si } w_{ij} \geq 0 \\ 0 & \text{si } w_{ij} < 0, \end{cases} \quad j = 1, \dots, m \quad (1.1)$$

$$(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{im}) \sim N_m(0, R).$$

Dans ce modèle, le vecteur des termes d'erreur $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{im})$ suit une distribution normale multivariée avec moyennes zéro et la matrice de corrélation R .

Contrairement au modèle SURE (*seemingly unrelated regression equation*), le modèle probit multivarié doit être contraint pour permettre l'identification. Pour comprendre le problème d'identification, multiplions le modèle par la matrice :

$$\lambda w_i = \lambda X_i \beta + \lambda \varepsilon_i, \quad (1.2)$$

$$\lambda \varepsilon_i \sim N_m(0, \lambda R \lambda).$$

$$\lambda = \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & & \\ \vdots & & \ddots & \\ 0 & & & \alpha_m \end{bmatrix}.$$

Nous voyons que les paramètres $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ ne sont pas identifiables, car quelle que soit leurs valeurs, ils ne permettent pas de modifier le signe de la variable latente et par conséquent :

$$y_i(w_i) = y_i(\lambda w_i). \quad (1.3)$$

Les choix du vecteur y_i sont les mêmes pour tous les modèles. Il n'y a donc aucune information dans les données qui nous permettrait d'estimer simultanément les paramètres β et la variance du terme d'erreur. Pour cette raison, on choisit de contraindre la matrice de variances-covariances à une matrice de corrélation.

Les hypothèses du modèle probit multivarié supposent que les facteurs non observables $\varepsilon_i = [\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{im}]$ sont distribués conjointement selon une normale multivariée :

$$\phi_m(\varepsilon_i) = \frac{1}{(2\pi)^{m/2} |R|^{1/2}} e^{\frac{1}{2} \varepsilon_i' R^{-1} \varepsilon_i}. \quad (1.6)$$

Sous cette hypothèse, la probabilité d'observer $y_i = Y_i$ est égale à :

$$p(y_i = Y_i | X_i, \beta, R) = \int_{A_{im}} \dots \int_{A_{i1}} \phi_m(w_i | X_i, \beta, R) dw_{i1} \dots dw_{im}. \quad (1.7)$$

Dans cette équation, A_{ij} prend une valeur comprise entre $(0, +\infty)$ si $y_{ij} = 1$ et une valeur comprise entre $(-\infty, 0]$ sinon. Étant donné qu'il n'existe pas de solution analytique pour cette expression, nous devons utiliser des méthodes numériques pour approximer cette probabilité.

1.3 Estimateur du maximum de vraisemblance simulé

Les modèles probits multivariés sont très difficiles à estimer par maximum de vraisemblance car il n'existe pas de forme analytique à la fonction de vraisemblance. Pour estimer ce modèle, il est nécessaire d'approximer la fonction de vraisemblance. Pour présenter cette méthode, considérons un modèle probit trivarié et imaginons que nous observons pour un individu i les choix $Y_i = (y_{i1} = 1, y_{i2} = 1, y_{i3} = 1)$. La probabilité d'observer ce choix est égale à :

$$p(y_i = Y_i) = p(y_{i1} = 1, y_{i2} = 1, y_{i3} = 1). \quad (1.8)$$

Si nous laissons tomber l'indice i est que nous développons cette probabilité à partir de la définition d'un probit multivarié, nous obtenons :

$$\begin{aligned} p(y_1 = 1, y_2 = 1, y_3 = 1) &= p(\varepsilon_1 < X_1\beta_1, \varepsilon_2 < X_2\beta_2, \varepsilon_3 < X_3\beta_3) \quad (1.9) \\ &= p(\varepsilon_3 < X_3\beta_3 | \varepsilon_2 < X_2\beta_2, \varepsilon_1 < X_1\beta_1) \times p(\varepsilon_2 < X_2\beta_2 | \varepsilon_1 < X_1\beta_1) \times p(\varepsilon_1 < X_1\beta_1) \end{aligned}$$

On voit qu'il est possible de calculer la probabilité de cette fonction de façon récursive à partir des probabilités conditionnelles. En utilisant la décomposition de Cholesky de la matrice de corrélation :

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} = \begin{bmatrix} c_{11} & 0 & 0 \\ c_{21} & c_{22} & 0 \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}, \quad (1.10)$$

nous pouvons échantillonner le vecteur des termes d'erreur à partir de valeurs tirés d'une normale centrée-réduite $e_1, e_2, e_3 \sim N(0,1)$. La probabilité devient donc :

$$\begin{aligned}
 p(y_1 = 1, y_2 = 1, y_3 = 1) &= p(\varepsilon_1 < X_1\beta_1, \varepsilon_2 < X_2\beta_2, \varepsilon_3 < X_3\beta_3) \quad (1.11) \\
 &= p\left(e_3 < \frac{X_3\beta_3 - c_{32}e_2 - c_{31}e_1}{c_{33}} \mid e_2 < \frac{X_2\beta_2 - c_{21}e_1}{c_{22}}, e_1 < \frac{X_1\beta_1}{c_{11}}\right) \\
 &\quad \times p\left(e_2 < \frac{X_2\beta_2 - c_{21}e_1}{c_{22}} \mid e_1 < \frac{X_1\beta_1}{c_{11}}\right) \times p\left(e_1 < \frac{X_1\beta_1}{c_{11}}\right)
 \end{aligned}$$

Étant donné que (e_1, e_2, e_3) ne sont pas corrélés entre eux, cette probabilité devient :

$$\begin{aligned}
 p(y_1 = 1, y_2 = 1, y_3 = 1) &= p(\varepsilon_1 < X_1\beta_1, \varepsilon_2 < X_2\beta_2, \varepsilon_3 < X_3\beta_3) \quad (1.12) \\
 &= p\left(e_3 < \frac{X_3\beta_3 - c_{32}e_2 - c_{31}e_1}{c_{33}}\right) \times p\left(e_2 < \frac{X_2\beta_2 - c_{21}e_1}{c_{22}}\right) \times p\left(e_1 < \frac{X_1\beta_1}{c_{11}}\right)
 \end{aligned}$$

Pour simuler cette probabilité, on tire les valeurs e_1, e_2 et e_3 de façon récursive à partir d'une normale centrée-réduite tronquée à $X_1\beta_1/c_{11}$, $(X_2\beta_2 - c_{21}e_1)/c_{22}$ et $(X_3\beta_3 - c_{32}e_2 - c_{31}e_1)/c_{33}$ respectivement. Ces valeurs peuvent être générées aléatoirement à partir d'une distribution uniforme tronquée. Malgré la simplicité de cette méthode, il faut un nombre important de tirages pour couvrir en totalité la probabilité simulée. Il existe par contre des méthodes qui permettent de réduire le nombre de tirages. Les deux méthodes fréquemment utilisés sont les tirages antithétiques et les tirages à partir d'une séquence Halton (Train, 2003). Cappellari et Jenkins (2006) ont montré que la méthode GHK nécessite 10 fois moins de tirages si elle utilise des tirages de type Halton au lieu de tirages pseudo-aléatoires.

Avec ces valeurs en poche, on peut maintenant évaluer les probabilités $p(e_3 < (X_3\beta_3 - c_{32}e_2 - c_{31}e_1)/c_{33})$ et $p(e_2 < (X_2\beta_2 - c_{21}e_1)/c_{22})$ avec la méthode de la fonction de répartition inverse et après avoir répété ce processus B fois, on peut calculer la probabilité moyenne de ces deux termes. Finalement, on multiplie ces deux probabilités avec le dernier terme $p(e_1 < X_1\beta_1/c_{11})$ pour obtenir $p(y_1 = 1, y_2 = 1, y_3 = 1)$.

On estime le modèle en maximisant la fonction log de vraisemblance à partir des probabilités simulées de l'échantillon complet :

$$LL(\theta) = \sum_{i=1}^n \ln p(y_i = Y_i). \quad (1.13)$$

Plusieurs difficultés sont associées à l'estimateur du maximum de vraisemblance simulé qui limite énormément la liberté du chercheur. La première difficulté est liée aux problèmes numériques découlant de la maximisation de la fonction de vraisemblance. Un mauvais choix des valeurs de départ est généralement à l'origine de ce problème. La convergence peut être lente voire impossible si certaines valeurs sont éloignées des maximums. La deuxième difficulté concerne l'estimation des paramètres qui peuvent converger à un maximum local au lieu du maximum global de la fonction de vraisemblance. Il n'y a donc aucune certitude que les paramètres aient convergé vers leur maximum global (Train, 2003).

1.4 Estimateur bayésien

À partir du théorème de Bayes, nous pouvons développer la distribution *a posteriori* du probit multivarié :

$$p(\beta, R|Y, X) \propto p(Y, X|\beta, R)p(\beta, R), \quad (1.14)$$

où β sont les coefficients de régression et R sont les paramètres de corrélation des termes résiduels du modèle, $Y = (y_1, y_2, \dots, y_n)$ est une matrice $n \times m$ des réponses observées, $X = (X_1, X_2, \dots, X_n)$ est une matrice $(n \times m) \times k$ des k variables explicatives, $p(Y|\beta, R)$ la fonction de vraisemblance et $p(\beta, R)$ la distribution *a priori*. Dans la suite de la présentation, la matrice X est mise de côté pour ne pas alourdir la notation :

$$p(\beta, R|Y) \propto p(Y|\beta, R)p(\beta, R). \quad (1.15)$$

Pour estimer les paramètres de cette distribution, une méthode possible est la simulation de cette distribution par l'échantillonneur de Gibbs. Cette méthode consiste à « briser » la distribution jointe (inconnue) en plusieurs blocs de distributions conditionnelles (connues) et de former une chaîne de Markov qui va générer une séquence de tirages de variables aléatoires chacun conditionné sur les précédents. Si la chaîne de Markov est bien formulée, alors la séquence de β tirages convergera vers la vraie distribution (Rossi, Allenby et McCulloch, 2005).

1.4.1 Les densités *a priori*

Habituellement, on suppose que β et R sont *a priori* indépendants et qu'ainsi $p(\beta, R) = p(\beta)p(R)$. On suppose aussi que la densité *a priori* $p(\beta)$ suit une distribution normale centrée sur $\underline{\beta}$ avec une matrice de variances-covariances diagonale \underline{V}_β . Cette densité est la densité *a priori* conjuguée propre. La nécessité de contraindre la matrice de variances-covariances en matrice de corrélation est problématique dans le

choix de l'*a priori* pour R car il n'existe pas d'*a priori* conjugué pour la matrice de corrélation. Une première solution est d'ignorer le problème d'identification en choisissant la distribution Wishart comme densité *a priori* de la matrice de variances-covariances Σ non-contrainte et de faire naviguer l'échantillonneur de Gibbs dans cet espace non contraint. Une fois les B tirages de l'échantillonneur de Gibbs terminés, on identifie les coefficients de régression $\hat{\beta}$ et on contraint la matrice de variances-covariances en matrice de corrélation avec la transformation :

$$R = D^{-1}\Sigma D^{-1} \text{ et } \beta = D^{-1}\hat{\beta}, \quad (1.16)$$

où

$$D^{-1} = \begin{bmatrix} 1/\sqrt{\Sigma_{11}} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\Sigma_{22}} & & \\ \vdots & & \ddots & \\ 0 & & & 1/\sqrt{\Sigma_{mm}} \end{bmatrix}$$

et où Σ_{jj} sont les éléments de la diagonale de la matrice de variances-covariances Σ . Cette méthode a été utilisée par McCulloch et Rossi (1994) et Edwards et Allenby (2003). Malgré la simplicité de cette méthode, certaines précautions doivent être prises concernant le choix des hyperparamètres. Premièrement, le choix des hyperparamètres est difficile car les *a priori* doivent être placés sur des paramètres non identifiés. Deuxièmement, Koop (2003) a montré que les paramètres non identifiés ont tendance à s'éloigner de plus en plus des moyennes de la densité *a posteriori* créant ainsi des écart-types de plus en plus élevés pour les coefficients estimés. De plus, si les densités *a priori* sont non informatives, rien ne peut empêcher les coefficients tirés d'atteindre des valeurs infinies du point de vue numérique créant ainsi d'autres pro-

blèmes. Une densité *a priori* informative permet en quelque sorte de structurer le modèle et d'empêcher ce type de résultats.

Plusieurs méthodes ont été proposées pour échantillonner la distribution *a posteriori* à partir du modèle identifié et ainsi permettre l'utilisation de densités *a priori* non informatives. Par exemple, les méthodes de Nandram et Chen (1996), Nobile (2000), Chib et Greenberg (1998), Daniels et Kass (1999), Barnard, McCulloch et Meng (2000), Wong, Carter et Kohn (2003) et Liechty, Liechty et Müller (2004) ont été proposées pour le tirage de la matrice de corrélation R . En particulier, Chib et Greenberg (1998) ont échantillonné la matrice de corrélations en utilisant l'algorithme Metropolis-Hastings avec marche aléatoire et une densité *a priori* uniforme jointe $p(R) \propto 1$. Malgré l'utilisation de cette densité *a priori* non informative, la méthode est peu efficiente car il n'est pas garanti que la matrice échantillonnée soit définie positive, ce qui requiert une étape supplémentaire de vérification et la convergence de l'échantillonneur de Gibbs peut être très lente lorsque le nombre d'équations du modèle probit multivarié est élevé. De plus, Barnard *et al.* (2000) démontrent que l'augmentation du nombre de dimensions de la matrice de corrélation favorise les valeurs proches de zéro au détriment des valeurs près de ± 1 . Pour éviter ces problèmes, Barnard *et al.* proposent une *a priori* sur R tel que chaque r_{ij} est distribuée de manière uniforme sur l'intervalle $[-1,1]$. Si on suppose que les densités *a priori* des paramètres β et Σ du modèle non identifié sont *a priori* indépendantes alors :

$$\begin{aligned}
 p(\beta, \Sigma) &= p(\beta)p(\Sigma), \\
 p(\Sigma) &= p(R, D) \times |J: \Sigma \rightarrow D, R| = p(R)p(D|R), \\
 p(\beta, \Sigma) &= p(\beta)p(R)p(D|R),
 \end{aligned}
 \tag{1.17}$$

où $|J: \Sigma \rightarrow D, R|$ est le déterminant de la jacobienne de la transformation $\Sigma \rightarrow D, R$. De plus, si la densité *a priori* de Σ est Wishart inverse avec $\underline{\nu} = m + 1$ degrés de liberté, alors :

$$p(R) \propto |R|^{\frac{m(m-1)}{2}-1} (\prod_{j=1}^m |r_{jj}|)^{-(m+1)/2} \quad (1.18)$$

et

$$p(D|R) = \prod_{j=1}^m p(d_j|R),$$

avec $d_j^2 \sim IG((m+1)/2, r_j/2)$ où r_j est le $j^{\text{ème}}$ élément de la diagonale de R^{-1} et IG est gamma inverse.

Barnard *et al.* montrent qu'il est possible d'échantillonner cette distribution à partir d'une Wishart inverse avec $\underline{\nu} = M + 1$ degré de liberté et de transformer la matrice de covariance obtenue en matrice de corrélation avec la relation $R = D^{-1}\Sigma D^{-1}$. Malgré cet avantage, la densité *a posteriori* qu'ils obtiennent est difficile à échantillonner. Pour y parvenir, ils utilisent l'échantillonneur de Gibbs de type *Griddy* (*Griddy Gibbs sampler*) développé par Ritter et Tanner (1992). Par contre, cette technique est très inefficace d'un point de vue numérique car elle exige de tirer les éléments de la matrice de corrélation un par un (Talhouk, Doucet et Murphy, 2012).

À partir de la densité *a priori* de Barnard *et al.*, Talhouk, Doucet et Murphy (2012) proposent d'utiliser l'approche de l'expansion des paramètres (PXDA, *Parameter Expection Data Augmentation*) dans un cadre de tirage par augmentation des données (Liu et Wu (1999); Dyk et Meng (2001)) pour faciliter l'échantillonnage de la distribution *a posteriori*. Cette méthode consiste à reparamétriser la matrice de corrélations en matrice de variances-covariances de telle sorte qu'elle soit toujours définie-

positive et donc inversible. Les difficultés associées à l'estimation de la matrice de corrélation R peuvent donc être surmontées en transformant cette dernière en matrice de covariances non contrainte Σ . L'élimination de cette contrainte permet entre autres à l'échantillonneur de Gibbs de se déplacer plus librement dans la distribution *a posteriori*, de diminuer la corrélation entre les tirages et d'augmenter du même coup la vitesse de convergence de la chaîne de Markov. Ainsi, l'utilisation de cette méthode réduit le nombre de tirages nécessaires pour obtenir des résultats et permet de traiter le problème du choix des hyperparamètres car les hyperparamètres sont directement placés sur les paramètres identifiés de la matrice de corrélation.

1.4.2 Distribution conditionnelle *a posteriori* du modèle probit multivarié

Dans sa forme actuelle, la distribution *a posteriori* ne permet pas de former aisément une chaîne de Markov car la variable dépendante prend uniquement la valeur 0 ou 1. Pour éviter ce problème, on utilise la méthode DA « data augmentation » développée par Tanner et Wong (1987) et appliquée par Albert et Chib (1993) aux modèles à choix discrets. La méthode consiste à introduire, dans le modèle, une variable artificielle dont le seul but est de faciliter la simulation de la densité *a posteriori*. Dans le cas des modèles à choix discrets, on utilise les variables latentes comme variables augmentées. Nous pouvons « augmenter » le modèle en introduisant la matrice de variables latentes W dans l'équation de Bayes :

$$p(\beta, R, W | Y) = p(\beta, R) p(W | \beta, R) p(Y | \beta, R, W), \quad (1.19)$$

où

$$p(Y | \beta, R, W) = \prod_{i=1}^n \prod_{j=1}^m I(w_{ij} \geq 0) I(y_{ij} = 1) + I(w_{ij} < 0) I(y_{ij} = 0).$$

La probabilité $p(Y|\beta, R, W)$ est égale à 0 ou 1 car connaître w_{ij} nous permet de connaître y_{ij} avec certitude. La fonction de vraisemblance augmentée est donc une distribution normale multivariée tronquée car la probabilité d'observer W est zéro pour certain Y . Sous cette forme, il est difficile d'effectuer des tirages. La méthode développée par Geweke (1991) permet d'échantillonner une distribution multivariée tronquée à partir de distributions conditionnelles univariées. Les distributions conditionnelles univariées des variables augmentées $w_{ij}|w_{i(-j)}$ sont :

$$w_{ij}|w_{i(-j)} \sim TN(\hat{w}_{ij}, \hat{R}_{jj}), \quad (1.20)$$

où

$$\hat{w}_{ij} = X'_{ij}\beta_j + R'_j R_j^{-1} R_j u_{i(-j)}$$

et

$$\hat{R}_{jj} = 1 - R'_j R_j^{-1} R_j.$$

Le vecteur R_j est formée de la $j^{\text{ème}}$ colonne de la matrice de corrélation R , la matrice R_j^{-1} est égale à l'inverse de la matrice de corrélation R^{-1} dont la $j^{\text{ème}}$ colonne et la $j^{\text{ème}}$ ligne ont été supprimées et $u_{i(-j)}$ est une matrice $(m \times (m-1))$ des erreurs $(w_i - X_i\beta)$ dont la $j^{\text{ème}}$ colonne a été supprimée. Enfin, $w_{i(-j)}$ est le vecteur comprenant tous les éléments de la ligne i sauf celui de la $j^{\text{ème}}$ colonne.

L'algorithme 1 présenté plus bas montre les différentes étapes pour estimer le modèle probit multivarié de manière bayésienne. Après avoir choisi les hyperparamètres pour les densités *a priori* du modèle, la première étape consiste à échantillonner les $n \times m$ valeurs des variables latentes du modèle conditionnellement aux coefficients β et R . La deuxième étape échantillonne les m paramètres $d_j^2 \sim IG((m+1)/2, r_j/2)$ pour former la matrice diagonale :

$$\tilde{D} = \begin{bmatrix} d_1^2 & 0 & \dots & 0 \\ 0 & d_1^2 & & \\ \vdots & & \ddots & \\ 0 & & & d_1^2 \end{bmatrix}. \quad (1.21)$$

À partir de cette matrice, on transforme les variables latentes, les coefficients de régression et la matrice de corrélation dans l'espace non contraint :

$$Z = W \times \tilde{D} \quad (1.22)$$

$$\tilde{\beta} = \beta \times \tilde{D} \quad (1.23)$$

$$\Sigma = \tilde{D} R \tilde{D}. \quad (1.24)$$

On échantillonne les coefficients de régression non identifiés $\tilde{\beta}$ et la matrice de variances-covariances Σ à partir de leur distribution conditionnelle présentée au Tableau 1.1. On calcule la matrice :

$$D^{-1} = \begin{bmatrix} 1/\sqrt{\Sigma_{11}} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\Sigma_{22}} & & \\ \vdots & & \ddots & \\ 0 & & & 1/\sqrt{\Sigma_{mm}} \end{bmatrix}. \quad (1.25)$$

et on transforme les coefficients dans l'espace contraint :

$$\beta = D^{-1} \tilde{\beta} \quad (1.26)$$

$$R = D^{-1} \Sigma D^{-1}. \quad (1.27)$$

Finalement, les étapes 1 à 7 sont recommencées B fois pour simuler de tirages de la distribution *a posteriori*.

Algorithme 1

1. Échantillonner $w_{ij} \sim TN(w_{ij} | w_{i,-j}, \beta, R)$ pour $i = (1, \dots, n)$ et $j = (1, \dots, m)$.
2. Échantillonner \tilde{D} , une matrice diagonale avec $d_{ii} \sim IG((m+1)/2, r^{ii}/2)$.
3. Transformer les variables latentes $Z = W \times \tilde{D}$ et les paramètres $\tilde{\beta} = \beta \times \tilde{D}$.
4. Échantillonner $\Sigma | \tilde{\beta}, Z$ avec le vecteur $\tilde{\beta}$ et la matrice Z de l'étape 3 (voir Tableau 1.1).
5. Échantillonner $\tilde{\beta} | Z, R$ avec le vecteur $\tilde{\beta}$ de l'étape 4 et la matrice Z de l'étape 3 (voir Tableau 1.1).
6. Trouver D^{-1} , la matrice diagonale de l'équation 1.25.
7. Transformer les variables $\beta = D^{-1} \tilde{\beta}$ et $R = D^{-1} \Sigma D^{-1}$.
8. Recommencer B fois les étapes 1 à 7 pour chaque tirage de l'échantillonneur de Gibbs.

Tableau 1.1 Distribution conditionnelle *a posteriori* du modèle probit multivarié

Densité conditionnelle	Loi de probabilité	Paramètres de localisation/forme	Paramètres d'échelle
$\tilde{\beta} Z, R$	$N(\tilde{\beta}, \tilde{V}_{\beta})$	$\tilde{\beta} = \tilde{V}_{\beta} \left(\sum_{i=1}^N X_i' \Sigma^{-1} z_i + \underline{V}_{\beta}^{-1} \underline{\beta} \right)$	$\tilde{V}_{\beta} = \left(\sum_{i=1}^N X_i' \Sigma^{-1} X_i + \underline{V}_{\beta}^{-1} \right)^{-1}$
$\Sigma \tilde{\beta}, Z$	$IW(\bar{\Sigma}, \bar{\nu})$	$\bar{\Sigma} = \sum_{i=1}^N (z_i - X_i' \tilde{\beta})' (z_i - X_i' \tilde{\beta}) + \underline{\Sigma}$	$\bar{\nu} = \underline{\nu} + n$

CHAPITRE II

L'ALGORITHME MCMC

2.1 Introduction

À notre connaissance, le programme le plus populaire pour estimer un probit multivarié de façon bayésienne est la fonction *rmvpGibbs* développé par Peter Rossi qu'on peut utiliser avec le logiciel *R*. Malgré les qualités intrinsèques du programme et le fait qu'il permet une estimation rapide du modèle probit multivarié lorsque le nombre d'équations est faible, il n'a pas été conçu pour traiter rapidement un modèle comportant un nombre élevé d'équations. Pour cette raison, il est difficile d'augmenter le nombre d'équations et le nombre d'observations sans que le temps de traitement computationnel devienne rapidement une limitation.

Dans la littérature, plusieurs méthodes ont été proposées pour augmenter l'efficacité numérique des estimateurs bayésiens des modèles discrets multivariés. Ces méthodes peuvent se classer en deux champs distincts. D'un côté, il y a la recherche théorique qui permet de réduire le nombre total de tirages nécessaires pour estimer ces modèles pour un niveau égal de précision numérique. Par exemple, Liu et Wu (1999) et Talhouk, Doucet et Murphy (2012) pour le modèle probit multivarié et Dyk et Meng (2001) pour le modèle probit multinomial ont développé des techniques qui permettent à la chaîne de Markov de converger vers la vraie distribution beaucoup

plus rapidement. De l'autre côté, il y a la recherche pratique qui cherche à réduire le temps de traitement à chaque itération. Par exemple, Chopin (2011) a proposé une méthode qui permet d'échantillonner les variables latentes plus efficacement.

Ce chapitre est divisé en trois sections. La première et la deuxième section discutent de la méthode de Chopin pour échantillonner les variables latentes plus efficacement et d'une stratégie que nous avons développée pour augmenter la vitesse d'échantillonnage des coefficients de régression. La troisième section compare la performance de ces différentes approches.

2.2 Échantillonnage des variables latentes

L'échantillonnage des variables latentes est une procédure relativement longue dans l'échantillonneur de Gibbs. Étant donné qu'il faut tirer m valeurs aléatoires pour n observations, chaque itération de l'échantillonneur de Gibbs nécessite $n \times m$ tirages aléatoires pour les variables latentes du modèle. Si la convergence de la chaîne de Markov exige B itérations, il faut alors $n \times m \times B$ tirages aléatoires d'une distribution normale tronquée univariée. Tous ces tirages peuvent être très exigeants d'un point de vue numérique si n , m et B sont grands. Par exemple, imaginons un cas extrême où le nombre d'observations est de 50 000, le nombre de dimensions est de 10 et le nombre d'itérations de l'échantillonneur de Gibbs est de 50 000. Avec ces paramètres, il faut 25 milliards de tirages à partir d'une distribution normale tronquée univariée pour estimer le modèle. Étant donné ce nombre imposant, le choix de la méthode pour générer ces tirages aléatoires peut faire une différence importante dans le temps de traitement de l'échantillonneur de Gibbs.

Il existe plusieurs méthodes pour générer des tirages aléatoires d'une distribution normale tronquée univariée. Par exemple, la méthode la plus simple, mais la plus inefficace est la méthode de rejet (*accept-reject method*). Cette méthode consiste à effectuer un tirage aléatoire d'une distribution normale dans un premier temps et de vérifier dans un deuxième temps si ce tirage se retrouve dans la zone non tronquée de la distribution normale. Si ce tirage se retrouve dans cette zone, il est accepté, sinon il est rejeté et on effectue un nouveau tirage. On remarque que si la zone d'acceptation est très petite, le taux d'acceptation est extrêmement faible. L'algorithme pourrait avoir besoin de plusieurs centaines voire des milliers de tirages pour trouver une seule valeur acceptable.

Une deuxième méthode possible est la méthode de la fonction de répartition inverse. Cette méthode est la méthode utilisée dans le programme de Rossi cité plus haut. Si on désire effectuer un tirage aléatoire d'une densité $f(\varepsilon)$ entre les bornes a et b , on tire tout d'abord une valeur aléatoire μ d'une distribution uniforme. À partir de cette valeur, on calcule la moyenne pondérée $\bar{\mu} = (1 - \mu)F(a) + \mu F(b)$ où $F(a)$ et $F(b)$ sont les valeurs de la fonction de répartition pour a et b . Finalement, on calcule l'inverse de la fonction de répartition $\varepsilon = F^{-1}(\bar{\mu})$ à partir de cette valeur pondérée. Cette méthode a un taux d'acceptation de 100 %, par contre elle nécessite de calculer, $F(a)$ et $F(b)$ et $F^{-1}(\bar{\mu})$ en plus de tirer de la loi uniforme. C'est une procédure relativement inefficace.

La méthode que nous avons choisie pour générer des tirages aléatoires d'une distribution normale tronquée univariée est celle de Chopin (2011). Sans entrer dans les détails, la méthode consiste à discrétiser une distribution normale centrée-réduite en transformant celle-ci en un histogramme de Q rectangles de densité identique. La

première étape consiste à numéroter chacun des rectangles d'un nombre allant de 1 à Q et à identifier les deux rectangles qui englobent les bornes de la troncature. Une fois cette étape complétée, on choisit de façon aléatoire un rectangle se trouvant entre ces deux bornes et on effectue un deuxième tirage pour trouver une valeur aléatoire à l'intérieur du rectangle tiré précédemment. Dans de rares cas, il est possible que le rectangle tiré soit une des bornes de la distribution tronquée. Si cette situation survient, l'algorithme doit tenir compte de la densité tronquée de ces deux rectangles.

