Assurances et gestion des risques
Insurance and Risk Management

Modélisation statistique du coût d’une assurance parentale pour les étudiants de cycles supérieurs au Québec

Jean-Philippe Boucher

Volume 79, numéro 3-4, 2011–2012

URI : https://id.erudit.org/iderudit/1091875ar
DOI : https://doi.org/10.7202/1091875ar

Aller au sommaire du numéro

Éditeur(s)

Faculté des sciences de l'administration, Université Laval

ISSN

1705-7299 (imprimé)
2371-4913 (numérique)

Découvrir la revue

Citer cet article

Boucher, J.-P. (2011). Modélisation statistique du coût d’une assurance parentale pour les étudiants de cycles supérieurs au Québec. *Assurances et gestion des risques / Insurance and Risk Management*, *79*(3-4), 201–222. https://doi.org/10.7202/1091875ar

Résumé de l'article

Le régime québécois d’assurance parentale (RQAP) existe depuis 2006. Ce système d’assurance matérialise les préoccupations à tendance sociale-démocrate de la population québécoise. Ce système couvre pratiquement la totalité des travailleurs au Québec, mais ne s’applique pas aux étudiants. Sachant qu’ils sont dans l’âge de procréer, il pourrait être intéressant de se questionner sur la possibilité d’étendre le régime actuel d’assurance parentale aux étudiants de maîtrise et de doctorat afin de leur donner une protection sociale dans le cas d’une naissance. À l’aide de données provenant d’un sondage auprès de la population étudiante d’études de cycles supérieurs, le présent papier propose une approche probabiliste et statistique pour prédire le nombre de naissances de parents étudiant à la maîtrise ou au doctorat au Québec. Par la suite, à l’aide de la modélisation de la natalité, nous évaluerons le coût de l’introduction d’une assurance parentale pour étudiants de cycles supérieurs au Québec. En estimant certaines informations et en posant certaines hypothèses, nous avons pu calculer un coût d’assurance d’environ 6.57 $ millions de dollar pour ce régime. Sachant que de nombreux régimes d’assurances collectives au Québec majorent le régime d’assurance parentale du Québec lorsqu’un employé quitte pour un congé parental, nous croyons que le modèle proposé pourrait aussi être utilisé pour prédire les coûts de cette couverture d’assurance.

Assurances et gestion des risques, vol. 79(3-4), octobre 2011-janvier 2012, 201-222
Insurance and Risk Management, vol. 79(3-4), October 2011-January 2012, 201-222

Modélisation statistique du coût d'une assurance parentale pour les étudiants de cycles supérieurs au Québec

par Jean-Philippe Boucher

RÉSUMÉ

Le régime québécois d'assurance parentale (RQAP) existe depuis 2006. Ce système d'assurance matérialise les préoccupations à tendance sociale-démocrate de la population québécoise. Ce système couvre pratiquement la totalité des travailleurs au Québec, mais ne s'applique pas aux étudiants. Sachant qu'ils sont dans l'âge de procréer, il pourrait être intéressant de se questionner sur la possibilité d'étendre le régime actuel d'assurance parentale aux étudiants de maîtrise et de doctorat afin de leur donner une protection sociale dans le cas d'une naissance. À l'aide de données provenant d'un sondage auprès de la population étudiante d'études de cycles supérieurs, le présent papier propose une approche probabiliste et statistique pour prédire le nombre de naissances de parents étudiant à la maîtrise ou au doctorat au Québec. Par la suite, à l'aide de la modélisation de la natalité, nous évaluerons le coût de l'introduction d'une assurance parentale pour étudiants de cycles supérieurs au Québec. En estimant certaines informations et en posant certaines hypothèses, nous avons pu calculer un coût d'assurance d'environ 6.57 $ millions de dollar pour ce régime. Sachant que de nombreux régimes d'assurances collectives au Québec majorent le régime d'assurance parentale du Québec lorsqu'un employé quitte pour un congé parental, nous croyons que le modèle proposé pourrait aussi être utilisé pour prédire les coûts de cette couverture d'assurance.

