UNIVERSITÉ DU QUÉBEC À MONTRÉAL

INTRODUCTION DES EFFETS FIXES DANS UN MODÈLE BINOMIAL NÉGATIF : APPLICATION À LA CONSOMMATION DE MÉDICAMENTS AU CANADA

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉCONOMIQUE

PAR
BOUBEKEUR SEDDIK AMMOUR

AVRIL 2010

REMERCIEMENTS

Je voudrais exprimer ma gratitude aux professeurs Pierre Ouellette et Philip Merrigan du département des sciences économiques à l'Université du Québec à Montréal, pour m'avoir assisté et accepté de diriger mon mémoire. Leurs disponibilités et commentaires éclairés m'ont amplement aidé tout au long de mon travail. Je les remercie infiniment pour leur précieuse aide. Je remercie également le Centre interuniversitaire québécois de statistiques sociales (CIQSS) pour la mise à disposition des données.

Je ne cesserai jamais de remercier mes parents pour m'avoir encouragé et soutenu dans la poursuite de mes études. Je tiens à remercier aussi ma femme Fatima, qui a été d'une grande patience avec moi.

Mes sincères remerciements à tous ceux qui ont contribué de loin ou de près à la réalisation de ce mémoire.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

Tableau

LISTE DES ABRÉVIATIONS

ENSP	Enquête nationale sur la santé de la population au Canada
FE	*Fixed effect*
FENB	*Fixed effect negative binomial model*
GEE	*Generalised Estimating Equations*
GLM	*Generalized linear models*
HHG	*Hausman, Hall and Griliches*
NB / NEGBIN	*Negative binomial model*
RE	*Random effect*
STD.DEV	*Standard deviation*

RÉSUMÉ

Dans ce travail, nous modélisons le nombre de médicaments consommés par les Canadiens en estimant deux modèles de données de comptage en panel. Les données incluent des observations effectuées sur 17 276 individus entre 1994 et 2007. Les modèles de Poisson et binomial négatif à effets fixes sont donc estimés afin de contrôler pour les effets individuels non observés. Allison et Waterman (2002) ont démontré que le modèle binomial négatif conditionnel à effets fixes proposé par Hausman, Hall and Griliches (1984) ne contrôle pas vraiment pour les effets fixes étant donné que le modèle inclut des régresseurs individuels fixes. Malgré cette argumentation, l'estimation de ce modèle, effectuée avec le progiciel Stata en utilisant les données de l'enquête nationale sur la santé de la population au Canada (ENSP), fournit des résultats presque identiques à ceux obtenus avec l'estimation du modèle de Poisson à effets fixes. En effet, une partie de la corrélation de certaines variables explicatives avec l'hétérogénéité inobservée n'existe plus lorsque nous estimons le modèle à effets fixes, comparativement aux modèles à effets aléatoires. Cela fait radicalement changer les effets individuels et prouve alors l'importance de contrôler pour les effets fixes. Wooldridge propose un modèle similaire au modèle binomial négatif à effets fixes. Celui-ci permet de contrôler et d'interpréter les effets du modèle binomial négatif à effets fixes en estimant les effets fixes à l'aide des moyennes des variables explicatives, qui seront introduites dans le modèle binomial négatif. Nous n'aurons donc plus besoin d'utiliser la commande Xtnbreg avec l'option FE (qui s'est révélée non performante) car le modèle alternatif de Wooldridge s'estime en utilisant les options RE ou GEE.

Mots clés : régression de Poisson, binomial négatif, effets fixes, effets aléatoires.

INTRODUCTION

Les modèles de données de comptage sont devenus monnaie courante dans plusieurs branches de la littérature empirique. C'est le cas en économie de la santé (avec le nombre de visites de santé), en management (avec le nombre de brevets), et en organisation industrielle (avec le nombre des entrants sur le marché). Les modèles de base sont les modèles de Poisson et binomial négatif, ce dernier étant proposé comme alternative pour corriger le problème d'équidispersion caractérisant le modèle de Poisson. Ces deux modèles ont été étendus à plusieurs applications afin de faciliter certaines utilisations des données qui représentent certains mécanismes tels que les effets de « hurdle » (« hurdle effects »), de zéro inflation (« zero inflated) » et la méthode de sélection d'échantillons (« sample selection »). Les modèles de base pour les données de panel, avec effets fixes et aléatoires, ont également été étendus aux modèles de Poisson et binomial négatif. Plusieurs autres améliorations et extensions de ces modèles (comme sur les plans bivarié et multivarié) ont été documentées (voir par exemple Cameron et Trivedi (1998), Hilbe (2007) et Greene (2007)).

Les différents modèles traitant des données de panel à effets fixes pour les données de comptage ont été introduits par Hausman, Hall and Griliches (abrégé ci-après HHG; 1984). Un modèle à effets fixes traite les effets individuels particuliers en tant que variables inobservées corrélées avec les régresseurs. En fait, l'introduction des effets fixes dans un modèle consiste en l'addition de variables binaires (indicatrices) pour chaque individu ou groupe de panel en vue d'expliquer l'effet individuel. Bien que cette approche, aussi appelée estimation non conditionnelle, fonctionne, le nombre de régresseurs dans le modèle peut exiger beaucoup de mémoire et rendre le calcul prohibitif. Guimarães et Portugal (2009) montrent, en utilisant le progiciel Stata, comment il est possible d'ajouter un très grand nombre de variables binaires dans un modèle sans avoir besoin de plus de mémoire. Pour quelques modèles (par exemple, les modèles linéaires, de Poisson et logit), une autre méthode fonctionne sans nécessiter d'ajouter les variables binaires comme régresseurs : il s'agit de la méthode de vraisemblance conditionnelle. Cette approche donnera les mêmes résultats pour les variables d'intérêt tout en évitant l'estimation des coefficients de variables binaires. Pour le modèle binomial négatif, il est prouvé que la vraisemblance conditionnelle[1] élimine également les effets fixes; toutefois, comme cela a été montré par Allison et Waterman (2002) et Guimarães (2007), cette méthode n'est plus performante dans ce cas.

Allison (2009) confirme également que l'estimation conditionnelle du modèle binomial négatif à effets fixes (abrégé ci-après FENB) n'est plus valide et qu'ainsi, la meilleure alternative pour estimer le modèle FENB est d'effectuer l'estimation inconditionnelle qui requiert l'inclusion de variables binaires pour chaque groupe de panel. En effet, la méthode de vraisemblance conditionnelle ne donne pas les mêmes résultats que l'ajout de variables binaires au modèle binomial négatif.

Les modèles de données de panel sont construits afin, d'une part, de contrôler toutes les variables stables dans le modèle, et d'autre part, de prendre en compte la corrélation des variables explicatives avec le terme d'erreur dans les groupes de panel. L'avantage du modèle conditionnel à effets fixes est qu'un grand nombre d'individus peut être ajusté et en même temps conditionné hors du modèle (Hilbe, 2007). Dans le cas particulier où les données ne présentent pas la variabilité requise par le modèle FENB, les modèles à effets aléatoires, ou même l'estimation par la méthode de « Generalised Estimating Equations » (GEE) peuvent être plus appropriés.

Ce travail comprend six chapitres. Dans le premier chapitre, nous dressons un bref survol des modèles à effets fixes des données de comptage. Nous exposons dans les deuxième et troisième chapitres le principe de base pour l'estimation des modèles

[1] L'estimation conditionnelle du modèle binomial négatif dans le progiciel Stata par le biais de la commande « XTnbreg » s'est révélée non adéquate.

de Poisson et binomial négatif, respectivement *sans* et *avec* panel. Nous présentons ensuite les principaux problèmes qui résultent de ces modèles en panel ainsi que les différentes alternatives proposées respectivement dans les quatrième et cinquième chapitres. Dans le dernier chapitre, nous ferons état des résultats de l'estimation des deux modèles de Poisson et binomial négatif et ce, à partir des données longitudinales issues de l'enquête nationale sur la santé au Canada (ENSP) effectuée de 1994 à 2007. Plus précisément, nous estimerons un modèle déterminant le nombre de médicaments consommés par les Canadiens, en l'examinant et le comparant en fonction de la méthode utilisée. Nous conclurons par l'interprétation des résultats.

CHAPITRE I

CONTEXTE AUTOUR DES MODÈLES À EFFETS FIXES DES DONNÉES DE COMPTAGE

L'utilisation de données longitudinales ou de panel est récemment devenue plus fréquente, particulièrement en économie. Elle concerne des échantillons de données micro-économiques observées en coupes transversales répétées sur les mêmes unités. Dans les données de panel, les observations du même individu ou de la même unité sont considérées comme dépendantes alors que chaque individu est indépendant des autres. L'avantage de l'analyse de ce type de données est qu'elle permet de modéliser l'hétérogénéité entre les groupes de panels ou l'hétérogénéité individuelle. Cela constitue un nouveau domaine de l'économétrie sur lequel se fondent les recherches récentes qui font apparaître de nouvelles techniques liées à ce domaine. Hausman, Hall et Griliches (1984) ont introduit une méthode de traitement de ces données en incluant un effet individuel commun à tous les individus dans le modèle qui permet la dépendance entre ces derniers.

Les modèles de données de panel se distinguent alors par l'inclusion de « l'effet individuel » ou de « l'hétérogénéité individuelle non observée », qui permet de différencier le comportement de chacun des individus. On distingue les modèles à effets fixes (FE) et les modèles à effets aléatoires (RE); la différence entre ces deux types de modèles réside dans la corrélation (cas des modèles FE) ou la non-corrélation (cas des modèles RE) entre l'effet individuel non observé et les régresseurs du modèle[2]. L'estimation de ces deux types de modèles pose certains

[2] FE : Fixed Effect Model; RE : Random Effect Model.

problèmes. Pour les modèles à effets fixes, le problème de paramètres incidents rend l'estimateur du maximum de vraisemblance non convergent; les modèles à effets aléatoires, quant à eux, exigent une hypothèse très restrictive sur l'hétérogénéité en ce qui concerne sa corrélation nulle avec les variables observées incluses dans le modèle (Greene, 2005).

Les modèles de données de panel se sont étendus aux modèles de choix discrets ainsi qu'aux modèles non linéaires. Nous nous intéressons ici aux modèles de données de dénombrement, qui peuvent être estimés par la méthode des moindres carrés. Toutefois, ces données de comptage sont en fait discrètes et elles présentent une distribution asymétrique, ce qui rend l'estimation par une régression non linéaire plus appropriée. Généralement, le modèle de Poisson est le plus utilisé parmi les modèles non linéaires; mais ce modèle a été critiqué à cause de l'hypothèse d'équidispersion supposant l'égalité entre la moyenne et la variance de la variable dépendante, conditionnelles aux régresseurs, d'où la nécessité de procéder à un test de « surdispersion » pour analyser les données avant modélisation. Plusieurs auteurs tels que Hausman, Hall et Griliches ont proposé des alternatives prenant en compte la surdispersion – la plus connue étant le modèle binomial négatif. En fait, la distribution de Poisson est obtenue comme une restriction paramétrique du modèle binomial négatif (Greene, 2005).

Les méthodes à effets fixes apportent une solution au problème de l'omission de certaines variables importantes, qui conduit à une estimation biaisée de l'effet des autres variables. Le rôle important de l'effet fixe réside dans la capacité d'attirer et de contrôler, dans la modélisation, toutes les caractéristiques non observées et stables dans le temps et ce, sans avoir à les mesurer. Cela élimine ainsi une grande partie du biais dans l'estimation (Allison, 2005). La méthode à effets fixes est une méthode d'estimation intra-individuelle (« within subject »). Cette dernière ne donne pas d'estimation pour les coefficients des variables qui n'ont pas de variation intrasujet (c'est-à-dire les variables qui ne changent pas dans le temps), comme le sexe ou le lieu de naissance. Toutes ces variables sont contrôlées par la régression à effets fixes même si elles ne sont pas mesurées (Allison, 2005).

Selon Greene (2001), l'utilisation du modèle non linéaire à effets fixes des données de panel est en pratique problématique en raison d'une part, du problème de paramètres incidents, et d'autre part, de la difficulté d'estimer le modèle avec probablement un nombre important de paramètres et ce, même si l'on a recours à la technologie récente. Les difficultés de calculs (informatiques), de biais et de non-convergence des estimateurs du maximum de vraisemblance quand T (la taille du panel) est petit, ont rendu le modèle à effets fixes contestable. Greene (2001) montre comment surmonter facilement les problèmes de calcul, mais cela requiert des modifications des algorithmes utilisés dans les logiciels existants.

La loi de Poisson est le modèle de base pour l'analyse des données de comptage, bien qu'il se révèle souvent inadéquat[3]. En effet, la régression de Poisson est souvent très restrictive. Un certain nombre d'alternatives sont donc suggérées, qui diffèrent selon les causes de la surdispersion. Dans les données de comptage, la surdispersion peut résulter de l'hétérogénéité inobservée. Dans de tels cas, une approche de mélange (« mixture models »), dans laquelle le paramètre de Poisson est une variable aléatoire, mène généralement à l'utilisation du modèle binomial négatif. En bref, les hypothèses relatives à l'hétérogénéité inobservée et à la dépendance des événements mènent au modèle binomial négatif (Cameron et Trivedi, 2005).

Hausman, Hall et Griliches (1984) ont développé le modèle binomial négatif conventionnel pour des données de panel. Allison et Waterman (2002) ont, de leur côté, critiqué le travail de HHG en avançant que le modèle FENB conditionnel n'est pas une méthode à effets fixes « valide » car il ne contrôle pas tous les régresseurs

[3] Cameron et Trivedi. 2005. « Microeconometrics Methods and Applications ».

fixes et invariants dans le temps. Toujours selon Allison et Waterman, HHG ont introduit l'effet fixe dans la partie hétérogène du modèle et non pas dans la moyenne conditionnelle. Cette partie est de fait conditionnée hors de la distribution pour produire le modèle de HHG. Allison et Waterman (2002) ont développé le modèle binomial négatif inconditionnel, qui inclut des variables binaires, pour représenter les effets fixes; ce modèle contrôle effectivement pour tous les effets individuels stables. Toutefois, les estimateurs des paramètres ne sont pas convergents en utilisant l'approche de variables binaires dans de courts panels; cela est dû au problème de paramètres incidents. Le modèle de Poisson à effets fixes, quant à lui, n'amène pas de problème de paramètres incidents.

CHAPITRE II

LES MODÈLES DE BASE DE RÉGRESSION DE POISSON ET BINOMIAL NÉGATIF

Les modèles de base de données de comptage sont les modèles de Poisson et binomial négatif. Ce dernier est le choix standard parmi ces modèles en raison de l'hypothèse d'équidispersion qui caractérise la distribution de Poisson.

2.1 Le modèle de régression de Poisson[4]

La distribution de Poisson est l'hypothèse de base de plusieurs modélisations des données de comptage en économétrie.