Cet algorithme est plus rapide pour deux raisons. La première est que le taux d'acceptation est de 100 % si le rectangle choisi ne contient pas une des deux bornes. La deuxième raison c'est que l'algorithme calcule uniquement deux additions et trois multiplications inversement aux autres méthodes qui peuvent nécessiter le calcul de logarithmes et de racines carrées.

2.3 Échantillonnage des coefficients de régression

Pour augmenter la vitesse de l'algorithme, nous proposons la solution suivante pour échantillonner la distribution conditionnelle *a posteriori* des coefficients de corrélation. Nous avons vu au premier chapitre que les paramètres de la densité conditionnelle *a posteriori* des β des modèles probits multivariés étaient donnés par :

$$\bar{\beta} = \bar{V}_{\beta} \left(\sum_{i=1}^N X_i' \Sigma^{-1} z_i + \underline{V}_{\beta}^{-1} \underline{\beta} \right) \quad \text{et} \quad \bar{V}_{\beta} = \left(\sum_{i=1}^N X_i' \Sigma^{-1} X_i + \underline{V}_{\beta}^{-1} \right)^{-1}. \quad (2.1)$$

L'échantillonnage de la densité conditionnelle *a posteriori* peut être extrêmement long si le nombre d'observations et le nombre de paramètres sont élevés. En effet, le calcul de :

$$\sum_{i=1}^N X_i' \Sigma^{-1} z_i \quad \text{et} \quad \sum_{i=1}^N X_i' \Sigma^{-1} X_i \quad (2.2)$$

peut nécessiter plusieurs milliards d'opérations dans des cas relativement lourds. Le nombre d'opérations va dépendre de la taille de l'échantillon n , le nombre d'équations m et le nombre de coefficients de régression k . Par exemple, le nombre d'opérations nécessaires à chaque itération pour calculer la matrice $\sum_{i=1}^N X_i' \Sigma^{-1} X_i$ est égal à 500 millions pour un modèle à 50 000 observations, 10 équations et 10 paramètres par équation. De plus, si le nombre d'itérations est égal à 50 000, il faut un total d'opérations de 25 000 milliards.

Il est possible de réduire le nombre d'opérations par une simple transformation algébrique du problème. Supposons le modèle SURE à deux équations suivant :

$$\begin{aligned} w_1 &= x_1 \beta_1 + x_2 \beta_2 + \varepsilon_1 \\ w_2 &= x_3 \beta_3 + x_4 \beta_4 + x_5 \beta_5 + \varepsilon_2 \end{aligned} \quad (2.3)$$

où $w_1, w_2, x_1, x_2, x_3, x_4, x_5, \varepsilon_1$ et ε_2 sont des vecteurs $n \times 1$ et $\beta_1, \beta_2, \beta_3, \beta_4$ et β_5 sont des scalaires. Si nous regroupons l'ensemble des k variables dans une seule matrice, nous obtenons :

$$\tilde{X} = [x_1, x_2, x_3, x_4, x_5]. \quad (2.4)$$

En effectuant la multiplication matricielle $\tilde{X}' \tilde{X}$, le résultat obtenu est une matrice $k \times k$ où chaque élément de cette matrice est égal au produit croisé de chacune des variables explicatives du modèle :

$$\tilde{X}'\tilde{X} = \begin{bmatrix} x_1x_1 & x_1x_2 & x_1x_3 & x_1x_4 & x_1x_5 \\ x_2x_1 & x_2x_2 & x_2x_3 & x_2x_4 & x_2x_5 \\ x_3x_1 & x_3x_2 & x_3x_3 & x_3x_4 & x_3x_5 \\ x_4x_1 & x_4x_2 & x_4x_3 & x_4x_4 & x_4x_5 \\ x_5x_1 & x_5x_2 & x_5x_3 & x_5x_4 & x_5x_5 \end{bmatrix}. \quad (2.5)$$

Finalement, on peut facilement montrer que $\sum_{i=1}^N X_i' \Sigma^{-1} X_i$ est égale à :

$$\sum_{i=1}^n X_i' \Sigma^{-1} X_i = \begin{bmatrix} \sigma_{11}x_1x_1 & \sigma_{11}x_1x_2 & \sigma_{12}x_1x_3 & \sigma_{12}x_1x_4 & \sigma_{12}x_1x_5 \\ \sigma_{11}x_2x_1 & \sigma_{11}x_2x_2 & \sigma_{12}x_2x_3 & \sigma_{12}x_2x_4 & \sigma_{12}x_2x_5 \\ \sigma_{21}x_3x_1 & \sigma_{21}x_3x_2 & \sigma_{22}x_3x_3 & \sigma_{22}x_3x_4 & \sigma_{22}x_3x_5 \\ \sigma_{21}x_4x_1 & \sigma_{21}x_4x_2 & \sigma_{22}x_4x_3 & \sigma_{22}x_4x_4 & \sigma_{22}x_4x_5 \\ \sigma_{21}x_5x_1 & \sigma_{21}x_5x_2 & \sigma_{22}x_5x_3 & \sigma_{22}x_5x_4 & \sigma_{22}x_5x_5 \end{bmatrix}$$

où les σ_{jl} sont les éléments de la matrice de variances-covariances inversée Σ^{-1} . Étant donné qu'on peut calculer la matrice $\tilde{X}'\tilde{X}$ avant de débiter la chaîne de Markov, ces transformations permettent de réduire de façon significative le nombre d'opérations qui doivent être effectuées par l'échantillonneur de Gibbs à chaque itération. Cette transformation nous permet de réduire le nombre d'opérations par un facteur égal au nombre d'observations. Dans notre exemple à 500 millions d'opérations, cette transformation réduit ce nombre à $k \times k$ opérations, c'est-à-dire 10 000. L'évaluation de \bar{V}_β est donc 50 000 fois plus rapide avec cette méthode.

2.4 Comparaison des programmes bayésiens

Cette partie compare la performance des différentes approches décrites plus haut. Nous avons testé ces approches sur neuf modèles différents à 2 000, 10 000 et 50 000 observations et 3, 5 et 8 dimensions à partir d'un seul algorithme que nous avons mo-

difié pour chacun des cas. Le premier cas est celui qui se rapproche le plus de l'algorithme de Rossi où l'échantillonnage des variables latentes est faite à partir de la méthode de la fonction de répartition inverse et l'échantillonnage des coefficients de régression est fait à partir de la méthode n'utilisant pas la décomposition matricielle décrite dans la section précédente. Le deuxième cas mesure le temps traitement lorsque la décomposition matricielle est incluse dans l'algorithme de Rossi et le troisième cas mesure le temps de traitement lorsque la décomposition matricielle et la méthode de Chopin sont incluses dans l'algorithme de Rossi. Le nombre de variables explicatives par équation est égal à cinq et le temps de traitement a été mesuré après 5 000 itérations. Tous les tests ont été effectués avec un processeur Intel de 2,4 GHz.

Tableau 2.1 Temps de traitement entre différents programmes

Nombre de dimensions	Nombre d'observations	Rossi	beta	beta + Chopin
3	2 000	0:01:48	0:00:17	0:00:13
3	10 000	0:08:42	0:01:32	0:01:09
3	50 000	0:45:02	0:07:52	0:05:48
5	2 000	0:03:48	0:00:27	0:00:19
5	10 000	0:18:23	0:02:29	0:01:46
5	50 000	1:40:23	0:12:20	0:09:16
8	2 000	0:10:00	0:00:44	0:00:32
8	10 000	0:55:23	0:04:01	0:02:48
8	50 000	4:37:03	0:21:06	0:15:08

La comparaison des résultats au Tableau 2.1 montrent que la modification de l'algorithme pour échantillonner les coefficients de régression permet de réduire considérablement le temps de traitement pour effectuer 5 000 itérations. Lorsque le nombre d'observations est de 50 000 et le nombre d'équations est égal à huit, notre

programme est 13 fois plus rapide que la version proche de celle de Rossi. Le temps de traitement nécessaire pour 5 000 itérations est de 21 minutes comparativement à 4 heures 37 minutes pour le programme de Rossi. Les résultats montrent que la performance du programme avec la décomposition matricielle lorsque le nombre de coefficients de régression augmente et montrent que la transformation de Chopin a permis quant à elle de réduire le temps de traitement d'environ 20 % peu importe la spécification.

CHAPITRE III

PROBIT MULTIVARIÉ AVEC EFFETS INDIVIDUELS

3.1 Introduction

Depuis quelques années, la multiplication des enquêtes longitudinales produites par Statistiques Canada permet aux chercheurs d'incorporer des effets fixes ou aléatoires à leurs modèles. Les modèles avec effets fixes ou aléatoires permettent de contrôler certains facteurs non observés constants dans le temps chez les individus. Pour profiter de cette possibilité, nous proposons de généraliser le modèle probit multivarié pour y inclure des effets aléatoires et de dériver l'estimateur bayésien nécessaire pour l'estimation de ce modèle. Dans ce chapitre nous dérivons les distributions conditionnelles *a posteriori* pour estimer le modèle probit multivarié avec effets individuels hiérarchiques (effets aléatoires) et dérivons l'algorithme MCMC de ce modèle en incorporant la méthode *PXDA* de (Talhouk, Doucet et Murphy, 2012). Étant donné que ce modèle est une généralisation du modèle SURE avec effets individuels hiérarchiques, nous développons dans un premier temps les distributions conditionnelles *a posteriori* de ce modèle. Par la suite, nous développons les distributions conditionnelles *a posteriori* du modèle probit multivarié avec effets individuels hiérarchiques, présentons l'échantillonneur de Gibbs et testons la performance de ce modèle au moyen d'une étude Monte Carlo.

3.2 Modèle multivarié (SURE) avec individuels hiérarchiques

Considérons le modèle multivarié avec effets individuels hiérarchiques suivant :

$$\begin{aligned}
 y_{it} &= \alpha + x_{it}\beta + v_{it}, \\
 v_{it} &= u_i + \varepsilon_{it}, \\
 u_i &\perp \varepsilon_{it}, \\
 v_{it} &\sim N(0, \Sigma_v), \varepsilon_{it} \sim N(0, \Sigma_\varepsilon), u_i \sim N(0, \Sigma_\alpha) \\
 i &= 1, \dots, n \text{ et } t = 1, \dots, T
 \end{aligned} \tag{3.1}$$

où

$$y_{it} = \begin{bmatrix} y_{it1} \\ \vdots \\ y_{itm} \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix}, x_{it} = \begin{bmatrix} x_{it1} & \dots & 0 \\ \vdots & \ddots & \\ 0 & & x_{itm} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}, \varepsilon_{it} = \begin{bmatrix} \varepsilon_{it1} \\ \vdots \\ \varepsilon_{itm} \end{bmatrix}.$$

Étant donné l'hypothèse sur l'indépendance des termes d'erreur, $u \perp \varepsilon$, la matrice de variance covariance Σ_v est égale à la somme des matrices Σ_ε et Σ_α :

$$\Sigma_v = \Sigma_\varepsilon + \Sigma_\alpha = \begin{bmatrix} \sigma_{\alpha_{11}}^2 + \sigma_{\varepsilon_{11}}^2 & \dots & \sigma_{\alpha_{1m}} + \sigma_{\varepsilon_{1m}} \\ \vdots & \ddots & \\ \sigma_{\alpha_{1m}} + \sigma_{\varepsilon_{1m}} & & \sigma_{\alpha_{mm}}^2 + \sigma_{\varepsilon_{mm}}^2 \end{bmatrix}. \tag{3.2}$$

Si nous empilons les équations par rapport à t , nous obtenons :

$$\begin{aligned}
 y_i &= (i_T \otimes \alpha) + x_i \beta + v_i, \\
 v_i &\sim N(0, \Xi), \\
 i &= 1, \dots, n,
 \end{aligned} \tag{3.3}$$

où i_T est un vecteur-colonne de taille T dont l'ensemble des éléments sont égaux à 1.

Finalement, si nous empilons les équations par rapport à i , nous obtenons :

$$y = (i_{Tn} \otimes \alpha) + X\beta + v, \quad (3.4)$$

où

$$E(vv') = (I_n \otimes \Sigma_v) = \begin{bmatrix} \Sigma_{v_1} & & \\ & \ddots & \\ & & \Sigma_{v_n} \end{bmatrix} \quad (3.5)$$

et

$$\Sigma = (i_T i_T' \otimes \Sigma_\alpha) + (I_t \otimes \Sigma_\varepsilon). \quad (3.6)$$

La matrice i_{Tn} est une matrice $(T \times n)$ dont tous les éléments sont égaux à 1.

Il existe deux façons d'inclure les effets individuels dans ce modèle. La première consiste à ajouter une variable dichotomique pour chaque individu (effets fixes), c'est-à-dire d'incorporer au modèle $(n - 1) \times m$ variables supplémentaires. Cependant, avec ce modèle, le nombre de paramètres à estimer est trop élevé pour le nombre d'observations total. Pour estimer ce modèle, il faut incorporer une structure supplémentaire qui permettra de réduire le nombre de paramètres. En économétrie bayésienne, on utilise les densités *a priori* hiérarchiques. On suppose que les effets individuels de toutes les équations proviennent d'une distribution normale multivariée de moyennes α et covariances Σ_α . La densité *a priori* hiérarchique est donnée par :

$$p(\alpha_i | \alpha, \Sigma_\alpha) \propto \exp\left(-\frac{1}{2}(\alpha_i - \alpha)' \Sigma_\alpha^{-1} (\alpha_i - \alpha)\right). \quad (3.7)$$

On suppose que la densité *a priori* de α est distribuée normalement :

$$p(\alpha; \underline{\alpha}, \underline{V}_\alpha) \propto \exp\left(-\frac{1}{2}(\alpha - \underline{\alpha})' \underline{V}_\alpha (\alpha - \underline{\alpha})\right) \quad (3.8)$$

et que la densité *a priori* de Σ_α est distribuée Wishart-Inverse :

$$p(\Sigma_\alpha; \underline{v}_\alpha, \underline{\Sigma}_\alpha) \propto |\Sigma_\alpha|^{-(\underline{v}_\alpha + m + 1)/2} \exp\left(-\frac{1}{2} \text{tr}(\underline{\Sigma}_\alpha \Sigma_\alpha^{-1})\right). \quad (3.9)$$

Finalement, de façon identique au modèle probit multivarié, on suppose que les densités *a priori* de β et Σ_ε sont respectivement normale et Wishart-Inverse :

$$p(\beta) \propto \exp\left(-\frac{1}{2}(\beta - \underline{\beta})' \underline{V}_\beta^{-1} (\beta - \underline{\beta})\right), \quad (3.10)$$

$$p(\Sigma_\varepsilon; \underline{v}_\varepsilon, \underline{\Sigma}_\varepsilon) \propto |\Sigma_\varepsilon|^{-(\underline{v}_\varepsilon + m + 1)/2} \exp\left(-\frac{1}{2} \text{tr}(\underline{\Sigma}_\varepsilon \Sigma_\varepsilon^{-1})\right). \quad (3.11)$$

Les différentes formes que peut prendre la fonction de vraisemblance sont :

$$\begin{aligned} & p(y|\alpha, \beta, \Sigma_\alpha, \Sigma_\varepsilon) \quad (3.12) \\ &= \prod_{i=1}^n \prod_{t=1}^T (2\pi)^{-1/2} |\Sigma_v|^{-1/2} \exp\left(-\frac{1}{2}(y_{it} - \alpha - x_{it}\beta)' \Sigma_v^{-1} (y_{it} - \alpha - x_{it}\beta)\right) \\ &= \prod_{i=1}^n (2\pi)^{-T/2} |\Sigma|^{-T/2} \exp\left(-\frac{1}{2}(y_i - (i_T \otimes \alpha) - x_i\beta)' \Sigma^{-1} (y_i - (i_T \otimes \alpha) - x_i\beta)\right) \\ &= (2\pi)^{-nT/2} |\Sigma|^{-nT/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n (y_i - (i_T \otimes \alpha) - x_i\beta)' \Sigma^{-1} (y_i - (i_T \otimes \alpha) - x_i\beta)\right) \end{aligned}$$

En multipliant la fonction de vraisemblance avec l'ensemble des densités *a priori*, on trouve les densités *a posteriori* conditionnelles présentées au Tableau 3.1. Les différentes étapes pour dériver les distributions conditionnelles du modèle se trouvent à l'appendice A.

Tableau 3.1 Distribution conditionnelle *a posteriori* du modèle SURE avec effets individuels hiérarchiques

Densité conditionnelle	Loi de probabilité	Paramètres de localisation/forme	Paramètres d'échelle
$\Sigma_\varepsilon \{\tilde{\alpha}_i\}, \tilde{\beta}, \tilde{\alpha}, \tilde{\Sigma}_\alpha, Z$	$IW(\tilde{\Sigma}_\varepsilon, \tilde{\nu}_\varepsilon)$	$\tilde{\Sigma}_\varepsilon = \sum_{i=1}^n \sum_{t=1}^T (z_{it} - \tilde{\alpha}_i - x_{it}\tilde{\beta})'(z_{it} - \tilde{\alpha}_i - x_{it}\tilde{\beta}) + \underline{\Sigma}_\varepsilon$	$\tilde{\nu}_\varepsilon = \underline{\nu}_\varepsilon + nT$
$\tilde{\beta} \{\tilde{\alpha}_i\}, \Sigma_\varepsilon, Z$	$N(\tilde{\beta}, \tilde{V}_\beta)$	$\tilde{\beta} = \tilde{V}_\beta \left(\sum_{i=1}^n \sum_{t=1}^T x'_{it} \Sigma_\varepsilon^{-1} (z_{it} - \tilde{\alpha}_i) + \underline{V}_\beta^{-1} \underline{\beta} \right)$	$\tilde{V}_\beta = \left(\sum_{i=1}^n \sum_{t=1}^T x'_{it} \Sigma_\varepsilon^{-1} x_{it} + \underline{V}_\beta^{-1} \right)^{-1}$
$\tilde{\alpha}_i \tilde{\beta}, \tilde{\alpha}, \tilde{\Sigma}_\alpha, \Sigma_\varepsilon, Z$	$N(\tilde{\alpha}_i, \tilde{A})$	$\tilde{\alpha}_i = \tilde{A} \left(\Sigma_\varepsilon^{-1} \sum_{t=1}^T (z_{it} - x_{it}\tilde{\beta}) + \tilde{\Sigma}_\alpha^{-1} \tilde{\alpha} \right)$	$\tilde{A} = (T\Sigma_\varepsilon^{-1} + \tilde{\Sigma}_\alpha^{-1})^{-1}$
$\tilde{\alpha} \{\tilde{\alpha}_i\}, \tilde{\beta}, \tilde{\Sigma}_\alpha, \Sigma_\varepsilon, Z$	$N(\tilde{\alpha}, \tilde{V}_\alpha)$	$\tilde{\alpha} = \tilde{V}_\alpha \left(\tilde{\Sigma}_\alpha^{-1} \sum_{i=1}^n \tilde{\alpha}_i + \underline{V}_\alpha^{-1} \underline{\alpha} \right)$	$\tilde{V}_\alpha = (n\tilde{\Sigma}_\alpha^{-1} + \underline{V}_\alpha^{-1})^{-1}$
$\tilde{\Sigma}_\alpha \{\tilde{\alpha}_i\}, \tilde{\beta}, \tilde{\alpha}, \Sigma_\varepsilon, Z$	$IW(\tilde{\Sigma}_\alpha, \tilde{\nu}_\alpha)$	$\tilde{\Sigma}_\alpha = \sum_{i=1}^n (\tilde{\alpha}_i - \tilde{\alpha})' (\tilde{\alpha}_i - \tilde{\alpha}) \tilde{\Sigma}_\alpha^{-1} + \underline{\Sigma}_\alpha$	$\tilde{\nu}_\alpha = \underline{\nu}_\alpha + n$

3.3 Modèle probit multivarié avec effets individuels hiérarchiques

En combinant les densités conditionnelles *a posteriori* du modèle multivarié à celles du probit multivarié, il est possible d'estimer un modèle probit multivarié avec effets individuels. L'ensemble des distributions conditionnelles *a posteriori* de ce modèle sont présentées au Tableau 3.2 et l'algorithme 2 présente les différentes étapes pour estimer le modèle probit multivarié avec effets individuels aléatoires.

La méthode de Talhouk *et al.* (2012) nous permet de choisir des *a priori* non informatifs pour les coefficients de régression, les moyennes et la matrice de variances-covariances des effets individuels. Les hyperparamètres sont :

$$\underline{\beta} = 0, \underline{V}_\beta^{-1} = 10^{-5}I, \underline{\alpha} = 0, \underline{V}_\alpha^{-1} = 10^{-5}I, \underline{\Sigma}_\alpha^{-1} = I, \underline{\nu}_\alpha = m + 1, \underline{\Sigma}_\varepsilon^{-1} = I, \underline{\nu}_\varepsilon = m + 1.$$

Ces hyperparamètres permettent que les coefficients de corrélations r_{ij} de la matrice de corrélation $p(R)$ soient distribués sur l'intervalle $[-1,1]$ après la tion $\Sigma = DRD$.

Tableau 3.2 Distribution conditionnelle *a posteriori* du modèle probit multivarié avec effets individuels aléatoires

Densité conditionnelle	Loi de probabilité	Paramètres de localisation/forme	Paramètres d'échelle
$w_{itj} w_{it(-j)}, \alpha_i, \beta, R, y_{itj}$	$TN(\hat{w}_{itj}, \hat{R}_{jj})$	$\hat{w}_{itj} = \alpha_{ij} + X'_{ijt}\beta_j + R'_j R_j^{-1} R_j u_{-j}$	$\hat{R}_{jj} = R_{jj} - R'_j R_j^{-1} R_j$
$\Sigma_\varepsilon \{\tilde{\alpha}_i\}, \tilde{\beta}, \tilde{\alpha}, \tilde{\Sigma}_\alpha, Z$	$IW(\tilde{\Sigma}_\varepsilon, \tilde{v}_\varepsilon)$	$\tilde{\Sigma}_\varepsilon = \sum_{i=1}^n \sum_{t=1}^T (z_{it} - \tilde{\alpha}_i - x_{it}\tilde{\beta})'(z_{it} - \tilde{\alpha}_i - x_{it}\tilde{\beta}) + \underline{\Sigma}_\varepsilon$	$\tilde{v}_\varepsilon = \underline{v}_\varepsilon + nT$
$\tilde{\beta} \{\tilde{\alpha}_i\}, \Sigma_\varepsilon, Z$	$N(\tilde{\beta}, \tilde{V}_\beta)$	$\tilde{\beta} = \tilde{V}_\beta \left(\sum_{i=1}^n \sum_{t=1}^T x'_{it} \Sigma_\varepsilon^{-1} (z_{it} - \tilde{\alpha}_i) + \underline{V}_\beta^{-1} \underline{\beta} \right)$	$\tilde{V}_\beta = \left(\sum_{i=1}^n \sum_{t=1}^T x'_{it} \Sigma_\varepsilon^{-1} x_{it} + \underline{V}_\beta^{-1} \right)^{-1}$
$\tilde{\alpha}_i \tilde{\beta}, \tilde{\alpha}, \tilde{\Sigma}_\alpha, \Sigma_\varepsilon, Z$	$N(\tilde{\alpha}_i, \tilde{A})$	$\tilde{\alpha}_i = \tilde{A} \left(\Sigma_\varepsilon^{-1} \sum_{t=1}^T (z_{it} - x_{it}\tilde{\beta}) + \tilde{\Sigma}_\alpha^{-1} \tilde{\alpha} \right)$	$\tilde{A} = (T\Sigma_\varepsilon^{-1} + \tilde{\Sigma}_\alpha^{-1})^{-1}$
$\tilde{\alpha} \{\tilde{\alpha}_i\}, \tilde{\beta}, \tilde{\Sigma}_\alpha, \Sigma_\varepsilon, Z$	$N(\tilde{\alpha}, \tilde{V}_\alpha)$	$\tilde{\alpha} = \tilde{V}_\alpha \left(\tilde{\Sigma}_\alpha^{-1} \sum_{i=1}^n \tilde{\alpha}_i + \underline{V}_\alpha^{-1} \underline{\alpha} \right)$	$\tilde{V}_\alpha = (n\tilde{\Sigma}_\alpha^{-1} + \underline{V}_\alpha^{-1})^{-1}$
$\tilde{\Sigma}_\alpha \{\tilde{\alpha}_i\}, \tilde{\beta}, \tilde{\alpha}, \Sigma_\varepsilon, Z$	$IW(\tilde{\Sigma}_\alpha, \tilde{v}_\alpha)$	$\tilde{\Sigma}_\alpha = \sum_{i=1}^n (\tilde{\alpha}_i - \tilde{\alpha})' (\tilde{\alpha}_i - \tilde{\alpha}) \tilde{\Sigma}_\alpha^{-1} + \underline{\Sigma}_\alpha$	$\tilde{v}_\alpha = \underline{v}_\alpha + n$

Les algorithmes pour le modèle SURE et le modèle probit multivarié sont presque identiques à l'exception de l'étape 2 qui ne doit pas être effectuée pour le modèle SURE.

Algorithme 2

1. Choisir les valeurs de départ pour $W, \beta, R, \{\alpha_i\}_{i=1}^n, \alpha, \Sigma_\alpha$
2. Tirer les variables latentes à partir d'une distribution normale tronquée avec la méthode de Geweke décrite à la section 1.4. :

$$w_{itj} | w_{it(-j)}, \alpha_i, \beta, R, y_{itj}, X_{itj} \sim TN(\hat{w}_{itj}, \hat{R}_{jj})$$

$$i = 1, \dots, n, j = 1, \dots, m \text{ et } t = 1, \dots, T$$

où

$$\hat{w}_{itj} = \alpha_{ij} + X'_{ijt} \beta_j + R'_j R_j^{-1} R_j u_{-j}$$

$$\hat{R}_{jj} = R_{jj} - R'_j R_j^{-1} R_j$$

3. Tirer la matrice diagonale \tilde{D} permettant de transposer le modèle dans l'espace non contraint. Les éléments de la diagonale sont égaux à :

$$d_{ii} \sim IG((m+1)/2, r^{ii}/2).$$

4. À partir de la matrice \tilde{D} , transposer les différentes variables du modèle vers l'espace non contraint :

- $Z = \tilde{D} \cdot W$
- $\tilde{\beta} = \tilde{D} \cdot \beta$
- $\tilde{\alpha}_i = \tilde{D} \cdot \alpha_i$
- $\tilde{\alpha} = \tilde{D} \cdot \alpha$
- $\tilde{\Sigma}_\alpha = \tilde{D} \cdot \Sigma_\alpha \cdot \tilde{D}$.

5. Tirer la matrice des variances-covariances des erreurs à partir d'une distribution Wishart inverse :

$$\Sigma_\varepsilon | \{\tilde{\alpha}_i\}, \tilde{\beta}, \tilde{\alpha}, \tilde{\Sigma}_\alpha, Z \sim IW(\bar{\Sigma}_\varepsilon, \bar{\nu}_\varepsilon)$$

où

$$\bar{\Sigma}_\varepsilon = \sum_{i=1}^n \sum_{t=1}^T (z_{it} - \tilde{\alpha}_i - x_{it}\tilde{\beta})' (z_{it} - \tilde{\alpha}_i - x_{it}\tilde{\beta}) + \underline{\Sigma}_\varepsilon$$

et

$$\bar{\nu}_\varepsilon = \underline{\nu}_\varepsilon + nT$$

6. Tirer le vecteur des coefficients de régression non contraint $\tilde{\beta}$ à partir d'une distribution normale multivariée :

$$\tilde{\beta} | \{\tilde{\alpha}_i\}, \Sigma_\varepsilon, Z \sim N(\bar{\beta}, \bar{V}_\beta)$$

où

$$\bar{\beta} = \bar{V}_\beta \left(\sum_{i=1}^n \sum_{t=1}^T x_{it}' \Sigma_\varepsilon^{-1} (z_{it} - \tilde{\alpha}_i) + \underline{V}_\beta^{-1} \underline{\beta} \right)$$

et

$$\bar{V}_\beta = \left(\sum_{i=1}^n \sum_{t=1}^T x_{it}' \Sigma_\varepsilon^{-1} x_{it} + \underline{V}_\beta^{-1} \right)^{-1}$$

7. Tirer les effets individuels non contraints $\{\alpha_i\}_{i=1}^n$ à partir d'une distribution normale :

$$\tilde{\alpha}_i | \tilde{\beta}, \tilde{\alpha}, \tilde{\Sigma}_\alpha, \Sigma_\varepsilon, Z \sim N(\bar{\alpha}_i, \bar{A})$$

où

$$\bar{\alpha}_i = \bar{A} \left(\Sigma_\varepsilon^{-1} \sum_{t=1}^T (z_{it} - x_{it}\tilde{\beta}) + \tilde{\Sigma}_\alpha^{-1} \tilde{\alpha} \right)$$

et

$$\bar{A} = (T\Sigma_\varepsilon^{-1} + \tilde{\Sigma}_\alpha^{-1})^{-1}$$

8. Tirer la moyenne des effets individuels non contraints $\tilde{\alpha}$ à partir d'une distribution normale :

$$\tilde{\alpha} | \{\tilde{\alpha}_i\}, \tilde{\beta}, \tilde{\Sigma}_\alpha, \Sigma_\varepsilon, Z \sim N(\bar{\alpha}, \bar{V}_\alpha)$$

où

$$\bar{\alpha} = \bar{V}_\alpha \left(\tilde{\Sigma}_\alpha^{-1} \sum_{i=1}^n \tilde{\alpha}_i + \underline{V}_\alpha^{-1} \underline{\alpha} \right)$$

et

$$\bar{V}_\alpha = (n\tilde{\Sigma}_\alpha^{-1} + \underline{V}_\alpha^{-1})^{-1}.$$

9. Tirer la matrice de variances-covariances des effets individuels non contraints $\tilde{\Sigma}_\alpha$ à partir d'une distribution Wishart inverse :

$$\tilde{\Sigma}_\alpha | \{\tilde{\alpha}_i\}, \tilde{\beta}, \tilde{\alpha}, \Sigma_\varepsilon, Z \sim IW(\bar{\Sigma}_\alpha, \bar{v}_\alpha)$$

où

$$\bar{\Sigma}_\alpha = \sum_{i=1}^n (\tilde{\alpha}_i - \tilde{\alpha})' (\tilde{\alpha}_i - \tilde{\alpha}) \tilde{\Sigma}_\alpha^{-1} + \underline{\Sigma}_\alpha$$

et

$$\bar{v}_\alpha = \underline{v}_\alpha + n.$$

10. Trouver D^{-1} , la matrice diagonale de l'équation 1.25 à partir de la matrice de variances-covariances :

$$D^{-1} = \begin{bmatrix} 1/\sqrt{\Sigma_{11}} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\Sigma_{11}} & & \\ \vdots & & \ddots & \\ 0 & & & 1/\sqrt{\Sigma_{mm}} \end{bmatrix},$$

où Σ_{jj} est le $j^{\text{ème}}$ élément de la diagonale de Σ_ε .