L'auteur :

Jean-Philippe Boucher : Quantact / Département de Mathématiques, Université du Québec à Montréal.

L'auteur aimerait remercier le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et le Fonds de recherche du Québec – Nature et technologies (FRQNT) pour leur support financier. L'auteur aimerait aussi souligner l'aide et le support financier de la Fédération étudiante universitaire du Québec (FEUQ).

Mots clés : Naissances, assurance parentale, données de comptage, modèle de Poisson non-homogène, temps d'attente, somme composée.

ABSTRACT

The Quebec Parental Insurance Plan (QPIP) exists since 2006. This system covers almost all workers in Quebec, but does not include the student population. Knowing they are in the childbearing age, it might be interesting to evaluate about the possibility of extending the current system of parental insurance for graduate students. Using data from a survey of the graduate students population, this paper proposes a probabilistic and a statistical approaches to predict the number of births from parents that are in completing graduate studies. Using the modeling of the birth rate, we will evaluate the cost of the introduction of parental insurance for graduate students in Quebec. Under certain assumptions, we are able to calculate a cost of insurance of about 6.57$ million for this plan. Knowing that many group insurance plans increase the QPIP benefits for their employees, we believe that the proposed model could also be used to predict the costs of such insurance coverage.

Keywords: Births, parental insurance, count data, non-homogeneous Poisson, waiting times, compound sum.

1. INTRODUCTION

Le régime québécois d'assurance parentale (RQAP) matérialise les préoccupations à tendance sociale-démocrate de sa population. Le RQAP a pour but de favoriser la conciliation travail-famille en assurant une sécurité financière aux nouveaux parents lors de la naissance ou l'adoption d'un enfant. Il permet également aux parents qui désirent avoir des enfants de conserver un lien avec le monde du travail sans avoir à se limiter dans la poursuite de leurs buts sur le marché du travail. Les prestations du RQAP peuvent être versées selon deux régimes : le régime de base et le régime particulier. Les deux régimes comprennent trois types de prestation, c'est-à-dire les congés de maternité, de paternité et parentaux dont la durée et le pourcentage de couverture salariale diffèrent. Ces différences apportent une certaine adaptabilité du régime face aux réalités du marché du travail.

La population qui entreprend des études de cycles supérieurs est très hétérogène en ce qui a trait à leur manière de financer leurs études de deuxième ou de troisième cycle. Les financements peuvent provenir de plusieurs sources : de bourses d'excellence et de récompenses au mérite, de revenus d'emploi à temps partiel ou à temps plein, du soutien du directeur de recherche, d'aides pour les stages, colloques et activités de formation et de l'aide financière aux études du gouvernement du Québec (AFE). Les bourses et souti ens proviennent eux-mêmes de multiples organismes, dont les organismes subventionnaires gouvernementaux (ex.: CRSNG, CRSH, FQRNT), de fonds universitaires, de sociétés et d'industries. Selon une étude effectuée récemment ([CNCS(2007)]) l'importance relative des sources de revenu des étudiants varie grandement selon l'âge de l'étudiant, le niveau d'étude, le secteur d'étude, l'appartenance à un groupe de recherche, le statut de citoyenneté et l'établissement d'enseignement. Le gouvernement québécois a implanté certaines mesures pour aider financièrement les parents bénéficiaires de l'aide financière aux études (AFE). Les mesures actuelles de conciliation famille-études offertes par l'AFE permettent aux étudiants avec des enfants à charge de continuer à recevoir des prêts et bourses malgré une inscription à temps partiel. En règle générale, le programme de prêts et bourses est offert aux étudiants à temps plein. D'autres mesures augmentent les dépenses admises pour l'obtention d'aide financière aux étudiants bénéficiaires de prêts et bourses du gouvernement du Québec. Il s'agit, entre autres, de la couverture des frais de subsistance pour un enfant, des médicaments et des frais de garde pour une place subventionnée en centre de la petite enfance. Une prolongation de la période d'admissibilité à une bourse, pour permettre de subvenir aux frais liés à un enfant à charge peut également, dans certains cas, faciliter l'adaptation lors de la naissance d'un enfant pendant les études. Il pourrait être intéressant de se questionner sur la possibilité d'étendre le régime actuel d'assurance parentale puisque, malgré ces mesures mises en place par l'aide financière aux études, il laisse une partie de sa population, notamment les étudiants boursiers d'autres organismes subventionnaires ou bourses provenant de l'industrie, vivre avec des ressources financières inadaptées lors de la naissance d'un enfant. Dans ce questionnement sur une possible assurance parentale pour étudiants, un des principaux problèmes est le calcul d'un coût total d'indemnisation.