La régression de Poisson découle du modèle suivant :

$$P(Y = Y_i | X_i) = \frac{e^{-\mu_i} \mu_i^{Y_i}}{Y_i!}, \quad Y = 0,1,... \quad i = 1,..., N, \tag{1}$$

$$\mu_i = e^{X_i \beta}, \tag{2}$$

où P est la probabilité, Y_i est une variable de comptage observée (un nombre d'événements) pour l'individu i, X_i est un vecteur de K variables explicatives linéairement indépendantes observées pour l'individu i, et β est un vecteur de paramètres de dimension appropriée K × 1. La forme de la fonction exponentielle

[4] Greene, William. Avril 2007. « Functional Form and Heterogeneity in Models for Count Data », Greene, William. Mai 2007. « Fixed and Random Effects Models for Count Data ».

assure la non-négativité du paramètre de la moyenne μ. La moyenne et la variance conditionnelles du modèle de Poisson sont égales au paramètre μ_i.

$$E[Y_i | X_i] = Var[Y_i | X_i] = \mu_i. \tag{3}$$

La fonction log-vraisemblance du modèle est donnée par l'équation suivante :

$$\ln L(\beta | Y_i) = \sum_{i=1}^{N} [-\mu_i + Y_i(X_i\beta) - \ln\Gamma(1+Y_i)]. \tag{4}$$

où Γ est la fonction de Gamma.

Les paramètres sont choisis de façon à maximiser la valeur de la fonction log-vraisemblance. Les conditions de premier ordre sont :

$$\frac{\partial \ln L}{\partial \beta_j} = \sum_{i=1}^{N} \left(Y_i - e^{X_i\beta}\right) X_i = 0, \qquad j = 1,2,\ldots,K. \tag{5}$$

Une fois les paramètres connus, les effets marginaux dans le modèle de Poisson sont calculés ainsi :

$$\frac{\partial E(Y_i | X_i)}{\partial \beta_j} = \mu_i \beta.$$

Le modèle de régression de Poisson est trop restrictif pour les données de comptage, ce qui a incité les économètres à recourir à des modèles alternatifs comme le modèle binomial négatif, qui permet la surdispersion.

2.2 La surdispersion

Dans le modèle de Poisson, la surdispersion se produit lorsque la variance est supérieure à la moyenne. Elle est considérée comme un problème parce que son omission peut entraîner une sous-estimation des écarts-types des estimateurs; une variable peut ainsi apparaître à tort significative. La surdispersion est causée par la corrélation positive entre les observations ou par un excès de variation entre lesdites observations; elle se produit également lorsque des hypothèses de la distribution des données sont violées. On fait aussi référence à la surdispersion apparente, mais cette dernière se manifeste en cas de changements dans le modèle, comme une omission de variables explicatives importantes, une spécification non adéquate de la fonction de lien pour le modèle, etc. (Hilbe, 2007).

Dans les données de comptage, la variance excède généralement la moyenne, ce qui constitue un rejet du modèle de Poisson. Il est donc important de contrôler pour la surdispersion. Dans le cas de données tronquées ou censurées, la surdispersion mène au problème de non-convergence. Dans des processus de données simples, la surdispersion conduit à une sous-estimation des écarts-types ainsi qu'à une surestimation de la statistique t-Student par la méthode du maximum de vraisemblance. Par conséquent, l'utilisation de l'estimation robuste de la variance est utile (Cameron et Trivedi, 2005).

D'après Hilbe (2007), lorsqu'une surdispersion est suspectée, il faut d'abord déterminer s'il s'agit d'une possibilité de surdispersion apparente, que l'on pourra corriger ou non, selon le cas qui se présente, en ajoutant des variables explicatives ou en utilisant la fonction de lien adéquate, etc. Toutefois, si la surdispersion persiste, différentes méthodes peuvent être employées, chacune se fondant sur la raison de ce problème. Parmi celles-ci, on trouve la méthode « Scale Standard Error Post Hoc », la statistique de déviance et la statistique de Khi-deux, les estimateurs de matrices variance-covariance robustes, la méthode de Bootstrap ou Jackknife pour calculer les variances, les modèles binomial négatif (NB), NB hétérogène, NB-P (NB de type-P), Poisson généralisé, GEE, de même que les modèles inconditionnel et conditionnel à effets fixes ou encore le modèle à effets aléatoires (Hilbe, 2007).

2.3 La famille des modèles binomiaux négatifs[5]

Le modèle de régression binomial négatif a l'utilité de traiter la surdispersion des données de comptage. Ce modèle est construit tel un modèle de mélange qui est utile pour ajuster la surdispersion de la distribution de Poisson. La vraisemblance du modèle binomial négatif est fondée sur le modèle de mélange Poisson-gamma. On distingue deux types de modèle de régression : NB1 (surdispersion constante) et NB2 (surdispersion variable).

2.3.1 Modèle binomial négatif NB1 à surdispersion constante

Comme auparavant, la fonction de probabilité de Poisson est donnée par l'équation :

$$f(Y_i) = e^{-\mu_i} \frac{\mu_i^{Y_i}}{Y_i!}.$$

Le paramètre μ_i est une variable aléatoire suivant une distribution gamma $\Gamma(\gamma_i, \delta)$, où $\gamma_i = e^{X_i\beta+\alpha}$.

La moyenne et la variance du paramètre de Poisson sont alors données par:

$$E(\mu_i) = \gamma_i / \delta = e^{X_i\beta+\alpha} / \delta, \tag{6}$$

[5] Hardin James W. and Joseph M. Hilbe. January 2007. «Generalized Linear Models and Extensions».

$$V(\mu_i) = \gamma_i / \delta^2 = e^{X_i\beta+\alpha} / \delta^2 .$$

La distribution du mélange Poisson-gamma se calcule comme suit :

$$\begin{aligned}
f(Y_i|X_i) &= \int_0^\infty e^{-\mu_i} \frac{\mu_i^{Y_i}}{Y_i!} \frac{\delta^{\gamma_i}}{\Gamma(\gamma_i)} e^{-\mu_i\delta} \mu_i^{\gamma_i - 1} d\gamma_i \\
&= \frac{\delta^{\gamma_i}}{\Gamma(Y_i+1)\Gamma(\gamma_i)} \int_0^\infty e^{-\mu_i(1+\delta)} \mu_i^{(Y_i+\gamma_i)-1} d\gamma_i \\
&= \frac{\delta^{\gamma_i}}{\Gamma(Y_i+1)\Gamma(\gamma_i)} \frac{\Gamma(Y_i+\gamma_i)}{(\delta+1)^{Y_i+\gamma_i}} \\
&= \frac{\Gamma(Y_i+\gamma_i)}{\Gamma(Y_i+1)\Gamma(\gamma_i)} \left(\frac{\delta}{\delta+1}\right)^{\gamma_i} \left(\frac{1}{\delta+1}\right)^{Y_i} \quad Y_i = 0,1,\ldots \quad i = 1,\ldots,N. \qquad (7)
\end{aligned}$$

Nous obtenons une forme de distribution binomiale négative (appelée NB1) dont les deux premiers moments sont donnés par les deux équations suivantes:

$$\begin{aligned}
E(Y_i|X_i) &= \gamma_i/\delta = e^{X_i\beta+\alpha}/\delta , \\
V(Y_i|X_i) &= \gamma_i(1+\delta)/\delta^2 = e^{X_i\beta+\alpha}(1+\delta)/\delta^2 .
\end{aligned} \qquad (8)$$

Le ratio de la surdispersion (variance divisée par la moyenne) vaut $\frac{(1+\delta)}{\delta} > 1$. Il est constant pour toutes les observations, ce qui traduit ainsi une surdispersion constante. Ce modèle, appelé NB1, autorise la surdispersion avec un cas limite du modèle de Poisson lorsque δ tend vers l'infini.

2.3.2 Modèle binomial négatif NB2 à surdispersion variable

Le modèle binomial négatif de type 2 (appelé NB2) est généralement construit à partir du mélange du modèle de Poisson avec une hétérogénéité gamma, où l'erreur de gamma a une moyenne égale à 1.

Ce modèle est appliqué en introduisant une hétérogénéité (un effet individuel non observé) dans l'espérance conditionnelle de Poisson qui devient alors :

$$E(Y_i|X_i,\varepsilon_i)=e^{X_i\beta+\alpha+\varepsilon_i}=\mu_i u_i, \tag{9}$$

où $$u_i=e^{\varepsilon_i},$$

La densité de Poisson conditionnelle à u_i est la suivante :

$$f(Y_i|u_i)=e^{-\mu_i u_i}\frac{(\mu_i u_i)^{Y_i}}{Y_i!}, \tag{10}$$

Sa moyenne et sa variance conditionnelle sont égales à $\mu_i u_i$. ($u_i=e^{\varepsilon_i}$) est supposé suivre une distribution gamma $\Gamma(\vartheta,\vartheta)$ avec $E(u_i)=1$ et $V(u_i)=1/\vartheta$. La fonction de densité de u_i est égale à :

$$f(u_i)=\frac{\vartheta^{\vartheta}\exp(-\vartheta h_i)h_i^{\vartheta-1}}{\Gamma(\vartheta)},\quad u_i\geq 0, \vartheta>0. \tag{11}$$

En intégrant sur u_i le mélange des densités de Poisson et gamma, on trouve :

$$
\begin{aligned}
P(Y_i|X_i) &= \int_0^{\infty} \frac{e^{-\mu_i u_i}(\mu_i u_i)^{Y_i}}{\Gamma(1+Y_i)} \frac{\vartheta^{\vartheta} e^{-\vartheta u_i} u_i^{\vartheta-1}}{\Gamma(\vartheta)} du_i, \\
&= \frac{\vartheta^{\vartheta}\mu_i^{Y_i}}{\Gamma(1+Y_i)\Gamma(\vartheta)} \int_0^{\infty} e^{-u_i(\mu_i+\vartheta)} u_i^{(\theta+Y_i)-1} du_i, \\
&= \frac{\vartheta^{\vartheta}\mu_i^{Y_i}}{\Gamma(1+Y_i)\Gamma(\vartheta)} \frac{\Gamma(\vartheta+Y_i)}{(\mu_i+\vartheta)^{\vartheta+Y_i}}, \qquad (12) \\
&= \frac{\Gamma(\vartheta+Y_i)}{\Gamma(1+Y_i)\Gamma(\vartheta)} \left(\frac{\vartheta}{\mu_i+\vartheta}\right)^{\vartheta} \left(\frac{\mu_i}{\mu_i+\vartheta}\right)^{Y_i} \quad Y_i = 0,1,\ldots \quad i=1,\ldots,N \quad \vartheta > 0.
\end{aligned}
$$

Nous obtenons donc une autre forme de la distribution binomiale négative (appelée NB2), dont la moyenne et la variance sont respectivement données par les deux équations suivantes :

$$
\begin{aligned}
E(Y_i|X_i) &= \mu_i = e^{X_i\beta+\alpha}, \\
V(Y_i|X_i) &= \mu_i\left(1+\frac{1}{\vartheta}\mu_i\right).
\end{aligned} \qquad (13)
$$

Le ratio de surdispersion est égal à $\left(1+\frac{1}{\vartheta}\mu_i\right)$. Il varie selon les individus, ce qui prouve une surdispersion variable. Le cas limite de ce modèle NB2 est le modèle de Poisson lorsque ϑ tend vers l'infini.

Les effets marginaux sont calculés ainsi :

$$
\frac{\partial E(Y_i|X_i)}{\partial X_i} = \mu_i\beta.
$$

Cameron et Trivedi suggèrent une notation des formes de la distribution binomiale négative suivant la forme de la variance. Ainsi, le modèle binomial négatif de type 1 noté NB1 a la même espérance conditionnelle μ_i que le modèle NB2, mais la variance conditionnelle est $Var(Y_i|X_i) = \mu_i\left(1+\left(\frac{1}{\vartheta}\right)\right)$. Ce modèle NB1 est obtenu en remplaçant ϑ par $\vartheta\mu_i$ dans la densité de NB2 dans l'équation (12).

Enfin, Cameron et Trivedi ont considéré une classe plus générale de distributions binomiales négatives comportant la même moyenne, mais dont la variance est caractérisée par un exposant p; $\mu_i + \left(\frac{1}{\vartheta}\right)\mu_i^p$. Ce genre de distribution peut être produit avec un facteur d'hétérogénéité suivant une distribution gamma avec les deux paramètres égaux à $\vartheta\mu_i^{2-p}$; $\Gamma\left(\vartheta\mu_i^{2-p}, \vartheta\mu_i^{2-p}\right)$. Si p = 1, c'est le modèle NB1. Si p = 2, c'est le modèle NB2.

À noter que la loi de Poisson est le cas limite des distributions binomiales négatives quand le paramètre ϑ tend vers l'infini.

CHAPITRE III

LES MODÈLES DE RÉGRESSION DE POISSON ET BINOMIAL NÉGATIF EN PANEL

Les modèles familiers des données de panel sont les modèles à effets fixes et effets aléatoires. Dans notre étude, nous ne considérons que les modèles à effets fixes. Dans le cas d'un panel, l'indice i (relatif à l'individu i) est ajouté pour la constante; il en va évidemment de même pour les deux indices i (individus) et t (temps), ajoutés pour les variables X, Y, et d'autres paramètres.

3.1 Le modèle de Poisson à effets fixes[6]

Le modèle est spécifié en incluant l'effet fixe (variable indicatrice) dans l'espérance conditionnelle. Ainsi, la loi pour Y_{it} devient :

$$P(Y_{it}|X_{it}) = \frac{e^{-\mu_{it}}\mu_{it}^{Y_{it}}}{\Gamma(1+Y_{it})}, \tag{14}$$

et

$$\ln \mu_{it} = \alpha_i + X_{it}\beta .$$

[6] Greene, William. Avril 2007. « Functional Form and Heterogeneity in Models for Count Data »,
Greene, William, Mai 2007. « Fixed and Random Effects Models for Count Data »,
Dionne Georges et Charles Vanasse. 1996. « Une évaluation empirique de la nouvelle tarification de l'assurance automobile (1992) au Québec ».

où α_i est l'effet fixe (effet individuel spécifique), X_{it} est un vecteur de variables explicatives observées au temps t pour l'individu i, et β est un vecteur de paramètres de dimension appropriée.

La maximisation de la fonction log-vraisemblance inconditionnelle équivaut à la maximisation de :

$$\ln L = \sum_{i=1}^{N} \sum_{t=1}^{T} \ln P(Y_{it} | X_{it}), \qquad \text{où } i = 1,\dots,N. \tag{15}$$

Elle produit une estimation de (α_i, β), $i = 1,\dots,N$ identique à la maximisation de la fonction log-vraisemblance conditionnelle à $\sum_{t=1}^{T} Y_{it}$.