11. Transposer de nouveau le modèle dans l'espace contraint

- $\beta = D^{-1} \cdot \tilde{\beta}$
- $R = D^{-1} \cdot \Sigma_{\varepsilon} \cdot D^{-1}$
- $\alpha_i = D^{-1} \cdot \tilde{\alpha}_i$ pour $i = 1, \dots, n$
- $\alpha = D^{-1} \cdot \tilde{\alpha}$
- $\Sigma_{\alpha} = D^{-1} \cdot \tilde{\Sigma}_{\alpha} \cdot D^{-1}$.

12. Recommencer B fois les étapes 2 à 11 en utilisant les valeurs des itérations précédentes.

3.4 Simulations du probit multivarié avec effets individuels hiérarchiques

Pour tester la performance de ce modèle, nous avons effectué trois simulations à partir d'une modèle probit multivarié avec effets hiérarchiques. Nous avons simulé 25 banques de données longitudinales avec un nombre de période égale à $t = 8$ et un nombre d'individus égal à $n = 4\,000$. Le modèle est le suivant :

$$\begin{aligned}
 w_{i1} &= \alpha_{i1} + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \beta_3 \cdot x_{i3} + \varepsilon_{i1} \\
 w_{i2} &= \alpha_{i1} + \beta_4 \cdot x_{i4} + \beta_5 \cdot x_{i5} + \beta_6 \cdot x_{i6} + \varepsilon_{i2} \\
 w_{i3} &= \alpha_{i3} + \beta_7 \cdot x_{i7} + \beta_8 \cdot x_{i8} + \beta_9 \cdot x_{i9} + \varepsilon_{i3} \\
 w_{i4} &= \alpha_{i4} + \beta_{10} \cdot x_{i10} + \beta_{11} \cdot x_{i11} + \beta_{12} \cdot x_{i12} + \varepsilon_{i4} \\
 w_{i5} &= \alpha_{i5} + \beta_{13} \cdot x_{i13} + \beta_{14} \cdot x_{i14} + \beta_{15} \cdot x_{i15} + \varepsilon_{i5}
 \end{aligned} \tag{3.13}$$

$$y_j = \begin{cases} 1 & \text{si } w_j \geq 0 \\ 0 & \text{si } w_j < 0 \end{cases}, \quad j = 1, 2, 3, 4, 5$$

$$x_k \sim \text{uniform}[0, 1]$$

$$(\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}, \varepsilon_{i4}, \varepsilon_{i5}) \sim N_5(0, R)$$

$$(\alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \alpha_{i4}, \alpha_{i5}) \sim N_5(\alpha_i, \Sigma_\alpha)$$

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} & r_{14} & r_{15} \\ & 1 & r_{23} & r_{24} & r_{25} \\ & & 1 & r_{34} & r_{35} \\ & & & 1 & r_{45} \\ & & & & 1 \end{bmatrix} \text{ et } \Sigma_\alpha = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ & & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ & & & \sigma_4^2 & \sigma_{45} \\ & & & & \sigma_5^2 \end{bmatrix}$$

$$\underline{\beta} = 0, \underline{V}_\beta^{-1} = 10^{-5}I, \underline{\alpha} = 0, \underline{V}_\alpha^{-1} = 10^{-5}I, \underline{\Sigma}_\alpha^{-1} = I, \underline{\nu}_\alpha = m + 1, \underline{\Sigma}_\varepsilon^{-1} = I, \underline{\nu}_\varepsilon = m + 1.$$

Les Tableaux 3.3 à 3.7 présentent les coefficients utilisés pour cette simulation ainsi que les résultats moyens obtenues pour les 25 simulations de 10 000 tirages de Gibbs chacun. Pour chacun des paramètres estimés, nous avons calculé le MCSE (*Monte Carlo standard error*) avec la méthode des « *batch mean* » (Flegal et Jones, 2010). Les MCSE obtenues sont très faibles et démontrent que le nombre de tirages est suffisant pour obtenir une précision numérique élevée de l'estimateur. Le premier modèle ($t = 8$ et $n = 4\,000$) est le modèle de référence. Le deuxième modèle a été estimé à partir des mêmes données simulées, mais en éliminant la moitié des données ($t = 8$ et $n = 2\,000$). Le troisième modèle est aussi estimé avec les données du premier modèle, mais en éliminant la moitié des périodes ($t = 4$ et $n = 4\,000$). Le deuxième et troisième modèle compte un total de 16 000 observations comparativement aux 32 000 observations du premier modèle. Pour chaque paramètre θ , l'erreur quadratique moyenne ou *mean-squared error* (MSE) est donnée par

$$MSE(\bar{\theta}|\theta) = E(\bar{\theta} - \theta)^2$$

où $\bar{\theta}$ est l'estimateur du paramètre θ . Comme prévu, l'erreur quadratique moyenne augmente pour l'ensemble des paramètres estimés lorsque le nombre d'observations diminue. Dans l'ensemble, les résultats des trois simulations sont tout à fait raisonnables : les moyennes estimées sont proches des vraies valeurs et l'écart-type est relativement faible. Le nombre total d'observation semble avoir un impact sur l'erreur quadratique moyenne, par contre, Contrairement à notre intuition, le nombre de cycle longitudinaux passant de ($t = 8$ et $n = 2\,000$) à ($t = 4$ et $n = 4\,000$) ne change pas l'erreur quadratique moyenne des variances et covariances des effets individuels hiérarchiques. Seul le nombre d'observations affecte l'erreur quadratique moyenne. Des tests supplémentaires sont nécessaires pour vérifier ce résultat.

Tableau 3.3 Résultats des simulations, coefficients de régression

	vraies valeurs	t = 8, n = 4 000				t = 8, n = 2 000				t = 4, n = 4 000			
		moy.	écart-t.	MCSE	MSE	moy.	écart-t.	MCSE	MSE	moy.	écart-t.	MCSE	MSE
β_1	-1,3	-1,30	0,04	0,001	0,04	-1,30	0,06	0,001	0,06	-1,29	0,05	0,001	0,05
β_2	0,8	0,81	0,03	0,000	0,04	0,80	0,05	0,000	0,05	0,79	0,05	0,000	0,05
β_3	0,3	0,30	0,03	0,000	0,03	0,29	0,04	0,000	0,04	0,30	0,05	0,000	0,05
β_4	0,9	0,90	0,04	0,000	0,03	0,88	0,05	0,000	0,05	0,90	0,05	0,001	0,05
β_5	0,3	0,30	0,03	0,000	0,03	0,29	0,05	0,000	0,05	0,28	0,06	0,000	0,06
β_6	-1,5	-1,50	0,03	0,001	0,03	-1,49	0,04	0,002	0,04	-1,50	0,05	0,001	0,05
β_7	1,0	0,98	0,04	0,001	0,04	0,98	0,05	0,001	0,05	1,01	0,06	0,001	0,06
β_8	-0,6	-0,60	0,04	0,000	0,04	-0,61	0,05	0,000	0,05	-0,59	0,05	0,000	0,05
β_9	1,3	1,30	0,04	0,001	0,04	1,30	0,05	0,001	0,05	1,29	0,06	0,001	0,06
β_{10}	-1,1	-1,10	0,03	0,001	0,03	-1,09	0,05	0,001	0,05	-1,11	0,04	0,001	0,04
β_{11}	0,1	0,10	0,03	0,000	0,03	0,11	0,04	0,000	0,04	0,10	0,04	0,000	0,04
β_{12}	0,7	0,70	0,03	0,000	0,03	0,69	0,06	0,000	0,06	0,71	0,04	0,000	0,04
β_{13}	-0,2	-0,20	0,03	0,000	0,03	-0,21	0,04	0,000	0,04	-0,20	0,04	0,000	0,04
β_{14}	1,1	1,09	0,04	0,001	0,04	1,09	0,04	0,001	0,04	1,12	0,05	0,001	0,05
β_{15}	0,7	0,69	0,03	0,000	0,03	0,71	0,05	0,000	0,05	0,70	0,05	0,000	0,05

Tableau 3.4 Résultats des simulations : corrélations du terme d'erreur

vraies valeurs		t = 8, n = 4 000				t = 8, n = 2 000				t = 4, n = 4 000			
		moy.	écart-t.	MCSE	MSE	moy.	écart-t.	MCSE	MSE	moy.	écart-t.	MCSE	MSE
r_{12}	0,6	0,56	0,01	0,000	0,04	0,54	0,02	0,000	0,06	0,54	0,02	0,000	0,06
r_{13}	0,4	0,33	0,02	0,000	0,03	0,33	0,02	0,000	0,04	0,32	0,02	0,000	0,04
r_{14}	-0,4	-0,40	0,01	0,000	0,02	-0,40	0,02	0,000	0,02	-0,39	0,02	0,000	0,02
r_{15}	-0,2	-0,20	0,02	0,000	0,02	-0,20	0,02	0,000	0,02	-0,20	0,03	0,000	0,03
r_{23}	-0,5	-0,45	0,01	0,000	0,02	-0,45	0,02	0,000	0,03	-0,45	0,02	0,000	0,03
r_{24}	0,0	0,02	0,01	0,000	0,02	0,01	0,02	0,000	0,03	0,01	0,02	0,000	0,03
r_{25}	-0,3	-0,29	0,02	0,000	0,02	-0,30	0,03	0,000	0,03	-0,30	0,03	0,000	0,03
r_{34}	-0,6	-0,56	0,01	0,000	0,01	-0,56	0,02	0,000	0,02	-0,55	0,02	0,000	0,03
r_{35}	0,1	0,07	0,02	0,000	0,02	0,07	0,02	0,000	0,03	0,08	0,02	0,000	0,02
r_{45}	0,5	0,54	0,01	0,000	0,01	0,53	0,01	0,000	0,01	0,54	0,02	0,000	0,02

Tableau 3.5 Résultats des simulations : moyenne des effets individuels hiérarchiques

vraies valeurs		t = 8, n = 4 000				t = 8, n = 2 000				t = 4, n = 4 000			
		moy.	écart-t.	MCSE	MSE	moy.	écart-t.	MCSE	MSE	moy.	écart-t.	MCSE	MSE
α_1	0,6	0,59	0,03	0,000	0,03	0,59	0,03	0,000	0,04	0,58	0,03	0,000	0,04
α_2	-1,5	-1,50	0,02	0,001	0,02	-1,50	0,04	0,001	0,04	-1,50	0,03	0,001	0,03
α_3	1,5	1,50	0,02	0,001	0,02	1,51	0,03	0,001	0,03	1,50	0,03	0,001	0,03
α_4	-0,6	-0,60	0,02	0,000	0,02	-0,60	0,03	0,000	0,03	-0,60	0,02	0,000	0,02
α_5	1,1	1,10	0,03	0,000	0,03	1,09	0,04	0,001	0,04	1,10	0,02	0,001	0,02

Tableau 3.6 Résultats des simulations : variances des effets individuels hiérarchiques

vraies valeurs		t = 8, n = 4 000				t = 8, n = 2 000				t = 4, n = 4 000			
		moy.	écart-t.	MCSE	MSE	moy.	écart-t.	MCSE	MSE	moy.	écart-t.	MCSE	MSE
a_{11}	2,6	2,54	0,07	0,003	0,07	2,55	0,08	0,003	0,08	2,53	0,11	0,003	0,11
a_{22}	0,8	0,81	0,04	0,000	0,04	0,80	0,05	0,000	0,04	0,81	0,05	0,000	0,05

a33	0,7	0,75	0,04	0,000	0,04	0,76	0,06	0,000	0,06	0,75	0,05	0,000	0,05
a44	1,5	1,48	0,06	0,001	0,06	1,46	0,07	0,001	0,07	1,5	0,07	0,001	0,07
a55	2,0	1,97	0,07	0,001	0,07	1,99	0,09	0,002	0,09	1,99	0,12	0,002	0,12

Tableau 3.7 Résultats des simulations : covariances des effets individuels hiérarchiques

	vraies valeurs	t = 8, n = 4 000				t = 8, n = 2 000				t = 4, n = 4 000			
		moy.	écart-t.	MCSE	MSE	moy.	écart-t.	MCSE	MSE	moy.	écart-t.	MCSE	MSE
a12	0,4	0,41	0,03	0,000	0,03	0,41	0,05	0,000	0,05	0,42	0,04	0,000	0,04
a13	0,0	0,01	0,03	0,000	0,03	0,01	0,05	0,000	0,05	0,02	0,04	0,000	0,04
a14	-0,7	-0,70	0,04	0,000	0,04	-0,70	0,06	0,000	0,06	-0,71	0,05	0,000	0,05
a15	0,0	0,03	0,04	0,000	0,04	0,07	0,07	0,000	0,07	0,04	0,05	0,000	0,05
a23	-0,6	-0,61	0,03	0,000	0,03	-0,60	0,04	0,000	0,04	-0,63	0,03	0,000	0,04
a24	-0,2	-0,15	0,03	0,000	0,03	-0,10	0,03	0,000	0,03	-0,21	0,03	0,000	0,03
a25	0,0	-0,03	0,03	0,000	0,03	-0,00	0,04	0,000	0,04	-0,01	0,03	0,000	0,03
a34	0,1	0,10	0,03	0,000	0,03	0,08	0,04	0,000	0,04	0,09	0,03	0,000	0,03
a35	-0,1	-0,05	0,03	0,000	0,03	-0,02	0,03	0,000	0,04	-0,09	0,04	0,000	0,04
a45	-0,9	-0,85	0,03	0,000	0,03	-0,9	0,06	0,001	0,06	-0,9	0,03	0,001	0,03

CHAPITRE IV

MARCHÉ DU TRAVAIL ET MALADIES CHRONIQUES

4.1 Introduction

Les maladies chroniques sont la principale cause de décès et d'incapacités à travers le monde. Selon l'Organisation Mondiale de la Santé (OMS), les maladies chroniques sont responsables de 60 % des décès à travers le monde, même si dans la plupart des cas, ces décès étaient évitables.⁴ Les maladies chroniques ne sont pas uniquement responsables de nombreux décès, elles sont également responsables d'une perte importante de bien-être liée à la souffrance physique et à la détresse psychologique. Les coûts directs constitués des dépenses de santé tels que le salaire des médecins, les frais d'hospitalisation et les médicaments représentaient au Canada plus de 10 % du PIB en 2006, c'est-à-dire 4 867 \$ par habitant. De ces dépenses, 15,2 % étaient consacrées aux maladies cardiovasculaires, 9,4 % aux maladies pulmonaires et 10,1 % aux cancers.⁵ Les coûts indirects sont également colossaux. La baisse de la participation au marché du travail, l'absentéisme, la perte de productivité, les retraites hâtives

⁴ http://www.who.int/topics/chronic_disease/fr/

⁵ Les soins de santé au Canada 2008, Institut canadien d'information sur la santé, http://secure.cihi.ca/cihiweb/products/HCIC_2008_f.pdf

et les décès prématurés sont tous des coûts supplémentaires engendrés par les maladies chroniques. Si rien n'est fait, la progression des taux de morbidité associé à chacune des maladies chroniques risque de faire exploser les coûts directs et indirects dans un proche avenir.

Tableau 4.1 Facteurs de risque des maladies chroniques

Naturels	Âge, sexe, génétique
Comportementaux	Tabac, activités physiques, alimentation, alcool
Intermédiaires	Cholestérol, diabète, hypertension, obésité
Sociodémographiques	Niveau de scolarité, pauvreté, emploi, composition familiale
Environnementaux	Climat, pollution
Culturels	Pratiques, normes et valeurs
Urbains	Logement, accès aux produits et services de santé

Sources : Santé Canada

À l'exception des facteurs de risque naturels, le Tableau 4.1 montre qu'il est possible de réduire le taux de morbidité des maladies chroniques et les coûts associés en modifiant les comportements des individus ou en transformant leur environnement. Dans le cas du tabagisme, les gouvernements ont mis en place depuis plusieurs années des mesures pour tenter de réduire l'usage du tabac. Ils ont par exemple augmenté les taxes sur les produits du tabac, sensibilisé la population sur les dangers du tabagisme et banni la fumée secondaire dans certains lieux publics. On sait aujourd'hui que le tabagisme est un des principaux facteurs de risque de plusieurs cancers, des maladies cardiovasculaire et pulmonaires. Au Canada, le nombre de fumeurs est passé de 35%

à 18 % entre 1985 et 2009.⁶ Malgré tous ces efforts, le tabagisme est un problème toujours présent chez certains groupes de la société comme les jeunes et les personnes à faibles revenus.

Problème relativement nouveau, l'obésité est devenue en quelques années une question de santé publique prioritaire. Depuis 25 ans, le nombre d'individus obèses ou en surpoids s'est accru de manière considérable au Canada. Entre 1978 et 2005, le pourcentage de canadiens obèses a presque doublé, passant de 14 % à 24 %.⁷ L'obésité étant un des principaux facteurs de risque pour les maladies cardio-vasculaires, le diabète, l'hypertension, l'arthrite et plusieurs cancers, l'augmentation du poids moyen de la population menace d'engorger le système de santé. L'impact sur la santé de l'obésité pourrait être aussi grave que celui du tabagisme. Pour cette raison, plusieurs initiatives ont été mises en place par les gouvernements pour freiner cette épidémie. Par exemple, une alimentation riche en fruits et légumes et une activité physique plus fréquente sont devenues le leitmotiv de la plupart des campagnes de prévention en santé. Même si ces campagnes sont nécessaires, plusieurs questions doivent être posées. Quelle sont les bénéfices de telles campagnes ? Sont-elles efficaces ? Si oui, sont-elles suffisantes ? Une analyse coût-bénéfice est nécessaire si nous voulons répondre à ces questions, mais ceci exige de connaître l'ensemble des coûts directs et indirects des maladies chroniques.

⁶Santé Canada (2009). L'Enquête de surveillance de l'usage du tabac au Canada.

⁷ Agence de la santé publique du Canada (2009) Obésité au Canada.

4.1.1 La santé et ses facteurs de risque

Nous avons reproduit graphiquement la corrélation existant entre la santé, le revenu, le niveau d'éducation et les habitudes de vie à partir des données de l'enquête sur la santé des communautés canadiennes (ESCC) de 2007. L'indicateur de santé utilisé dans ces graphiques est la perception moyenne des individus à propos de leur propre santé (excellente, très bonne, bonne, passable ou mauvaise). Cette moyenne a été calculée en attribuant aux différentes réponses un nombre allant de zéro à quatre où quatre signifie une excellente santé. Pour une plus grande lisibilité, les graphiques ont été lissés avec une moyenne mobile utilisant les deux années précédentes et les deux années suivantes.

Les Figures 4.1 et 4.2 présentent l'état de santé selon les revenus et le niveau de scolarité pour chacune des tranches d'âge. On remarque qu'il existe une forte corrélation entre ces variables et le niveau de santé. En particulier, on constate une moyenne de santé auto-déclarée plus élevée chez les riches que chez les pauvres. Les individus les plus riches doivent avoir au moins 70 ans avant de ressentir un niveau de santé inférieur aux individus les plus pauvres âgés de 30 ans. On remarque également que la santé des plus riches varie très peu dans le temps, alors que celle des plus pauvres diminue rapidement avant de remonter un peu vers la soixantaine. Les conditions de pauvreté dans lesquelles certains individus vivent pourraient expliquer une partie de la différence observée. Par exemple, des revenus plus élevés permettraient d'acquérir des soins supplémentaires non couverts par les assurances et de mitiger une partie des effets négatifs de la maladie. Également, des revenus plus élevés permettraient d'investir dans la santé préventive comme la consommation de fruits et légumes et l'achat de matériel pour la pratique de sports.

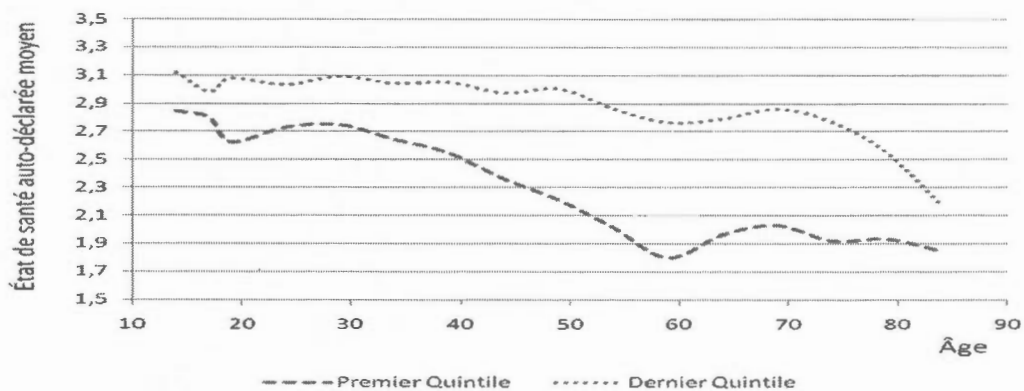


Figure 4.1 Santé auto-déclarée moyenne - revenu (ESCC 2007)

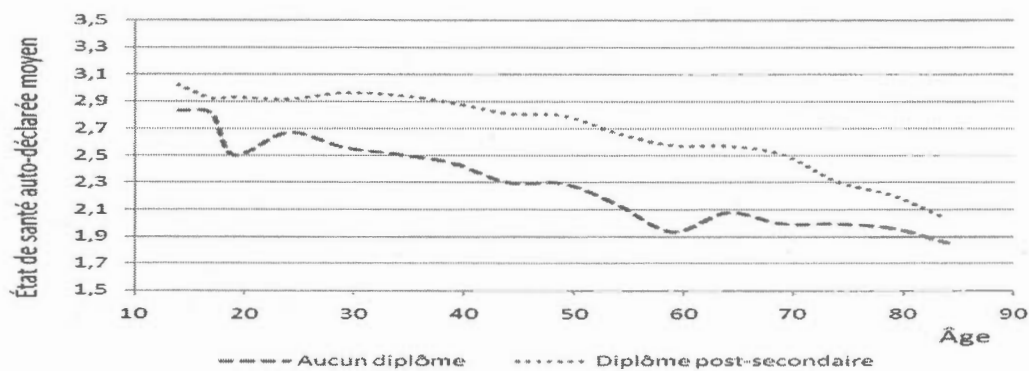


Figure 4.2 Santé auto-déclarée moyenne - scolarité (ESCC 2007)

De son côté, l'éducation pourrait avoir un impact sur la santé en cantonnant les individus dans certains secteur d'emploi. En effet, les emplois qui exigent un niveau faible de scolarité sont souvent monotones, précaires, dures physiquement et se déroulent parfois dans un environnement dangereux. Tous ces facteurs peuvent avoir un impact négatif sur la santé mentale et physique d'un individu. De façon différente,

l'éducation permettrait également aux individus de produire ou d'acquérir plus efficacement de la santé. Par exemple, les individus scolarisés pourraient davantage profiter des campagnes de prévention et des indications du médecin. La Figure 4.2 trace le niveau de santé pour des individus avec un diplôme postsecondaire et les individus sans diplôme secondaire. Contrairement aux revenus, la différence entre les deux groupes est relativement constante dans le temps. Cette différence débute durant l'enfance et se poursuit tout au long de la vie. Si on compare les Figures 4.1 et 4.2, on voit que les courbes sont quasiment identiques pour les individus pauvres et les individus sans diplôme, alors qu'il y a des différences entre les riches et les individus diplômés. L'éducation ne peut donc être la seule raison des différences observées entre les différents groupes sociodémographiques. La forte corrélation entre la santé, l'éducation et la richesse tend à démontrer que ces trois variables sont intimement liées. Malheureusement, cette corrélation ne dit rien sur les liens de causalité qui sous-tendent ces relations.

La situation socio-économique étant fortement corrélée avec les habitudes de consommation dommageable pour la santé, celles-ci pourraient expliquer en partie les différences de santé observées entre les différents groupes. Les Figures 4.3 et 4.4 et montrent la forte corrélation entre les revenus, le tabagisme et l'activité physique. La Figure 4.3 indique que les individus les plus pauvres fument davantage que les riches et pendant plus longtemps. Par ailleurs, on découvre à la Figure 4.4 que le niveau d'activité physique change très rapidement à la sortie de l'école secondaire. Il est probable que le revenu joue un rôle important dans cette observation car les individus pauvres n'ont plus accès à l'équipement sportif à la sortie du secondaire. D'autre part, il est possible que les préférences des individus qui dictent leurs habitudes de consommation tout au long de leur vie se soient formées très jeunes au contact de leurs

parents. Si les parents influencent également la scolarité de l'individu, ceci expliquerait en partie la corrélation observée entre la santé, l'éducation, le revenu et les habitudes de vie.

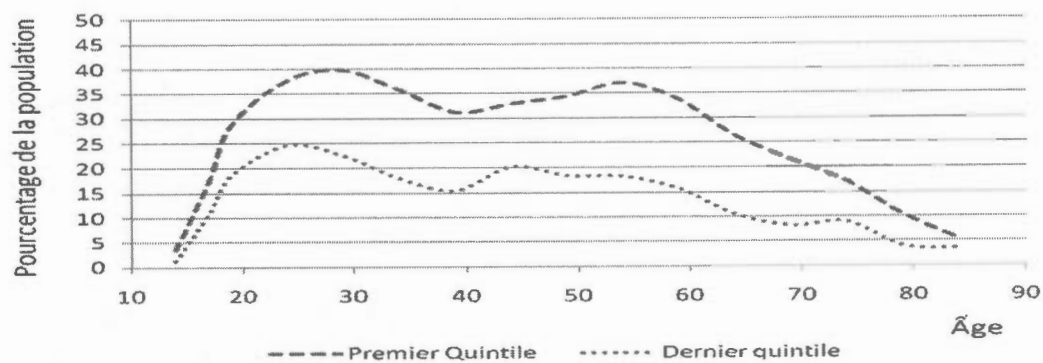


Figure 4.3 Habitudes de vie selon le revenu - tabac (ESCC 2007)

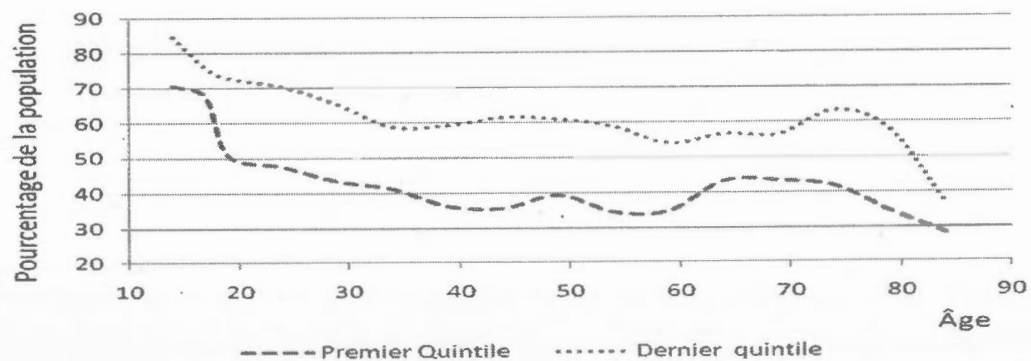


Figure 4.4 Habitudes de vie selon le revenu - activité physique (ESCC 2007)

Le lien entre les facteurs socio-économiques et la santé peuvent également s'expliquer dans l'autre sens. Habituellement, on considère la santé comme une forme de capital humain qui a des impacts importants sur le plan professionnel et social des individus tout au long de leur vie. Par exemple, on observe fréquemment chez les individus qui ont eu des problèmes de santé durant leur enfance des salaires plus faibles à l'âge adulte (Smith, 2009). Cette différence de salaire s'explique généralement par le niveau de scolarité plus faible causé par les retards scolaires durant l'enfance. À l'âge adulte, on observe également des salaires plus faibles chez les individus avec des problèmes de santé (Cai, 2009). Cette fois-ci, le rôle de la santé est direct car il influence directement la productivité du travailleur. Les travailleurs en mauvaise santé pourraient avoir des problèmes de concentration ou de fatigue. L'effet de la santé sur la situation socio-économique pourrait également s'expliquer par la préférence pour le présent. L'espérance de vie réduite des individus en mauvaise santé diminuerait leur investissement en capital humain au profit d'une consommation immédiate.