À l'aide de données provenant d'un sondage fait auprès de la population étudiante de cycles supérieurs, le présent papier propose une approche statistique pour prédire le nombre de naissances des étudiants de maîtrise et de doctorat au Québec. À notre connaissance, il s'agit de la première recherche évaluant ainsi l'impact de la poursuite des études de cycles supérieurs sur le taux de natalité. Par la suite, à l'aide de la modélisation de la natalité, nous évaluerons le coût de l'introduction d'une assurance parentale pour étudiants de cycles d'études supérieures au Québec. Sachant que de nombreux régimes d'assurances collectives au Québec majorent le régime d'assurance parentale du Québec lorsqu'un employé quitte pour un congé parental, nous croyons que le modèle proposé pourrait aussi être utilisé pour prédire les coûts de cette couverture d'assurance.

Dans la section 2, nous décrivons les données utilisées pour l'étude, soit les taux de fécondité de la population québécoise provenant de l'Institut de la Statistique du Québec et un sondage effectué par le Conseil national des cycles supérieurs (CNCS) en 2006, sur plus de 1500 étudiants de cycles supérieurs au Québec. Dans la section suivante, nous effectuons une revue de la littérature pour la modélisation de la natalité d'une population globale, mais proposons aussi une nouvelle approche pour le type de données utilisées. En effet, en nous basant sur des modèles probabilistes de comptage, nous proposons une manière d'estimer le taux de natalité observée pour une population sondée. Appliqué au sondage du CNCS, ce modèle nous permettra de comparer le taux de natalité de la population étudiante à celle de la population générale au Québec. Également, le calcul du coût de l'assurance parentale, selon divers scénarios, sera proposé à la section 4. La dernière partie de l'article conclut.

2. SOURCES EXTERNES ET DONNÉES UTILISÉES

Les prestations du RQAP peuvent être versées selon deux régimes : le régime de base et le régime particulier. Les deux régimes comprennent trois types de prestation, c'est-à-dire les congés de maternité, de paternité et parentaux dont la durée maximale et le pourcentage de couverture salariale diffèrent. Par exemple, le congé parental qui peut être choisi par la mère ou par le père, couvre 70 % du salaire lors des 7 premières semaines et 55 % du salaire ensuite pour le régime de base. Le régime particulier propose une couveture à 75 % du salaire, mais est limitée à 25 semaines au lieu de 32 (25+7) semaines. Ces différences, illustrées dans le tableau 1, apportent une certaine adaptabilité du régime face aux réalités du marché du travail.

Au niveau de la population étudiante au Québec, le site du Ministère de l'Éducation et des Loisirs du Québec (MELS) fournit les informations de base au niveau du nombre d'étudiants universitaires, ventilé par l'âge de ces derniers. Le tableau 2 indique le nombre d'étudiants de cycles supérieurs, par âge, pour l'année 2006.

TABLEAU 1
CARACTÉRISTIQUES DU RÉGIME DE BASE ET DU RÉGIME PARTICULIER DU RQAP

Types de prestation	Régime de base		Régime particulier	
	Nombre maximal de semaines de prestation	Pourcentage du revenu hebdo. moyen	Nombre maximal de semaines de prestation	Pourcentage du revenu hebdo. moyen
Congé de maternité	18	70 %	15	75 %
Congé de paternité	5	70 %	3	75 %
Congé parental	7	70 %	25	75 %
	25	55 %		