Du fait que les N paramètres α_i ne peuvent être estimés directement, puisque T est petit et N est grand, Hausman, Hall et Griliches (1984) proposent un estimateur de maximum de vraisemblance conditionnelle pour le modèle de Poisson à effets fixes. Celui-ci est conditionnel à $\sum_{t=1}^{T} Y_{it}$ et provient de la vraisemblance conditionnelle pour un individu i :

$$P\left(Y_{i1}, Y_{i2}, \dots, Y_{iT} \middle| \sum_{t=1}^{T} Y_{it} \right) = \frac{P(Y_{i1}, Y_{i2}, \dots, Y_{it})}{P\left(\sum_{t=1}^{T} Y_{it} \right)}. \tag{16}$$

Si Y_{it} suit une loi de Poisson de paramètre μ_{it}, alors $\sum_{t=1}^{T} Y_{it}$ suit une loi de Poisson de paramètre $\sum_{t=1}^{T} \mu_{it}$,

$$P\left(Y_{i1},Y_{i2},\ldots,Y_{iT}\left|\sum_{t=1}^{T}Y_{it}\right.\right)=\frac{\prod_{T}\frac{e^{-\mu_{it}}\mu_{it}^{Y_{it}}}{Y_{it}!}}{\frac{e^{-\sum_t \mu_{it}}\left(\sum_t \mu_{it}\right)^{-\sum_t Y_{it}}}{\left(\sum_t Y_{it}\right)!}}=\frac{\left[\sum_t Y_{it}\right]!}{\prod_T Y_{it}!}\prod_T\left[\frac{\mu_{it}}{\sum_t \mu_{it}}\right]^{Y_{it}}. \quad (17)$$

Étant donné que $\mu_{it} = e^{\alpha_i + X_{it}\beta}$, on aura $\sum_{t=1}^{T}\mu_{it} = e^{\alpha_i}.\left[\sum_{t=1}^{T}e^{X_{it}\beta}\right]$.

D'où l'équation suivante :

$$P\left(Y_{i1},\ldots,Y_{iT}\left|\sum_{t=1}^{T}Y_{it},X_i\right.\right)=\frac{\Gamma\left[\left(\sum_{i=1}^{T}Y_{it}\right)+1\right]}{\prod_{t=1}^{T}\Gamma\left(Y_{it}+1\right)}\prod_{t=1}^{T}\left(\frac{e^{X_{it}\beta}}{\sum_{t=1}^{T}e^{X_{it}\beta}}\right)^{Y_{it}}. \quad (18)$$

On remarque que la fonction log-vraisemblance n'introduit pas la constante α_i.

En bref, le modèle de Poisson est une méthode à effets fixes qui ne présente pas de problèmes connus dans l'estimation – mis à part la contrainte d'équidispersion –, contrairement au modèle binomial négatif que nous allons analyser dans la section suivante.

3.2 Les modèles binomiaux négatifs à effets fixes (FENB)

3.2.1 Le modèle NEGBIN P[7]

Selon la notation de Cameron et Trivedi (2005), en ajoutant les indices *i* et *t* de telle façon que l'hétérogénéité $u_i = e^{\varepsilon_i}$ soit supposée suivre une distribution Gamma $\Gamma(\vartheta_i, \vartheta_i)$, la spécification du modèle NB2 dans l'équation (12) à effets fixes est alors donnée par :

$$P(Y_{it} | \mu_{it}, \vartheta_i) = \frac{\Gamma(\vartheta_i + Y_{it})}{\Gamma(1 + Y_{it})\Gamma(\vartheta_i)} \left(\frac{\vartheta_i}{\vartheta_i + \mu_{it}} \right)^{\vartheta_i} \left(\frac{\mu_{it}}{\vartheta_i + \mu_{it}} \right)^{Y_{it}}, \qquad Y_{it} = 0,1,... \quad (19)$$

et l'on obtient :

$$E(Y_{it} | X_{it}) = \mu_{it},$$

$$Var(Y_{it} | X_{it}) = \mu_{it} \left(1 + \left(\frac{1}{\vartheta_i} \right) \mu_{it} \right).$$

Si l'on remplace ϑ_i par $\vartheta_i \mu_{it}$, le modèle obtenu donne la densité du modèle NB1 à effets fixes :

$$P[Y_{it} | \mu_{it}, \vartheta_i] = \frac{\Gamma(\vartheta_i \mu_{it} + Y_{it})}{\Gamma(1 + Y_{it})\Gamma(\vartheta_i \mu_{it})} \left(\frac{1}{1 + \vartheta_i} \right)^{\vartheta_i \mu_{it}} \left(\frac{\vartheta_i}{1 + \vartheta_i} \right)^{Y_{it}}, \qquad Y_{it} = 0,1,... \quad (20)$$

On arrive alors aux équations suivantes :

[7] Greene, William. Avril 2007. « Functional Form and Heterogeneity in Models for Count Data ».

$$E\left(Y_{it}\middle|X_{it}\right)=\mu_{it},$$

$$Var\left(Y_{it}\middle|X_{it}\right)=\mu_{it}\left(1+\left(\frac{1}{\vartheta_i}\right)\right).$$

Note : Nous remplaçons ϑ_i par 1/δ et $\vartheta_i\mu_{it}$ par γ_{it} afin d'obtenir la densité du modèle NB1 dans l'équation (7).

Les auteurs notent que d'autres exposants seraient possibles; le remplacement de ϑ_i par $\vartheta_i\mu_{it}^{2-P}$ nous permettrait d'obtenir le modèle communément appelé NEGBIN P, ou NBP :

$$P\left[Y_{it}\middle|\mu_{it},\vartheta_i\right]=\frac{\Gamma\left(\vartheta_i\mu_{it}^{2-P}+Y_{it}\right)}{\Gamma\left(1+Y_{it}\right)\Gamma\left(\vartheta_i\mu_{it}^{2-P}\right)}\left(\frac{\mu_{it}}{\mu_{it}+\vartheta_i\mu_{it}^{2-P}}\right)^{\vartheta_i\mu_{it}^{2-P}}\left(\frac{\vartheta_i\mu_{it}^{2-P}}{\mu_{it}+\vartheta_i\mu_{it}^{2-P}}\right)^{Y_{it}},Y_i=0,1,.....\quad(21)$$

On obtient alors :

$$E\left(Y_{it}\middle|X_{it}\right)=\mu_{it},$$

$$Var\left(Y_i\middle|X_i\right)=\mu_i+\frac{1}{\vartheta_i}\mu_i^P=\mu_i\left[1+\frac{1}{\vartheta_i}\mu_i^{P-1}\right].$$

Les effets marginaux sont calculés ainsi :

$$\frac{\partial E\left(Y_i\middle|X_i\right)}{\partial X_i}=\mu_i\beta.$$

Les modèles NB1 et NB2 sont des cas spéciaux correspondant respectivement à $p = 1$ et $p = 2$.

3.2.2 Le modèle binomial négatif HHG et FENB[8]

Le modèle binomial négatif NB1 a été proposé par Hausman, Hall et Griliches (1984). C'est une généralisation de la distribution de Poisson avec un paramètre additionnel permettant à la variance d'excéder la moyenne. Différentes méthodes existent pour formuler la distribution binomiale négative; le choix de celle à employer peut avoir des conséquences sur la régression (Allison et Waterman, 2002).

Dans le modèle HHG, la fonction de densité de la distribution binomiale négative est donnée par l'équation suivante:

$$f(Y_{it}|\gamma_{it},\theta_i)=\frac{\Gamma(\gamma_{it}+Y_{it})}{\Gamma(1+Y_{it})\Gamma(\gamma_{it})}\left(\frac{1}{1+\theta_i}\right)^{\gamma_{it}}\left(\frac{\theta_i}{1+\theta_i}\right)^{Y_{it}}. \tag{22}$$

Note : Il s'agit du même modèle NB1 que dans l'équation (7); δ a été remplacé par $1/\theta_i$.

Le paramètre θ_i est supposé constant dans le temps pour chaque individu, alors que γ_{it} dépend des variables explicatives car $\ln\gamma_{it}=X_{it}\beta$. On remarque que cet effet fixe n'est pas introduit dans la fonction de régression.

Le paramètre $\delta=1/\theta_i$ est généralement considéré comme un paramètre de surdispersion car l'équation (22) devient une fonction de Poisson lorsque $\delta\rightarrow\infty$.

La moyenne et la variance de Y_{it} sont respectivement données par les deux équations suivantes :

[8] Allison, Paul et Waterman, Richard. 2002. « Fixed Effects Negative Binomial Regression Models », Hausman, J. A., B. H. Hall et Z. Griliches. 1984. « Econometric Models For Count Data with an Application to the Patents - R and D Relationship ».

$$
\begin{aligned}
E(Y_{it}) &= \theta_i \gamma_{it}, \\
Var(Y_{it}) &= (1+\theta_i)\theta_i \gamma_{it}.
\end{aligned} \tag{23}
$$

On suppose que pour un individu i donné, Y_{it} est indépendant dans le temps. Cette hypothèse implique que $\sum_{t=1}^{T} Y_{it}$ suit une loi binomiale négative avec les paramètres θ_i et $\sum_{t=1}^{T} \gamma_{it}$. En conditionnant sur $\sum_{t=1}^{T} y_{it}$, la fonction de vraisemblance pour un seul individu est donnée dans l'équation suivante par Allison et Waterman (2002):

$$
P\left(Y_{iT} \middle| \sum_{t=1}^{T} y_{it}, X_i \right) = \frac{\Gamma\left[\left(\sum_{i=1}^{T} Y_{it}\right)+1\right]\Gamma\left(\sum_{t=1}^{T}\gamma_{it}\right)}{\Gamma\left(\sum_{t=1}^{T} Y_{it} + \sum_{t=1}^{T}\gamma_{it}\right)} \prod_{t=1}^{T} \frac{\Gamma(Y_{it}+\gamma_{it})}{\Gamma(Y_{it}+1)\Gamma(\gamma_{it})}. \tag{24}
$$

Cela représente la formulation du modèle FENB conditionnel rapportée par HHG. Remarquons que le paramètre θ_i (ou les effets fixes), ne figurent pas dans cette fonction de vraisemblance. C'est la formulation utilisée dans des progiciels connus comme Stata, SAS et LIMDEP. Le ratio de la surdispersion est égal à $1+\theta_i > 1$. Ce ratio est variable à travers les individus i. Le modèle de HHG permet les deux types de surdispersion que les spécifications des modèles de Poisson à effets fixes et NB1 ne permettent pas.

Le modèle FENB permet l'introduction d'une constante dans l'estimation. L'explication est que le modèle HHG n'est pas obtenu en indiquant les effets fixes dans l'espérance conditionnelle $\mu_{it} = \exp(\alpha_i + X_{it}\beta)$ comme dans le cas du modèle de Poisson. Le modèle FENB est obtenu en construisant les effets fixes dans le modèle tels des effets spécifiques individuels θ_i dans le modèle NB1 dans l'équation (22)[9].

Les deux types de modèle binomial négatif ont les espérances et variances conditionnelles suivantes :

Pour NB1(HHG) : $E(Y_{it}|X_{it}) = \theta_i \gamma_{it} = \theta_i e^{X_{it}\beta}$,

$$Var(Y_{it}|X_{it}) = \theta_i \gamma_{it}(1+\theta_i). \tag{25}$$

Pour NB2 : $E(Y_{it}|X_{it}) = \mu_{it} = e^{\alpha_i + X_{it}\beta}$,

$$Var(Y_{it}|X_{it}) = \mu_{it}\left(1+\left(\frac{1}{\vartheta_i}\right)\mu_{it}\right). \tag{26}$$

Note : La vraisemblance conditionnelle du modèle NB2 ne peut être calculée si l'on suppose l'indépendance dans le temps des observations pour chaque individu. Cela est dû à l'inexistence d'une statistique exhaustive de l'effet fixe qui dépend seulement des données (voir plus loin dans la section-V sur les alternatives proposées).

3.3 Fonctions de vraisemblance conditionnelle et inconditionnelle

En théorie, l'hypothèse de base pour construire la vraisemblance d'un modèle est l'indépendance des observations. Cependant, cette hypothèse n'est pas réaliste dans le cas des études longitudinales (Hilbe, 2007). Nous dressons dans ce qui suit les fonctions de vraisemblances conditionnelles et inconditionnelles des modèles de Poisson et binomial négatif.

---

[9] Greene, William. Avril 2007. « Functional Form and Heterogeneity in Models for Count Data ».

3.3.1 Les modèles binomial négatif et Poisson inconditionnels à effets fixes[10]

L'estimation inconditionnelle des modèles de Poisson et FENB peut être obtenue par la méthode des modèles linéaires généralisés (GLM). La fonction log-vraisemblance inconditionnelle du modèle de Poisson à effets fixes équivaut à :

$$\ln L(\beta|Y_i) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[-\exp(\alpha_i + X_{it}\beta) + Y_{it}(\alpha_i + X_{it}\beta) - \ln\Gamma(1+Y_{it})\right] \qquad (27)$$

Le modèle binomial négatif est utilisé pour prendre en considération la surdispersion absente dans le modèle de Poisson.

Le log-vraisemblance non conditionnelle du modèle FENB est donnée par les équations ci-après :

$$\text{NB2 : } \ln L(\beta,\vartheta_i,\mu_{it}|Y_i) = \ln\prod_{i=1}^{N}\prod_{t=1}^{T}\left[\frac{\Gamma(\vartheta_i + Y_{it})}{\Gamma(1+Y_{it})\Gamma(\vartheta_i)}\left(\frac{\vartheta_i}{\vartheta_i+\mu_{it}}\right)^{\vartheta_i}\left(\frac{\mu_{it}}{\vartheta_i+\mu_{it}}\right)^{Y_{it}}\right], \qquad (28)$$

$$\text{NB1 : } \ln L(\beta,\vartheta_i,\mu_{it}|Y_i) = \ln\prod_{i=1}^{N}\prod_{t=1}^{T}\left[\frac{\Gamma(\vartheta_i\mu_{it} + Y_{it})}{\Gamma(1+Y_{it})\Gamma(\vartheta_i\mu_{it})}\left(\frac{1}{1+\vartheta_i}\right)^{\vartheta_i\mu_{it}}\left(\frac{\vartheta_i}{1+\vartheta_i}\right)^{Y_{it}}\right], \qquad (29)$$

$$\text{NB1-HHG : } \ln L(\beta,\vartheta_i,\mu_{it}|Y_i) = \ln\prod_{i=1}^{N}\prod_{t=1}^{T}\left[\frac{\Gamma(\gamma_{it} + Y_{it})}{\Gamma(1+Y_{it})\Gamma(\gamma_{it})}\left(\frac{1}{1+\theta_i}\right)^{\gamma_{it}}\left(\frac{\theta_i}{1+\theta_i}\right)^{Y_{it}}\right] \qquad (30)$$

Le problème de paramètres incidents est fondamental pour le modèle binomial négatif inconditionnel à effets fixes : il y a en effet risque de biais et de non-convergence pour les estimateurs des paramètres du modèle lorsque le nombre

[10] Hilbe, Joseph M. 2007. Negative Binomial Regression.

d'individus est grand et que la taille T du panel est petite. Par conséquent, ce modèle binomial négatif inconditionnel à effets fixes est utilisé seulement lorsque le nombre d'individus est petit. S'il y a plus de vingt individus, il est suggéré d'utiliser le modèle binomial négatif conditionnel à effets fixes. Le modèle de Poisson ne présente pas de problème de paramètres incidents car l'estimateur conditionnel de l'effet fixe est égal à l'estimateur inconditionnel (Hilbe, 2007).