4.1.1 Santé et participation au marché du travail

Il existe une large littérature mesurant l'impact de la santé sur l'offre de travail. Ces travaux ont montré qu'il existe une forte corrélation entre la santé et le travail : les individus malades travaillent moins que les individus en santé. Par contre, les résultats divergent sur la magnitude de cet impact. Une des raisons de ces différences est la mesure de santé. Currie et Madrian (1999) ont recensé huit façons de mesurer la santé : la santé auto-déclarée, les limitations au travail, les limitations dans la vie de tous les jours, la présence de conditions chroniques, l'utilisation de soins de santé, les évaluations cliniques tel que l'alcoolisme ou la santé mentale, l'état nutritionnel (in-

dice de masse corporelle) et l'espérance de vie. Selon eux, l'estimation de l'effet de la santé sur l'offre de travail est très sensible à la mesure de santé utilisée. L'endogénéité des mesures subjectives comme la santé auto-déclarée ou les limitations serait responsable de cette sensibilité : les individus auraient des raisons psychologiques ou sociales d'altérer leur déclaration de santé pour que celle-ci corresponde mieux à leur situation sur le marché du travail. Pour éviter ce problème, certains chercheurs ont préféré se tourner vers des mesures plus objectives de la santé comme la présence de conditions physiques ou l'utilisation de soins de santé. Même si ces mesures sont beaucoup moins subjectives que la mesure de santé auto-déclarée, l'endogénéité reste néanmoins un problème car il existe des facteurs explicatifs non observés communs à la santé et à la participation au marché du travail.

Pour éliminer cette source d'endogénéité, les chercheurs ont ajouté des variables instrumentales au modèle (Norton et Han, 2008) ou construit un modèle endogène avec structure récursive bivariée (Brown, Pagán et Bastida, 2005) ou multivariée (Zhang, Zhao et Harris, 2009). Zhang *et al.* (2009) souligne que les estimés d'un modèle bivarié sont non convergents s'il existe une comorbidité importante entre les maladies chroniques. Ils donnent comme exemple la forte corrélation entre le diabète et les maladies cardiovasculaires. Dans ce cas, si les maladies cardio-vasculaires sont omises, l'estimé du diabète ne convergera pas vers la vraie valeur omises du modèle. Le modèle qu'ils ont proposé est un modèle probit multivarié avec structure récursive à cinq dimensions. L'estimation conjointe de la participation au marché du travail, des maladies cardiaques, du diabète, de la santé mentale leur ont permis d'estimer l'effet de la santé sur la participation au marché du travail sans le biais occasionné (sous certaines restrictions) par l'endogénéité de la mesure de santé.

Currie et Madrian (1999) ont également souligné que la sensibilité des résultats pouvait dépendre du contexte social des individus. Par exemple, Costa (1996) ont montré qu'une santé physique déficiente avait un impact beaucoup plus important sur le travail il y a cent ans qu'aujourd'hui. Si l'époque influence l'effet de la santé sur la participation au marché du travail, il est probable que le sexe, l'âge, l'éducation et plusieurs autres variables influencent aussi cette relation. Il est donc important de contrôler pour ces variables. Par contre, certains facteurs de risque tels que l'environnement et les comportements passés sont difficilement observables. Si ces facteurs ne sont pas observables mais sont corrélés à la fois avec la santé et la participation au marché du travail, il est probable que les résultats obtenus seront biaisés.

Dans ce chapitre, nous examinerons l'impact de cinq maladies chroniques (les maladies de cœur, les maladies pulmonaires, l'arthrite, les maux de dos et le diabète) sur la participation au marché du travail en utilisant un modèle probit multivarié avec structure réursive endogène. Pour tenir compte des facteurs de risque non observables, nous incorporons au modèle des effets aléatoires spécifiques aux individus.

4.2 Les données

Les données utilisées pour estimer l'impact des maladies chroniques sur la probabilité de travailler sont les données de l'enquête nationale sur la santé des populations (ENSP). L'ENSP est une enquête longitudinale qui collecte des données sur les mêmes individus depuis 1994. Elle contient des renseignements sur des sujets très variés, dont l'activité physique, la taille et le poids, l'usage du tabac, l'exposition à la fumée secondaire, la consommation d'alcool, l'état de santé général, les problèmes de santé chroniques, les blessures et l'utilisation des services de santé, ainsi que les ca-

ractéristiques socioéconomiques, le revenu et la situation professionnelle de la population.

À partir des neuf cycles disponibles de l'ENSP, nous avons formé deux sous-échantillons d'hommes et femmes âgés de 30 à 65 ans. Les individus âgés de moins de 30 ans ont été exclus à cause du faible taux de morbidité associé à cette tranche d'âge et les individus âgés de 65 ans et plus ont été exclus en raison des départs à la retraite. Nous avons également éliminé les femmes enceintes des deux échantillons et certains problèmes de santé liés à la grossesse, telle que le diabète gestationnel ainsi que la prise de poids, ne sont que provisoires dans bien des cas. Finalement, nous avons exclus de notre échantillon les observations comportant des données manquantes. Les échantillons finaux de l'ENSP sont non balancés. Le minimum de périodes observées par individu est de trois et le maximum est de neuf.

L'équation structurelle du modèle est la participation au marché du travail. Selon les variables disponibles dans la banque de données, les individus sont considérés sur le marché du travail s'ils avaient un emploi la semaine précédant l'enquête ou étaient à la recherche d'un emploi depuis les quatre dernières semaines. Ainsi, la variable dichotomique de la participation au marché du travail est égale à 1 si les individus sont employés ou à la recherche de travail et 0 sinon. Les variables explicatives de l'équation ont été choisies en se basant sur la littérature en économie du travail. Ces variables sont l'âge, le niveau d'éducation, l'immigration, l'état matrimonial, le nombre d'enfants et la région canadienne de résidence. À ces variables, nous avons incorporé cinq états de santé : l'arthrite, les maux de dos, les maladies pulmonaires, les maladies cardio-vasculaires et le diabète. Les valeurs des variables dichotomiques associées à chacune de ces maladies chroniques sont les réponses auto-déclarées (oui/non)

des répondants sur leurs problèmes de santé de longue durée. Par problème de santé de longue durée, Statistiques Canada entend un état qui dure ou qui devrait durer 6 mois ou plus et qui a été diagnostiqué par un professionnel de la santé. Pour permettre l'identification du modèle, les facteurs de risque présentés au Tableau 4.1 sont utilisés comme variables explicatives des maladies chroniques. Les statistiques descriptives des variables utilisées dans ce modèle sont présentées au Tableau 4.2.

Le poids des individus est mesuré avec l'indice de masse corporel (IMC). L'IMC est une mesure qui permet d'estimer la corpulence d'une personne en utilisant sa taille et sa masse. On considère que les individus ont un poids normal si leur IMC se situe entre 18 et 25. Les individus avec un IMC se trouvant entre 25 et 30 sont considérés en surpoids et les individus avec un IMC supérieur à 30 sont considérés obèses. Selon les données utilisées, 30,5 % des hommes ont un poids normal, alors que ce nombre monte à 52 % chez les femmes. Même si les femmes sont plus nombreuses à avoir un poids normal, elles sont également plus nombreuses à souffrir d'embonpoint sévère.

L'usage du tabac est mesuré par le nombre d'années d'utilisation du tabac. Seul les fumeurs réguliers sont considérés fumeurs. Nous ne tenons pas compte des fumeurs occasionnels. Certains facteurs de risque important ont été écartés du modèle. On pense particulièrement à l'alcool et l'activité physique. Le problème avec ces deux variables c'est qu'elles sont endogènes dans le modèle. Par exemple, il est fort probable que les individus diminuent leur consommation d'alcool lorsqu'ils apprennent qu'ils ont une maladie chronique ou qu'ils diminuent leur niveau d'activité physique si la maladie chronique est handicapante. Ce problème n'est pas aussi important pour l'obésité et le nombre d'année de tabagisme car ces deux mesures sont des stocks et non des flux.

Tableau 4.2 Statistiques descriptives

variable	définition des variables	Homme	Femme
participation	1 si participe au marché du travail, 0 sinon	0,887	0,750
âge 30-39	1 si âgé de 30 à 39 ans, 0 sinon	0,288	0,295
âge 40-49	1 si âgé de 40 à 49 ans, 0 sinon	0,336	0,325
âge 50-54	1 si âgé de 50 à 54 ans, 0 sinon	0,151	0,150
âge 55-59	1 si âgé de 54 à 59 ans, 0 sinon	0,128	0,124
âge 60-64	1 si âgé de 60 à 64 ans, 0 sinon	0,097	0,106
Ontario	1 si réside en Ontario, 0 sinon	0,372	0,371
Atlantique	1 si réside dans une province de l'Atlantique, 0 sinon	0,077	0,084
Québec	1 si réside au Québec, 0 sinon	0,261	0,247
Manitoba et Saskatchewan	1 si réside au Manitoba ou en Saskatchewan, 0 sinon	0,067	0,068
Alberta et Colombie-Britannique	1 si réside en Alberta ou en Colombie-Britannique, 0 sinon	0,224	0,230
couple	1 si en couple, 0 sinon	0,780	0,734
enfant	1 si enfant(s) âgé(s) de 5 ans ou moins	0,210	0,188
immigrant	1 si né à l'extérieur du Canada, 0 sinon	0,183	0,187
secondaire	1 si aucun secondaire, 0 sinon	0,068	0,066
postsecondaire	1 si diplôme secondaire ou postsecondaire, 0 sinon	0,325	0,338
universitaire	1 si diplôme universitaire, 0 sinon	0,606	0,596
arthrite	1 si souffre d'arthrite, 0 sinon	0,132	0,198
dos	1 si a des maux de dos autres que l'arthrite, 0 sinon	0,175	0,175
poumon	1 si atteint d'une MPOC, 0 sinon	0,019	0,027
cœur	1 si souffre d'une maladie cardio-vasculaire, 0 sinon	0,048	0,028
diabète	1 si est atteint du diabète, 0 sinon	0,047	0,033
asthme	1 si fait de l'asthme, 0 sinon	0,066	0,082
IMC25-	1 si IMC est égal à 25 ou moins, 0 sinon	0,305	0,521
IMC25-30	1 si IMC est entre 26 et 30, 0 sinon	0,492	0,297
IMC30-35	1 si IMC est entre 30 et 35, 0 sinon	0,158	0,121
IMC35-40	1 si IMC est entre 35 et 40, 0 sinon	0,031	0,041
IMC40+	1 si IMC est supérieur à 40, 0 sinon	0,013	0,021
tabac0	1 si n'a jamais été un fumeur régulier 0 sinon	0,391	0,492
tabac10	1 si fumeur régulier pendant 10 ans ou moins, 0 sinon	0,093	0,109
tabac20	1 si a été un fumeur régulier 11 et 20 ans, 0 sinon	0,184	0,147

tabac30	1 si a été un fumeur régulier 21 et 30 ans, 0 sinon	0,180	0,139
tabac30+	1 si a été un fumeur régulier plus de 30 ans, 0 sinon	0,153	0,113
cycle1	1 si les données proviennent du cycle 1, 0 sinon	0,120	0,126
cycle2	1 si les données proviennent du cycle 2, 0 sinon	0,131	0,133
cycle3	1 si les données proviennent du cycle 3, 0 sinon	0,127	0,128
cycle4	1 si les données proviennent du cycle 4, 0 sinon	0,121	0,123
cycle5	1 si les données proviennent du cycle 5, 0 sinon	0,112	0,109
cycle6	1 si les données proviennent du cycle 6, 0 sinon	0,106	0,106
cycle7	1 si les données proviennent du cycle 7, 0 sinon	0,104	0,100
cycle8	1 si les données proviennent du cycle 8, 0 sinon	0,093	0,091
cycle9	1 si les données proviennent du cycle 9, 0 sinon	0,087	0,085
diabète-parent	1 si un des parents a eu le diabète, 0 sinon	0,567	0,618
diabète-nsp	1 si ne sait pas si un des parents a eu le diabète, 0 sinon	0,424	0,372
cœur-parent	1 si un des parents a eu une maladie cardiaque, 0 sinon	0,541	0,571
cœur-nsp	1 si ne sait pas si un des parents a eu une maladie cardiaque, 0 sinon	0,423	0,370

Tableau 4.3 Participation au marché du travail des individus selon qu'ils sont atteints ou non d'une maladie chronique

	Femme			Homme		
	avec la maladie	sans la maladie	réduction	avec la maladie	sans la maladie	réduction
arthrite	59,18 %	78,30 %	24,42 %	71,64 %	90,30 %	20,67 %
dos	66,44 %	75,71 %	12,24 %	80,56 %	89,12 %	9,60 %
poumon	55,34 %	74,59 %	25,82 %	59,91 %	88,12 %	32,02 %
diabète	47,14 %	74,96 %	37,12 %	69,32 %	88,52 %	21,69 %
cœur	41,97 %	75,08 %	44,10 %	61,48 %	89,05 %	30,96 %

Les proportions d'hommes et de femmes de 30 à 64 ans sur le marché du travail selon chacune des maladies chroniques utilisées dans le modèle sont présentées au Tableau 4.3. On remarque que les individus avec des problèmes cardiaques sont beaucoup moins nombreux à participer au marché du travail comparativement aux individus

avec des maux de dos. De plus, on remarque des différences significatives entre les hommes et les femmes au niveau du diabète. Les femmes semblent plus affectées par le diabète que les hommes. À la lumière des résultats du Tableau 4.3, il est probable que les maladies chroniques aient un impact négatif sur la participation au marché du travail. Néanmoins, il faut faire très attention dans l'analyse de ces résultats puisque plusieurs facteurs pourraient expliquer la corrélation entre les maladies chroniques et la participation au marché du travail. L'âge est certainement un de ces facteurs car les individus les plus âgés sont à la fois ceux qui ont le plus de maladies chroniques et ceux qui travaillent le moins. Également, ces différences pourraient s'expliquer par les taux de comorbidité. Si le diabète et les maladies cardiaques sont fortement corrélés, il est possible que la faible participation au marché du travail chez les diabétiques soit expliquée en partie par les maladies cardiaques. Finalement, il est possible que des facteurs non observables soient corrélés à la fois avec la participation au marché du travail et les maladies chroniques. Ces facteurs pourraient à la fois augmenter la probabilité de développer une maladie chronique et réduire la probabilité de participer au marché du travail. Dans ce cas, ce n'est pas la maladie chronique qui cause la faible participation au marché du travail, mais le facteur non observable.

Tableau 4.4 Comorbidités entre les maladies chroniques

Variables	Femme	Homme
arthrite-dos	0,19	0,16
arthrite-poumon	0,10	0,10
arthrite-cardiaque	0,13	0,10
arthrite-diabète	0,09	0,10
dos-poumon	0,07	0,06
dos-cardiaque	0,07	0,06
dos-diabète	0,02	0,01
poumon-cardiaque	0,08	0,06
poumon-diabète	0,05	0,03
cardiaque-diabète	0,10	0,14

Le Tableau 4.4 présente les taux de comorbidité entre les différentes maladies chroniques. On observe que les maladies chroniques sont toutes corrélées de façon positive. Trois raisons peuvent expliquer ces corrélations. Premièrement, certaines maladies chroniques sont elles-mêmes des facteurs de risque pour d'autres maladies chroniques. Selon la littérature médicale, le diabète est par exemple un facteur de risque de l'hypertension, des maladies cardiaques, des accidents cardio-vasculaires et de l'arthrite. Deuxièmement, il peut exister des facteurs de risque communs à deux ou plusieurs maladies chroniques. Par exemple, l'obésité est un facteur de risque aux maladies cardiaques, au diabète, à l'arthrite et aux maux de dos. En affectant plusieurs maladies chroniques, l'obésité est en partie responsable de la comorbidité observée. Finalement, de façon identique à l'obésité qui est observable, il est possible que des facteurs de risque non observés tels que la génétique ou les comportements passés ont un impact simultané sur plusieurs maladies chroniques.

4.3 Modèle probit endogène avec structure récursive

Pour éliminer l'endogénéité des maladies chroniques dans l'équation de participation au marché du travail, nous utilisons un probit multivarié endogène avec structure récursive. Pour voir comment ce modèle peut éliminer le problème des variables endogènes, supposons le modèle multivarié à m équations suivant :

$$\begin{aligned}
 w_{i1} &= f_1(x_i) + \varepsilon_{i1}, \\
 w_{i2} &= f_2(y_{i1}, x_i) + \varepsilon_{i2}, \\
 &\vdots \\
 w_{im} &= f_m(y_{i1}, y_{i2}, \dots, y_{i,m-1}, x_i) + \varepsilon_{im}
 \end{aligned} \tag{4.1}$$

$$y_{ij} = \begin{cases} 1 & \text{si } w_{ij} \geq 0 \\ 0 & \text{si } w_{ij} < 0 \end{cases}, \quad j = 1, \dots, m$$

où les variables dépendantes sont déterminées de façon récursive. La première équation de ce modèle est complètement déterminée par des facteurs exogènes. La deuxième équation est déterminée par des facteurs exogènes ainsi que par le résultat de la première équation. La troisième équation est déterminée par des facteurs exogènes ainsi que les résultats de la première et deuxième équation et ainsi de suite. En estimant un modèle simultané récursif, il est possible d'éliminer l'endogénéité des variables de santé dans l'équation de la participation au marché du travail. En effet, supposons le probit bivarié suivant⁸ :

⁸ Exemple tiré de Greene (2003) page 710.

$$\begin{aligned}w_{i1} &= x'_{i1}\beta_1 + \gamma y_{i2} + \varepsilon_{i1}, \\w_{i2} &= x'_{i2}\beta_2 + \varepsilon_{i2},\end{aligned}\tag{4.2}$$

$$y_{ij} = \begin{cases} 1 & \text{si } w_{ij} \geq 0 \\ 0 & \text{si } w_{ij} < 0 \end{cases}, \quad j = 1, 2$$

$$\varepsilon_i \sim N(0, R) \quad \text{et} \quad R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

où ρ est la corrélation entre les deux équations. La probabilité d'observer ($y_{i1} = 1, y_{i2} = 1$) est donnée par :

$$p(y_{i1} = 1, y_{i2} = 1) = p(y_{i1} = 1 | y_{i2} = 1) p(y_{i2} = 1)\tag{4.3}$$

Étant donné que les probabilités marginales et conditionnelles sont :

$$p(y_{i1} = 1 | y_{i2} = 1) = \frac{\Phi_2(x'_{i1}\beta_1 + \gamma y_{i2}, x'_{i2}\beta_2, \rho)}{\Phi(x'_{i2}\beta_2)} \quad \text{et} \quad p(y_{i2} = 1) = \Phi(x'_{i2}\beta_2),\tag{4.4}$$

où Φ_2 est la fonction de répartition normale multivariée d'ordre 2, alors la probabilité jointe :

$$p(y_{i1} = 1, y_{i2} = 1) = \Phi_2(x'_{i1}\beta_1 + \gamma y_{i2}, x'_{i2}\beta_2, \rho)\tag{4.5}$$

ainsi que les autres probabilités :

$$\begin{aligned}p(y_{i1} = 1, y_{i2} = 0) &= \Phi_2(x'_{i1}\beta_1, -x'_{i2}\beta_2, -\rho) \\p(y_{i1} = 0, y_{i2} = 1) &= \Phi_2(-(x'_{i1}\beta_1 + \gamma y_{i2}), x'_{i2}\beta_2, -\rho) \\p(y_{i1} = 0, y_{i2} = 0) &= \Phi_2(-x'_{i1}\beta_1, -x'_{i2}\beta_2, \rho)\end{aligned}\tag{4.6}$$

sont identiques à celles du probit bivarié dont l'ensemble des variables sont exogènes.

Définissons le niveau de santé h des composantes humaines associées aux cinq maladies chroniques de l'individu i à la période t par :

$$\begin{aligned} h_{it1} &= \alpha_{i1} + x'_{it1}\beta_1 + \varepsilon_{it1} \\ &\vdots \\ h_{it5} &= \alpha_{i5} + x'_{it5}\beta_5 + \varepsilon_{it5} \end{aligned} \quad (4.7)$$

où x_{itj} est le vecteur des facteurs de risque spécifiques à la maladie chronique j , β_j est le vecteur des coefficients de régression des facteurs de risque, α_{ij} est l'effet individuel et ε_{itj} est le terme d'erreur. On suppose qu'une maladie chronique est diagnostiquée ($H_{itj} = 1$) si le niveau de santé de la composante humaine associé à cette maladie est négatif :

$$H_{itj} = \begin{cases} 1 & \text{si } h_{itj} \leq 0 \\ 0 & \text{si } h_{itj} > 0 \end{cases} \quad (4.8)$$

Définissons également la propension à travailler de l'individu i à la période t par :

$$w_{it} = \alpha_{i6} + x'_{it6}\beta_6 + \gamma_1 H_{it1} + \gamma_2 H_{it2} + \gamma_3 H_{it3} + \gamma_4 H_{it4} + \gamma_5 H_{it5} + \varepsilon_{it6},$$

où w_{it} est la propension à travailler, x_{it6} est le vecteur des variables explicatives exogène, β_6 est le vecteur des coefficients de régression exogènes, γ_j le vecteur des coefficients de régression endogènes, α_{i6} est l'effet individuel de la propension à travailler et ε_{it6} est le terme d'erreur. La décision de participer au marché du travail est définie par la relation suivante :

$$y_{it} = \begin{cases} 1 & \text{si } w_{it} \leq 0 \\ 0 & \text{si } w_{it} > 0 \end{cases} \quad (4.9)$$

On suppose que les termes d'erreur sont distribués de façon indépendante et identique selon une distribution normale multivariée avec moyennes zéro et la matrice de corrélation suivante :

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{16} \\ r_{12} & 1 & \cdots & r_{26} \\ \vdots & \vdots & \ddots & \vdots \\ r_{16} & r_{26} & \cdots & 1 \end{bmatrix}$$

Finalement, on suppose que les effets individuels sont distribués de façon indépendante et identique selon une distribution normale multivariée avec moyenne $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_6]$ et la matrice de variances-covariances suivante :

$$\Sigma_\alpha = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{16} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{26} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{16} & \sigma_{26} & \cdots & \sigma_6^2 \end{bmatrix}$$

4.4 Identification du modèle

L'identification du modèle probit multivarié est toujours sujette à débat. Maddala (1983) affirme qu'une variable d'exclusion est nécessaire pour l'identification de ce modèle. Ajouter une variable d'exclusion au modèle permet que le nombre de probabilités indépendantes soit plus grand que le nombre de paramètres à estimer. Cependant, Wilde (2000) a démontré que le modèle probit multivarié avec structure endo-

gène était identifié si chaque équation contenait au moins un régresseur exogène variable :

« The existence of one varying exogenous regressor in each regression is sufficient to avoid small variation identification problems in multiple equation probit models with endogenous dummy regressors».

Selon Méango et Mourifié (2013), le critère proposé par Wilde n'est pas suffisant pour s'assurer de l'identification du modèle. La raison est que, même si le nombre de probabilités indépendantes est égal ou plus grand que le nombre de paramètres à identifier, cela n'est pas suffisant pour s'assurer de l'unicité de la solution étant donné que le système est non linéaire.

La façon habituelle d'identifier une équation structurelle c'est d'utiliser de l'information *a priori*. Pour notre modèle, nous supposons que les antécédents familiaux (maladies cardiovasculaires et diabète) ainsi que le nombre d'années d'utilisation du tabac n'a pas d'impact sur la probabilité de participer au marché du travail. Dans le cas des antécédents familiaux, il est raisonnable d'affirmer que les antécédents de santé des parents n'ont pas d'impact sur la probabilité de participer au marché du travail de leur enfant. Le cas du tabagisme est par contre moins évident. Contrairement à l'obésité qui peut avoir un effet sur la participation au marché du travail à cause d'une baisse de la productivité ou par une certaine forme de discrimination, il n'y a pas lieu de croire que le nombre d'années de tabagisme a un effet direct sur la participation au marché du travail. Des études ont montré que les fumeurs ont une productivité plus faible sur le marché du travail que les non-fumeurs. Cette productivité plus faible est causée principalement par les pauses plus longues prises par les fumeurs (Chaloupka et Warner, 1999). Par contre, ces études montrent également que la productivité des anciens fumeurs est la même que celle des non-

fumeurs. Il n'y a donc pas de lien de causalité entre le nombre d'années de tabagisme et la productivité. Si le nombre d'années de tabagisme est corrélé négativement avec la participation au marché du travail c'est à cause de facteurs non observés corrélés à la fois avec le nombre d'année de tabagisme et la participation au marché du travail ou par des effets indirects via les maladies chroniques.

4.5 Les résultats

Cette section présente les résultats de nos estimations. Pour nos deux groupes d'hommes et de femmes, nous avons estimé un modèle probit multivarié avec effets individuels hiérarchiques (effets aléatoires). Les hyperparamètres choisis sont les hyperparamètres non informatifs suivants présentés au chapitre 3 :

$$\begin{aligned}
 \underline{\beta} &= 0, \\
 \underline{V}_{\beta}^{-1} &= 10^{-5}I, \\
 \underline{\alpha} &= 0, \\
 \underline{V}_{\alpha}^{-1} &= 10^{-5}I, \\
 \underline{\Sigma}_{\alpha}^{-1} &= I, \\
 \underline{v}_{\alpha} &= 0, \\
 \underline{\Sigma}_{\varepsilon}^{-1} &= I, \\
 \underline{v}_{\varepsilon} &= m + 1 = 7.
 \end{aligned}$$

Pour comparer nos résultats, nous avons également estimé un modèle probit multivarié sans effet individuel hiérarchique. Les résultats de ce dernier modèle sont présentés en annexe. Le Tableau 4.5 présente le nombre d'observations et le nombre de paramètres des deux modèles estimés ainsi que le nombre de tirages et de *burn-in* choi-

sis. Pour chacun des paramètres estimés, nous avons calculé le MCSE (*Monte Carlo standard error*) avec la méthode des « *batch mean* » (Flegal et Jones, 2010). Les MCSE obtenues sont très faibles et démontrent que le nombre de tirages est suffisant pour obtenir une précision numérique élevée de l'estimateur.

Tableau 4.5 Informations sur les modèles estimés

	Modèle sans effet individuel hiérarchiques		Modèle avec effets individuels hiérarchiques	
	homme	femme	homme	femme
Nombre d'observations	25 626	27 976	25 626	27 976
Nombre de paramètres	144	144	123	123
Nombres de tirages	100 000	100 000	100 000	100 000
Burn-in	20 000	20 000	20 000	20 000

La prochaine section présente de façon détaillée les résultats des équations de santé. Pour chacune des maladies chroniques, nous présentons les paramètres de la distribution *a posteriori* (moyenne et écart-type) et le HPDI (*highest posterior density interval*). Le HPDI est utilisé ici de façon *ad hoc* pour la comparaison de modèle. Par exemple, si le HDPI n'inclut pas zéro, c'est un signe que le régresseur a un impact sur la variable dépendante. Ce n'est pas la façon la plus formelle de comparer des modèles en économétrie bayésienne, mais cette approche a l'avantage de la simplicité. Cette approche est similaire à l'approche fréquentiste des tests d'hypothèse. Par contre, il faut faire attention : les intervalles de confiance fréquentistes n'ont pas la même interprétation que les HDPI bayésiens (Koop, 2003). Aux HDPI, nous avons ajouté une colonne pour indiquer si zéro est non inclus dans le HDPI à 95 % (**) ou à 90 % (*). La colonne MSCE présente les erreurs-type Monte Carlo associés à chacun

des estimés et la dernière colonne intitulée ME présente les effets marginaux calculés à partir des distributions *a posteriori* prédictives.

4.5.1 L'arthrite

L'arthrite est une inflammation aiguë ou chronique des articulations dont les symptômes peuvent se manifester progressivement ou soudainement. Le terme arthrite est utilisé pour décrire plus de 100 troubles médicaux qui touchent à la fois les articulations et les tissus qui les entourent. Habituellement, la plupart des formes d'arthrite sont caractérisées par la douleur et la raideur dans une ou plusieurs articulations et autour de celles-ci. L'évolution, la gravité et la localisation des symptômes varient selon la forme particulière de la maladie. L'arthrite touche environ 4,2 millions de Canadiens de tous âges⁹.