TABLEAU 2
POPULATION ÉTUDIANTE AUX CYCLES SUPÉRIEURS AU QUÉBEC PAR GROUPES D'ÂGE ET SEXE

Âge	Maîtrise		Doctorat	
	Homme	Femme	Homme	Femme
20 – 24	4161	6032	267	420
25 – 29	6460	8441	2653	2461
30 – 34	3836	3791	1760	1376
35 – 39	2545	2545	864	587
40 – 44	1746	2058	533	400
45 – 49	1079	1604	323	333
50 – 59	710	1066	273	307
> 60	97	98	46	56
Total	20634	25635	6719	5940

2.1 Taux de fécondité, Institut de la Statistique du Québec

On peut définir le taux de fécondité, ou plus précisément l'*indice synthétique de fécondité des femmes*, comme étant le rapport entre le nombre total de naissances vivantes et le nombre total de femmes. De manière plus générale, on peut définir le taux de natalité comme le rapport entre le nombre total de naissances vivantes et la population. Selon la population choisie, nous évaluerons ainsi le taux de natalité pour les hommes et le taux de natalité pour les femmes. Dans ce papier, nous prendrons l'hypothèse que le taux de fécondité québécois possède certaines similitudes au taux de natalité.

L'analyse graphique de la courbe de l'année 2006 de la figure 1 nous permet de faire quelques observations. Nous pouvons constater que le taux de fécondité tend à augmenter avec l'âge, jusqu'à atteindre un point maximum vers 30 ans, pour ensuite diminuer jusqu'à l'âge de 45 ans. L'analyse des autres courbes de la figure 1 nous confirme que ce type de forme peut être observé à chaque année calendaire. À partir de l'analyse de ces taux de fécondité, il convient de nous questionner sur la forme du taux de fécondité (ou taux de natalité) des étudiants de cycles supérieurs. En effet, pour les étudiants, nous pouvons penser qu'il est probable d'observer le même genre de croissance et de décroissance du taux.

Parallèlement, puisque par exemple une personne de 20 ans en 2001 est âgé de 21 ans en 2002, et de 22 ans en 2003, l'analyse de taux de natalité peut aussi s'effectuer sous l'angle de cohorte, comme cela est fait maintenant dans l'analyse de l'assurance-vie (voir le livre de Delwarde et Denuit (2005) pour une revue des modèles de mortalité sous un angle de cohorte). Cette analyse pourrait nous assurer de l'évolution de la forme du taux de fécondité au cours des prochaines années. La figure 2 illustre l'impact de l'effet de cohorte sur le taux de fécondité, où chaque courbe contenant 8 points (les années 2001 à 2008) représente l'évolution du taux de fécondité des femmes nées la même année au cours des années 2001 à 2008. Bien qu'il semble bel et bien y avoir un effet de cohorte, ou encore un effet de génération sur le taux de fécondité (les femmes de 30 ans en 2008 ont plus d'enfants que les femmes de 30 ans en avaient en 2001, par exemple), la forme du taux de fécondité reste assez similaire : une courbe unimodale, symétrique ayant une forme de cloche.

2.2 Taux de natalité, sondage effectué par le CNCS

La modélisation de la natalité constitue une partie importante de l'analyse de l'assurance parentale pour les étudiants. Les diverses agences gouvernementales ou instituts statistiques collectent le taux de fécondité de la population, tel qu'illustré à la figure ??. Toutefois,