Rappelons par ailleurs que l'effet fixe du modèle de Poisson est introduit dans le paramètre de la moyenne alors que le modèle binomial négatif construit l'effet fixe comme une hétérogénéité introduite dans la distribution du Gamma, ce qui rend difficile à interpréter le problème de paramètres incidents du modèle binomial négatif. Cela a aussi pour conséquence que les estimateurs seront non convergents en présence d'un grand nombre d'effet fixes; cependant, la raison de la non-convergence demeure toujours un sujet de discussion (Hilbe, 2007).

3.3.2 Les modèles binomial négatif et Poisson conditionnels à effets fixes[11]

La fonction log-vraisemblance est conditionnelle à la somme totale de Y_i pour chaque panel (dans le temps), $\sum_{t=1}^{T} Y_{it}$.

La fonction log-vraisemblance conditionnelle du modèle de Poisson à effets fixes est donnée par :

[11] Hilbe, Joseph M. 2007. Negative Binomial Regression.

$$\ln L(\beta,\mu_{it}|Y_i)=\ln\prod_{i=1}^{N}\left\{\frac{\Gamma\left[\left(\sum_{i=1}^{T}Y_{it}\right)+1\right]}{\prod_{t=1}^{T}\Gamma(Y_{it}+1)}\prod_{t=1}^{T}\left(\frac{\exp(X_{it}\beta)}{\sum_{t=1}^{T}\exp(X_{it}\beta)}\right)^{Y_{it}}\right\}, \tag{31}$$

$$=\sum_{i=1}^{N}\left\{\ln\Gamma\left[\left(\sum_{i=1}^{T}Y_{it}\right)+1\right]-\sum_{t=1}^{T}\ln\Gamma(Y_{it}+1)+\sum_{t=1}^{T}Y_{it}\left(X_{it}\beta-\ln\left[\sum_{t=1}^{T}\exp(X_{it}\beta)\right]\right)\right\}.$$

La fonction log-vraisemblance conditionnelle du modèle FENB-HHG est égale à :

$$\ln L(\beta,\mu_{it}|Y_i)=\ln\prod_{i=1}^{N}\left\{\frac{\Gamma\left[\left(\sum_{i=1}^{T}Y_{it}\right)+1\right]\Gamma\left(\sum_{t=1}^{T}\gamma_{it}\right)}{\Gamma\left(\sum_{t=1}^{T}Y_{it}+\sum_{t=1}^{T}\gamma_{it}\right)}\prod_{t=1}^{T}\frac{\Gamma(Y_{it}+\gamma_{it})}{\Gamma(Y_{it}+1)\Gamma(\gamma_{it})}\right\}, \tag{32}$$

$$=\sum_{i=1}^{N}\left\{\begin{array}{l}\ln\Gamma\left[\left(\sum_{i=1}^{T}Y_{it}\right)+1\right]+\ln\Gamma\left(\sum_{t=1}^{T}\exp(X_{it}\beta)\right)-\ln\Gamma\left(\sum_{t=1}^{T}Y_{it}+\sum_{t=1}^{T}\exp(X_{it}\beta)\right)+\\ \sum_{t=1}^{T}\{\ln\Gamma(Y_{it}+\exp(X_{it}\beta))-\ln\Gamma(Y_{it}+1)-\ln\Gamma(\exp(X_{it}\beta))\}\end{array}\right\}.$$

Comme cela a bien été noté, le paramètre d'hétérogénéité n'apparaît pas dans cette fonction log-vraisemblance. Allison et Waterman (2002) ont remarqué que le modèle FENB proposé par HHG n'est pas « une vraie méthode à effets fixes ». En effet, il ne contrôle pas pour toutes les variables explicatives stables (invariantes dans le temps), (Guimarães, 2007). Le modèle multinomial négatif est proposé par Allison et Waterman (2002) comme alternative au modèle binomial négatif. Cependant, ce dernier produit les mêmes estimateurs que le modèle de Poisson conditionnel. D'autres alternatives consistent à se référer au modèle binomial négatif non conditionnel, mais elles nécessitent un ajustement de l'écart-type par la dispersion de Pearson Khi-deux ou la déviance. Si ces approches ne sont pas satisfaisantes, l'ultime recours est d'utiliser d'autres modèles de panels, soit les modèles à effets aléatoires ou les modèles GEE (Hilbe, 2007).

CHAPITRE IV

PROBLÈMES SOULEVÉS

Dans cette section, nous retraçons brièvement les problèmes rencontrés avec l'utilisation du modèle FENB. Nous présenterons deux raisons principales qui font de ce modèle un sujet de discussion et incitent par conséquence à la recherche d'autres alternatives. La non-validité de la méthode de vraisemblance conditionnelle du modèle FENB est le problème le plus marqué dans ce contexte. Le problème de paramètres incidents, quant à lui, est commun à plusieurs modèles à effets fixes.

4.1 Le modèle binomial négatif à effets fixes (FENB) n'est pas une vraie méthode à effets fixes[12]

Avec sa vraisemblance conditionnelle, le modèle FENB ne répond pas réellement au besoin de la méthode à effets fixes. Selon Allison et Waterman (2002), le problème de base est que dans la vraisemblance conditionnelle de HHG (équation (24)), le paramètre θ_i qui n'apparaît pas ne correspond pas aux constantes α_i (soit à l'effet fixe tel qu'il apparaît dans la régression du modèle de Poisson) dans la relation exponentielle de γ_{it} tel que $\gamma_{it} = \exp(X_{it}\beta)$ (Allison et Waterman 2002). Autrement dit, en suivant le raisonnement de HHG donné par Allison et Waterman (2002), si nous posons $\theta_i = \exp(\omega_i)$, les équations (22) et (23) impliquent que :

$$E(Y_{it}) = \exp(\omega_i + X_{it}\beta), \tag{33}$$

[12] Allison, Paul et Waterman Richard. 2002. « Fixed Effects Negative Binomial Regression Models ».

$$Var(Y_{it}) = (1 + e^{\omega_i})E(Y_{it}).$$

Par la suite, nous remarquons que le modèle permet une constante arbitraire ω_i pour chaque individu. Si ω_i est considéré comme étant la somme des effets représentant les variables explicatives omises, alors ces variables peuvent constituer des effets fixes additionnels.

En prenant l'exemple donné par Allison et Waterman (2002), on spécifie à partir de l'équation (22) que :

$$\gamma_{it} = \exp(\omega_i + X_{it}\beta + \lambda z_i),$$

où ω_i est une constante individuelle spécifique et z_i est un vecteur de régresseurs invariants dans le temps. La fonction de vraisemblance conditionnelle sur le total des événements pour chaque individu $\sum_{t=1}^{T} Y_{it}$, ne fait apparaître ni ω_i ni λz_i. Toujours selon Allison et Waterman (2002), si l'on utilise la vraisemblance conditionnelle de HHG dans l'équation (24), le problème est que nous pouvons estimer des modèles de régression avec une constante ou des régresseurs invariants dans le temps. Nous devons retenir ici que le modèle binomial négatif n'est pas une vraie méthode à effets fixes car il autorise d'autres effets fixes additionnels en plus des régresseurs invariants dans le temps.

4.2 Le problème de paramètres incidents dans les modèles à effets fixes[13]

Soit $\mu_{it} = \exp(\alpha_i + X_{it}\beta + \lambda z_i)$. S'il existe une statistique exhaustive[14] du paramètre de nuisance α_i, et si l'on conditionne la densité $f(Y_{it})$ sur cette statistique

[13] Cameron, A. Colin et Pravin K. Trivedi. 2005. « Microeconometrics Methods and Applications »,

exhaustive, le paramètre α_i sera éliminé. La densité conditionnelle dépendra seulement des paramètres communs β et λ, dont l'estimation est convergente.

Précédemment, nous avons vu que les effets fixes n'apparaissent pas dans la vraisemblance conditionnelle binomiale négative du moment que les vraisemblances sont conditionnelles à $\sum_{t=1}^{T} Y_{it}$, qui est exhaustif pour les termes d'effets fixes. Mais l'estimation des effets fixes $\alpha_1,...,\alpha_N$ et des autres paramètres du modèle β et λ mène généralement à une estimation non convergente de tous les paramètres dans le cas des panels courts. C'est ce que l'on appelle le problème de paramètres incidents.

Des approches ont été proposées. Ces approches éliminent les effets fixes par le biais de quelques ajustements et permettent ainsi une estimation convergente des autres paramètres du modèle (Cameron et Trivedi, 2005). β et λ sont des paramètres communs, mais $\alpha_1,...,\alpha_N$ sont des paramètres incidents si le panel est court; à ce moment-là, chaque α_i dépend en effet de T observations fixes et il existe infiniment plus de α_i lorsque N tend vers l'infini. Les paramètres incidents sont des estimateurs non convergents lorsque N tend vers l'infini, puisque seules les T observations sont utilisées pour estimer chaque paramètre. En fait, le problème de paramètres incidents induit la contamination de l'estimation des paramètres communs. En général, les paramètres communs sont également des estimateurs non convergents, quoiqu'ils soient finis en nombre et estimés en utilisant $NT \rightarrow \infty$ observations (Cameron et Trivedi, 2005).

[14] La statistique t est appelée une statistique exhaustive pour α si la distribution de Y conditionnelle à t ne dépend pas de α.

CHAPITRE V

ALTERNATIVES PROPOSÉES

Nous citons dans ce qui suit quelques méthodes et alternatives proposées suite aux problèmes soulevés par l'utilisation du modèle FENB.

5.1 Le modèle multinomial négatif[15]

Si l'on suppose, comme dans le cas de HHG, que les événements ou observations sont indépendants à travers le temps pour chaque individu, alors la vraisemblance conditionnelle du modèle NB2 de Cameron et Trivedi (1998) ne peut être calculée parce que $\sum_{t=1}^{T} Y_{it}$ ne suivra pas une distribution binomiale négative. Techniquement, il n'existe pas de statistique exhaustive du paramètre de l'effet fixe qui soit seulement fonction des données. Une autre approche consiste à supposer que Y_{it} suit une distribution multinomiale négative (Allison et Waterman, 2002), qui est la généralisation multivariée de la distribution binomiale négative. La fonction de densité pour un individu i est donnée par :

$$f\left(Y_{i1},\dots,Y_{iT}\middle|\theta_i,\mu_{i1},\dots,\mu_{iT}\right)=\frac{\Gamma\left(\theta_i+\sum_{t=1}^{T}Y_{it}\right)}{\Gamma(\theta_i)Y_{i1}!\dots Y_{iT}!}\left(\frac{\theta_i}{\sum_{t=1}^{T}\mu_{it}+\theta_i}\right)^{\theta_i}\prod_{t=1}^{T}\left(\frac{\mu_{it}}{\sum_{t=1}^{T}\mu_{it}+\theta_i}\right)^{Y_{it}}, \tag{34}$$

[15] Allison, Paul et Waterman Richard. 2002. « Fixed Effects Negative Binomial Regression Models ».

$$\ln \mu_{it} = \delta_i + X_{it}\beta.$$

En raison des propriétés de la distribution multivariée, la distribution marginale de chaque Y_{it} est la distribution binomiale négative NB2 définie dans l'équation (19). Ainsi, la distribution de $\sum_{t=1}^{T} Y_{it}$ est une binomiale négative avec les paramètres $\sum_{t=1}^{T} \mu_{it}$ et θ_i. Par opposition au modèle HHG, cette distribution ne suppose pas l'indépendance dans le temps des Y_{it} pour un individu donné.

La vraisemblance conditionnelle à $\sum_{t=1}^{T} Y_{it}$ est égale à :

$$\begin{aligned}
P(Y_i = y|\mathrm{x}_i) &= \int_{h_i} \Pr ob(Y = y_i|\mathrm{x}_i) f(h_i) dh_i, \\
&= \int_{h_i} \frac{\exp(-h_i\mu_i)(h_i\mu_i)^{y_i}}{\Gamma(1+y_i)} \frac{\theta^\theta \exp(-\theta h_i) h_i^{\theta-1}}{\Gamma(\theta)} dh_i, \\
&= \frac{\theta^\theta \mu_i^{y_i}}{\Gamma(1+y_i)\Gamma(\theta)} \int_{h_i} \exp(-h_i(\mu_i+\theta))(h_i)^{\theta+y_i-1} dh_i, \\
&= \frac{\theta^\theta \mu_i^{y_i}}{\Gamma(1+y_i)\Gamma(\theta)} \frac{\Gamma(\theta+y_i)}{(\mu_i+\theta)^{\theta+y_i}}, y = 0,1,\ldots \quad i = 1,\ldots,N. \quad \theta > 0, \qquad (35)
\end{aligned}$$

Elle est proportionnelle à la vraisemblance conditionnelle du modèle de Poisson dans l'équation (18), ce qui explique que le modèle multinomial négatif à effets fixes mène au même estimateur conditionnel de β que le modèle de Poisson à effets fixes (Allison et Waterman, 2002). Enfin, l'approche multinomiale négative n'accomplit rien en ce qui concerne la surdispersion.

En résumé, la distribution multinomiale négative peut être générée en mélangeant les deux distributions de la variable aléatoire gamma et des variables aléatoires indépendantes de Poisson. Par contre, la distribution binomiale négative est issue du mélange des distributions des variables aléatoires de gamma et Poisson (Allison et Waterman, 2002). Le conditionnement sur le total des événements pour chaque individu supprime toute l'hétérogénéité inobservée comme les effets fixes et les hétérogénéités inobservées intrinsèques dans la distribution multinomiale négative.

5.2 L'approche conventionnelle à la surdispersion[16]

L'une des alternatives est de se référer au modèle de Poisson à effets fixes et d'estimer β en prenant garde d'ajuster l'écart-type des estimés à la surdispersion. Multiplier l'écart-type par la statistique de Pearson sur le degré de liberté (ou la déviance) est d'ailleurs souvent utilisé (Allison et Waterman, 2002).

Une autre approche consiste à estimer le modèle binomial négatif inconditionnel en spécifiant le modèle de régression NB2 avec des variables binaires[17] afin d'estimer les effets fixes. Selon Allison et Waterman (2002), d'après l'application sur les données de brevets, les coefficients estimés sont similaires à ceux obtenus avec la spécification de Poisson; toutefois, l'écart-type du modèle binomial négatif est plus grand que celui obtenu avec le modèle de Poisson (Allison et Waterman, 2002).