Les principaux facteurs de risque non modifiables sont l'âge, le sexe et la génétique. Les femmes âgées sont les personnes les plus à risque de développer de l'arthrite. Les facteurs modifiables sont l'obésité, le tabagisme et l'emploi où certains gestes sont répétés. Une bonne alimentation et de l'exercice physique peuvent réduire la probabilité de développer cette maladie.

Les paramètres de la distribution *a posteriori* associés à cette maladie sont détaillés au Tableau 4.6a. Première constatation, l'âge, le poids et le tabagisme sont des facteurs de risque importants pour l'arthrite. En moyenne, les femmes âgées de 60 à 65 ans ont 35 points de pourcentage de chance de plus de développer des problèmes

⁹Agence de la santé publique du Canada, <http://www.phac-aspc.gc.ca/cd-mc/musculo/arthrite-arthritis-fra.php>

d'arthrite que les femmes âgées de 30 à 40 ans. Chez les hommes, cette différence est de 25 points. Les femmes avec un IMC supérieur à 25, les hommes avec un IMC supérieur à 30 et les grands fumeurs ont plus de chance d'avoir de l'arthrite. L'éducation ne semble pas jouer un rôle sur l'apparition de cette maladie.

Tableau 4.6a Impact des facteurs de risque sur l'arthrite

Arthrite	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	-1,839	0,289	(-2,39; -1,29)	**	0,0199	-
age40-50	0,872	0,055	(0,76; 0,98)	**	0,0008	8,82
age50-55	1,760	0,069	(1,63; 1,89)	**	0,0011	16,26
age55-60	2,341	0,076	(2,19; 2,49)	**	0,0011	24,57
age60-65	2,736	0,082	(2,57; 2,90)	**	0,0013	35,45
secondaire	-0,061	0,081	(-0,22; 0,10)		0,0014	-0,68
universitaire	-0,051	0,087	(-0,22; 0,12)		0,0015	-0,55
IMC25-30	0,303	0,047	(0,21; 0,40)	**	0,0008	3,27
IMC30-35	0,435	0,068	(0,30; 0,57)	**	0,0011	4,14
IMC35-40	0,727	0,097	(0,54; 0,91)	**	0,0016	8,06
IMC40+	1,065	0,127	(0,82; 1,31)	**	0,0021	13,04
tabac10-	0,217	0,095	(0,03; 0,40)	**	0,0020	2,28
tabac20-	0,305	0,088	(0,13; 0,48)	**	0,0020	2,85
tabac30-	0,459	0,084	(0,29; 0,62)	**	0,0020	4,38
tabac30+	0,689	0,089	(0,52; 0,87)	**	0,0020	8,04

Arthrite	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	-2,172	0,617	(-3,36; 0,00)	**	0,0370	-
age40-50	0,681	0,134	(0,02; 0,82)	**	0,0069	5,73
age50-55	1,430	0,221	(0,28; 1,61)	**	0,0119	11,26
age55-60	1,851	0,264	(0,48; 2,06)	**	0,0143	16,36

age60-65	2,172	0,294	(0,65; 2,40)	**	0,0159	25,06
secondaire	-0,033	0,186	(-0,92; 0,16)		0,0095	-1,15
universitaire	-0,056	0,201	(-1,01; 0,15)		0,0103	-1,38
IMC25-30	0,051	0,057	(-0,06; 0,16)		0,0012	0,46
IMC30-35	0,312	0,093	(0,01; 0,47)	**	0,0034	2,41
IMC35-40	0,636	0,160	(0,02; 0,89)	**	0,0064	6,16
IMC40+	0,878	0,235	(0,02; 1,26)	**	0,0090	9,31
tabac10-	0,354	0,123	(0,00; 0,57)	**	0,0042	3,25
tabac20-	0,262	0,101	(0,00; 0,45)	**	0,0035	1,91
tabac30-	0,410	0,110	(0,01; 0,59)	**	0,0045	2,96
tabac30+	0,514	0,129	(0,00; 0,71)	**	0,0056	4,83

4.5.2 Les maux de dos

Les maux de dos ont pour origine des problèmes de vertèbres, d'articulations ou de disques qui peuvent occasionner des spasmes musculaires, des hernies discales et des sciatiques par exemple. Les maux de dos sont rarement la conséquence d'une affection grave et disparaissent en général spontanément, par contre ils reviennent souvent et, pour cette raison, ils sont considérés comme des maladies chroniques. Les causes des maux de dos sont le vieillissement, l'usure des articulations du dos, les mauvaises postures, l'obésité, le tabagisme et le type d'emploi.

Les paramètres de la distribution *a posteriori* associés à cette maladie sont détaillés au Tableau 4.6b. Contrairement à l'arthrite, l'âge a peu d'impact sur l'apparition de maux de dos et il y a peu de différence entre les hommes et les femmes. Le poids continue d'être un facteur de risque important ainsi que le tabagisme. L'effet du tabac sur les maux de dos peut sembler étrange, mais le tabac est un facteur de risque important des maux de dos (discopathie dégénérative).

Tableau 4.6b Impact des facteurs de risque sur les maux de dos

Dos	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	-0,832	0,236	(-1,31;-0,32)	**	0,0137	-
age40-50	0,169	0,036	(0,10;0,24)	**	0,0005	2,60
age50-55	0,258	0,049	(0,16;0,35)	**	0,0006	3,42
age55-60	0,389	0,053	(0,29;0,49)	**	0,0007	5,47
age60-65	0,306	0,059	(0,19;0,42)	**	0,0008	4,92
secondaire	0,060	0,066	(-0,07;0,19)		0,0010	0,87
universitaire	0,109	0,068	(-0,02;0,24)	*	0,0011	1,61
IMC25-30	0,031	0,036	(-0,04;0,10)		0,0005	0,49
IMC30-35	0,188	0,051	(0,09;0,29)	**	0,0007	2,63
IMC35-40	0,390	0,073	(0,25;0,53)	**	0,0010	6,33
IMC40+	0,390	0,100	(0,19;0,59)	**	0,0013	6,61
tabac10-	0,178	0,065	(0,05;0,31)	**	0,0011	2,82
tabac20-	0,175	0,060	(0,05;0,29)	**	0,0011	2,31
tabac30-	0,185	0,060	(0,07;0,30)	**	0,0012	2,46
tabac30+	0,389	0,066	(0,26;0,52)	**	0,0013	6,44

Dos	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	-0,918	0,328	(-1,55;-0,00)	**	0,0189	-
age40-50	0,171	0,050	(0,02;0,25)	**	0,0020	2,57
age50-55	0,244	0,053	(0,14;0,34)	**	0,0010	3,22
age55-60	0,339	0,059	(0,23;0,45)	**	0,0012	4,72
age60-65	0,361	0,072	(0,23;0,51)	**	0,0020	5,98
secondaire	0,030	0,188	(-0,92;0,19)		0,0099	-0,30
universitaire	0,039	0,207	(-1,01;0,21)		0,0110	-0,19
IMC25-30	-0,021	0,041	(-0,10;0,06)		0,0007	-0,36
IMC30-35	0,137	0,060	(0,00;0,25)	**	0,0016	1,80

IMC35-40	0,094	0,094	(-0,09;0,28)		0,0017	1,42
IMC40+	0,401	0,153	(0,01;0,69)	**	0,0043	6,78
tabac10-	0,319	0,090	(0,01;0,47)	**	0,0033	5,05
tabac20-	0,227	0,071	(0,01;0,35)	**	0,0024	2,79
tabac30-	0,274	0,076	(0,00;0,40)	**	0,0029	3,40
tabac30+	0,409	0,096	(0,00;0,55)	**	0,0042	6,61

4.5.3 Les maladies pulmonaires obstructives chroniques

Les maladies pulmonaires obstructives chroniques (MPOC) sont des maladies chroniques caractérisées par l'essoufflement, la toux chronique et la production accrue de crachat. On considère la bronchite chronique et l'emphysème comme des maladies pulmonaires obstructives chroniques. Les MPOC évoluent lentement au cours des années et l'aggravation de ces maladies se conjugue avec des épisodes plus fréquents, un débit aérien de plus en plus limité et le décès prématuré. La progression de ces maladies entraîne chez les personnes atteintes un essoufflement qui limite l'activité et réduit la qualité de vie.

La cigarette et la fumée secondaire sont les causes principales des MPOC. L'environnement joue également un rôle important. L'exposition aux poussières (charbons, céréalières) et la pollution atmosphérique augmentent le risque de développer cette maladie.

Les paramètres de la distribution *a posteriori* associés à cette maladie sont détaillés au Tableau 4.6c. L'âge a peu d'impact sur les maladies pulmonaires, contrairement à l'éducation qui semble jouer un rôle un peu plus important. Dans ce cas-ci, l'éducation diminue la probabilité de développer une maladie pulmonaire. Il est pos-

sible que l'environnement de travail (poussière) explique en partie ce résultat car les individus moins scolarisés se retrouvent en plus grand nombre dans ce type d'emploi. À l'exception de l'asthme, le tabagisme est le facteur de risque le plus important expliquant les maladies pulmonaires. Par contre, les effets marginaux du tabac sont relativement faibles par rapport aux autres maladies chroniques telles que les maux de dos et l'arthrite. L'effet marginal d'avoir fumé plus de 30 ans est de moins de 3 % peu importe le sexe. Ce résultat ne semble pas intuitif. Par contre, si les effets du tabagisme sur les maladies pulmonaires se font sentir à plus long terme, l'exclusion des individus âgés de 65 ans et plus pourrait expliquer ce résultat.

Tableau 4.6c Impact des facteurs de risque sur les maladies pulmonaires

Poumon	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	-0,881	0,714	(-2,24;0,14)	*	0,0382	-
age40-50	-0,132	0,079	(-0,29;0,02)	**	0,0020	-0,37
age50-55	-0,032	0,101	(-0,22;0,17)		0,0026	-0,08
age55-60	0,093	0,106	(-0,12;0,30)		0,0028	0,25
age60-65	0,135	0,112	(-0,09;0,36)		0,0029	0,43
secondaire	-0,338	0,106	(-0,54;-0,13)	**	0,0021	-1,18
universitaire	-0,375	0,113	(-0,60;-0,15)	**	0,0025	-1,28
tabac10-	0,165	0,135	(-0,11;0,43)		0,0042	0,47
tabac20-	0,142	0,117	(-0,09;0,37)		0,0035	0,34
tabac30-	0,520	0,106	(0,31;0,73)	**	0,0032	1,37
tabac30+	0,786	0,116	(0,57;1,02)	**	0,0033	3,07
asthme	1,061	0,085	(0,89;1,23)	**	0,0016	4,79

Poumon	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	-0,521	0,429	(-1,57;0,11)	*	0,0252	-
age40-50	-0,158	0,104	(-0,36;0,04)	*	0,0036	-0,34
age50-55	-0,124	0,147	(-0,40;0,19)		0,0056	-0,08
age55-60	0,099	0,142	(-0,18;0,38)		0,0050	0,46
age60-65	0,295	0,149	(0,01;0,59)	**	0,0052	1,23
secondaire	-0,094	0,204	(-0,92;0,19)		0,0094	-1,23
universitaire	-0,009	0,231	(-1,01;0,29)		0,0112	-1,12
tabac10-	0,351	0,187	(-0,00;0,70)	**	0,0079	0,80
tabac20-	0,558	0,162	(0,00;0,83)	**	0,0076	1,00
tabac30-	0,683	0,179	(0,00;0,97)	**	0,0087	1,29
tabac30+	1,047	0,232	(0,00;1,37)	**	0,0119	3,37
asthme	1,075	0,215	(0,01;1,32)	**	0,0107	3,91

4.5.4 Les maladies cardio-vasculaires

On appelle les maladies cardiovasculaires toutes les maladies de l'appareil circulatoire que constituent le cœur et les vaisseaux sanguins. Les deux principales maladies cardiaques sont les cardiopathies ischémiques et les maladies cérébrovasculaires. Les cardiopathies coronariennes (ou cardiopathies ischémiques) correspondent à un groupe de pathologies cardiaques engendrées par des problèmes d'irrigation sanguine du muscle cardiaque. L'insuffisance en sang oxygéné provoque les symptômes d'angine de poitrine et le blocage complet d'une artère du cœur provoque ce qu'on appelle un infarctus (crise cardiaque). Ce dernier est la forme de maladie cardiaque la plus répandue au Canada. Les maladies cérébrovasculaires sont un problème de circulation sanguine dans le cerveau. Un blocage complet à long terme d'une artère dans le cerveau provoque un accident vasculaire cérébral (AVC). Les autres maladies cardio-

vasculaires affectent principalement les jambes ou sont des problèmes liés à une infection ou anomalie congénitale du cœur.

La liste des facteurs qui contribuent aux maladies cardiovasculaire est impressionnante. Elle comprend le tabagisme, le manque d'exercice, l'embonpoint, le sexe et l'âge de la personne, l'hypertension, les mauvaises habitudes alimentaires, le sodium et le stress. La génétique joue aussi un rôle important.

Les paramètres de la distribution *a posteriori* associés à cette maladie sont détaillés au Tableau 4.6d. L'âge est un facteur de risque important des maladies cardiovasculaires surtout chez les hommes. La probabilité de développer une maladie cardiovasculaire est de 15 points de pourcentage plus élevée chez les hommes âgés de 60 à 65 ans que chez ceux âgés de 30 à 40 ans. Cette même probabilité augmente que de 7 % chez les femmes. L'impact de l'obésité et du tabac sur les maladies cardiovasculaires sont semblables entre les hommes et les femmes et cet impact est plutôt faible. Comme pour les maladies pulmonaires, les impacts de l'obésité et du tabagisme pourraient survenir après l'âge de 65 ans. Finalement, l'éducation ne semble pas être un facteur important.

Tableau 4.6d Impact des facteurs de risque sur les maladies cardio-vasculaires

Cœur	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	-0,862	0,349	(-1,58;-0,23)	**	0,0197	-
age40-50	0,381	0,116	(0,15;0,61)	**	0,0041	0,90
age50-55	0,870	0,133	(0,61;1,14)	**	0,0045	1,98
age55-60	1,300	0,136	(1,03;1,56)	**	0,0046	3,41

age60-65	1,718	0,144	(1,44;1,99)	**	0,0048	7,07
secondaire	-0,114	0,116	(-0,34;0,11)		0,0028	-0,34
universitaire	-0,083	0,128	(-0,33;0,17)		0,0033	-0,25
IMC25-30	0,062	0,088	(-0,11;0,24)		0,0028	0,17
IMC30-35	0,177	0,121	(-0,06;0,42)	*	0,0038	0,42
IMC35-40	0,483	0,164	(0,16;0,80)	**	0,0049	1,44
IMC40+	0,627	0,204	(0,23;1,02)	**	0,0058	2,15
tabac10-	0,194	0,167	(-0,14;0,52)		0,0067	0,53
tabac20-	0,305	0,150	(0,01;0,60)	**	0,0058	0,73
tabac30-	0,521	0,137	(0,25;0,78)	**	0,0053	1,26
tabac30+	0,807	0,136	(0,54;1,08)	**	0,0048	2,67
cœur-parent	-0,764	0,155	(-1,08;-0,46)	**	0,0057	-3,31
cœur-nsp	-0,384	0,216	(-0,81;0,04)	**	0,0079	-0,89

Hommes

Cœur	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	-2,070	0,619	(-3,20;-0,00)	**	0,0377	-
age40-50	0,886	0,203	(0,02;1,17)	**	0,0101	2,67
age50-55	1,681	0,299	(0,28;2,03)	**	0,0154	4,98
age55-60	2,121	0,346	(0,47;2,50)	**	0,0179	7,25
age60-65	2,738	0,422	(0,64;3,14)	**	0,0223	15,44
secondaire	-0,111	0,198	(-0,92;0,16)		0,0093	-1,37
universitaire	-0,067	0,223	(-1,01;0,22)		0,0109	-1,29
IMC25-30	0,122	0,092	(-0,05;0,30)		0,0030	0,41
IMC30-35	0,167	0,127	(-0,07;0,42)		0,0043	0,46
IMC35-40	0,324	0,195	(-0,04;0,70)	*	0,0063	1,13
IMC40+	0,519	0,284	(-0,01;1,08)	**	0,0090	2,08
tabac10-	-0,072	0,195	(-0,47;0,30)		0,0080	-0,17
tabac20-	0,406	0,168	(0,00;0,73)	**	0,0074	1,15

tabac30-	0,483	0,163	(0,00;0,78)	**	0,0073	1,33
tabac30+	0,705	0,190	(0,00;1,01)	**	0,0090	2,72
parent-cœur	-1,069	0,279	(-1,52;-0,01)	**	0,0132	-7,13
nsp-parent	-1,036	0,318	(-1,61;-0,02)	**	0,0144	-2,75

4.5.5 Le diabète

Le diabète est une maladie chronique qui se manifeste lorsque le corps est incapable de produire suffisamment d'insuline ou de l'utiliser correctement. Le corps a besoin d'insuline pour transformer le sucre en énergie. Le diabète peut entraîner de graves complications et le décès prématuré. Ceux qui sont atteints du diabète peuvent toutefois prendre des mesures pour contrôler et réduire les risques de complications.

Sans connaître les causes exactes du diabète de type 1, les scientifiques pensent que les facteurs génétiques et l'exposition à des virus en font partie. Les facteurs de risque qui contribuent au diabète de type 2 comprennent l'obésité, l'âge, l'inactivité physique, l'hypertension, la génétique et l'appartenance à certains groupes ethniques.

Les paramètres de la distribution *a posteriori* associés à cette maladie sont détaillés au Tableau 4.6e. L'âge et la génétique sont les facteurs de risque les plus importants du diabète. Le poids est également un facteur de risque, mais beaucoup moins important. Si on compare ces résultats avec ceux du modèle probit multivarié en annexe, on remarque que les effets marginaux sont beaucoup plus faibles. Il est possible que l'effet du poids soit capté par l'effet individuel qui a une moyenne relativement élevée dans cette équation.

Tableau 4.6e Impact des facteurs de risque sur le diabète

Diabète	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	-3,998	0,789	(-6,00;-2,79)	**	0,0651	-
age40-50	0,998	0,153	(0,69;1,30)	**	0,0062	1,54
age50-55	2,183	0,205	(1,79;2,60)	**	0,0086	3,29
age55-60	3,038	0,225	(2,60;3,47)	**	0,0094	5,39
age60-65	4,026	0,252	(3,54;4,52)	**	0,0108	11,71
secondaire	-0,396	0,182	(-0,77;-0,05)	**	0,0064	-0,74
universitaire	-0,352	0,200	(-0,74;0,04)	**	0,0073	-0,65
IMC25-30	0,589	0,143	(0,31;0,87)	**	0,0061	1,06
IMC30-35	0,919	0,169	(0,59;1,25)	**	0,0069	1,45
IMC35-40	1,484	0,211	(1,07;1,90)	**	0,0079	2,99
IMC40+	1,785	0,257	(1,28;2,29)	**	0,0092	4,11
diabète-parent	-3,492	0,550	(-4,62;-2,46)	**	0,0271	-14,83
nsp-parent	-2,930	0,609	(-4,21;-1,80)	**	0,0299	-3,45

Diabète	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	-4,311	1,266	(-6,37;-0,00)	**	0,0683	-
age40-50	1,825	0,382	(0,02;2,27)	**	0,0208	2,80
age50-55	3,549	0,640	(0,27;4,15)	**	0,0358	5,48
age55-60	4,915	0,850	(0,47;5,63)	**	0,0477	9,22
age60-65	6,062	1,032	(0,63;6,89)	**	0,0579	18,28
secondaire	-0,155	0,247	(-0,92;0,26)		0,0113	-1,30
universitaire	0,132	0,302	(-1,01;0,59)		0,0149	-0,88
IMC25-30	0,145	0,163	(-0,16;0,49)		0,0077	0,24
IMC30-35	0,485	0,221	(0,00;0,91)	**	0,0104	0,70
IMC35-40	0,865	0,295	(0,02;1,38)	**	0,0134	1,60

IMC40+	0,977	0,376	(0,02;1,65)	**	0,0155	1,97
diabète-parent	-5,241	1,235	(-7,11;-0,03)	**	0,0658	-21,17
nsp-parent	-4,367	1,116	(-6,23;-0,03)	**	0,0592	-5,16

4.5.6 Les coefficients de corrélation

Le Tableau 4.7 présente les paramètres de la distribution *a posteriori* des coefficients de corrélation des termes résiduels. Comparativement au modèle sans effet individuel, les corrélations sont plutôt faibles dans ce modèle. L'effet individuel semble capter une grande partie des facteurs non observables. Pour pouvoir observer une corrélation du terme résiduel, il faut un choc exogène qui doit affecter simultanément deux ou plusieurs maladies. L'information pourrait être un facteur important pouvant agir comme un choc exogène : consulter un médecin pour un problème cardiaque augmente nécessairement la probabilité d'apprendre du même coup que nous sommes atteints de diabète.

Tableau 4.7 Corrélation des facteurs de risque exogènes non observés

Corrélation	Femmes				
	Moyenne	Écart-type	(HPDI 95%)	MCSE	
cœur-diabète	0,303	0,095	(0,11;0,49)	**	0,0049
cœur-arthrite	-0,055	0,058	(-0,17;0,06)		0,0023
cœur-dos	0,133	0,047	(0,04;0,23)	**	0,0017
cœur-poumon	0,085	0,088	(-0,08;0,27)		0,0042
diabète-arthrite	0,019	0,076	(-0,13;0,17)		0,0033
diabète-dos	-0,001	0,068	(-0,13;0,13)		0,0030
diabète-poumon	0,074	0,102	(-0,14;0,26)		0,0053

arthrite-dos	0,128	0,025	(0,08;0,18)	**	0,0005
arthrite-poumon	0,022	0,051	(-0,08;0,12)		0,0018
dos-poumon	0,073	0,044	(-0,01;0,16)	*	0,0014

Hommes

Corrélation	Moyenne	Écart-type	(HPDI 95%)		MCSE
cœur-diabète	0,525	0,128	(0,31;0,99)	**	0,0069
cœur-arthrite	0,102	0,165	(-0,03;0,99)	*	0,0089
cœur-dos	0,138	0,155	(0,01;0,99)	**	0,0083
cœur-poumon	0,049	0,183	(-0,15;0,99)		0,0098
diabète-arthrite	0,293	0,145	(0,11;0,99)	**	0,0076
diabète-dos	0,113	0,167	(-0,05;0,99)		0,0090
diabète-poumon	0,240	0,180	(-0,06;0,99)	*	0,0097
arthrite-dos	0,142	0,149	(0,06;0,99)	**	0,0081
arthrite-poumon	0,198	0,152	(0,04;0,99)	**	0,0081
dos-poumon	0,097	0,164	(-0,04;0,99)		0,0088

4.5.7 Variances et covariances des effets individuels hiérarchiques

Le Tableau 4.8a présente les paramètres de la distribution *a posteriori* des variances des effets individuels hiérarchiques. On remarque d'abord que les variances des effets individuels sont relativement élevées comparativement aux variances du terme résiduel qui sont contraintes à un. Les hétérogénéités générées par les effets aléatoires sont donc plus grandes que celles produites par les termes résiduels. Ceci démontre que les chocs exogènes ont peu d'impact sur les maladies chroniques comparative-ment aux facteurs de risque non observables spécifiques aux individus. Ce résultat est intuitif car les maladies chroniques sont des maladies de longue durée d'évolution

généralement lente. Les facteurs de risque doivent être présents pendant un certain temps avant d'avoir un impact observable sur la santé.

Tableau 4.8a Variances des effets individuels hiérarchiques

Variance	Femmes			
	Moyenne	Écart-type	(HPDI 95%)	MCSE
cœur	4,104	0,403	(3,37;5,01)	** 0,0164
diabète	15,446	1,886	(12,27;19,53)	** 0,0888
arthrite	4,254	0,210	(3,85;4,68)	** 0,0042
dos	1,882	0,086	(1,72;2,06)	** 0,0016
poumon	2,319	0,222	(1,93;2,79)	** 0,0081

Variance	Hommes			
	Moyenne	Écart-type	(HPDI 95%)	MCSE
cœur	4,867	0,951	(0,03;5,93)	** 0,0510
diabète	23,062	5,251	(0,03;31,06)	** 0,2892
arthrite	3,937	0,716	(0,03;4,54)	** 0,0386
dos	1,821	0,323	(0,03;2,06)	** 0,0175
poumon	2,296	0,471	(0,03;2,90)	** 0,0250

Le Tableau 4.8b présente les paramètres de la distribution *a posteriori* des covariances des effets individuels hiérarchiques. À l'exception des maux de dos et du diabète chez les deux sexes, et de l'arthrite et du diabète chez les hommes, la covariance entre les maladies chroniques sont toutes positives. Ces covariances montrent qu'il existe des facteurs de risque spécifiques aux individus qui ont un impact sur plusieurs maladies chroniques à la fois. Ce résultat est intuitif pour les mêmes raisons que nous avons données pour expliquer la plus grande hétérogénéité de l'effet individuel com-

parativement au terme résiduel. Les facteurs de risque qui affectent plusieurs maladies chroniques doivent être présents pendant une certaines périodes avant d'avoir un effet sur la santé.

Tableau 4.8b Covariances des effets individuels hiérarchiques

Covariance	Femmes			
	Moyenne	Écart-type	(HPDI 95%)	MCSE
cœur-diabète	2,294	0,430	(1,49;3,19)	** 0,0195
cœur-arthrite	1,096	0,151	(0,82;1,41)	** 0,0053
cœur-dos	0,594	0,103	(0,39;0,79)	** 0,0037
cœur-poumon	0,655	0,149	(0,37;0,95)	** 0,0059
diabète-arthrite	0,600	0,298	(0,01;1,19)	** 0,0134
diabète-dos	0,070	0,205	(-0,34;0,47)	0,0092
diabète-poumon	0,608	0,306	(0,02;1,23)	** 0,0145
arthrite-dos	1,186	0,076	(1,04;1,34)	** 0,0015
arthrite-poumon	0,739	0,113	(0,52;0,97)	** 0,0037
dos-poumon	0,373	0,079	(0,23;0,54)	** 0,0024

Covariance	Hommes			
	Moyenne	Écart-type	(HPDI 95%)	MCSE
cœur-diabète	3,106	0,932	(0,00;4,68)	** 0,0445
cœur-arthrite	0,623	0,188	(0,00;0,94)	** 0,0082
cœur-dos	0,340	0,118	(0,00;0,55)	** 0,0049
cœur-poumon	0,614	0,212	(0,00;1,00)	** 0,0100
diabète-arthrite	0,999	0,376	(0,00;1,71)	** 0,0179
diabète-dos	0,024	0,231	(-0,45;0,48)	0,0104
diabète-poumon	0,397	0,467	(-0,53;1,35)	0,0229
arthrite-dos	1,046	0,197	(0,00;1,25)	** 0,0103
arthrite-poumon	0,680	0,181	(0,00;0,98)	** 0,0085
dos-poumon	0,366	0,107	(0,00;0,55)	** 0,0047

À titre indicatif et pour mieux comparer les impacts des termes résiduels sur la comorbidité des maladies chroniques à ceux des effets individuels, nous avons produit le Tableau 4.8c. À partir des moyennes normalisées de la distribution *a posteriori* associées aux effets individuels, nous avons produit une matrice de corrélation que nous avons comparée ensuite avec celles des termes résiduels. La corrélation peut être produite par des chocs exogènes qui se manifesteront dans le terme d'erreur ou par des facteurs non observés spécifiques aux individus qui se manifesteront dans l'effet individuel. Dans le cas de l'arthrite chez les femmes, la corrélation est plus importante dans l'effet individuel que dans celle du terme résiduel. Chez les hommes, la corrélation du terme résiduel est importante dans le cas du diabète et des maladies pulmonaires. Les corrélations élevées du terme résiduel entre le diabète et les autres maladies chroniques pourraient s'expliquer par la visite chez le médecin. En effet, dans les premiers stades de la maladie, le diabète provoque peu ou pas de symptômes chez l'individu malade. Il est fort probable que l'individu apprend qu'il atteint de diabète par un test préventif lorsqu'il visite le médecin pour des problèmes liés à d'autres maladies chroniques.