FIGURE 1
TAUX DE FÉCONDITÉ DE LA POPULATION DE 2001 À 2008 SELON L'ÂGE DE LA MÈRE

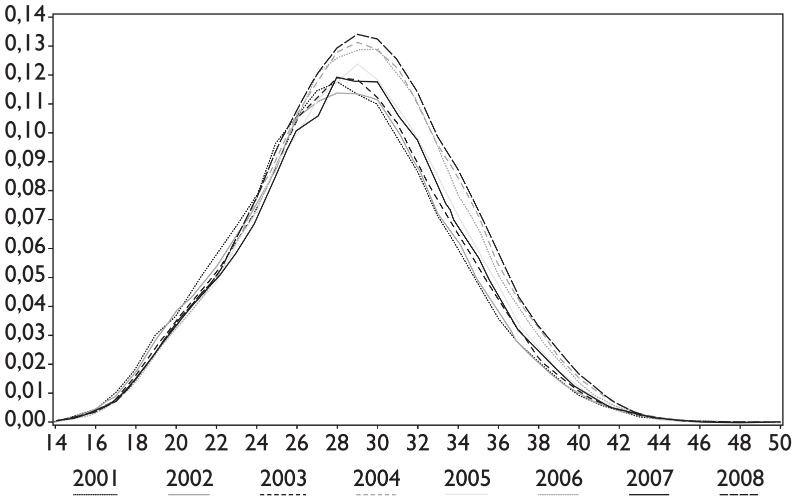
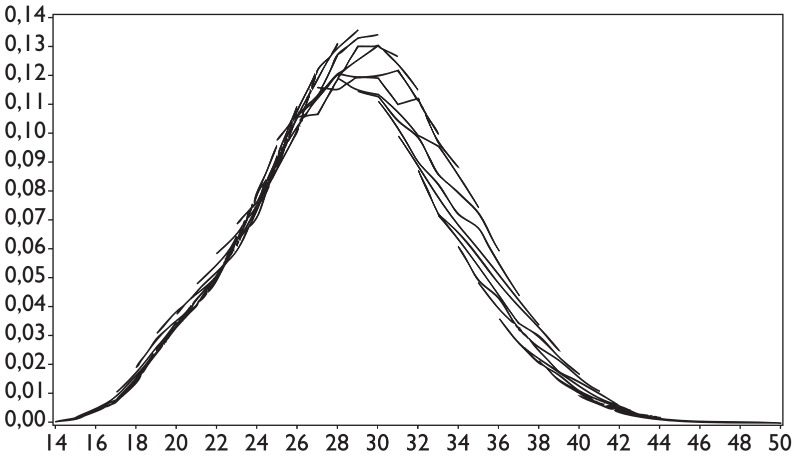


FIGURE 2
TAUX DE FÉCONDITÉ (SOUS FORME DE COHORTE) DE LA POPULATION DE 2001 À 2008 SELON L'ÂGE DE LA MÈRE



aucune statistique n'est collectée au niveau de l'occupation des parents (aux études ou non) lors de la naissance. Ainsi, il n'est pas possible de connaître le nombre d'enfants nés de parents aux études lors de la naissance. Il convient de trouver une façon d'estimer le plus adéquatement possible ce taux de natalité des étudiants.

Nous utilisons un sondage du CNCS en 2006, effectué sur 1531 étudiants étant inscrits dans un programme de maîtrise ou de doctorat. Le sondage ne contient pas l'information précise concernant les naissances d'enfants, ou plus précisément concernant le nombre d'enfants nés pendant que le parent était aux études. Cette information aurait pu nous permettre de modéliser directement le taux de natalité. Toutefois, nous avons accès au nombre d'enfants à charge et à l'âge de l'étudiant. À partir du taux de fécondité de la population générale, et des informations obtenues par le sondage effectué par le CNCS, il est toutefois possible d'avoir un estimé du taux de natalité, tel que nous le montrerons à la section suivante.

3. MODÈLE PROBABILISTE DU NOMBRE DE NAISSANCES

3.1 Modèle de Poisson non-homogène

Beaucoup de papiers et de livres ont été publiés en démographie pour l'analyse de la natalité. Le livre de Alho et Spencer (2005) est une excellente référence. Toutefois, les modèles utilisés en démographie ne se basent pas sur des bases de données contruites comme celle du CNCS. En effet, les analyses sont davantage sur des populations entières et les projections à long terme, que sur de petits groupes d'individus pour des prédictions à court terme.