[16] Allison, Paul et Waterman Richard. 2002. « Fixed Effects Negative Binomial Regression Models ».

[17] Variables muettes ou indicatrices ou binaires (dummy variables). La variable muette prend une valeur égale à 1 dans le cas de certaines observations pour indiquer, par exemple, la présence d'un effet; sinon, elle prend la valeur 0.

Selon Allison et Waterman (2002), deux problèmes potentiels peuvent être rencontrés avec la méthode binomiale négative inconditionnelle. En premier lieu, tant que le problème de paramètres incidents persiste, il reste à déterminer si les coefficients estimés sont convergents. Jusqu'ici, aucune preuve ne va dans ce sens. En second lieu, l'estimation des coefficients d'un grand nombre de variables binaires devient pratiquement impossible dans le cas d'échantillons de grande taille.

5.3 Un estimateur conditionnel approximatif (« The Projected Score Method »)[18]

La vraisemblance conditionnelle n'est pas calculable pour le modèle NB2 du moment que les observations sont indépendantes dans le temps pour chaque individu et du fait qu'il n'existe pas de statistique exhaustive complète pour les paramètres incidents, qui dépend seulement des données. Waterman et Lindsay (1996a) introduisent une méthode approximative qui exploite les propriétés de l'inférence conditionnelle, et ce, même dans les cas de manque d'approches de traitement. Cette méthodologie est appelée la méthode projetée de score (*The Projected Score Method*)[19].

5.4 Le test de score pour le modèle binomial négatif à effets fixes FENB[20]

Guimarães (2007) réaffirme les résultats d'Allison et de Waterman (2002) selon lesquels le modèle conditionnel FENB n'est pas réellement un modèle à effets fixes (« True Fixed Effect »), et montre que le modèle contrôlera seulement différents

[18] Allison, Paul et Waterman Richard. 2002. « Fixed Effects Negative Binomial Regression Models ».

[19] Pour plus de détails, voir Richard Waterman et Bruce Lindsay. 1996. « Projected score methods for approximating conditional scores ».

[20] Guimarães, Paulo. 2007. « The Fixed Effects Negative Binomial Model Revisited ».

effets spécifiques sous un ensemble d'hypothèses particulières. Il propose également un test de score qui peut aider à déterminer si ces hypothèses sont vérifiées. En effet, il dérive un test de score qui peut être utile pour confirmer si les effets fixes ont été retirés après l'estimation par le maximum de vraisemblance conditionnel. Des simulations suggèrent que le test de score s'applique bien lorsque le nombre de périodes de temps n'est pas trop petit (Guimarães, 2007).

5.5 Synthèse des méthodes de panels

La méthode de vraisemblance conditionnelle du modèle binomial négatif proposée par Hausman, Hall et Griliches (1984) ne satisfait pas l'hypothèse de contrôle pour toutes les variables explicatives stables présentes dans la méthode à effets fixes. Cela vient du fait que l'effet fixe est introduit dans la partie hétérogène du modèle plutôt que dans la moyenne conditionnelle, comme observé dans le cas du modèle de Poisson. Allison et Waterman (2002) ont développé une alternative, soit le modèle inconditionnel binomial négatif, qui estime les effets fixes par l'introduction de variables binaires. Cependant, cette approche est confrontée au problème de paramètres incidents dans le cas de panels courts.

Le tableau 5 ci-après résume les particularités des modèles de Poisson et binomial négatif à effets fixes.

Le problème de paramètres incidents et la non-fiabilité de la méthode conditionnelle du modèle FENB sont encore sujets à discussion. Wooldridge (2006) propose une alternative qui permet de contrôler les effets fixes du modèle binomial négatif. Celle-ci repose sur la régression binomiale négative à effets aléatoires, mais elle introduit les moyennes des variables explicatives qui varient dans le temps pour représenter les effets fixes. Nous verrons dans le chapitre suivant l'application de cette méthode alternative de Wooldridge.

Tableau 5. Comparaison entre les modèles de Poisson et binomial négatif

Modèle de Poisson à effets fixes	Modèle FENB
- Propriété d'équidispersion. - Les estimateurs du modèle conditionnel sont égaux à ceux du modèle inconditionnel. - Pas de problème de paramètres incidents. - Contrôle pour les effets fixes.	- Propriété de surdispersion. - Les estimateurs du modèle conditionnel sont différents de ceux du modèle inconditionnel. - Problème de paramètres incidents pour le modèle inconditionnel. - Le modèle conditionnel ne contrôle pas pour les effets fixes et introduit des régresseurs stables.
Si le nombre d'individus est grand, le modèle inconditionnel est pratiquement impossible à estimer.	

CHAPITRE VI

APPLICATION À LA CONSOMMATION DE MÉDICAMENTS AU CANADA

6.1 Les données

L'enquête nationale sur la santé de la population (ENSP) concerne la population canadienne et a pour but de mesurer l'état et les déterminants de la santé. C'est une enquête longitudinale dans laquelle les données sont collectées auprès d'un même panel de répondants; elle est reprise tous les deux ans pendant plus d'une décennie. L'ENSP, établie sur sept cycles allant de 1994-1995 à 2006-2007, permet de collecter à la fois des données transversales sur un échantillon initial de plus de 20 000 ménages, et des données longitudinales sur un échantillon initial de 17 276 répondants longitudinaux choisis au cours du premier cycle. La population cible qui nous intéresse parmi les trois populations visées de l'ENSP est formée des membres des ménages de tous les territoires et provinces, à l'exception des résidents des réserves indiennes, des bases des forces armées du Canada et de certaines régions éloignées du Québec et de l'Ontario[21]. Le questionnaire de l'ENSP est établi selon deux composantes : une composante générale dans laquelle l'entrevue est adressée à tous les membres de chaque ménage, et une composante santé dans laquelle des questions approfondies sont posées à une personne choisie dans chaque ménage. Ce questionnaire couvre l'état de santé, l'utilisation des services de santé, les facteurs de risque ainsi que la situation démographique et socio-économique. Il vise une population âgée de 12 ans et plus. Notre variable d'intérêt, « la consommation de

[21] Statistique Canada. Aperçu de l'Enquête nationale sur la santé de la population 1996-1997 - Catalogue 82-567.

médicaments », est l'une des mesures de la question portant sur l'utilisation des services de santé[22].

Notre objectif est la modélisation de la demande de médicaments au Canada grâce à l'utilisation des données de l'ENSP. Le sous-échantillon longitudinal cible qui a été utilisé couvre les répondants des ménages de toutes les provinces.

La consommation de médicaments est mesurée à partir de deux types de questions concernant leur utilisation. La première question porte sur les différents médicaments pris au cours du dernier mois par un répondant âgé de 12 ans et plus. Les réponses sont obtenues pour 22 codes de médicaments (A-V) à partir de la question suivante : *« Au cours du dernier mois, avez-vous pris un des médicaments suivants (...) » :*

Les antibiotiques (antib),
Les analgésiques (analg),
Les antidépresseurs (antidep),
Les hormonothérapies (hormon),
Les hypotenseurs (hypot),
et ainsi de suite pour un total de 22 codes de médicaments.

Il s'agit d'analyser une variable dérivée indiquant des réponses « Oui » ou « Non » pour chacun des 22 médicaments précédemment cités, c'est-à-dire que cela indique si la personne a consommé ou non lesdits médicaments au cours du dernier mois (pris ou non sur ordonnance)[23].

---

[22] Statistique Canada. Guide du fichier de microdonnées à grande diffusion de l'ENSP 1996-1997.

[23] Statistique Canada. Enquête nationale sur la santé de la population, novembre 2006. Volet ménages.

Une fois les réponses à la première question obtenues, on demande dans la deuxième question le nombre de médicaments différents pris par les répondants pendant les dernières 48 heures:

> *« Pensez maintenant à hier et à avant-hier. Durant ces 2 journées, combien de médicaments différents avez-vous pris? »*[24]

Cette question posée dans chaque cycle constitue la variable d'intérêt longitudinale dans notre étude : il s'agit de la variable dépendante. Le nombre total de médicaments consommés par les individus est notée « med ». Le tableau 6.1 définit et décrit les variables considérées dans la modélisation. Notons que la question sur le nombre de médicaments consommés n'est posée qu'aux répondants à une des 22 questions concernant le dernier mois. Nous attribuons 0 médicament consommés au cours des 48 dernières heures à ceux qui ont répondu « Non » aux 22 questions.

Cycle 6 (2004-2005). Documentation des variables dérivées et des variables longitudinales constantes (Spécifications) Cycles 1 à 6.

[24] Statistique Canada. Enquête nationale sur la santé de la population (ENSP), novembre 2006. Cycle 6 (2004-2005). Dictionnaire de données. Fichier maître: Longitudinal carré (Arrondi).

Tableau 6.1 Définition et description des données

Variables		*Définition*	*%*
med	*quantitative*	Nombre de médicaments consommés par l'individu pendant les derniers 48 heures	1,17*
Caractéristiques de l'individu			
homme	*indicatrice*	l'individu est un homme	49,50
âge	*indicatrice*	L'âge de l'individu	42,94 *
âge2	*indicatrice*	L'âge au carré	-
tm	*quantitative*	Taille de ménage de l'individu	2,80 *
Comportements influant sur la santé			
f	*indicatrice*	Individu fumeur	21,30
alnb	*indicatrice*	Individu non buveur	19,11
alo	*indicatrice*	Individu buveur occasionnel	16,80
lim	indicatrice	Individu avec limitation d'activité	19,87
État de santé			
pesf	*indicatrice*	Perception état de santé faible de l'individu	9,13
pess	*indicatrice*	Perception état de santé supérieure de l'individu	64,02
imc	*quantitative*	Indice de masse corporelle de l'individu	43,02*
Classe de Provinces de l'individu			
atl	*indicatrice*	Individu habitant à la province de la Atlantique	8,08
qc	*indicatrice*	Individu habitant à la province de la Québec	24,76
ont	*indicatrice*	Individu habitant à la province de l'Ontario	37,67
prairies	*indicatrice*	Individu habitant à la province de la Prairies	16,75
cb	*indicatrice*	Individu habitant à la province de la Colombie-Britannique,	12,77
Revenu et travail			
work	*indicatrice*	Individu qui a travaillé au cours des 12 derniers mois	53,09
incinf	*indicatrice*	Individu avec un revenu inférieur	11,17
incimoy	*indicatrice*	Individu avec un revenu inférieur moyen	23,28
incsmoy	*indicatrice*	Individu avec un revenu supérieur moyen	29,03
Assurance médicaments dans les provinces			
asstn	*indicatrice*	Assurance médicaments de l'individu à la province de Terre-Neuve	1,30

assipe	*indicatrice*	Assurance médicaments de l'individu à la province de l'Île-du-Prince-Édouard	0,34
assne	*indicatrice*	Assurance médicaments de l'individu à la province de la Nouvelle-Écosse	2,55
assnb	*indicatrice*	Assurance médicaments de l'individu à la province du Nouveau-Brunswick	2,07
assqc	*indicatrice*	Assurance médicaments de l'individu à la province de Québec	21,73
assont	*indicatrice*	Assurance médicaments de l'individu à la province d'Ontario	30,94
assmb	*indicatrice*	Assurance médicaments de l'individu à la province du Manitoba	2,82
assab	*indicatrice*	Assurance médicaments de l'individu à la province de la Saskatchewan	7,70
asssk	*indicatrice*	Assurance médicaments de l'individu à la province de l'Alberta	2,36
asscb	*indicatrice*	Assurance médicaments de l'individu à la province de la Colombie-Britannique	9,84
Changement d'assurance médicament			
ccb	*indicatrice*	Assurance médicaments de l'individu à la province CB à partir de 2002	2,53
csk	*indicatrice*	Assurance médicaments de l'individu à la province SK à partir de 2002	0,67
cqc	*indicatrice*	Assurance médicaments de l'individu à la province QC à partir de 1997	18,71
cne	*indicatrice*	Assurance médicaments de l'individu à la province NÉ à partir de 2000	1,31

Le nombre d'observation N varie entre 95643 et 120904 (à cause des valeurs manquantes).
* (moyenne)
Source : Calculs de l'auteur à partir des microdonnées de l'ENSP.

6.2 Résultats empiriques de l'estimation des modèles de Poisson et binomial négatif en panel

À la différence des données en coupe transversale, la modélisation des données de panel permet différents types d'hétérogénéité individuelle. En effet, dans le cas de l'estimation de l'effet des déterminants de la consommation de médicaments, d'autres effets inobservables peuvent influencer la consommation, par exemple le comportement et l'attitude des individus en termes de consommation de médicaments, ou encore l'idée qu'ils se font d'un médicament. L'estimation devient non convergente si l'on omet ces effets. Les modèles à effets fixes et ceux à effets aléatoires représentent ces effets individuels en introduisant un indicateur spécifique pour chaque individu. Nous présenterons dans la section suivante une comparaison des résultats d'estimation des modèles à effets fixes et de ceux à effets aléatoires.

6.2.1 Estimation des modèles de Poisson et binomial négatif à effets aléatoires

Le modèle à effets aléatoires, à l'inverse du modèle à effets fixes, ne contrôle pas les variables individuelles inobservées et invariantes dans le temps car l'effet individuel est supposé non corrélé avec les variables explicatives. C'est pourquoi les effets des variables explicatives invariantes dans le temps peuvent être estimés avec l'effet aléatoire.

Le modèle à effets aléatoires est une alternative au modèle à effets fixes. Ce modèle est fondé sur la même équation qui est utilisée dans le modèle à effets fixes. Par exemple, pour le modèle de Poisson à effets aléatoires, avec le paramètre μ_i, on obtient :

$$P(Y_i = y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \mu_i ,$$

$$\ln \mu_i = x_i' \beta + \alpha_i + \varepsilon_i = \ln \lambda_i + \ln u_i ,$$

où ε_i est une erreur de spécification qui peut représenter une forme d'hétérogénéité. Ce terme aléatoire permet de régler le problème de surdispersion. Le modèle à effets aléatoires suppose qu'il existe un groupe d'effets éventuels avec une répartition précise autour d'un effet global moyen, et qu'il produit une matrice de variance-covariance robuste.

La différence avec le modèle à effets fixes réside dans le fait qu'au lieu de traiter les indicateurs α_i comme des indicateurs individuels fixes, il est supposé que α_i constitue un ensemble de variables aléatoires avec une distribution de probabilité spécifique, de même qu'avec une corrélation probable avec les variables explicatives. La différence apparente entre les deux modèles est que la méthode à effets aléatoires peut inclure des régresseurs invariants dans le temps. C'est ce que l'on peut remarquer dans le tableau 6.2.1, qui représente les résultats d'estimation des modèles de Poisson et binomial négatif à effets aléatoires. La variable « homme », qui est fixe dans le temps, est estimée dans les deux modèles.