Tableau 4.8c Comparaison des corrélations des termes résiduels et des effets individuels hiérarchiques

	Femmes		Hommes	
	Termes résiduels	Effets individuels	Termes résiduels	Effets individuels
cœur-diabète	0,303 **	0,288 **	0,525 **	0,293 **
cœur-arthrite	-0,055	0,262 **	0,102 *	0,142 **
cœur-dos	0,133 **	0,214 **	0,138 **	0,114 **
cœur-poumon	0,085	0,212 **	0,049	0,184 **
diabète-arthrite	0,019	0,074 **	0,293 **	0,105 **
diabète-dos	-0,001	0,013	0,113	0,004
diabète-poumon	0,074	0,102 **	0,240 *	0,055
arthrite-dos	0,128 **	0,419 **	0,142 **	0,391 **

arthrite-poumon	0,022	0,235 **	0,198 **	0,226 **
dos-poumon	0,073 *	0,179 **	0,097	0,179 **

4.5.8 La participation au marché du travail

Dans cette section, nous présentons les résultats concernant l'équation structurelle de la participation au marché du travail. Les paramètres de la distribution *a posteriori*, concernant les variables explicatives exogènes de l'équation de travail, sont présentés au Tableau 4.9. Premièrement, l'âge est le facteur le plus important dans la décision de participer au marché du travail chez les deux sexes. L'éducation joue également un rôle important. L'effet des variables de cycle montre que plus le temps avance, plus la participation au marché du travail chez les femmes augmente. Entre 1994 et 2010, il y a une différence de sept points de pourcentage chez les femmes et seulement deux points chez les hommes. Nous retrouvons également des résultats classiques en économie du travail concernant l'état matrimonial, le nombre d'enfants et le pays d'origine. En particulier, les enfants et le concubinage ont un effet négatif sur la participation au marché du travail des femmes et un effet positif pour celle des hommes. Les hommes issus de l'immigration ont une plus grande probabilité de participer au marché du travail alors que l'effet est contraire pour les femmes. La région d'origine a également un impact sur la participation au marché du travail. Par rapport à l'Ontario, les individus provenant des provinces de l'Atlantique et du Québec ont une probabilité plus faible d'être sur le marché du travail. Finalement, le poids a un effet direct sur la participation des femmes au marché du travail alors que cet effet est peu concluant chez les hommes.

Tableau 4.9 Coefficients de la participation au marché du travail (facteurs exogènes)

Participation	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	1,395	0,100	(1,19;1,59)	**	0,0050	-
Atlantique	-0,185	0,074	(-0,33;-0,04)	**	0,0023	-2,07
Québec	-0,321	0,081	(-0,48;-0,16)	**	0,0025	-2,71
Man. & sas	0,121	0,086	(-0,05;0,29)	*	0,0026	1,26
Alb. et C.B.	-0,053	0,073	(-0,19;0,09)		0,0021	-0,62
couple	-0,245	0,042	(-0,33;-0,16)	**	0,0010	-2,74
enfant	-0,439	0,030	(-0,50;-0,38)	**	0,0005	-5,63
immigrant	-0,236	0,080	(-0,39;-0,08)	**	0,0021	-2,63
age40-50	0,054	0,045	(-0,03;0,14)		0,0011	0,60
age50-55	-0,390	0,067	(-0,52;-0,26)	**	0,0022	-3,65
age55-60	-1,150	0,084	(-1,31;-0,99)	**	0,0031	-11,16
age60-65	-2,353	0,105	(-2,56;-2,14)	**	0,0040	-27,03
secondaire	0,476	0,069	(0,34;0,61)	**	0,0023	5,35
universitaire	0,778	0,074	(0,63;0,92)	**	0,0025	8,75
IMC25-30	-0,010	0,039	(-0,09;0,07)		0,0007	-0,11
IMC30-35	-0,101	0,056	(-0,21;0,01)	**	0,0011	-1,02
IMC35-40	-0,163	0,083	(-0,33;-0,00)	**	0,0017	-1,79
IMC40+	-0,218	0,115	(-0,44;0,01)	**	0,0025	-2,48
cycle 2	0,097	0,044	(0,01;0,18)	**	0,0008	1,10
cycle 3	0,217	0,047	(0,12;0,31)	**	0,0008	2,15
cycle 4	0,278	0,048	(0,19;0,37)	**	0,0009	2,77
cycle 5	0,378	0,052	(0,28;0,48)	**	0,0009	3,81
cycle 6	0,434	0,054	(0,33;0,54)	**	0,0010	4,34
cycle 7	0,509	0,056	(0,40;0,62)	**	0,0010	5,20
cycle 8	0,583	0,060	(0,47;0,70)	**	0,0011	6,02
cycle 9	0,575	0,063	(0,45;0,70)	**	0,0011	6,51

Participation	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	2,459	0,447	(0,00;2,80)	**	0,0254	-
Atlantique	-0,310	0,108	(-0,50;-0,00)	**	0,0047	-2,62
Québec	-0,425	0,128	(-0,64;-0,00)	**	0,0058	-2,65
Man. & sas	0,317	0,127	(0,00;0,56)	**	0,0052	2,59
Alb. et C.B.	0,035	0,098	(-0,16;0,22)		0,0037	0,28
couple	0,495	0,103	(0,02;0,63)	**	0,0050	4,16
enfant	0,061	0,057	(-0,04;0,18)		0,0018	1,45
immigrant	0,247	0,106	(0,00;0,46)	**	0,0039	2,13
age40-50	-0,133	0,070	(-0,27;0,01)	**	0,0025	-1,13
age50-55	-0,698	0,123	(-0,90;-0,29)	**	0,0057	-5,04
age55-60	-1,492	0,218	(-1,75;-0,50)	**	0,0114	-10,59
age60-65	-2,517	0,363	(-2,86;-0,70)	**	0,0195	-20,94
secondaire	0,336	0,143	(0,15;0,92)	**	0,0072	3,68
universitaire	0,476	0,141	(0,29;1,01)	**	0,0070	4,96
IMC25-30	0,237	0,068	(0,01;0,35)	**	0,0027	2,04
IMC30-35	0,126	0,077	(-0,02;0,27)	*	0,0023	0,90
IMC35-40	-0,090	0,119	(-0,33;0,14)		0,0031	-0,73
IMC40+	-0,197	0,190	(-0,57;0,17)		0,0049	-1,61
cycle 2	-0,012	0,065	(-0,14;0,12)		0,0016	-0,07
cycle 3	-0,067	0,069	(-0,20;0,07)		0,0018	-0,49
cycle 4	-0,113	0,073	(-0,25;0,03)	*	0,0021	-0,82
cycle 5	-0,032	0,074	(-0,18;0,11)		0,0019	-0,20
cycle 6	0,087	0,078	(-0,06;0,24)		0,0020	0,68
cycle 7	0,137	0,082	(-0,01;0,30)	**	0,0023	1,07
cycle 8	0,120	0,085	(-0,04;0,29)	*	0,0022	0,93
cycle 9	0,233	0,095	(0,01;0,41)	**	0,0030	2,05

Le Tableau 4.10a présente les paramètres des distributions *a posteriori* de l'impact des maladies chroniques sur la participation au marché du travail et le Tableau 4.10b présente les probabilités conditionnelles calculées à partir de la distribution *a poste-*

riori prédictive. Les résultats montrent que les impacts des maladies chroniques sont très différents entre les hommes et les femmes. L'impact des maladies cardiovasculaires et des maux de dos est en moyenne plus important chez les femmes que chez les hommes. Chez ces derniers, c'est l'impact de l'arthrite et du diabète qui est plus important. L'effet des maladies pulmonaires sur les deux sexes est très semblable.

Tableau 4.10a Coefficients des maladies chroniques sur la participation au marché du travail

	Femmes				
	Moyenne	Écart-type	(HPDI 95%)		MCSE
cœur	-1,042	0,171	(-1,37;-0,71)	**	0,0088
diabète	0,194	0,259	(-0,26;0,70)		0,0142
arthrite	-0,234	0,161	(-0,53;0,10)	*	0,0081
dos	-0,897	0,165	(-1,18;-0,57)	**	0,0081
poumon	-0,695	0,221	(-1,07;-0,22)	**	0,0111

	Hommes				
	Moyenne	Écart-type	(HPDI 95%)		MCSE
cœur	-0,683	0,255	(-1,12;0,00)	**	0,0130
diabète	-0,443	0,219	(-0,84;0,01)	**	0,0108
arthrite	-0,710	0,269	(-1,13;0,00)	**	0,0147
dos	-0,842	0,292	(-1,26;0,01)	**	0,0171
poumon	-1,110	0,307	(-1,57;-0,00)	**	0,0164

Tableau 4.10b Probabilités conditionnelles - changements en point de pourcentage de la probabilité de participer au marché du travail selon les maladies chroniques

	Femmes			
	Moyenne	Écart-type	(HPDI 95%)	
cœur	-18,792	3,743	(-25,7;-11,7)	**
diabète	1,925	4,066	(-6,3;9,4)	

arthrite	-3,979	2,648	(-9,1;1,2)	*
dos	-15,489	3,174	(-21,4;-9,3)	**
poumon	-12,007	4,260	(-19,8;-3,9)	**

	Hommes			
	Moyenne	Écart-type	(HPDI 95%)	
cœur	-7,112	3,001	(-12,9;0,0)	**
diabète	-4,176	2,350	(-8,9;0,3)	**
arthrite	-7,600	3,165	(-13,2;0,1)	**
dos	-8,558	3,538	(-14,0;0,3)	**
poumon	-12,718	4,166	(-19,4;-0,00)	**

Les paramètres des distributions *a posteriori* des corrélations des termes résiduels et des covariances des effets individuels entre la participation au marché du travail et les maladies chroniques sont présentés au Tableau 4.11a, 4.11b et 4.11c. La première remarque importante à propos de ces résultats concerne le sens des corrélations. D'un côté, on observe que les coefficients de corrélation des termes résiduels sont plutôt en zone positive alors que les covariances des effets individuels sont plutôt en zone négative. Ces résultats pourraient s'expliquer par les différents facteurs de risque captés par les termes résiduels et les effets individuels. Les modes de vie à risque comme l'alcoolisme et l'abus de drogue ou encore un environnement socio-économique difficile peuvent augmenter la probabilité de développer une maladie chronique tout en réduisant la probabilité de participer au marché du travail. Si l'effet individuel permet de capter ces effets, les covariances entre la participation au marché du travail et les maladies chroniques seront négatives. La corrélation positive des termes résiduels entre les maladies chroniques et la participation au marché du travail s'explique différemment. Pour que cette corrélation soit positive, il doit y avoir un choc exogène corrélié à la fois avec le travail et avec le diagnostic d'une maladie chronique. Une expli-

cation possible est la perte de productivité et la hausse de l'absentéisme qui inciteraient les travailleurs à consulter un médecin. Des difficultés à respirer, des maux de dos, des problèmes d'articulations inciteront les travailleurs à consulter alors que ces mêmes maux seront supportés par l'individu hors de marché du travail. Les résultats obtenus pour les maux de dos semblent aller dans ce sens.

Tableau 4.11a Coefficients de corrélation entre les maladies chroniques et la participation au marché du travail

Corrélation	Femmes			
	Moyenne	Écart-type	(HPDI 95%)	MCSE
cœur	0,393	0,089	(0,22;0,56)	** 0,0051
diabète	-0,136	0,162	(-0,44;0,15)	0,0094
arthrite	0,069	0,093	(-0,12;0,24)	0,0047
dos	0,486	0,090	(0,30;0,64)	** 0,0045
poumon	0,276	0,113	(0,03;0,46)	** 0,0059

Corrélation	Hommes			
	Moyenne	Écart-type	(HPDI 95%)	MCSE
cœur	0,090	0,201	(-0,25;0,38)	0,0044
diabète	0,047	0,223	(-0,39;0,37)	0,0051
arthrite	0,146	0,186	(-0,15;0,40)	0,0042
dos	0,288	0,203	(-0,04;0,55)	** 0,0057
poumon	0,362	0,202	(0,00;0,60)	** 0,0065

Tableau 4.11b Variances et covariances des effets individuels hiérarchiques entre les maladies chroniques et la participation au marché du travail

Covariance	Femmes				
	Moyenne	Écart-type	(HPDI 95%)	MCSE	
variance	2.229	0.115	(2.02;2.47)	**	0.0039
cœur	-0.220	0.131	(-0.47;0.03)	**	0.0063
diabète	-1.181	0.349	(-1.81;-0.48)	**	0.0187
arthrite	-0.465	0.122	(-0.71;-0.23)	**	0.0060
dos	-0.056	0.083	(-0.23;0.10)		0.0041
poumon	-0.347	0.108	(-0.57;-0.15)	**	0.0048

Covariance	Hommes				
	Moyenne	Écart-type	(HPDI 95%)	MCSE	
variance	2.442	0.322	(2.12;2.83)	**	0.0259
cœur	-0.629	0.216	(-1.05;-0.20)	**	0.0116
diabète	-0.753	0.548	(-1.91;0.17)	*	0.0285
arthrite	-0.420	0.172	(-0.77;-0.10)	**	0.0096
dos	-0.155	0.109	(-0.37;0.03)	*	0.0057
poumon	-0.475	0.146	(-0.76;-0.19)	**	0.0087

4.5.9 Les effets marginaux directs et indirects

Le Tableau 4.12 présente les effets marginaux directs, indirects et totaux des variables exogènes sur la participation au marché du travail. Les effets directs sont les effets des variables exogènes apparaissant dans l'équation structurelle de la participation au marché du travail et les effets indirects sont les effets des variables exogènes des équations de santé sur la participation au marché du travail. Contrairement à l'estimateur classique par maximum de vraisemblance, l'estimateur bayésien permet de caractériser facilement les effets marginaux directs et indirects. Pour calculer ces

effets, il suffit simplement d'échantillonner deux fois les distributions *a posteriori* prédictives en conditionnant l'ensemble des observations aux deux valeurs de la variable dichotomique étudiée et de calculer la différence entre les deux probabilités prédites de la participation au marché du travail. Sans surprise, l'âge est le facteur exogène qui a le plus d'impact sur la participation au marché du travail. La participation au marché du travail pour un homme de 60 à 64 ans est réduite de 27 points de pourcentage comparativement à un homme de 30 à 40 ans. De ces 27 points, près de six points de pourcentage sont expliqués par l'âge via les maladies chroniques. Chez les femmes âgées de 60 à 65 ans, la réduction est de 30 points dont 3 points via les maladies chroniques.

Tableau 4.12 Effets marginaux directs et indirects de la participation au marché du travail en points de pourcentage

	Femmes					
	Direct		Indirect		Total	
Atlantique	-2,07	(1,44)	-	-	-2,07	(1,44)
Québec	-2,71	(1,22)	-	-	-2,71	(1,22)
Man. & Sas	1,26	(1,45)	-	-	1,26	(1,45)
Alb. et C.B.	-0,62	(1,41)	-	-	-0,62	(1,41)
couple	-2,74	(0,91)	-	-	-2,74	(0,91)
enfant	-5,63	(1,04)	-	-	-5,63	(1,04)
immigrant	-2,63	(1,48)	-	-	-2,63	(1,48)
age40-50	0,60	(0,81)	-0,83	(0,53)	-0,22	(0,81)
age50-55	-3,65	(1,24)	-1,42	(0,87)	-5,07	(1,12)
age55-60	-11,16	(2,43)	-2,28	(1,30)	-13,44	(2,17)
age60-65	-27,03	(3,98)	-3,13	(2,02)	-29,82	(3,41)
secondaire	5,35	(1,57)	0,07	(0,37)	5,41	(1,55)
universitaire	8,75	(1,95)	-0,05	(0,39)	8,69	(1,92)
IMC25-30	-0,11	(0,78)	-0,21	(0,28)	-0,32	(0,76)
IMC30-35	-1,02	(0,99)	-0,59	(0,39)	-1,62	(0,97)

IMC35-40	-1,79	(1,57)	-1,43	(0,75)	-3,21	(1,53)
IMC40+	-2,48	(2,28)	-1,77	(1,07)	-4,23	(2,21)
cycle 2	1,10	(0,92)	-	-	1,10	(0,92)
cycle 3	2,15	(0,91)	-	-	2,15	(0,91)
cycle 4	2,77	(0,98)	-	-	2,77	(0,98)
cycle 5	3,81	(1,09)	-	-	3,81	(1,09)
cycle 6	4,34	(1,14)	-	-	4,34	(1,14)
cycle 7	5,20	(1,25)	-	-	5,20	(1,25)
cycle 8	6,02	(1,43)	-	-	6,02	(1,43)
cycle 9	6,51	(1,54)	-	-	6,51	(1,54)
tabac10-	-	-	-0,66	(0,40)	-0,66	(0,22)
tabac20-	-	-	-0,62	(0,35)	-0,62	(0,19)
tabac30-	-	-	-0,92	(0,43)	-0,91	(0,21)
tabac30+	-	-	-2,09	(0,77)	-2,08	(0,33)
parent-cœur	-	-	0,55	(0,19)	0,55	(0,19)
nsp-parent	-	-	0,15	(0,09)	0,15	(0,09)
diabète-parent	-	-	-0,34	(0,65)	-0,34	(0,65)
nsp-parent	-	-	-0,07	(0,15)	-0,07	(0,15)
asthme	-	-	-0,55	(0,20)	-0,55	(0,20)

	Hommes					
	Direct		Indirect		Total	
Atlantique	-2,62	(1,68)	-	-	-2,62	(1,68)
Québec	-2,65	(1,31)	-	-	-2,65	(1,31)
Man. & Sas	2,59	(1,80)	-	-	2,59	(1,80)
Alb. et C.B.	0,28	(1,44)	-	-	0,28	(1,44)
couple	4,16	(1,51)	-	-	4,16	(1,51)
enfant	1,45	(1,00)	-	-	1,45	(1,00)
immigrant	2,13	(1,69)	-	-	2,13	(1,69)
age40-50	-1,13	(1,13)	-1,24	(0,62)	-2,35	(1,18)
age50-55	-5,04	(1,69)	-2,21	(1,02)	-7,13	(1,70)
age55-60	-10,59	(2,98)	-3,34	(1,45)	-13,72	(2,96)
age60-65	-20,94	(5,11)	-5,58	(2,35)	-26,60	(4,75)

secondaire	3,68	(5,93)	-0,07	(0,46)	3,62	(5,81)
universitaire	4,96	(6,20)	-0,15	(0,46)	4,82	(6,12)
IMC25-30	2,04	(1,06)	-0,06	(0,19)	1,98	(1,05)
IMC30-35	0,90	(1,07)	-0,53	(0,32)	0,36	(1,06)
IMC35-40	-0,73	(1,98)	-0,97	(0,62)	-1,70	(1,99)
IMC40+	-1,61	(3,31)	-2,01	(1,11)	-3,59	(3,34)
cycle 2	-0,07	(1,11)	-	-	-0,07	(1,11)
cycle 3	-0,49	(0,96)	-	-	-0,49	(0,96)
cycle 4	-0,82	(1,04)	-	-	-0,82	(1,04)
cycle 5	-0,20	(1,13)	-	-	-0,20	(1,13)
cycle 6	0,68	(1,16)	-	-	0,68	(1,16)
cycle 7	1,07	(1,21)	-	-	1,07	(1,21)
cycle 8	0,93	(1,24)	-	-	0,93	(1,24)
cycle 9	2,05	(1,60)	-	-	2,05	(1,60)
tabac10-	-	-	-1,01	(0,57)	-1,01	(0,35)
tabac20-	-	-	-0,77	(0,40)	-0,77	(0,24)
tabac30-	-	-	-1,01	(0,47)	-1,01	(0,28)
tabac30+	-	-	-2,02	(0,86)	-2,01	(0,52)
parent-cœur	-	-	0,65	(0,26)	0,65	(0,26)
nsp-parent	-	-	0,25	(0,10)	0,25	(0,10)
diabète-parent	-	-	1,20	(0,63)	1,20	(0,63)
nsp-parent	-	-	0,29	(0,15)	0,29	(0,15)
asthme	-	-	-0,59	(0,21)	-0,59	(0,21)

L'éducation quant à elle augmente la probabilité de participer au marché du travail, mais l'effet de l'éducation via la santé est très faible. Le poids et le tabagisme ont également un impact sur la probabilité de participer au marché du travail. Les femmes avec un IMC supérieur à 30 et les hommes avec un IMC supérieur à 35 ont une probabilité plus faible de travailler que les individus avec un poids normal. L'effet moyen de l'obésité sur la participation au marché du travail semble plus important chez les femmes que chez les hommes. Chez les individus avec un IMC de 40 et plus,

la réduction est de 4,3 points de pourcentage chez les femmes et de 3,6 chez les hommes. Le tabagisme a des effets équivalents sur la probabilité de participer au marché du travail sur les deux sexes. Chez les individus qui ont fumé plus de 20 ans, la participation au marché du travail est réduite de 1 % alors qu'elle est réduite de 2 % chez les individus qui ont fumé plus de 30 ans.

4.5.10 Comparaison de modèle

Le Tableau 4.13 présente les probabilités conditionnelles de différents modèles obtenues à partir des distributions *a posteriori* prédictive. Les quatre modèles utilisés sont le probit, le probit avec effets individuels hiérarchiques, le probit multivarié et le probit multivarié avec effets individuels hiérarchiques. Les résultats des deux modèles probit sont détaillés en annexe ainsi que ceux du probit multivarié. Pour l'ensemble des modèles, le probit multivarié est le modèle dont la distribution des estimés des maladies chroniques est la plus négative. À l'exception des maladies cardiovasculaires et du diabète chez les hommes, les moyennes des probabilités conditionnelles sont supérieures à 20 %. Comparativement aux modèles univariés, il semble que l'endogénéité des maladies chroniques dans l'équation de participation au marché du travail ait un impact.

Les résultats du modèle incorporant des effets individuels sont par contre très différents. Chez les femmes, les paramètres de position de la distribution *a posteriori* prédictive concernant les maux de dos et les maladies cardio-vasculaires sont plus négatifs que ceux des hommes. Il semble que ces maladies aient un impact plus important sur la participation au marché du travail chez les femmes que les hommes. Chez les hommes, c'est le diabète et l'arthrite qui semblent avoir un impact plus important sur

la participation au marché du travail. L'impact des maladies pulmonaires est semblable pour les deux sexes.

Tableau 4.13 Comparaisons de modèles - changements en point de pourcentage de la probabilité de participer au marché du travail selon les maladies chroniques

Femmes						
	Probit			Probit multivarié		
	Moyenne	Écart-type		Moyenne	Écart-type	
cœur	-9,3	1,65	**	-23,77	4,49	**
diabète	-5,07	1,59	**	-24,89	4,33	**
arthrite	-4,74	0,74	**	-24,75	2,78	**
dos	-2,08	0,65	**	-32,51	3,12	**
poumon	-5,34	1,51	**	-25,05	3,76	**

Femmes						
	Probit avec effets individuels hiérarchiques			Probit multivarié avec effets individuels hiérarchiques		
	Moyenne	Écart-type		Moyenne	Écart-type	
cœur	-14,76	1,62	**	-18,79	3,74	**
diabète	-9,91	1,36	**	1,93	4,07	
arthrite	-6,94	0,67	**	-3,98	2,65	*
dos	-4,59	0,67	**	-15,49	3,17	**
poumon	-9,36	1,49	**	-12,01	4,26	**

Hommes						
	Probit			Probit multivarié		
	Moyenne	Écart-type		Moyenne	Écart-type	
cœurs	-5,64	1,03	**	-12,29	3,67	**
diabète	-3,6	1,02	**	-4,95	2,95	**
arthrite	-5,05	0,61	**	-24,86	2,71	**
dos	-2,22	0,5	**	-26,99	3,01	**
poumon	-5,42	1,36	**	-27,23	3,75	**

	Probit			Probit multivarié		
	avec effets individuels hiérarchiques			avec effets individuels hiérarchiques		
	Moyenne	Écart-type		Moyenne	Écart-type	
cœur	-8,75	0,94	**	-7,11	3	**
diabète	-3,56	0,83	**	-4,18	2,35	**
arthrite	-7,89	0,59	**	-7,6	3,17	**
dos	-4,13	0,49	**	-8,56	3,54	**
poumon	-10,91	1,51	**	-12,72	4,17	**

Ces différences de résultats entre les deux modèles multivariés sont intéressantes. Regardons de plus près l'impact de l'arthrite sur la participation au marché du travail. Dans le modèle probit multivarié, l'impact de l'arthrite est très important tant chez les hommes que chez les femmes. L'arthrite réduit la participation au marché du travail de 25 points de pourcentage. En comparant ces résultats au modèle à effets individuels, on remarque que l'impact moyen de l'arthrite est beaucoup plus faible. Une explication possible est la corrélation entre le mode de vie des individus et le travail. Si les individus ont un mode de vie qui augmente la probabilité de développer de l'arthrite et qui réduit du même coup la probabilité de travailler, l'arthrite aura peu d'impact sur la participation au marché du travail. En effet, les données de panel permettent d'observer l'historique de la participation au marché du travail des individus. Si les individus ne travaillaient pas avant le diagnostic et ne travaillent pas après le diagnostic, l'arthrite ne sera pas responsable de ce fait. Les paramètres de l'effet individuel hiérarchique de la distribution *a posteriori* captera alors cet effet et réduira le biais occasionné par les facteurs non observés.

4.5.11 Analyse de sensibilité

Pour vérifier la robustesse des résultats obtenus dans ce chapitre, une analyse de sensibilité a été effectuée avec plusieurs modèles différents. Le Tableau 4.14 présente les résultats de cette analyse de sensibilité. La première colonne (base) présente les estimés (moyenne et écart-type) des variables endogènes obtenus dans ce chapitre. Ce sont les résultats obtenus au Tableau 4.10.b. Étant donné que l'équation structurelle de la participation au marché du travail est l'équation qui risque de poser le plus de problème, nous avons estimé plusieurs fois le modèle en supprimant ou ajoutant des variables exogènes à cette équation. Ces modifications ont pour but de vérifier si les résultats dépendent des choix *a priori* sur les variables d'exclusion. La colonne « sans poids » présente les estimés obtenus sans les quatre variables de poids (IMC25-30, IMC30-35, IMC35-40 et IMC40+) dans l'équation de travail. Les résultats dans la colonne « tabac » sont ceux obtenus après avoir introduit les variables tabac (tabac10, tabac20, tabac30 et tabac30+) dans l'équation de travail de base. La colonne « antécédents » présente les résultats obtenus lorsque les variables d'antécédents familiaux (cœur-parent, cœur-nsp, diabète-parent et diabète-nsp) sont ajoutées aux variables tabac dans le modèle précédent. La dernière colonne présente les résultats obtenus une fois que l'ensemble des variables exogènes du modèle ont été ajouté à toutes les équations. Les résultats montrent que les estimés des variables endogènes du modèle ne sont pas très sensibles aux choix des variables d'exclusion.

Tableau 4.14 Analyse de sensibilité

Probit multivarié : femmes

	Base	Sans poids	Tabac	Antécédents	Complet
cœur	-0,238 (0,04)	-0,255 (0,04)	-0,237 (0,04)	-0,213 (0,05)	-0,207 (0,05)
diabète	-0,249 (0,04)	-0,209 (0,05)	-0,253 (0,05)	-0,257 (0,04)	-0,240 (0,04)
arthrite	-0,248 (0,03)	-0,218 (0,03)	-0,244 (0,03)	-0,244 (0,03)	-0,253 (0,03)
dos	-0,325 (0,03)	-0,329 (0,03)	-0,335 (0,03)	-0,332 (0,03)	-0,342 (0,03)
poumon	-0,251 (0,04)	-0,262 (0,04)	-0,247 (0,04)	-0,253 (0,04)	-0,251 (0,04)

Probit multivarié : hommes

	Base	Sans poids	Tabac	Antécédents	Complet
cœur	-0,123 (0,04)	-0,104 (0,04)	-0,113 (0,04)	-0,145 (0,04)	-0,189 (0,03)
diabète	-0,050 (0,03)	-0,025 (0,03)	-0,046 (0,03)	-0,006 (0,04)	0,015 (0,04)
arthrite	-0,249 (0,03)	-0,244 (0,03)	-0,251 (0,03)	-0,251 (0,03)	-0,257 (0,02)
dos	-0,270 (0,03)	-0,277 (0,03)	-0,275 (0,03)	-0,264 (0,03)	-0,267 (0,03)
poumon	-0,272 (0,04)	-0,281 (0,04)	-0,270 (0,04)	-0,266 (0,04)	-0,231 (0,03)

Probit multivarié avec effets individuels hiérarchiques : femmes

	Base	Sans poids	Tabac	Antécédents	Complet
cœur	-0,188 (0,04)	-0,182 (0,04)	-0,166 (0,05)	-0,156 (0,05)	-0,159 (0,04)
diabète	0,019 (0,04)	0,017 (0,04)	0,025 (0,04)	0,043 (0,04)	0,012 (0,03)
arthrite	-0,040 (0,03)	-0,052 (0,02)	-0,030 (0,03)	-0,035 (0,03)	-0,030 (0,02)
dos	-0,155 (0,03)	-0,172 (0,03)	-0,142 (0,06)	-0,140 (0,04)	-0,175 (0,03)
poumon	-0,120 (0,04)	-0,128 (0,04)	-0,091 (0,05)	-0,093 (0,05)	-0,113 (0,05)

Probit multivarié avec effets individuels hiérarchiques : hommes

	Base	Sans poids	Tabac	Antécédents	Complet
cœur	-0,071 (0,03)	-0,077 (0,02)	-0,070 (0,03)	-0,055 (0,03)	-0,071 (0,03)
diabète	-0,042 (0,02)	-0,048 (0,03)	-0,035 (0,03)	-0,023 (0,02)	-0,044 (0,02)
arthrite	-0,076 (0,03)	-0,095 (0,02)	-0,082 (0,03)	-0,074 (0,03)	-0,097 (0,03)
dos	-0,086 (0,04)	-0,098 (0,02)	-0,059 (0,04)	-0,063 (0,04)	-0,066 (0,05)
poumon	-0,127 (0,04)	-0,135 (0,04)	-0,117 (0,04)	-0,113 (0,05)	-0,102 (0,03)

CONCLUSION

Les modèles probits multivariés permettent d'étudier des modèles où plusieurs réponses discrètes sont corrélées entre elles. Ce modèle permet une grande flexibilité au niveau de la structure de corrélation des termes résiduels en supposant que les variables latentes suivent une distribution normale multivariée. L'estimation de ce modèle n'est pas possible par maximum de vraisemblance car il n'existe pas de forme analytique de la fonction de vraisemblance. Par ailleurs, l'approche par maximum de vraisemblance simulée est difficile dans des cas complexes car elle nécessite de simuler une intégrale à plusieurs dimensions, ce qui est très exigeant du point de vue numérique.