Le livre de Alho et Spencer (2005) propose néanmoins certains modèles probabilistes pouvant être utiliser dans notre étude. En effet, en utilisant un processus de Poisson non-homogène de fonction d'intensité $\lambda(t)$, où $\lambda(t)$ représente le taux de natalité, il peut être possible de faire une utilisation adéquate d'une base de données contenant l'âge et le nombre d'enfants à charge d'une personne. En effet, selon un modèle de Poisson non-homogène, pour un individu i âgé de T années, il peut être montré que la distribution de probabilité d'avoir un nombre d'enfants à charge égal à n peut être exprimé comme :

$$\Pr(N_i(T) = n) = \frac{m_i(T)^n \exp(-m_i(T))}{n!} \tag{1}$$

où N_i est la variable aléatoire modélisant le nombre d'enfants à charge de l'individu i, et $m_i(T) = \int_0^T \lambda_i(t)dt$. L'indice i permet d'utiliser une fonction de lien, tel que $\exp(X_i'\beta)$, et d'ainsi introduire dans le modèle certaines caractéristiques de l'assuré, tel que son sexe, son état civil, son domaine d'étude, etc. Cette forme d'analyse de la natalité partage d'ailleurs de nombreuses similitudes avec les bases de données classiques en assurance de dommage. En effet, le format des bases de données en assurance automobile, par exemple, est assez similaire (voir Boucher et al.(2007) par exemple), et l'analyse du nombre de réclamations en assurance peut fortement s'apparenter à l'analyse du nombre de naissances. Nous proposons que la fonction d'intensité $\lambda(t)$ soit approximée paramétriquement. La section 4 propose d'ailleurs une forme paramétrique gaussienne pour $\lambda_i(T)$.

3.2 Naissances multiples

Sachant que la sondage ne contient pas le nombre d'accouchements, mais bien le nombre de naissances (ce qui ne correspond pas dans le cas de naissances multiples), il convient d'adapter le modèle. En effet, dans le cas de naissances multiples, la prestation d'assurance reste la même. Un modèle particulier semble convenir à cette situation. Récemment, Boucher et al. (2007) ont utilisé une nouvelle distribution pour le nombre de réclamations en assurance : la *binomiale négative X*. Ce modèle est basé sur une somme composée correspondant à :

$$N = \sum_{j=1}^{Y} X_j \tag{2}$$

où les X_j sont des variables aléatoires entières, indépendentes and identiquement distribuées, et où Y et les X_j sont indépendants. Lorsque Y est supposé Poisson de moyenne μ et X_j est logarithmique avec paramètre η, il en résulte que N est binomial négatif (en utilisant l'hypothèse standard que $\sum_{j=1}^{Y} X_j = 0$ si $Y = 0$). Ce modèle semble être intuitivement valide car la variable Y se veut définir le nombre d'accouchements, alors que chaque X_j modélise le nombre de naissances pour chaque accouchement j. En conséquence, la variable aléatoire N définit le nombre total d'enfants.

Un modèle avec régresseurs a été proposé par Santos Silva et Windmeijer (2001) qui ont défini le modèle de régression $NegBin_x$ de la manière suivante : le paramètre η_i de la distribution logarithmique est exprimé par les régresseurs de la manière suivante :

$$\exp(\mathbf{x}_i'\kappa) = \frac{\eta_i}{1-\eta_i}$$

et la distribution de Poisson utilise la fonction $\lambda_i = \exp(\mathbf{x}_i'\hat{\mathbf{a}})$ pour sa moyenne. Conséquemment, N_i est binomial négatif avec paramètres $\lambda_i / \log(1 + \exp(\mathbf{x}_i'\kappa))$ et $\exp(\mathbf{x}_i'\kappa)$. La fonction de probabilité est définie comme :

$$\Pr(N_i = n) = \frac{\Gamma\left(n + \dfrac{\lambda_i}{\log(1+\exp(\mathbf{x}_i'\kappa))}\right)\exp(-\lambda_i)}{\Gamma(n+1)\Gamma\left(\dfrac{\lambda_i}{\log(1+\exp(\mathbf{x}_i'\kappa))}\right)(1+\exp(-\mathbf{x}_i'\kappa))^n}. \quad (3)$$

3.3 Temps d'attente entre deux accouchements

Un autre correctif se doit toutefois d'être ajouté au modèle du nombre d'enfants. En effet, on pourrait croire qu'un temps d'attente entre chaque accouchement se doit d'être ajouté. Sauf pour le cas des adoptions et dans certaines situations exceptionnelles pour la modélisation du nombre d'enfants pour un homme, on peut s'attendre à ce qu'il y ait un temps d'attente minimal entre deux accouchements. En ce sens, la théorie du temps d'attente entre les événements se doit d'être utilisée pour obtenir un bon modèle.