Les coefficients et les écarts-types estimés pour les deux modèles sont presque identiques. Les effets du sexe, de la limitation d'activité, de l'état de santé, de la détention d'une assurance dans chaque province et du changement des coûts des assurances semblent influencer de façon notable le nombre espéré de médicaments consommés.

Cependant, le modèle à effets aléatoires ne considère pas l'effet de corrélation entre les variables explicatives et celles qui ne sont pas observées, dans l'estimation car cette corrélation, qui est causée par l'hétérogénéité individuelle inobservée, n'est pas contrôlée. Le modèle à effets fixes permet, quant à lui, de contrôler cette hétérogénéité individuelle, comme nous le verrons ci-après.

Tableau 6.2.1 Estimation des modèles de Poisson et binomial négatif à effets aléatoires

Variables	Poisson à effets aléatoires		Binomial négatif à effets aléatoires	
	Coefficients	Std.dev	Coefficients	Std.dev
homme	-0,2925***	0,0155	-0,4583***	0,0169
âge	0,0387***	0,0018	0,0568***	0,0019
âge2	-0,0002***	0,0000	-0,0003***	0,0000
tm	-0,0569***	0,0059	-0,0628***	0,0063
f	-0,0326**	0,0157	-0,0168	0,0168
alnb	0,0695***	0,0157	0,0697***	0,0165
alo	0,0698***	0,0145	0,0853***	0,0150
lim	0,3786***	0,0133	0,4052***	0,0139
pesf	0,2669***	0,0147	0,2683***	0,0152
pess	-0,2955***	0,0128	-0,3188***	0,0132
imc	0,0032***	0,0003	0,0030***	0,0003
atl	-0,0263	0,0416	0,0018	0,0435
qc	-0,2087***	0,0518	-0,2456***	0,0534
prairies	-0,1270***	0,0430	-0,1135***	0,0448
cb	-0,1602***	0,0574	-0,1809***	0,0590
work	-0,0915***	0,0156	-0,0631***	0,0164
incinf	-0,0915***	0,0219	-0,1075***	0,0229
incimoy	-0,0871***	0,0191	-0,1022***	0,0198
incsmoy	-0,0580***	0,0169	-0,0705***	0,0175
asstn	0,1349***	0,0410	0,1619***	0,0439
assipe	0,1869***	0,0404	0,2066***	0,0436
assne	0,0813**	0,0423	0,0979**	0,0447
assnb	0,1803***	0,0384	0,1950***	0,0412
assqc	0,0189	0,0531	0,0492	0,0538
assont	0,1011***	0,0333	0,1302***	0,0345
assmb	0,1438***	0,0399	0,1466***	0,0421
assab	0,1729***	0,0375	0,1806***	0,0396

asssk	0,0798**	0,0418	0,0888**	0,0439
asscb	0,1142**	0,0519	0,1348***	0,0529
ccb	0,1133***	0,0363	0,1060***	0,0365
csk	0,1771***	0,0446	0,1739***	0,0450
cqc	0,2392***	0,0363	0,2355***	0,0366
cne	0,1865***	0,0351	0,1747***	0,0353
_cons	-0,9914***	0,0628	13,545	51,706

Le nombre d'observation N = 51099.
Le nombre d'individus = 15928.
Seuils de signification : (1%)***, (5%)**, (10%)*.
Source : Calculs de l'auteur à partir des microdonnées de l'ENSP.

6.2.2 Estimation des modèles de Poisson et binomial négatif à effets fixes

Le tableau 6.2.2 ci-après présente l'estimation des modèles de Poisson et binomial négatif à effets fixes. L'inclusion d'un grand nombre de variables binaires pour estimer le modèle inconditionnel est pratiquement impossible étant donné que la base de données comporte 17 276 individus. Cependant, nous obtenons des estimations grâce aux modèles conditionnels à effets fixes. Rappelons que la vraisemblance conditionnelle binomiale négative utilisée par le progiciel Stata n'est plus une méthode adaptée parce qu'elle permet d'introduire dans l'estimation des régresseurs invariants dans le temps. Cela est bien visible dans le tableau 6.2.2, où l'on voit que le coefficient de la variable du sexe (qui ne varie pas dans le temps), variable appelée « homme », est estimé, alors que ce n'est pas le cas dans la régression de Poisson à effets fixes[25].

Comparativement aux résultats de l'estimation des modèles à effets aléatoires, nous remarquons que les coefficients sont un peu moins élevés (en valeur absolue);

[25] Notons que l'estimation de la binomiale négative à effets fixes est obtenue sans constante en raison de la non-convergence de cette régression avec l'inclusion de la constante. Cela dit, enlever la constante n'est pas une bonne idée dans l'économétrie. Dans le cas de la méthode inconditionnelle, si l'on ajoute des variables binaires pour chaque cas différent, alors il n'y aura plus besoin d'inclure la constante.

Tableau 6.2.2 Estimation des modèles de Poisson et binomial négatif à effets fixes

Variables	Poisson à effets fixes		Binomial négatif à effets fixes	
	Coefficients	Std.dev	Coefficients	Std.dev
homme	-	-	-0,5037*	0,3054
âge	0,0552***	0,0067	0,1018***	0,0050
âge2	0,0002***	0,0001	-0,0001***	0,0000
tm	-0,0130	0,0100	0,0041	0,0100
f	-0,0286	0,0261	-0,0194	0,0265
alnb	0,0132	0,0221	0,0206	0,0223
alo	0,0302*	0,0181	0,0359**	0,0183
lim	0,2098***	0,0164	0,2105***	0,0166
pesf	0,1386***	0,0170	0,1409***	0,0172
pess	-0,1818***	0,0151	-0,1726***	0,0152
imc	0,0014***	0,0004	0,0016***	0,0004
atl	0,0454	0,1054	0,2997***	0,1021
qc	-0,2738	0,1684	-0,1018	0,1561
prairies	-0,1957*	0,1152	0,0391	0,1110
cb	-0,0789	0,1268	0,1663	0,1233
work	-0,0368*	0,0211	-0,0290	0,0214
incinf	-0,0512*	0,0302	0,0135	0,0299
incimoy	-0,0401	0,0261	0,0123	0,0259
incsmoy	-0,0432**	0,0217	-0,0033	0,0217
asstn	0,2096***	0,0743	0,2065***	0,0751
assipe	0,0887	0,0777	0,0949	0,0786
assne	0,0250	0,0679	0,0248	0,0687
assnb	0,1576**	0,0720	0,1408**	0,0731
assqc	0,0894	0,0599	0,0788	0,0607
assont	0,0108	0,0428	0,0585	0,0436
assmb	0,0766	0,0673	0,0658	0,0683
assab	0,1649***	0,0668	0,1487**	0,0677
asssk	0,1095*	0,0611	0,0838	0,0625

asscb	0,0836	0,0656	0,0754	0,0664
ccb	-0,0934***	0,0383	-0,1159***	0,0385
csk	-0,0349	0,0471	-0,0502	0,0475
cqc	0,0593	0,0396	0,0359	0,0399
cne	-0,0336	0,0371	-0,0561	0,0374

Le nombre d'observation N = 34908.
Le nombre d'individus = 10014.
Seuils de signification : (1%)***, (5%)**, (10%)*.
Source : Calculs de l'auteur à partir des microdonnées de l'ENSP.

par ailleurs, certains deviennent non significatifs et d'autres ne vont pas dans le même sens. Les écarts-types, quant à eux, sont plus élevés. L'effet des variables qui ne changent pas dans le temps n'est pas estimé par la régression à effets fixes : la seule variable invariante dans le temps présente dans cette estimation est la variable du sexe (dont le coefficient n'est pas estimé dans le cas du modèle de Poisson).

La différence entre les deux approches – à effets aléatoires et à effets fixes – réside dans la corrélation entre l'effet individuel et la variable explicative. En effet, lorsqu'on estime le modèle à effets fixes, cette corrélation n'existe plus parce que l'on contrôle pour l'effet fixe. Cela signifie que la diminution de la valeur des coefficients observée dans l'estimation du modèle à effets fixes est un résultat intéressant : cela prouve que les variables correspondantes étaient corrélées avec l'effet fixe. Un cas notable est celui du « revenu » : ce dernier est un facteur endogène qui peut donc capter en réalité d'autres effets associés au revenu mais qui ne sont pas vraiment observés. Citons par exemple les dépenses liées à de mauvaises habitudes alimentaires. Ces effets vont être attribués au revenu alors qu'en réalité ils proviennent de l'effet fixe. Cela explique la diminution en valeur absolue de l'effet du revenu (c'est-à-dire l'effet des variables « incinf », « incimoy » et « incsmoy ») dans le cas du modèle à effets fixes. Par ailleurs, les variables « incinf » et « incimoy » changent de sens dans le cas du modèle binomial négatif. Avec le modèle de Poisson, le nombre moyen de médicaments à consommer dépend négativement du revenu faible. En effet, le manque de moyens financiers empêche les ménages de se permettre d'utiliser la médication (sauf en cas de nécessité) et de se procurer une assurance médicaments.

Les autres variables explicatives peuvent être interprétées de la même façon. Les fumeurs peuvent ne pas avoir encore observé de maladie, ou ils ne savent pas qu'ils sont / qu'ils deviendront malades, alors ils continuent à fumer et n'utilisent pas de médicaments. Cela explique le signe négatif du coefficient et la diminution de l'impact pour les fumeurs avec le modèle à effets fixes. Toutefois, l'effet des fumeurs devient non significatif dans le modèle à effets fixes (cela est visible dans l'estimation du modèle de Poisson). La même constatation peut être faite pour les coefficients des variables « lim », « imc », « pesf » et « pess », qui décroissent en valeur absolue par rapport aux modèles à effets aléatoires, ce qui explique qu'une partie de la corrélation avec l'hétérogénéité inobservée n'existe plus.

Bref, les effets fixes ont un impact sur l'effet des variables explicatives. Ainsi, ces résultats servent de preuve que cette méthode est extrêmement importante dans l'étude visant la détermination des effets fixes, tels que l'effet revenu et l'effet des indicateurs de santé. Notons que les résultats d'estimation des modèles à effets aléatoires de Poisson et binomial négatif sont similaires à ceux obtenus avec les modèles initiaux (sans panel) respectifs de Poisson et binomial négatif. La conclusion majeure que l'on peut tirer de cette analyse est que les résultats obtenus sont très différents si l'on s'ajuste aux effets fixes.

Même si l'on utilise l'argument selon lequel le modèle FENB n'est pas un vrai modèle à effets fixes, dans notre cas, ce modèle nous donne, de façon très remarquable, des estimations de coefficients et d'écarts-types semblables comparativement au modèle de Poisson à effets fixes.

6.3 Test de Hausman

Il est nécessaire de spécifier lequel des deux modèles, à effets fixes ou à effets aléatoires, est le plus approprié à nos données. Le choix entre l'effet fixe et l'effet aléatoire vient de l'existence d'une corrélation entre les effets individuels et les variables explicatives. Le test de Hausman s'appuie sur la différence entre les estimateurs à effets fixes et ceux à effets aléatoires. Ce test est utilisé pour les deux modèles de Poisson et binomial négatif (voir tableau 6.3). L'effet fixe est convergent lorsque l'effet individuel est corrélé avec la variable explicative; l'effet aléatoire est donc non convergent. Une différence statistiquement significative entre les deux estimateurs met en cause l'hypothèse de l'effet aléatoire (Wooldridge, 2006). Autrement dit, sous l'hypothèse nulle, qui suppose l'absence de corrélation entre l'effet individuel et les régresseurs, les estimateurs à effets fixes $\hat{\beta}_{FE}$ et ceux à effets aléatoires $\hat{\beta}_{RE}$ sont tous convergents, mais les estimateurs à effets fixes sont inefficients. Ainsi, le modèle à effets aléatoires est correctement spécifié. Sous l'hypothèse alternative, les effets aléatoires sont corrélés avec les régresseurs; dans ce cas, les estimateurs à effets fixes sont convergents alors que les estimateurs à effets aléatoires ne le sont pas (Greene, 2005).

En bref, sous l'hypothèse nulle H0, le modèle à effets aléatoires est approprié. Par contre, sous l'hypothèse alternative H1, le modèle à effets fixes est préféré.

Le test de Hausman est fondé sur la statistique suivante :

$$T_H = \left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right)' \left[Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})\right]^{-1} \left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right),$$

où $Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})$ représente la différence entre les variances des coefficients $\hat{\beta}_{FE}$ et $\hat{\beta}_{RE}$. Cette statistique suit une distribution de Khi-deux χ^2 à k (dimension de $\hat{\beta}$) degrés de liberté (dl=k).

Tableau 6.3 Résultats du test de Hausman

	$\chi^2_{(0,05)}$ (dl)	Prob > χ^2
Modèle de Poisson	χ^2 (31)= 1585,75	0,000
Modèle binomial négatif	χ^2 (32)= 2328.43	0,000

Source : Calculs de l'auteur à partir des microdonnées de l'ENSP.

Nous pouvons ainsi rejeter l'hypothèse nulle; le modèle à effets fixes est donc le plus approprié. Cela fait que les effets individuels sont corrélés avec les variables explicatives.

6.4 La méthode alternative de Wooldridge

Comme nous l'avons vu précédemment, l'importance de l'introduction des effets fixes dans les modèles de Poisson et binomial négatif réside dans le fait que certains estimateurs d'effets individuels changent radicalement étant donné l'inclusion des effets fixes. C'est notamment le cas pour l'effet de revenu et celui des fumeurs. Dans le cas des modèles non linéaires à effets fixes, il est difficile de calculer et d'interpréter les effets marginaux; en fait, cela n'a pas de sens du moment que ces modèles sont sans constante. Par conséquent, le calcul des effets marginaux interprétables dans le cas du modèle à effets fixes requiert un modèle qui contrôle les effets fixes et admet une constante. C'est le cas du modèle que nous propose Wooldridge[26] (2006).

L'approche de Wooldridge permet d'introduire parmi les variables explicatives, dans le modèle, des moyennes individuelles pour capter les effets fixes. Ce modèle s'estime avec l'option RE (« random effect »). En fait, le modèle à effets aléatoires introduit un terme qui fait en sorte que la corrélation soit la même en toutes périodes. Toutefois, une erreur de spécification dans la forme de corrélation dans le terme individuel va avoir un impact sur les écarts-types. Nous pouvons en outre utiliser la méthode GEE, qui permet de calculer les estimateurs avec une matrice de variance-covariance qui est robuste à toute forme de corrélation en termes individuels.