Cette limite numérique imposée par le maximum de vraisemblance peut être surpassée par une approche bayésienne. Néanmoins, le nombre d'équations, de paramètres et d'observations peut devenir rapidement un obstacle si l'algorithme utilisé n'est pas conçu pour estimer un modèle très complexe. En combinant la méthode PXDA de Talhouk, Doucet et Murphy (2012), la méthode de Chopin (2011) pour échantillonner des tirages à partir d'une distribution tronquée normale univariée et nos propres solutions pour échantillonner les coefficients de régression, nous avons créé un algorithme très efficace du point de vue numérique.

Le chapitre 3 présente les distributions conditionnelles *a posteriori* du probit multivarié avec effets individuels hiérarchiques et nous avons généralisé l'échantillonneur de Gibbs à la méthode PXDA de Talhouk, Doucet et Murphy (2012). À partir de nos

recherches exposées au chapitre 2 et 3, nous avons estimé l'impact de cinq maladies chroniques sur la participation au marché du travail à l'aide d'un probit multivarié avec structure réursive endogène et avec effets individuels hiérarchiques. Ce modèle a été estimé avec deux échantillons d'hommes et de femmes de 30 à 64 ans formés à partir des données l'Enquête nationale sur la santé des populations.

Les résultats montrent que l'âge, le poids et le tabagisme sont des facteurs de risque importants des maladies chroniques. L'âge a un impact très important sur l'arthrite, modéré sur les maux de dos, les maladies cardio-vasculaires et le diabète et a peu ou pas d'effets sur les maladies pulmonaires. L'âge est un facteur de risque sur l'arthrite plus important chez les femmes alors qu'il est un facteur de risque sur les maladies cardio-vasculaires plus important chez les hommes. Ces résultats sont conformes à la littérature médicale. À l'inverse, l'éducation a très peu d'impact sur les maladies chroniques. Il semblerait que la corrélation observée entre l'éducation et la santé provienne principalement des habitudes de vie qui sont corrélées à la fois avec l'éducation et la santé.

Le poids et le tabac sont quant à eux des facteurs de risque importants pour plusieurs maladies chroniques. Les individus avec un IMC supérieur à 25, c'est-à-dire ceux dont le poids est supérieur à la normale ont une plus grande probabilité d'avoir de l'arthrite, des maux de dos, du diabète et des problèmes cardio-vasculaires. L'impact devient de plus en plus grand au fur et à mesure que le poids augmente. L'obésité semble également affecter directement la participation au marché du travail des femmes.

Le nombre d'année de tabagisme a également un impact important sur plusieurs maladies chroniques, mais de façon beaucoup moins prononcée que le poids corporel.

Les individus qui ont fumé plus de trente ans ont une plus grande probabilité de développer des problèmes d'arthrite, des maux de dos, des maladies cardiaques et des maladies pulmonaires.

Ces modèles nous ont permis de contrôler pour les problèmes de l'endogénéité des maladies chroniques dans l'équation de participation au marché du travail et d'utiliser la comorbidité des différentes maladies comme source d'information sur les facteurs de risque non observables. Les paramètres de la distribution *a posteriori* concernant la corrélation des termes résiduels ont montré que les maladies chroniques étaient endogènes dans l'équation de travail et que l'estimation d'un modèle probit univarié sous-estimait l'impact des maladies chroniques sur la participation au marché du travail. Nous avons également montré que l'estimation d'un modèle probit multivarié sans effet individuel hiérarchique surestimait cet impact. À partir des effets marginaux indirects, nous avons pu montrer que l'obésité et le tabagisme avaient un impact négatif sur la participation au marché du travail via leurs effets sur les maladies chroniques. Les individus avec un IMC supérieur à 40 ont en moyenne une probabilité de participer au marché du travail réduite de quatre points de pourcentage et les individus qui ont fumé plus de 30 ans de leur vie voient cette probabilité réduite de deux points de pourcentage. Cela peut sembler peu, mais en termes de coût, cette réduction de la participation au marché du travail se chiffre à plusieurs milliards de dollars. Ces résultats montrent clairement que l'obésité et le tabagisme ont un coût indirect important.

Nous avons vu que les résultats changeaient énormément selon le modèle utilisé. En effet, il y a de grandes différences entre les résultats du probit, du probit multivarié et le probit multivarié avec effets individuels hiérarchiques. Ces grandes différences

entre les modèles nous poussent à la plus grande prudence. Nous savons que les maladies chroniques ont un impact sur la participation au marché du travail, mais il est encore difficile d'évaluer avec certitude cet impact.

En étudiant ce sujet, nous avons vu l'énorme complexité de la structure de corrélation entre la santé, l'emploi et les variables socio-économiques. Pour bien comprendre tous les impacts que peut avoir la santé sur la société, il est nécessaire de poursuivre la recherche sur ce sujet. Par exemple, est-ce que les habitudes de vie sont endogènes dans les équations de santé ? Est-il possible de contrôler ce problème avec des équations supplémentaires. Est-ce que un modèle dynamique utilisant pleinement la puissance des données de panel permettrait un meilleur éclairage sur ces questions ?

ANNEXE A

MODÈLE MULTIVARIÉ AVEC EFFETS INDIVIDUELS HIÉRARCHIQUES

Pour trouver les autres distributions conditionnelles *a posteriori* on utilise le théorème de Bayes en combinant la fonction de vraisemblance et les distributions *a priori* :

$$\begin{aligned}
 & p(\{\alpha_i\}, \alpha, \beta, \Sigma_\alpha, \Sigma_\varepsilon | y) \propto \\
 & \prod_{i=1}^N (2\pi)^{-T/2} |\Xi|^{-T/2} \exp\left(-\frac{1}{2}(y_i - (i_T \otimes \alpha) - x_i \beta)' \Xi^{-1} (y_i - (i_T \otimes \alpha) - x_i \beta)\right) \times \\
 & |\Sigma_\varepsilon|^{-(y_\varepsilon + m + 1)/2} \exp\left(-\frac{1}{2} \text{tr}(\Sigma_\varepsilon \Sigma_\varepsilon^{-1})\right) \times \exp\left(-\frac{1}{2}(\beta - \underline{\beta})' V_\beta^{-1} (\beta - \underline{\beta})\right) \times \\
 & \prod_{i=1}^n (2\pi)^{-1/2} |\Sigma_\alpha|^{-1/2} \exp\left(-\frac{1}{2}(\alpha_i - \alpha)' \Sigma_\alpha^{-1} (\alpha_i - \alpha)\right) \times \\
 & \exp\left(-\frac{1}{2}(\alpha - \underline{\alpha})' V_\alpha (\alpha - \underline{\alpha})\right) \times |\Sigma_\alpha|^{-(y_\alpha + m + 1)/2} \exp\left(-\frac{1}{2} \text{tr}(\Sigma_\alpha \Sigma_\alpha^{-1})\right).
 \end{aligned}$$

En posant $\tilde{y}_i = y_i - (i_T \otimes \alpha)$ alors :

$$\beta | \alpha, \sigma_\varepsilon^2, \sigma_\alpha^2, y \square N(\bar{\beta}, \bar{V}_\beta),$$

où

$$\bar{\beta} = \bar{V}_\beta \left(\sum_{i=1}^n x_i' \Xi^{-1} [y_i - (i_T \otimes \alpha)] + V_\beta^{-1} \underline{\beta} \right) \quad \text{et} \quad \bar{V}_\beta = \left(\sum_{i=1}^n x_i' \Xi^{-1} x_i + V_\beta^{-1} \right)^{-1}.$$

En éliminant tous les termes qui ne dépendent pas de α_i , on trouve la distribution conditionnelle suivante :

$$\begin{aligned} p(\alpha_i | \beta, \alpha, \sigma_\varepsilon^2, \sigma_\alpha^2, y) &\propto \\ \exp\left(-\frac{1}{2} \sum_{i=1}^T (y_{it} - \alpha_i - x_{it}\beta)' \Sigma_\varepsilon^{-1} (y_{it} - \alpha_i - x_{it}\beta)\right) &\times \exp\left(-\frac{1}{2} (\alpha_i - \alpha)' \Sigma_\alpha^{-1} (\alpha_i - \alpha)\right) = \\ \exp\left(-\frac{1}{2} \left[\sum_{i=1}^T (y_{it} - \alpha_i - x_{it}\beta)' \Sigma_\varepsilon^{-1} (y_{it} - \alpha_i - x_{it}\beta) + (\alpha_i - \alpha)' \Sigma_\alpha^{-1} (\alpha_i - \alpha) \right]\right). \end{aligned}$$

En posant $\hat{y}_{it} = y_{it} - x_{it}\beta$ et en réécrivant :

$$\begin{aligned} &\sum_{i=1}^T (\hat{y}_{it} - \alpha_i)' \Sigma_\varepsilon^{-1} (\hat{y}_{it} - \alpha_i) + (\alpha_i - \alpha)' \Sigma_\alpha^{-1} (\alpha_i - \alpha) = \\ &\sum_{i=1}^T \hat{y}_{it}' \Sigma_\varepsilon^{-1} \hat{y}_{it} - 2 \sum_{i=1}^T \alpha_i' \Sigma_\varepsilon^{-1} \hat{y}_{it} + \sum_{i=1}^T \alpha_i' \Sigma_\varepsilon^{-1} \alpha_i + \alpha_i' \Sigma_\alpha^{-1} \alpha_i - 2 \alpha_i' \Sigma_\alpha^{-1} \alpha + \alpha' \Sigma_\alpha^{-1} \alpha = \\ &\sum_{i=1}^T \hat{y}_{it}' \Sigma_\varepsilon^{-1} \hat{y}_{it} + \alpha_i' (T \Sigma_\varepsilon^{-1} + \Sigma_\alpha^{-1}) \alpha_i - 2 \alpha_i' \left(\Sigma_\varepsilon^{-1} \sum_{i=1}^T \hat{y}_{it} + \Sigma_\alpha^{-1} \alpha \right) + \alpha' \Sigma_\alpha^{-1} \alpha. \end{aligned}$$

De façon identique aux autres développements, on trouve :

$$a_i | \beta, \alpha, \sigma_\varepsilon^2, \sigma_\alpha^2, y \square N(\bar{a}_i, \bar{A}), \quad i = 1, \dots, n,$$

où

$$\bar{A} = (T \Sigma_\varepsilon^{-1} + \Sigma_\alpha^{-1})^{-1} \quad \text{et} \quad \bar{a} = \bar{A} \left(\Sigma_\varepsilon^{-1} \sum_{i=1}^T (y_{it} - x_{it}\beta) + \Sigma_\alpha^{-1} \alpha \right).$$

En éliminant tous les termes qui ne dépendent pas de α on trouve la distribution conditionnelle suivante :

$$p(\alpha | \beta, \{\alpha_i\}, \Sigma_\alpha, \Sigma_\varepsilon, y) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^N (\alpha_i - \alpha)' \Sigma_\alpha^{-1} (\alpha_i - \alpha) + (\alpha - \underline{\alpha})' V_\alpha^{-1} (\alpha - \underline{\alpha})\right).$$

En réécrivant :

$$\begin{aligned} & \sum_{i=1}^N (\alpha_i - \alpha)' \Sigma_\alpha^{-1} (\alpha_i - \alpha) + (\alpha - \underline{\alpha})' V_\alpha^{-1} (\alpha - \underline{\alpha}) = \\ & \sum_{i=1}^n \alpha_i' \Sigma_\alpha^{-1} \alpha_i - 2 \sum_{i=1}^n \alpha' \Sigma_\alpha^{-1} \alpha_i + \sum_{i=1}^n \alpha' \Sigma_\alpha^{-1} \alpha + \alpha' V_\alpha^{-1} \alpha - 2 \alpha' V_\alpha^{-1} \underline{\alpha} + \underline{\alpha}' V_\alpha^{-1} \underline{\alpha} = \\ & \sum_{i=1}^n \alpha_i' \Omega^{-1} \alpha_i + \alpha' (n \Sigma_\alpha^{-1} + V_\alpha^{-1}) \alpha - 2 \alpha' \left(\Sigma_\alpha^{-1} \sum_{i=1}^n \alpha_i + V_\alpha^{-1} \underline{\alpha} \right) + \underline{\alpha}' V_\alpha^{-1} \underline{\alpha}. \end{aligned}$$

De façon identique aux autres développements, on trouve :

$$\alpha | \{\alpha_i\}, \beta, \Sigma_\alpha, \Sigma_\varepsilon, y \square N(\bar{\alpha}, \bar{V}_\alpha),$$

où

$$\bar{V}_\alpha = (n \Sigma_\alpha^{-1} + V_\alpha^{-1})^{-1} \quad \text{et} \quad \bar{\alpha} = \bar{V}_\alpha \left(\Sigma_\alpha^{-1} \sum_{i=1}^n \alpha_i + V_\alpha^{-1} \underline{\alpha} \right).$$

En éliminant tous les termes qui ne dépendent pas de Σ_α , on trouve la distribution conditionnelle suivante :

$$p(\Sigma_\alpha | \{\alpha_i\}, \alpha, \beta, \Sigma_\varepsilon, y) \propto$$

$$|\Sigma_\alpha|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n (\alpha_i - \alpha)' \Sigma_\alpha^{-1} (\alpha_i - \alpha)\right) \times |\Sigma_\alpha|^{-(\nu_\alpha + m + 1)/2} \exp\left(-\frac{1}{2} \text{tr}(\underline{\Sigma}_\alpha \Sigma_\alpha^{-1})\right) =$$

$$|\Sigma_\alpha|^{-(n + \nu_\alpha + m + 1)/2} \exp\left(-\frac{1}{2} \text{tr}\left[\sum_{i=1}^n (\alpha_i - \alpha)' (\alpha_i - \alpha) \Sigma_\alpha^{-1} + \underline{\Sigma}_\alpha\right] \Sigma_\alpha^{-1}\right),$$

alors :

$$\Sigma_\alpha | \{\alpha_i\}, \alpha, \beta, \Sigma_\varepsilon, y \square IW(\bar{\Sigma}_\alpha, \bar{\nu}_\alpha),$$

où

$$\bar{\nu}_\alpha = \nu_\alpha + n \quad \text{et} \quad \bar{\Sigma}_\alpha = \sum_{i=1}^n (\alpha_i - \alpha)' (\alpha_i - \alpha) \Sigma_\alpha^{-1} + \underline{\Sigma}_\alpha.$$

En éliminant tous les termes qui ne dépendent pas de Σ_ε , on trouve la distribution conditionnelle suivante :

$$p(\Sigma_\varepsilon | \{\alpha_i\}, \alpha, \beta, \Sigma_\alpha, y) =$$

$$|\Sigma_\varepsilon|^{-(nT - \nu_\varepsilon - m - 1)/2} \exp\left(-\frac{1}{2} \left[\sum_{i=1}^n \sum_{t=1}^T (y_{it} - \alpha_i - x_{it} \beta)' \Sigma_\varepsilon^{-1} (y_{it} - \alpha_i - x_{it} \beta) + \text{tr}(\underline{\Sigma}_\varepsilon \Sigma_\varepsilon^{-1}) \right]\right)$$

$$|\Sigma_\varepsilon|^{-(nT - \nu_\varepsilon - m - 1)/2} \exp\left(-\frac{1}{2} \text{tr}\left[\sum_{i=1}^n \sum_{t=1}^T (y_{it} - \alpha_i - x_{it} \beta)' (y_{it} - \alpha_i - x_{it} \beta) + \underline{\Sigma}_\varepsilon\right] \Sigma_\varepsilon^{-1}\right),$$

alors

$$\Sigma_\varepsilon | \{\alpha_i\}, \alpha, \beta, \Sigma_\alpha, y \square IW[\bar{\Sigma}_\varepsilon, \bar{\nu}_\varepsilon],$$

où

$$\bar{\nu}_\varepsilon = \nu_\varepsilon + nT \quad \text{et} \quad \bar{\Sigma}_\varepsilon = \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \alpha_i - x_{it} \beta)' (y_{it} - \alpha_i - x_{it} \beta) + \underline{\Sigma}_\varepsilon.$$

ANNEXE B

PROBIT UNIVARIÉ

Tableau B.1 Paramètres de la distribution *a posteriori* du probit univarié

	Femmes				MCSE	ME(%)
	moyenne	écart-type	(HPDI 95%)			
constante	0,639	0,044	(0,55; 0,72)	**	0,0008	
Atlantique	-0,112	0,026	(-0,16; -0,06)	**	0,0005	-0,03
Québec	-0,184	0,029	(-0,24; -0,13)	**	0,0005	-0,04
Man. & Sas	0,125	0,031	(0,06; 0,19)	**	0,0006	0,03
Alb. et C.B.	-0,044	0,027	(-0,10; 0,01)	*	0,0005	-0,01
couple	-0,167	0,020	(-0,20; -0,13)	**	0,0004	-0,04
enfant	-0,330	0,019	(-0,37; -0,29)	**	0,0003	-0,11
immigrant	-0,123	0,028	(-0,18; -0,07)	**	0,0005	-0,03
age40-50	0,023	0,027	(-0,03; 0,07)		0,0005	0,01
age50-55	-0,263	0,032	(-0,32; -0,20)	**	0,0006	-0,06
age55-60	-0,716	0,031	(-0,78; -0,65)	**	0,0005	-0,19
age60-65	-1,429	0,034	(-1,49; -1,36)	**	0,0005	-0,48
secondaire	0,441	0,032	(0,38; 0,50)	**	0,0006	0,14
universitaire	0,672	0,031	(0,61; 0,73)	**	0,0005	0,20
cœur	-0,497	0,050	(-0,59; -0,40)	**	0,0008	-0,15
diabète	-0,348	0,044	(-0,44; -0,26)	**	0,0007	-0,10
arthrite	-0,253	0,022	(-0,30; -0,21)	**	0,0004	-0,07
dos	-0,171	0,022	(-0,21; -0,13)	**	0,0004	-0,05
poumon	-0,328	0,048	(-0,42; -0,23)	**	0,0008	-0,09
cycle 2	0,075	0,033	(0,01; 0,14)	**	0,0006	0,02
cycle 3	0,143	0,034	(0,08; 0,21)	**	0,0006	0,03
cycle 4	0,192	0,034	(0,13; 0,26)	**	0,0006	0,04
cycle 5	0,290	0,036	(0,22; 0,36)	**	0,0007	0,06

cycle 6	0,338	0,036	(0,27; 0,41)	**	0,0007	0,07
cycle 7	0,396	0,037	(0,32; 0,47)	**	0,0007	0,08
cycle 8	0,470	0,039	(0,39; 0,54)	**	0,0007	0,10
cycle 9	0,454	0,041	(0,37; 0,53)	**	0,0007	0,10
IMC25-30	0,026	0,021	(-0,02; 0,07)		0,0004	0,01
IMC30-35	-0,041	0,028	(-0,10; 0,01)	*	0,0005	-0,01
IMC35-40	-0,124	0,043	(-0,21; -0,04)	**	0,0007	-0,03
IMC40+	-0,241	0,054	(-0,35; -0,13)	**	0,0009	-0,07

	Hommes					
	moyenne	écart-type	(HPDI 95%)		MCSE	ME(%)
constante	1,219	0,058	(1,11; 1,34)	**	0,0014	
Atlantique	-0,166	0,034	(-0,23; -0,10)	**	0,0008	-0,03
Québec	-0,195	0,036	(-0,26; -0,12)	**	0,0009	-0,02
Man. & Sas	0,148	0,046	(0,06; 0,24)	**	0,0011	0,02
Alb. et C.B.	0,016	0,037	(-0,06; 0,09)		0,0011	0,00
couple	0,436	0,027	(0,38; 0,49)	**	0,0006	0,07
enfant	0,065	0,036	(-0,00; 0,14)	**	0,0014	0,02
immigrant	0,186	0,039	(0,11; 0,27)	**	0,0010	0,02
age40-50	-0,100	0,038	(-0,18; -0,03)	**	0,0012	-0,01
age50-55	-0,454	0,043	(-0,54; -0,37)	**	0,0012	-0,05
age55-60	-0,959	0,041	(-1,04; -0,88)	**	0,0010	-0,14
age60-65	-1,598	0,042	(-1,68; -1,52)	**	0,0010	-0,39
secondaire	0,215	0,037	(0,14; 0,29)	**	0,0009	0,04
universitaire	0,350	0,038	(0,27; 0,43)	**	0,0008	0,05
cœur	-0,495	0,042	(-0,58; -0,41)	**	0,0007	-0,09
diabète	-0,228	0,047	(-0,32; -0,13)	**	0,0009	-0,04
arthrite	-0,468	0,027	(-0,52; -0,42)	**	0,0006	-0,08
dos	-0,269	0,028	(-0,32; -0,21)	**	0,0006	-0,04
poumon	-0,591	0,066	(-0,72; -0,46)	**	0,0012	-0,11
cycle 2	0,015	0,046	(-0,08; 0,10)		0,0011	0,00
cycle 3	0,037	0,046	(-0,06; 0,13)		0,0012	0,00
cycle 4	0,014	0,047	(-0,08; 0,10)		0,0011	0,00
cycle 5	0,104	0,048	(0,01; 0,20)	**	0,0011	0,01
cycle 6	0,185	0,050	(0,09; 0,28)	**	0,0012	0,02
cycle 7	0,212	0,051	(0,11; 0,31)	**	0,0012	0,02

cycle 8	0,231	0,052	(0,13; 0,33)	**	0,0013	0,03
cycle 9	0,315	0,052	(0,21; 0,42)	**	0,0012	0,04
IMC25-30	0,148	0,028	(0,10; 0,20)	**	0,0007	0,02
IMC30-35	0,075	0,036	(0,00; 0,14)	**	0,0009	0,01
IMC35-40	-0,020	0,062	(-0,15; 0,10)		0,0014	0,00
IMC40+	-0,087	0,089	(-0,25; 0,08)		0,0018	-0,01

APPENDICE C

PROBIT UNIVARIÉ AVEC EFFETS INDIVIDUELS HIÉRARCHIQUES

Tableau B.1 Paramètres de la distribution *a posteriori* du probit univarié avec effets individuels hiérarchiques

Travail	Femmes				
	Moyenne	Écart-type	(HPDI 95%)		MCSE ME(%)
effet individuel	1,375	1,375	(1,20; 1,55)	**	0,0097 -
Atlantique	-0,218	0,077	(-0,36; -0,06)	**	0,0051 -0,03
Québec	-0,326	0,083	(-0,48; -0,17)	**	0,0055 -0,04
Man. & Sas	0,097	0,097	(-0,10; 0,28)		0,0061 0,01
Alb. et C.B.	-0,071	0,080	(-0,23; 0,09)		0,0047 -0,01
couple	-0,234	0,042	(-0,31; -0,15)	**	0,0020 -0,03
enfant	-0,463	0,032	(-0,52; -0,40)	**	0,0012 -0,09
immigrant	-0,229	0,080	(-0,39; -0,08)	**	0,0045 -0,04
age40-50	0,037	0,043	(-0,04; 0,12)		0,0020 0,01
age50-55	-0,436	0,058	(-0,55; -0,32)	**	0,0026 -0,06
age55-60	-1,257	0,064	(-1,38; -1,12)	**	0,0031 -0,20
age60-65	-2,497	0,077	(-2,65; -2,35)	**	0,0038 -0,48
secondaire	0,469	0,062	(0,34; 0,59)	**	0,0042 0,08
universitaire	0,794	0,062	(0,67; 0,91)	**	0,0044 0,13
cœur	-0,557	0,089	(-0,73; -0,39)	**	0,0030 -0,09
diabète	-0,312	0,092	(-0,49; -0,14)	**	0,0033 -0,05
arthrite	-0,299	0,043	(-0,38; -0,22)	**	0,0013 -0,05
dos	-0,135	0,040	(-0,21; -0,06)	**	0,0014 -0,02
poumon	-0,334	0,085	(-0,50; -0,16)	**	0,0029 -0,05
cycle 2	0,097	0,044	(0,01; 0,18)	**	0,0015 0,02
cycle 3	0,226	0,047	(0,13; 0,31)	**	0,0016 0,03
cycle 4	0,285	0,048	(0,19; 0,38)	**	0,0015 0,04
cycle 5	0,400	0,052	(0,30; 0,50)	**	0,0016 0,05

cycle 6	0,463	0,054	(0,36; 0,56)	**	0,0016	0,06
cycle 7	0,542	0,059	(0,42; 0,65)	**	0,0017	0,07
cycle 8	0,627	0,064	(0,51; 0,75)	**	0,0020	0,08
cycle 9	0,614	0,062	(0,49; 0,74)	**	0,0019	0,09
IMC25-30	-0,037	0,041	(-0,12; 0,04)		0,0018	-0,01
IMC30-35	-0,182	0,056	(-0,29; -0,07)	**	0,0025	-0,02
IMC35-40	-0,285	0,085	(-0,45; -0,12)	**	0,0034	-0,04
IMC40+	-0,358	0,113	(-0,58; -0,14)	**	0,0041	-0,06

Femmes						
Travail	Moyenne	Écart-type	(HPDI 95%)		MCSE	
variance	2,295	2,295	(2,10; 2,49)	**	0,0057	

Hommes						
Travail	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
effet individuel	2,494	2,494	(2,23; 2,75)	**	0,0163	-
Atlantique	-0,327	0,100	(-0,54; -0,14)	**	0,0073	-0,03
Québec	-0,399	0,109	(-0,61; -0,18)	**	0,0086	-0,02
Man. & Sas	0,388	0,118	(0,16; 0,62)	**	0,0092	0,03
Alb. et C.B.	0,076	0,104	(-0,12; 0,29)		0,0074	0,01
couple	0,560	0,061	(0,44; 0,68)	**	0,0037	0,05
enfant	0,062	0,060	(-0,06; 0,19)		0,0037	0,02
immigrant	0,332	0,112	(0,10; 0,57)	**	0,0078	0,03
age40-50	-0,175	0,065	(-0,31; -0,04)	**	0,0041	-0,01
age50-55	-0,834	0,082	(-0,99; -0,68)	**	0,0048	-0,06
age55-60	-1,742	0,087	(-1,91; -1,57)	**	0,0056	-0,15
age60-65	-2,883	0,103	(-3,08; -2,69)	**	0,0071	-0,40
secondaire	0,354	0,080	(0,20; 0,51)	**	0,0056	0,03
universitaire	0,517	0,079	(0,37; 0,69)	**	0,0056	0,05
cœur	-0,587	0,088	(-0,76; -0,41)	**	0,0036	-0,06
diabète	-0,395	0,101	(-0,59; -0,19)	**	0,0044	-0,04
arthrite	-0,550	0,057	(-0,66; -0,44)	**	0,0024	-0,05
dos	-0,258	0,052	(-0,36; -0,15)	**	0,0022	-0,02
poumon	-0,574	0,121	(-0,81; -0,34)	**	0,0045	-0,05

cycle 2	-0,001	0,066	(-0,13; 0,13)		0,0030	0,00
cycle 3	-0,071	0,068	(-0,20; 0,06)		0,0033	0,00
cycle 4	-0,118	0,068	(-0,25; 0,02)	**	0,0032	-0,01
cycle 5	-0,025	0,073	(-0,17; 0,11)		0,0033	0,00
cycle 6	0,108	0,078	(-0,05; 0,26)	*	0,0038	0,01
cycle 7	0,159	0,079	(-0,00; 0,31)	**	0,0034	0,01
cycle 8	0,145	0,085	(-0,03; 0,31)	*	0,0038	0,01
cycle 9	0,267	0,088	(0,09; 0,44)	**	0,0039	0,02
IMC25-30	0,265	0,055	(0,15; 0,37)	**	0,0029	0,02
IMC30-35	0,086	0,080	(-0,06; 0,25)		0,0044	0,01
IMC35-40	-0,165	0,124	(-0,40; 0,09)		0,0059	-0,01
IMC40+	-0,346	0,199	(-0,73; 0,04)	**	0,0088	-0,03