Le tableau 6.4 représente les résultats d'estimation de la régression binomiale négative obtenus avec la méthode GEE en introduisant les moyennes des variables explicatives (qui varient dans le temps). En comparant ces résultats aux estimations du modèle binomial négatif à effets aléatoires, nous réalisons que certains effets individuels ont un impact moins prononcé sur le nombre espéré de médicaments consommés. Ces effets correspondent aux variables qui comportent probablement des facteurs non observables. C'est le cas des variables « alo », « lim », « pesf », « pess », « incimoy », « incsmoy », pour lesquelles une partie fixe non mesurée (corrélée avec l'hétérogénéité non observée) semble être prise en compte par le modèle de Wooldridge, ce qui explique la diminution en valeur absolue des coefficients estimés. Autrement dit, cette approche proposée par Wooldridge contrôle effectivement pour les effets fixes. Les écarts-types robustes obtenus par cette méthode sont plus élevés que ceux issus du modèle binomial négatif à effets aléatoires.

[26] Wooldridge. 2006. « Cluster-Sample Methods In Applied Econometrics: An Extended Analysis ».

Tableau 6.4 Résultats d'estimation par la méthode alternative de Wooldridge

Variables	Méthode alternative de Wooldridge_ GEE		Binomial négatif à effets aléatoires	
	Coefficients	Semi-robust Std.dev	Coefficients	Std.dev
homme	-0,4479***	0,0181	-0,4583***	0,0169
âge	0,1019***	0,0032	0,0568***	0,0019
âge2	-0,0003***	0,0000	-0,0003***	0,0000
tm	-0,0119	0,0174	-0,0628***	0,0063
f	0,0251	0,0274	-0,0168	0,0168
alnb	0,0619***	0,0229	0,0697***	0,0165
alo	0,0603***	0,0203	0,0853***	0,0150
lim	0,2344***	0,0174	0,4052***	0,0139
pesf	0,1994***	0,0177	0,2683***	0,0152
pess	-0,1903***	0,0149	-0,3188***	0,0132
Imc	0,0006	0,0004	0,0030***	0,0003
atl	0,0064	0,0454	0,0018	0,0435
qc	-0,2049***	0,0545	-0,2456***	0,0534
prairies	-0,1214***	0,0453	-0,1135***	0,0448
cb	-0,2102***	0,0576	-0,1809***	0,0590
work	-0,0672***	0,0215	-0,0631***	0,0164
incinf	-0,0992***	0,0257	-0,1075***	0,0229
incimoy	-0,0483**	0,0239	-0,1022***	0,0198
incsmoy	-0,0515***	0,0199	-0,0705***	0,0175
asstn	0,1862***	0,1862	0,1619***	0,0439
assipe	0,2237***	0,0434	0,2066***	0,0436
assne	0,1599***	0,0446	0,0979**	0,0447
assnb	0,1604***	0,0413	0,1950***	0,0412
assqc	0,1242**	0,0576	0,0492	0,0538
assont	0,1093***	0,0352	0,1302***	0,0345
assmb	0,1276***	0,0422	0,1466***	0,0421

assab	0,1355***	0,0392	0,1806***	0,0396
asssk	0,0850*	0,0489	0,0888**	0,0439
asscb	0,1631***	0,0502	0,1348***	0,0529
ccb	-0,0481	0,0341	0,1060***	0,0365
csk	0,0205	0,0480	0,1739***	0,0450
cqc	0,0871**	0,0432	0,2355***	0,0366
cne	0,0021	0,0372	0,1747***	0,0353
_cons	-1,006***	0,0782	13,545	51,706

Le nombre d'observation N = 51099.
Le nombre d'individus = 15928.
Seuils de signification : (1%)***, (5%)**, (10%)*.
Source : Calculs de l'auteur à partir des microdonnées de l'ENSP.

6.5 Synthèse des résultats selon les méthodes

Nous venons de voir plusieurs méthodes qui ont été appliquées à une même banque de données sur la consommation de médicaments au Canada. Nous avons regroupé au tableau 6.5 les divers coefficients en omettant les écarts-types afin de faciliter la discussion.

Le test de Hausman nous permet de conclure que le modèle à effets fixes est le plus approprié à nos données sur la consommation de médicaments au Canada, ce qui prouve que les variables explicatives sont nécessairement corrélées avec l'hétérogénéité inobservée et que le modèle à effets fixes prenne en compte cette corrélation. Cependant le modèle binomial négatif s'est avéré non adéquat du point de vue de sa vraisemblance (considérée dans le progiciel Stata) qui n'introduit pas les effets fixes. De ce fait, le modèle binomial négatif ne contrôle pas pour les effets fixes. Dans notre cas, ce modèle estimé sans constante remet en cause encore davantage sa crédibilité, malgré que les résultats d'estimation sont similaires à ceux du modèle de Poisson à effets fixes. Le modèle de Wooldridge permet de contrôler pour les effets fixes et d'introduire une constante afin d'avoir des effets marginaux interprétables. Cela répond exactement à ce qu'on cherche pour modéliser la relation entre les variables déterminant l'état de santé et la consommation de médicaments.

Bref, les modèles à effets aléatoires sont mis à côté parce qu'ils ne contrôlent pas pour les effets fixes et cela n'est pas approprié à nos données sur la consommation de médicaments au Canada. Ceci dit, le modèle binomial négatif à effets aléatoires avec l'introduction des moyennes des variables explicatives dans la régression, selon l'approche de Wooldridge, est semblable à un modèle binomial négatif mais qui contrôle pour les effets fixes. Cela nous permet de calculer des effets marginaux interprétables pour un modèle contrôlant les effets fixes. Par ailleurs, le modèle de Poisson à effets fixes pourrait être pris en considération, mais ce modèle est caractérisé par la propriété d'équidispersion, ce qui n'est pas le cas dans nos données[27]. Il doit donc être rejeté. Étant donné l'importance d'interpréter des effets marginaux expliquant l'effet des différents indicateurs de santé (variables explicatives) sur le nombre espéré de médicaments consommés, on aura donc recourt au modèle proposé par Wooldridge.

---

[27] Les données de cette étude sont caractérisées par une surdispersion. En effet, la répartition du nombre de médicament n'est pas homogène et est aplatie vers la droite (voir appendice A, tableau 8.2).

Tableau 6.5 Synthèse des résultats selon la méthode d'estimation

Variables	Poisson RE	Binomial négatif RE	Poisson FE	Binomial négatif FE	Méthode alternative de Wooldridge GEE
	Coefficients	Coefficients	Coefficients	Coefficients	Coefficients
homme	-0,2925***	-0,4583***	-	-0,5037*	-0,4479***
âge	0,0387***	0,0568***	0,0552***	0,1018***	0,1019***
âge2	-0,0002***	-0,0003***	0,0002***	-0,0001***	-0,0003***
tm	-0,0569***	-0,0628***	-0,013	0,0041	-0,0119
f	-0,0326**	-0,0168	-0,0286	-0,0194	0,0251
alnb	0,0695***	0,0697***	0,0132	0,0206	0,0619***
alo	0,0698***	0,0853***	0,0302*	0,0359**	0,0603***
lim	0,3786***	0,4052***	0,2098***	0,2105***	0,2344***
pesf	0,2669***	0,2683***	0,1386***	0,1409***	0,1994***
pess	-0,2955***	-0,3188***	-0,1818***	-0,1726***	-0,1903***
imc	0,0032***	0,0030***	0,0014***	0,0016***	0,0006
atl	-0,0263	0,0018	0,0454	0,2997***	0,0064
qc	-0,2087***	-0,2456***	-0,2738	-0,1018	-0,2049***
prairies	-0,1270***	-0,1135***	-0,1957*	0,0391	-0,1214***
cb	-0,1602***	-0,1809***	-0,0789	0,1663	-0,2102***
work	-0,0915***	-0,0631***	-0,0368*	-0,029	-0,0672***
incinf	-0,0915***	-0,1075***	-0,0512*	0,0135	-0,0992***
incimoy	-0,0871***	-0,1022***	-0,0401	0,0123	-0,0483**
incsmoy	-0,0580***	-0,0705***	-0,0432**	-0,0033	-0,0515***
asstn	0,1349***	0,1619***	0,2096***	0,2065***	0,1862***
assipe	0,1869***	0,2066***	0,0887	0,0949	0,2237***
assne	0,0813**	0,0979**	0,025	0,0248	0,1599***
assnb	0,1803***	0,1950***	0,1576**	0,1408**	0,1604***
assqc	0,0189	0,0492	0,0894	0,0788	0,1242**
assont	0,1011***	0,1302***	0,0108	0,0585	0,1093***
assmb	0,1438***	0,1466***	0,0766	0,0658	0,1276***

assab	0,1729***	0,1806***	0,1649***	0,1487**	0,1355***
asssk	0,0798**	0,0888**	0,1095*	0,0838	0,0850*
asscb	0,1142**	0,1348***	0,0836	0,0754	0,1631***
ccb	0,1133***	0,1060***	-0,0934***	-0,1159***	-0,0481
csk	0,1771***	0,1739***	-0,0349	-0,0502	0,0205
cqc	0,2392***	0,2355***	0,0593	0,0359	0,0871**
cne	0,1865***	0,1747***	-0,0336	-0,0561	0,0021
_cons	-0,9914***	13,545			-1,006***

Seuils de signification : (1%)***, (5%)**, (10%)*.
Source : Calculs de l'auteur à partir des microdonnées de l'ENSP.

Le choix de la méthode GEE implique des coefficients robustes, contrairement aux modèles à effets aléatoires, où une erreur de spécification dans la forme de corrélation dans le terme individuel peut avoir un impact sur les écarts-types. Les coefficients ne sont pas tout à fait stables et changent de significativité selon les méthodes qui contrôlent ou non pour les effets fixes. En effet, l'estimation des modèles à effets fixes implique une diminution de la valeur des coefficients (en valeur absolue), comparativement aux modèles à effets aléatoires, parce que l'on contrôle pour l'effet fixe et donc, il n'y a plus de corrélation entre les variables explicatives et l'hétérogénéité inobservée. En ce qui concerne la significativité des coefficients, ceci est relatif aux écarts-types estimés selon chaque méthode et au nombre de degré de liberté. En fait, les écarts-types sont plus élevés dans la méthode de Wooldridge qui implique une matrice de variance-covariance robuste, par rapport aux méthodes à effets aléatoires et les méthodes à effets fixes. Ce qui explique le changement de significativité des coefficients d'une méthode à l'autre, comme par exemple le coefficient de l'assurance pour le Québec « assqc » qui devient significatif dans le modèle de Wooldridge avec GEE, et le coefficient de changement d'assurance médicament au Québec « cqc » qui reste significatif, alors qu'il ne l'était pas dans les modèles à effets fixes. Rappelons que dans le cas de la méthode de Wooldridge, le nombre de degré de liberté change à cause de l'introduction des moyennes des variables explicatives dans la régression.

Enfin, l'approche de Wooldridge est la solution ultime pour traduire l'effet des différentes variables explicatives de santé choisies sur la consommation de médicament. Cette approche permet de prendre en compte les effets inobservables et d'interpréter les effets marginaux. Les coefficients sont plus robustes et significatifs en prenant la méthode de GEE (au lieu de la simple méthode RE). Étant donné que cette approche de Wooldridge est la meilleure parmi les différentes alternatives que nous avons vues, nous calculons ci-après, les effets marginaux du modèle de Wooldridge, et nous pouvons voir que cette méthode nous offre des estimations plus près de l'intuition.

6.6 Interprétation des effets marginaux du modèle de Wooldridge

Le modèle à effets fixes (FENB), qui n'admet pas de constante, ne permet pas de faire de simulation; Wooldridge nous offre donc un modèle similaire à celui obtenu par GEE afin de simuler et d'interpréter les effets du modèle à effets fixes, qui sont calculés à la moyenne des variables explicatives du modèle. Les effets marginaux de ce modèle sont présentés dans le tableau 6.6.

Nous ferons ici état des premières constatations de cette simulation. L'âge a un effet positif décroissant dans le temps sur la consommation de médicament, c'est-à-dire qu'au fur et à mesure que la personne vieillisse, elle diminue son utilisation de médicaments. Par ailleurs, Les personnes de sexe masculin consommeront, *ceteris paribus*, 0,31 médicament de moins que les personnes de sexes féminins. De leur côté, les personnes ne consommant pas d'alcool et les consommateurs d'alcool occasionnels consommeront, *ceteris paribus*, 0,04 médicament de plus que les consommateurs d'alcool ce qui est un effet somme toute plutôt faible bien que significatif. De même que les personnes qui présentent une limitation d'activité comparativement aux personnes actives, la consommation de médicament est plus élevée, *ceteris paribus*, de 0,18. Le fait de fumer ne semble pas avoir d'effet significatif sur la consommation de médicaments. La taille du ménage a un très faible impact négatif sur la quantité moyenne de médicaments consommés.

L'état de santé joue un rôle primordial dans les dépenses liées aux médicaments. En effet, c'est le premier facteur qui incite une personne à recourir à un service de santé, et ultérieurement à la médication. Une personne ayant un état de santé médiocre a tendance à consommer davantage de médicaments qu'une personne en état de santé moyen. Quant à une personne en bon état de santé, elle a tendance à consommer moins de médicament qu'une personne en état de santé moyen. Cela paraît évident avec les résultats obtenus en ce qui a trait aux indicateurs « pesf » et « pess », qui ont respectivement un effet positif et négatif sur le nombre espéré de médicament consommés. Un autre indice important de l'état de santé est l'indice de masse corporelle (IMC) : ce dernier a un impact positif mais assez faible sur le nombre moyen de médicaments consommés. Notons que l'IMC est défini en tant que le rapport entre le poids de la personne et sa grandeur au carré (kg/m^2). Si l'IMC s'écarte de l'indice de poids normal pour la santé, la personne sera confrontée à plus de risques d'avoir des problèmes de santé. Notons également que cet indice est approuvé seulement pour les personnes âgées de plus de 20 ans et qu'il ne considère pas le cas des femmes enceintes[28]. L'effet positif de l'indice IMC témoigne alors de l'existence de problèmes de santé reliés au poids de la personne, autrement dit, reliés au fait que le poids est supérieur ou inférieur au poids normal.

Le revenu influence négativement, bien que très faiblement, la consommation de médicaments. Dans notre banque de données, la classe de référence est constituée des revenus élevés. Une personne qui a un faible revenu consommera, *ceteris paribus*, 0,07 médicament de moins qu'une personne qui a un revenu élevé. Par ailleurs, on constate que la même personne avec un revenu moyen faible ou moyen

[28] Dominick Latrémouille-Viau. 2007. « Les déterminants de la consommation de médicaments au Canada ».

élevé consommera, *ceteris paribus*, 0,03 de médicament de moins qu'une personne qui a un revenu élevé. Les personnes avec un revenu élevé bénéficient probablement d'une bonne couverture d'assurance en plus d'avoir les moyens financiers pour utiliser des médicaments hors assurance. Quant au travail, ce dernier a également un impact négatif sur le nombre moyen de médicament consommé, en conséquence, une personne qui a travaillé au cours de la dernière année consommera, *ceteris paribus*, 0,05 de médicament de moins qu'une personne qui n'a pas travaillé pendant cette période.

Tableau 6.6 Estimation des effets marginaux par la méthode alternative de Wooldridge

Variables	Méthode alternative de Wooldridge Effets marginaux
Caractéristiques de l'individu	
homme	-0,3122***
âge	0,0717***
âge2	-0,0002***
tm	-0,0084
Comportements influant sur la santé	
f	-0,0176
alnb	0,0444***
alo	0,0433***
lim	0,1771***
État de santé	
pesf	0,1519***
pess	-0,1374***
imc	0,0004
Classe de Provinces de l'individu	
atl	0,0045
qc	-0,1352***
Prairies	-0,0828***

cb	-0,1362***
Revenu et travail	
work	-0,0474***
incinf	-0,0676***
incimoy	-0,0336**
incsmoy	-0,0359***
Assurance médicaments par province	
asstn	0,1429***
assipe	0,1748***
assne	0,1211***
assnb	0,1214***
assqc	0,0913**
assont	0,0795***
assmb	0,0952***
assab	0,1011***
asssk	0,0622*
asscb	0,1232***
Changement d'assurance médicaments	
ccb	-0,0331
csk	0,0146
cqc	0,0633**
cne	0,0015

Le nombre d'observation N = 51099.
Le nombre d'individus = 15928.
Seuils de signification : (1%)***, (5%)**, (10%)*.
Source : Calculs de l'auteur à partir des microdonnées de l'ENSP.

L'introduction de la variable de classes de provinces par rapport à la variable de l'assurance (asstn, assipe, etc.) dans chaque province nous permettra de faire la liaison entre les personnes non assurées pour les médicaments et celles qui le sont. Les variables de changement (ccb, csk, cqc et cne) indiquent, quant à elles, l'impact sur les assurés de la modification de la distribution des coûts du régime de l'assurance médicaments. La province de contrôle choisie dans notre modèle est l'Ontario.

Compte tenu des résultats de simulation, il apparaît que partout au Canada, sauf dans les Maritimes où la différence est non significative, la consommation de médicaments est significativement moindre qu'en Ontario. Mais dans le cas où les personnes détiennent une assurance et qu'elles résident dans l'une des cinq régions, l'effet de cette assurance aura un important impact positif et significatif sur le nombre de médicaments consommés. Notons que cette assurance est sous un régime soit public ou privé; le régime privé ne diffère généralement pas selon les provinces mais le régime d'assurance public est différent pour chaque province. Par exemple, au Québec, l'assurance est obligatoire comparativement à d'autres provinces : on ne trouve pas d'accès général à l'assurance publique[29] dans les provinces de l'Atlantique. Bref, le degré d'accessibilité à l'assurance médicaments déterminera son effet sur la consommation de médicaments en fonction du régime appliqué dans chaque province.

En ce qui concerne l'impact des modifications à la réglementation de l'assurance-médicaments, le seul changement significatif de la distribution des coûts du régime d'assurance médicaments est observé dans la province de Québec. L'introduction du déductibles et de coassurances par prescription chez les individus bénéficiaires de l'assurance médicaments gouvernementale[29] se révèle avoir un impact positif sur l'utilisation des médicaments au cours des deux derniers jours.

On note que sauf pour le genre et l'état de santé, peu de variables semblent influencer la consommation de médicaments. Seul le statut d'assuré a un impact significatif et important sur la consommation.

---

[29] Dominick Latrémouille-Viau. 2007. « Les déterminants de la consommation de médicaments au Canada ».

Le lien avec l'état de santé est naturel d'autant plus que la consommation de médicaments à consommation restreinte exige une ordonnance d'un médecin qui doit s'assurer que l'état de santé de la personne exige une médication pharmaceutique.

Le genre a un impact important et cela relève sans doute de considération sociologique. Les hommes consultent moins que les femmes et cela suffit à expliquer un accès moindre aux médicaments.

Finalement, le fait que seul le statut d'assuré ait un impact important est une indication du succès du système de santé et du régime d'assurance-médicaments. Peu importe le statut social ou le pouvoir d'achat, les personnes malades assurées ont un accès égal aux médicaments que nécessite leur état de santé. On note que la nature du régime d'assurance-médicaments à un effet comme ne fait foi l'impact de la variable qui fait état d'un changement dans le régime québécois.

CONCLUSION

Malgré l'argument selon lequel le modèle FENB ne contrôle pas vraiment pour les effets fixes – ce qui conduit à une vraisemblance conditionnelle calculée par le progiciel Stata non appropriée –, nous avons estimé ce modèle en utilisant les données de l'ENSP en vue de déterminer le nombre de médicaments consommés par les Canadiens. Notre estimation fournit des résultats presque identiques aux résultats d'estimation du modèle de Poisson à effets fixes et ce, rappelons-le, en omettant la constante du modèle FENB. La comparaison avec les modèles à effets aléatoires amène à privilégier les modèles à effets fixes pour les données utilisées dans le cadre de notre étude. En effet, une partie de la corrélation de certaines variables explicatives avec l'hétérogénéité inobservée n'existe plus lorsqu'on estime le modèle à effets fixes. Cela est constaté par le biais de la décroissance de la valeur absolue des coefficients comparativement aux modèles à effets aléatoires. Enfin, dans le cas des données de l'ENSP, le contrôle pour les effets fixes est nécessaire afin d'estimer l'effet des déterminants de la santé et de certains facteurs économiques sur la consommation de médicaments. Ceci peut être traduit par l'importance des effets fixes dans la modélisation des données de santé.

Le tableau 7 résume l'effet du revenu sur la consommation de médicaments. Nous remarquons clairement pour la variable « incsmoy » que l'impact, moins élevé, est significatif avec un large écart-type pour le modèle de Wooldridge comparativement aux modèles à effets aléatoires.

Nous avons donc conclu que certains effets changent radicalement si l'on contrôle pour les effets fixes, ce qui traduit l'importance des effets fixes dans ce type de modèles. Wooldridge nous propose un modèle similaire qui permet de capter et d'interpréter les effets du modèle à effets fixes. Ce dernier modèle estime les effets fixes par les moyennes des variables explicatives du modèle en utilisant l'option RE ou l'approche GEE. Finalement, l'introduction des effets fixes ne permet pas de faire une bonne simulation; la solution ultime est donc de recourir au modèle de Wooldridge qui nous permet de calculer et d'interpréter les effets marginaux.

Tableau 7. Comparaison des estimations alternatives de l'effet revenu

Modèle	Variable	Coefficient	Écart-type
Poisson	incinf,.	-0,1249***	0,0198
	incimoy	-0,0923***	0,0175
	incsmoy	-0,0451***	0,0160
Binomial négatif	incinf	-0,1131***	0,0201
	incimoy	-0,0898***	0,0174
	incsmoy	-0,0456***	0,0158
Poisson à effets aléatoires	incinf	-0,0915***	0,0219
	incimoy	-0,0871***	0,0191
	incsmoy	-0,0580***	0,0169
Binomial négatif à effets aléatoires	incinf	-0,1075***	0,0229
	incimoy	-0,1022***	0,0198
	incsmoy	-0,0705***	0,0175
Poisson à effets fixes	incinf	-0,0512*	0,0302
	incimoy	-0,0401	0,0261
	incsmoy	-0,0432**	0,0217
Binomial négatif à effets fixes	incinf	0,0135	0,0299
	incimoy	0,0123	0,0259
	incsmoy	-0,0033	0,0217
Binomial négatif_GEE par la méthode de Wooldridge	incinf	-0,0992***	0,0257
	incimoy	-0,0483**	0,0239
	incsmoy	-0,0515***	0,0199

Seuils de signification : (1%)***, (5%)**, (10%)*.

Les recherches sont toujours en cours d'avancement quant aux modèles à effets fixes en général et la performance de la méthode conditionnelle du modèle binomial négatif à effets fixes en particulier. Les plus récentes recherches dans ce domaine que l'on peut citer sont : Guimarães et Portugal (2009), qui montrent comment il est possible d'ajouter un très grand nombre de variables binaires dans un modèle sans avoir besoin de plus de mémoire (en utilisant le progiciel Stata). De même, Guimarães (2007) démontre que le modèle FENB contrôlera seulement différents effets spécifiques sous un ensemble d'hypothèses particulières et propose également un test de score qui peut aider à déterminer si ces hypothèses sont vérifiées. Waterman et Lindsay (1996a), quant à eux, introduisent une méthode approximative (*The Projected Score Method*), qui exploite les propriétés de l'inférence conditionnelle.

APPENDICE A

Tableau 8.1 Estimation des modèles de base de Poisson et binomial négatif

Variables	Poisson		Binomial négatif	
	Coefficients	Std. dev	Coefficients	Std. dev
homme	-0,2496***	0,0110	-0,2808***	0,0112
âge	0,0458***	0,0015	0,0441***	0,0014
âge2	-0,0003***	0,0000	-0,0003***	0,0000
tm	-0,0628***	0,0050	-0,0654***	0,0049
f	-0,0597***	0,0127	-0,0610***	0,0129
alnb	0,0948***	0,0131	0,0859***	0,0137
alo	0,0852***	0,0135	0,0868***	0,0137
lim	0,4537***	0,0120	0,4618***	0,0125
pesf	0,3320***	0,0137	0,3440***	0,0151
pess	-0,3481***	0,0128	-0,3514***	0,0126
imc	0,0038***	0,0003	0,0030***	0,0003
atl	-0,0306	0,0377	-0,0395	0,0369
qc	-0,1716***	0,0503	-0,1950***	0,0488
prairies	-0,1063***	0,0399	-0,1119***	0,0386
cb	-0,1793***	0,0539	-0,1891***	0,0517
work	-0,1591***	0,0139	-0,1347***	0,0140
incinf	-0,1249***	0,0198	-0,1131***	0,0201
incimoy	-0,0923***	0,0175	-0,0898***	0,0174
incsmoy	-0,0451***	0,0160	-0,0456***	0,0158
asstn	0,1852***	0,0339	0,1724***	0,0342
assipe	0,2626***	0,0327	0,2606***	0,0333
assne	0,1299***	0,0377	0,1405***	0,0385
assnb	0,2094***	0,0314	0,2033***	0,0319
assqc	0,0726	0,0560	0,0732	0,0549
assont	0,1750***	0,0314	0,1504***	0,0310

assmb	0,1742***	0,0351	0,1737***	0,0349
assab	0,1789***	0,0333	0,1722***	0,0326
asssk	0,0884***	0,0381	0,0900***	0,0382
asscb	0,1865***	0,0503	0,1782***	0,0486
ccb	0,1061***	0,0412	0,1112***	0,0430
csk	0,1637***	0,0497	0,1761***	0,0529
cqc	0,2121***	0,0409	0,2042***	0,0407
cne	0,1769***	0,0396	0,1856***	0,0424
_cons	-1,0842***	0,0563	-1,0194***	0,0541

Le nombre d'observation N = 51099.
Le nombre d'individus = 15928.
Seuils de signification : (1%)***, (5%)**, (10%)*.
Source : Calculs de l'auteur effectués à partir des microdonnées de l'ENSP.

Tableau 8.2 Répartition du nombre de médicaments

"med"	Pourcentage %
0	55,54
1	22,06
2	11,13
3	5,18
4	3,0
5	1,92
>=6	1,18
Total	100

Source : Calculs de l'auteur effectués à partir des microdonnées de l'ENSP.

BIBLIOGRAPHIE

Allison Paul D. et Richard Waterman. 2002. « Fixed Effects Negative Binomial Regression Models ». *Sociological Methodology*, 32, 247-265.

Allison, Paul D. 2005. « Fixed Effects Regression Methods In SAS », Paper 184-31, *Statistics and Data Analysis*, University of Pennsylvania, Philadelphia.

Allison, Paul D. 2005. *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. Cary, NC: SAS Institute Inc.

Allison, Paul D. 2009. *Fixed Effects Regression Models*. Series: Quantitative Applications in the Social Sciences; 160 p. *Sage Publications*. ISBN 978-0-7619-2497-5.

Cameron, A. Colin et Pravin K. Trivedi. 2005. *Microeconometrics Methods and Applications*. Cambridge University Press.

Dominick Latrémouille-Viau. 2007. « *Les déterminants de la consommation de médicaments au Canada* ». Mémoire de maîtrise en économie. Université de Québec à Montréal.

Greene, William. 2005. *Économétrie*. Sous la dir. Didier Schlacther. Trad. de l'anglais par Nicolas Couderc, Stéphanie Monjon et Phu Nguyen. p 946. Van Pearson Education (Paris - FRA).

Greene, William. Avril 2007. « Functional Form and Heterogeneity in Models for Count Data ». Department of Economics, Stern School of Business, New York University.

Greene, William. Mai 2007. « Fixed and Random Effects Models for Count Data ». Department of Economics, Stern School of Business, New York University.

Guimarães, Paulo et Pedro Portugal. January 2009. « A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects ». *IZA Discussion Paper* No. 393.

Guimarães, Paulo. 2007. « The Fixed Effects Negative Binomial Model Revisited ». *Science direct. Economics Letters* 99, 63-66.

Hausman, J. A, B. H. Hall et Z. Griliches. 1984. « Econometric Models For Count Data with an Application to the Patents - R and D Relationship ». *Econometrica* 52, 909-938.

Hilbe, Joseph M. 2007. *Negative Binomial Regression*. Cambridge University Press.

Hardin James W. and Joseph M. Hilbe. January 2007. *Generalized Linear Models and Extensions*. 2nd Edition. Stata Press.

Noriszura, Ismail et Abdul Aziz, Jemain. 2007. « Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models ». Casualty Actuarial Society Forum.

Qi Hu. 2002. « *Analysis of panel patent data using Poisson, negative binomial and GMM estimation* ». Mémoire de maîtrise en économie. Dalian University of Technology.

Richard P. Waterman et Bruce G. Lindsay. 1996. « Projected score methods for approximating conditional scores ». *Biometrika*. Printed in Great Britain. Vol.83, No.1, 1-13.

Statistique Canada. *Aperçu de l'Enquête nationale sur la santé de la population* 1996-1997 - Catalogue 82-567.

Statistique Canada. *Guide du fichier de microdonnées à grande diffusion de l'ENSP* 1996-1997.

Statistique Canada. *Enquête nationale sur la santé de la population*, novembre 2006. Volet ménages. Cycle 6 (2004-2005). Documentation des variables dérivées et des variables longitudinales constantes (Spécifications) Cycles 1 à 6.

Statistique Canada. *Enquête nationale sur la santé de la population (ENSP)*, novembre 2006. Cycle 6 (2004-2005). Dictionnaire de données. Fichier maître: Longitudinal carré (Arrondi).

Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data. MIT Press*, Cambridge.

Wooldridge, Jeffrey. 2006, « Cluster-sample methods in applied econometrics, an extended analysis ». Mimeo Department of Economics, Michigan State University.