Hommes					
Travail	Moyenne	Écart-type	(HPDI 95%)		MCSE
variance	2,631	2,631	(2,34; 2,97)	**	0,0123

APPENDICE D

PROBIT MULTIVARIÉ

Tableau D.1 Paramètres de la distribution *a posteriori* de la participation au marché du travail

Participation	Femmes				MCSE	ME(%)
	Moyenne	Écart-type	(HPDI 95%)			
constante	0,865	0,042	(0,78;0,95)	**	0,0004	-
Atlantique	-0,082	0,023	(-0,13;-0,04)	**	0,0002	-0,58
Québec	-0,153	0,025	(-0,20;-0,10)	**	0,0002	-0,79
Man. & Sas	0,126	0,028	(0,07;0,18)	**	0,0002	0,73
Alb. et C.B.	-0,034	0,024	(-0,08;0,01)	*	0,0002	-0,24
couple	-0,220	0,018	(-0,25;-0,19)	**	0,0002	-1,38
enfant	-0,280	0,017	(-0,31;-0,25)	**	0,0002	-2,69
immigrant	-0,117	0,024	(-0,17;-0,07)	**	0,0002	-0,82
age40-50	0,090	0,024	(0,04;0,14)	**	0,0002	0,52
age50-55	-0,052	0,032	(-0,11;0,01)	*	0,0006	-0,29
age55-60	-0,345	0,037	(-0,42;-0,27)	**	0,0009	-2,15
age60-65	-0,931	0,044	(-1,02;-0,85)	**	0,0013	-9,54
secondaire	0,330	0,031	(0,27;0,39)	**	0,0003	2,91
universitaire	0,510	0,031	(0,45;0,57)	**	0,0004	4,10
IMC25-30	0,082	0,020	(0,04;0,12)	**	0,0002	0,54
IMC30-35	0,086	0,027	(0,03;0,14)	**	0,0003	0,49
IMC35-40	0,116	0,043	(0,03;0,20)	**	0,0006	0,72
IMC40+	0,079	0,054	(-0,03;0,19)	*	0,0008	0,51
cycle 2	0,062	0,029	(0,01;0,12)	**	0,0002	0,43
cycle 3	0,113	0,030	(0,06;0,17)	**	0,0002	0,65
cycle 4	0,148	0,031	(0,09;0,21)	**	0,0003	0,84
cycle 5	0,228	0,032	(0,17;0,29)	**	0,0003	1,27

cycle 6	0,268	0,032	(0,21;0,33)	**	0,0003	1,47
cycle 7	0,314	0,034	(0,25;0,38)	**	0,0003	1,71
cycle 8	0,366	0,035	(0,30;0,44)	**	0,0004	1,95
cycle 9	0,360	0,036	(0,29;0,43)	**	0,0004	2,06

Participation	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	1,351	0,054	(1,24;1,46)	**	0,0006	-
Atlantique	-0,108	0,030	(-0,17;-0,05)	**	0,0003	-0,37
Québec	-0,150	0,032	(-0,21;-0,09)	**	0,0003	-0,36
Man. & Sas	0,148	0,038	(0,07;0,22)		0,0004	0,43
Alb. et C.B.	0,023	0,032	(-0,04;0,09)	**	0,0003	0,08
couple	0,309	0,025	(0,26;0,36)	**	0,0004	1,10
enfant	0,085	0,031	(0,03;0,15)	**	0,0005	0,48
immigrant	0,161	0,033	(0,10;0,23)		0,0004	0,50
age40-50	-0,030	0,035	(-0,10;0,04)	**	0,0005	-0,08
age50-55	-0,263	0,041	(-0,34;-0,18)	**	0,0006	-0,68
age55-60	-0,630	0,043	(-0,71;-0,55)	**	0,0009	-1,91
age60-65	-1,148	0,048	(-1,25;-1,05)	**	0,0013	-6,46
secondaire	0,181	0,036	(0,11;0,25)	**	0,0003	0,67
universitaire	0,267	0,035	(0,20;0,34)	**	0,0004	0,95
IMC25-30	0,129	0,026	(0,08;0,18)	**	0,0003	0,43
IMC30-35	0,142	0,033	(0,08;0,21)	**	0,0006	0,39
IMC35-40	0,112	0,064	(-0,02;0,23)	*	0,0013	0,36
IMC40+	0,146	0,092	(-0,03;0,33)		0,0017	0,46
cycle 2	0,005	0,041	(-0,08;0,08)		0,0004	0,02
cycle 3	0,015	0,041	(-0,07;0,09)		0,0004	0,04
cycle 4	-0,013	0,042	(-0,10;0,07)	*	0,0004	-0,04
cycle 5	0,060	0,044	(-0,03;0,15)	**	0,0005	0,17
cycle 6	0,121	0,044	(0,04;0,21)	**	0,0005	0,35
cycle 7	0,148	0,045	(0,06;0,24)	**	0,0005	0,42
cycle 8	0,147	0,046	(0,06;0,24)	**	0,0005	0,42
cycle 9	0,222	0,048	(0,13;0,32)	1,00	0,0006	0,68

Tableau D.2 Impact des facteurs de risque sur l'arthrite

Arthrite	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	0,633	0,145	(0,35;0,91)	**	0,0029	-
age40-50	0,429	0,026	(0,38;0,48)	**	0,0002	9,93
age50-55	0,864	0,030	(0,81;0,92)	**	0,0002	18,31
age55-60	1,087	0,031	(1,03;1,15)	**	0,0002	25,81
age60-65	1,208	0,032	(1,14;1,27)	**	0,0002	34,88
secondaire	0,012	0,033	(-0,05;0,08)		0,0003	0,34
universitaire	0,035	0,034	(-0,03;0,10)		0,0004	0,89
IMC25-30	0,217	0,021	(0,18;0,26)	**	0,0001	5,41
IMC30-35	0,323	0,027	(0,27;0,37)	**	0,0002	7,12
IMC35-40	0,517	0,041	(0,44;0,60)	**	0,0002	13,52
IMC40+	0,725	0,051	(0,63;0,82)	**	0,0003	21,04
tabac10-	0,042	0,031	(-0,02;0,10)	*	0,0002	1,05
tabac20-	0,109	0,027	(0,06;0,16)	**	0,0002	2,35
tabac30-	0,225	0,026	(0,17;0,28)	**	0,0002	4,92
tabac30+	0,226	0,027	(0,17;0,28)	**	0,0002	5,89

Arthrite	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	0,950	0,173	(0,61;1,29)	**	0,0036	-
age40-50	0,307	0,030	(0,25;0,37)	**	0,0003	5,93
age50-55	0,640	0,035	(0,57;0,71)	**	0,0003	11,32
age55-60	0,828	0,035	(0,76;0,90)	**	0,0003	16,28
age60-65	0,899	0,037	(0,83;0,97)	**	0,0003	22,42
secondaire	0,150	0,037	(0,08;0,22)	**	0,0003	2,80
universitaire	0,126	0,037	(0,05;0,20)	**	0,0004	2,32
IMC25-30	0,100	0,025	(0,05;0,15)	**	0,0002	1,97
IMC30-35	0,281	0,030	(0,22;0,34)	**	0,0002	4,87
IMC35-40	0,501	0,053	(0,40;0,60)	**	0,0003	11,40

IMC40+	0,611	0,075	(0,46;0,76)	**	0,0005	15,15
tabac10-	0,154	0,038	(0,08;0,23)	**	0,0003	3,20
tabac20-	0,162	0,030	(0,10;0,22)	**	0,0002	2,73
tabac30-	0,245	0,029	(0,19;0,30)	**	0,0002	4,13
tabac30+	0,239	0,030	(0,18;0,30)	**	0,0002	5,11

Tableau D.3 Impact des facteurs de risque sur les maux de dos

Dos	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	0,637	0,144	(0,35;0,92)	**	0,0031	-
age40-50	0,083	0,022	(0,04;0,13)	**	0,0002	2,16
age50-55	0,107	0,029	(0,05;0,16)	**	0,0002	2,36
age55-60	0,188	0,030	(0,13;0,25)	**	0,0002	4,39
age60-65	0,073	0,034	(0,01;0,14)	**	0,0003	1,90
secondaire	0,122	0,034	(0,06;0,19)	**	0,0003	2,96
universitaire	0,169	0,035	(0,10;0,24)	**	0,0004	4,16
IMC25-30	0,080	0,021	(0,04;0,12)	**	0,0001	2,09
IMC30-35	0,160	0,027	(0,11;0,21)	**	0,0002	3,78
IMC35-40	0,281	0,041	(0,20;0,36)	**	0,0002	7,58
IMC40+	0,245	0,053	(0,14;0,35)	**	0,0003	6,82
tabac10-	0,116	0,028	(0,06;0,17)	**	0,0002	3,07
tabac20-	0,088	0,025	(0,04;0,14)	**	0,0002	1,94
tabac30-	0,079	0,026	(0,03;0,13)	**	0,0002	1,74
tabac30+	0,194	0,028	(0,14;0,25)	**	0,0002	5,29

Dos	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	0,562	0,165	(0,23;0,88)	**	0,0039	-
age40-50	0,109	0,025	(0,06;0,16)	**	0,0002	2,85
age50-55	0,119	0,031	(0,06;0,18)	**	0,0002	2,61
age55-60	0,148	0,032	(0,08;0,21)	**	0,0002	3,37
age60-65	0,098	0,036	(0,03;0,17)	**	0,0003	2,58

secondaire	0,040	0,035	(-0,03;0,11)		0,0003	0,95
universitaire	0,077	0,036	(0,01;0,15)	**	0,0003	1,90
IMC25-30	0,005	0,021	(-0,04;0,05)		0,0002	0,14
IMC30-35	0,129	0,028	(0,08;0,18)	**	0,0002	2,90
IMC35-40	0,058	0,052	(-0,05;0,16)		0,0003	1,48
IMC40+	0,321	0,074	(0,17;0,47)	**	0,0004	9,41
tabac10-	0,169	0,033	(0,10;0,23)	**	0,0003	4,46
tabac20-	0,163	0,026	(0,11;0,21)	**	0,0002	3,41
tabac30-	0,198	0,026	(0,15;0,25)	**	0,0002	4,19
tabac30+	0,271	0,028	(0,22;0,33)	**	0,0002	7,45

Tableau D.4 Impact des facteurs de risque sur les maladies pulmonaires

Poumon	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	0,425	0,231	(-0,03;0,88)	**	0,0040	-
age40-50	-0,041	0,045	(-0,13;0,05)		0,0006	-0,23
age50-55	0,007	0,054	(-0,10;0,11)		0,0007	0,03
age55-60	0,068	0,055	(-0,04;0,18)		0,0008	0,35
age60-65	-0,005	0,059	(-0,12;0,11)		0,0008	-0,02
secondaire	-0,157	0,052	(-0,26;-0,05)	**	0,0005	-0,98
universitaire	-0,177	0,055	(-0,29;-0,07)	**	0,0006	-1,09
tabac10-	0,118	0,057	(0,01;0,23)	**	0,0008	0,64
tabac20-	0,045	0,053	(-0,06;0,15)		0,0008	0,20
tabac30-	0,351	0,047	(0,26;0,44)	**	0,0006	1,75
tabac30+	0,503	0,047	(0,41;0,60)	**	0,0005	3,76
asthme	0,839	0,039	(0,76;0,92)	**	0,0004	8,28

Poumon	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	0,714	0,265	(0,21;1,23)	**	0,0043	-
age40-50	-0,060	0,055	(-0,16;0,05)		0,0009	-0,23
age50-55	-0,054	0,069	(-0,19;0,08)		0,0012	-0,18
age55-60	0,018	0,068	(-0,12;0,15)		0,0011	0,07
age60-65	0,081	0,070	(-0,06;0,22)		0,0010	0,39
secondaire	0,089	0,065	(-0,04;0,22)	*	0,0008	0,33

universitaire	0,131	0,066	(0,00;0,26)	**	0,0009	0,50
tabac10-	0,144	0,085	(-0,03;0,31)	*	0,0015	0,55
tabac20-	0,274	0,064	(0,15;0,40)	**	0,0012	0,88
tabac30-	0,398	0,061	(0,28;0,51)	**	0,0011	1,33
tabac30+	0,652	0,061	(0,53;0,77)	**	0,0010	3,94
asthme	0,903	0,050	(0,80;1,00)	**	0,0004	7,52

Tableau D.5 Impact des facteurs de risque sur les maladies cardio-vasculaires

Cœur	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	0,653	0,234	(0,18;1,10)	**	0,0047	-
age40-50	0,119	0,052	(0,02;0,22)	**	0,0008	0,62
age50-55	0,328	0,058	(0,21;0,44)	**	0,0008	1,63
age55-60	0,528	0,056	(0,42;0,64)	**	0,0008	3,00
age60-65	0,651	0,056	(0,54;0,76)	**	0,0008	5,46
secondaire	-0,070	0,049	(-0,17;0,03)	*	0,0006	-0,45
universitaire	-0,086	0,052	(-0,19;0,02)	*	0,0007	-0,53
IMC25-30	0,100	0,038	(0,02;0,17)	**	0,0005	0,59
IMC30-35	0,164	0,047	(0,07;0,25)	**	0,0005	0,84
IMC35-40	0,388	0,065	(0,26;0,52)	**	0,0006	2,72
IMC40+	0,355	0,080	(0,19;0,51)	**	0,0007	2,67
tabac10-	0,116	0,059	(-0,00;0,23)	**	0,0008	0,65
tabac20-	0,121	0,052	(0,02;0,22)	**	0,0007	0,60
tabac30-	0,221	0,048	(0,13;0,32)	**	0,0006	1,12
tabac30+	0,347	0,044	(0,26;0,43)	**	0,0005	2,42
parent-cœur	-0,310	0,044	(-0,39;-0,22)	**	0,0004	-2,69
nsp-cœur	-0,115	0,064	(-0,24;0,01)	**	0,0007	-0,60

Cœur	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	-0,474	0,271	(-1,04;0,03)	**	0,0081	-
age40-50	0,347	0,054	(0,24;0,46)	**	0,0009	2,62
age50-55	0,678	0,057	(0,56;0,79)	**	0,0009	4,87

age55-60	0,872	0,056	(0,77;0,98)	**	0,0008	7,17
age60-65	1,102	0,055	(1,00;1,21)	**	0,0007	14,73
secondaire	-0,043	0,049	(-0,14;0,05)		0,0006	-0,37
universitaire	-0,023	0,051	(-0,12;0,08)		0,0008	-0,21
IMC25-30	0,173	0,037	(0,10;0,25)	**	0,0005	1,43
IMC30-35	0,275	0,045	(0,19;0,36)	**	0,0005	1,90
IMC35-40	0,371	0,074	(0,22;0,52)	**	0,0007	3,39
IMC40+	0,413	0,108	(0,20;0,63)	**	0,0009	4,26
tabac10-	0,023	0,062	(-0,10;0,14)		0,0008	0,20
tabac20-	0,123	0,048	(0,03;0,22)	**	0,0006	0,84
tabac30-	0,276	0,043	(0,19;0,36)	**	0,0005	1,92
tabac30+	0,352	0,043	(0,27;0,44)	**	0,0006	3,41
parent-cœur	-0,334	0,047	(-0,43;-0,24)	**	0,0005	-5,25
nsp-cœur	-0,290	0,062	(-0,41;-0,17)	**	0,0009	-2,06

Tableau D.6 Impact des facteurs de risque sur le diabète

Diabète	Femmes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	0,359	0,238	(-0,13;0,81)	*	0,0048	-
age40-50	0,285	0,048	(0,19;0,38)	**	0,0007	1,86
age50-55	0,480	0,052	(0,38;0,58)	**	0,0007	2,79
age55-60	0,612	0,052	(0,51;0,71)	**	0,0007	3,94
age60-65	0,738	0,052	(0,64;0,84)	**	0,0007	6,95
secondaire	-0,079	0,051	(-0,18;0,02)	*	0,0007	-0,55
universitaire	-0,030	0,052	(-0,13;0,08)		0,0008	-0,23
IMC25-30	0,353	0,039	(0,28;0,43)	**	0,0005	2,53
IMC30-35	0,659	0,044	(0,57;0,74)	**	0,0005	4,57
IMC35-40	0,909	0,057	(0,80;1,02)	**	0,0006	9,28
IMC40+	1,025	0,067	(0,90;1,16)	**	0,0006	12,52
parent-diabète	-0,707	0,076	(-0,86;-0,56)	**	0,0007	-10,06
nsp-diabète	-0,507	0,088	(-0,68;-0,33)	**	0,0009	-2,55

Diabète	Hommes					
	Moyenne	Écart-type	(HPDI 95%)		MCSE	ME(%)
constante	-1,325	0,281	(-1,88;-0,79)	**	0,0072	-
age40-50	0,486	0,054	(0,38;0,59)	**	0,0008	3,90
age50-55	0,784	0,057	(0,67;0,89)	**	0,0008	5,68
age55-60	1,058	0,055	(0,95;1,16)	**	0,0008	9,17
age60-65	1,142	0,057	(1,03;1,25)	**	0,0007	14,62
secondaire	0,084	0,056	(-0,02;0,19)	*	0,0008	0,63
universitaire	0,094	0,057	(-0,02;0,20)	*	0,0010	0,71
IMC25-30	0,137	0,040	(0,06;0,21)	**	0,0005	1,11
IMC30-35	0,522	0,045	(0,43;0,61)	**	0,0005	4,05
IMC35-40	1,009	0,064	(0,88;1,14)	**	0,0005	13,85
IMC40+	1,234	0,086	(1,07;1,40)	**	0,0006	20,61
parent-diabète	-0,519	0,084	(-0,68;-0,35)	**	0,0006	-7,89
nsp-diabète	-0,324	0,093	(-0,50;-0,14)	**	0,0008	-2,25

Tableau D.7 Corrélation des facteurs de risque exogènes non observés

Corrélation	Femmes				
	Moyenne	Écart-type	(HPDI 95%)		MCSE
cœur-diabète	0,226	0,028	(0,17;0,28)	**	0,0007
cœur-arthrite	0,179	0,021	(0,14;0,22)	**	0,0004
cœur-dos	0,151	0,021	(0,11;0,19)	**	0,0004
cœur-poumon	0,155	0,032	(0,09;0,22)	**	0,0008
diabète-arthrite	0,040	0,021	(-0,00;0,08)	**	0,0004
diabète-dos	-0,027	0,021	(-0,07;0,01)		0,0004
diabète-poumon	0,045	0,034	(-0,02;0,11)	*	0,0009
arthrite-dos	0,316	0,012	(0,29;0,34)	**	0,0001
arthrite-poumon	0,171	0,022	(0,13;0,21)	**	0,0004
dos-poumon	0,139	0,021	(0,10;0,18)	**	0,0004

Corrélation	Hommes				MCSE
	Moyenne	Écart-type	(HPDI 95%)		
cœur-diabète	0,267	0,024	(0,22;0,31)	**	0,0005
cœur-arthrite	0,125	0,021	(0,08;0,17)	**	0,0004
cœur-dos	0,110	0,020	(0,07;0,15)	**	0,0003
cœur-poumon	0,142	0,035	(0,07;0,21)	**	0,0010
diabète-arthrite	0,085	0,022	(0,04;0,13)	**	0,0004
diabète-dos	-0,016	0,021	(-0,06;0,03)		0,0003
diabète-poumon	0,036	0,040	(-0,05;0,12)		0,0012
arthrite-dos	0,294	0,013	(0,27;0,32)	**	0,0002
arthrite-poumon	0,224	0,026	(0,17;0,28)	**	0,0006
dos-poumon	0,123	0,026	(0,07;0,17)	**	0,0006

Tableau D.8 Coefficients des maladies chroniques sur la participation au marché du travail

Participa- tion	Femmes				MCSE	ME(%)
	Moyenne	Écart-type	(HPDI 95%)			
cœur	-0,800	0,134	(-1,05;-0,53)	0,00	0,0061	-0,24
diabète	-0,839	0,133	(-1,08;-0,56)	0,00	0,0062	-0,25
arthrite	-0,827	0,080	(-0,98;-0,66)	0,00	0,0035	-0,25
dos	-1,060	0,084	(-1,21;-0,88)	0,00	0,0038	-0,33
poumon	-0,848	0,112	(-1,07;-0,62)	0,00	0,0046	-0,25

Participa- tion	Hommes				MCSE	ME(%)
	Moyenne	Écart-type	(HPDI 95%)			
cœur	-0,627	0,155	(-0,90;-0,30)	0,00	0,0077	-0,12
diabète	-0,278	0,160	(-0,59;0,02)	0,04	0,0075	-0,05
arthrite	-1,096	0,087	(-1,27;-0,92)	0,00	0,0037	-0,25
dos	-1,215	0,092	(-1,38;-1,03)	0,00	0,0042	-0,27
poumon	-1,175	0,124	(-1,42;-0,93)	0,00	0,0049	-0,27

Tableau D.9 Coefficients de corrélation entre les maladies chroniques et la participation au marché du travail

Corrélation	Femmes			
	Moyenne	Écart-type	(HPDI 95%)	MCSE
cœur	0,305	0,055	(0,19;0,41)	** 0,0025
diabète	0,286	0,060	(0,16;0,39)	** 0,0028
arthrite	0,459	0,040	(0,38;0,53)	** 0,0017
dos	0,575	0,041	(0,48;0,65)	** 0,0018
poumon	0,364	0,048	(0,27;0,46)	** 0,0020

Corrélation	Hommes			
	Moyenne	Écart-type	(HPDI 95%)	MCSE
cœur	0,191	0,072	(0,04;0,32)	** 0,0036
diabète	0,076	0,076	(-0,07;0,22)	0,0035
arthrite	0,494	0,040	(0,41;0,57)	** 0,0017
dos	0,600	0,042	(0,51;0,67)	** 0,0019
poumon	0,422	0,048	(0,33;0,52)	** 0,0020

RÉFÉRENCES

- Albert, James H., et Siddhartha Chib. 1993. «Bayesian Analysis of Binary and Polychotomous Response Data». *Journal of the American Statistical Association*, vol. 88, no 422, p. 669-679.
- Ashford, J., et R. Snowden. 1970. «Multivariate Probit Analysis». *Biometrics*, vol. 26, no 3, p. 535-545.
- Barnard, John, Robert McCulloch et Xiao-Li Meng. 2000. «Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage». *Statistica Sinica*, vol. 10, no 4, p. 1281-1311.
- Brown, Shelton H., José A. Pagán et Elena Bastida. 2005. «The Impact of Diabète on Employment: Genetic IVs in a Bivariate Probit». *Health Economics*, vol. 14, no 5, p. 537-544.
- Butler, J. S., et Robert Moffitt. 1982. «A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model». *Econometrica*, vol. 50, no 3, p. 761-764.
- Cai, Lixin. 2009. «Effects of Health on Wages of Australian Men». *Economic Record*, vol. 85, no 270, p. 290-306.
- Cappellari, Lorenzo, et Stephen P. Jenkins. 2006. «Calculation of Multivariate Normal Probabilities by Simulation, with Applications to Maximum Simulated Likelihood Estimation». *The STATA Journal*, vol. 6, p. 156-189.
- Chaloupka, Frank J. et Kenneth E. Warner. 2000. «The Economics of Smoking». In *Handbook of Health Economics*, A. J. Culyer et J. P. Newhouse, pages 1539-1627 Elsevier.
- Chib, Siddhartha, et Edward Greenberg. 1998. «Analysis of Multivariate Probit Models». *Biometrika*, vol. 85, no 2, p. 347-361.
- Chopin, Nicolas. 2011. «Fast Simulation of Truncated Gaussian Distributions». *Statistics and Computing*, vol. 21, no 2, p. 275-288.

- Costa, Dora L. 1996. «Health and Labor Force Participation of Older Men, 1900–1991». *The Journal of Economic History*, vol. 56, no 01, p. 62-89.
- Currie, Janet, et Brigitte C. Madrian. 1999. «Health, Health Insurance and the Labor Market». In *Handbook of Labor Economics*, C. Ashenfelter Orley et Card David, p. 3309-3416: Elsevier.
- Daniels, Michael J., et Robert E. Kass. 1999. «Nonconjugate Bayesian Estimation of Covariance Matrices and Its Use in Hierarchical Models». *Journal of the American Statistical Association*, vol. 94, no 448, p. 1254-1263.
- Dyk, David A. van, et Xiao-Li Meng. 2001. «The Art of Data Augmentation». *Journal of Computational and Graphical Statistics*, vol. 10, no 1, p. 1-50.
- Edwards, Yancy D., et Greg M. Allenby. 2003. «Multivariate Analysis of Multiple Response Data». *Journal of Marketing Research*, vol. 40, no 3, p. 321-334.
- Flegal, J.M. et G.L. Jones. 2010. «Batch Means and Spectral Variance Estimators in Markov Chain Monte Carlo». *Annals of Statistics*, vol. 38, no 2, p. 1034-1070.
- Geweke, J. 1992. «Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments». In *Bayesian Statistics 4*, J. M. Bernardo, J. Berger, A. P. Dawid et J. F. M. Smith, p. 169-193: Oxford University Press.
- Geweke, John. 1991. *Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints: Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*. Fairfax: Interface Foundation of North America, Inc., 571-578 p.
- Geweke, John, Michael Keane et David Runkle. 1994. «Alternative Computational Approaches to Inference in the Multinomial Probit Model». *The Review of Economics and Statistics*, vol. 76, no 4, p. 609-632.
- Hajivassiliou, Vassilis A., et Daniel L. McFadden. 1998. «The Method of Simulated Scores for the Estimation of LDV Models». *Econometrica*, vol. 66, no 4, p. 863-896.
- Keane, M.P. 1990. *Four Essays in Empirical Macro and Labor Economics*: Brown University p.
- Koop, G. 2003. *Bayesian Econometrics*: J. Wiley, 359 p.

- Lerman, S., et C. Manski. 1981. «On the Use of Simulated Frequencies to Approximate Choice Probabilities». *Structural analysis of discrete data with econometric applications*, vol. 10, p. 305-319.
- Liechty, John C., Merrill W. Liechty et Peter Müller. 2004. «Bayesian Correlation Estimation». *Biometrika*, vol. 91, no 1, p. 1-14.
- Liu, Jun S., et Ying Nian Wu. 1999. «Parameter Expansion for Data Augmentation». *Journal of the American Statistical Association*, vol. 94, no 448, p. 1264-1274.
- Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*: Cambridge University Press, 395 p.
- McCulloch, Robert, et Peter E. Rossi. 1994. «An Exact Likelihood Analysis of the Multinomial Probit Model». *Journal of Econometrics*, vol. 64, no 1-2, p. 207-240.
- McFadden, Daniel. 1989. «A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration». *Econometrica*, vol. 57, no 5, p. 995-1026.
- Meango, Romuald et Ismael Mourifié. 2013. «A Note on the Identification in Two Equations Probit Model with Dummy Endogenous Regressor». Working Paper. No. 503. University of Toronto.
- Nandram, Balgobin, et Ming-Hui Chen. 1996. «Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence». *Journal of Statistical Computation and Simulation*, vol. 54, no 1-3, p. 129-144.
- Nobile, Agostino. 2000. «Comment: Bayesian Multinomial Probit Models with a Normalization Constraint». *Journal of Econometrics*, vol. 99, no 2, p. 335-345.
- Norton, Edward C., et Euna Han. 2008. «Genetic Information, Obesity, and Labor Market Outcomes». *Health Economics*, vol. 17, no 9, p. 1089-1104.
- Ritter, Christian, et Martin A. Tanner. 1992. «Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler». *Journal of the American Statistical Association*, vol. 87, no 419, p. 861-868.
- Rossi, P.E., G.M. Allenby et R.E. McCulloch. 2005. *Bayesian Statistics and Marketing*, no vol. 13: Wiley, 348 p.

- Smith, James P. 2009. «The Impact of Childhood Health on Adult Labor Market Outcomes». *Review of Economics and Statistics*, vol. 91, no 3, p. 478-489.
- Talhouk, Aline, Arnaud Doucet et Kevin Murphy. 2012. «Efficient Bayesian Inference for Multivariate Probit Models with Sparse Inverse Correlation Matrices». *Journal of Computational and Graphical Statistics*.
- Tanner, Martin A., et Wing Hung Wong. 1987. «The Calculation of Posterior Distributions by Data Augmentation». *Journal of the American Statistical Association*, vol. 82, no 398, p. 528-540.
- Train, Kenneth. 2003. *Discrete Choice Methods With Simulation*: Cambridge University Press.
- Wilde, J. 2000. «Identification of Multiple Equation Probit Models with Endogenous Dummy Regressors». *Economics Letters*, vol. 69, no 3, p. 309-312.
- Wong, Frederick, Christopher K. Carter et Robert Kohn. 2003. «Efficient Estimation of Covariance Selection Models». *Biometrika*, vol. 90, no 4, p. 809-830.
- Zhang, Xiaohui, Xueyan Zhao et Anthony Harris. 2009. «Chronic Diseases and Labour Force Participation in Australia». *Journal of Health Economics*, vol. 28, no 1, p. 91-108.