En définissant $m(t) = \int_0^t \lambda(s)ds = \Lambda[t] - \Lambda[0]$, nous pouvons ainsi utiliser le développement suivant :

$$\begin{aligned} Y(T) &= 0 \qquad W_1 > T \\ \Pr(Y(T) = 0) &= e^{-(\Lambda[T]-\Lambda[0])} \end{aligned}$$

où W_1 correspond au temps d'attente avant le premier événement. En supposant W_k comme le temps d'attente entre le k – 1^e et le k^e événement et a comme le temps d'attente minimal entre deux accouchements, nous poursuivons ainsi :

$$\begin{aligned} Y(T) < k \qquad & W_1 + (a+W_2) + (a+W_3) + \ldots + (a+W_k) > T \\ & W_1 + W_2 + \ldots + W_k > T - (k-1)a \\ & S_k > T - (k-1)a \end{aligned} \quad (4)$$

Connaissant le temps d'attente du k^e événement, i.e S_k, pour un modèle de Poisson non-homogène (voir [Ross(1996)] par exemple), nous pouvons donc inscrire la fonction de probabilité du nombre d'événements k comme :

$$\Pr(Y(t) < k) = \Pr(S_k > T - (k-1)a)$$

$$\Pr(Y(t) < k) = \int_{T-(k-1)a}^{\infty} \frac{\lambda(t)}{\Gamma(k)} [m(t)]^{k-1} e^{-m(t)} dt \qquad (5)$$
$$= 1 - G\left(k; \Lambda[T-(k-1)a] - \Lambda[0]\right)$$

avec :

$$G(k, \lambda t) = \frac{1}{\Gamma(k)} \int_0^t \lambda^k \nu^{k-1} e^{-\lambda \nu} d\nu = \sum_{i=0}^{k} \frac{(\lambda t)^i e^{-\lambda t}}{i!}$$

Ainsi, la fonction de probabilité du nombre d'événements Y se produisant dans l'intervalle de temps $(0,T)$ s'écrit comme :

$$\begin{aligned} \Pr(Y(T) = k) &= \Pr(Y(T) < k+1) - \Pr(Y(T) < k) \\ &= \left(1 - G\left(k+1; \Lambda[T-k\,a] - \Lambda[0]\right)\right) - \left(1 - G\left(k; \Lambda[T-(k-1)a] - \Lambda[0]]\right)\right) \qquad (6) \\ &= G\left(k+1; \Lambda[T-k\,a] - \Lambda[0]\right) - G\left(k; \Lambda[T-(k-1)a] - \Lambda[0]]\right) \end{aligned}$$

Puisque la variable Y n'est plus poissonnienne, le modèle binomiale négative X ne peut plus être utilisé. Nous proposons plutôt l'approche suivante où $n \geq k$, i.e. que le nombre d'enfants est plus grand ou égal au nombre d'accouchements :

$$\begin{aligned} \Pr(N(T) = n) &= \sum_{k=0}^{n} Pr(X_1 + X_2 + ... + X_i = n \mid Y(T) = k) \Pr(Y(T) = k) \\ &= \sum_{k=0}^{n} Pr(X_{1'} + X_{2'} + ... + X_{n'} = n - k) \Pr(Y(T) = k), \\ &= \sum_{k=0}^{n} \frac{(\gamma k)^{n-k} e^{-\gamma k}}{(n-k)!} \Pr(Y(T) = k) \qquad (7) \\ &= \sum_{k=0}^{n} \frac{(\gamma k)^{n-k} e^{-\gamma k}}{(n-k)!} \begin{pmatrix} G\left(k+1; \Lambda[T-k\,a] - \Lambda[0]\right) \\ -G\left(k; \Lambda[T-(k-1)a] - \Lambda[0]]\right) \end{pmatrix} \end{aligned}$$

où X_i' n'est qu'une translation de X_i tel que $X_i' = X_i + 1$ puisque chaque naissance crée au moins un nouvel enfant. Ainsi, en supposant que le taux de natalité à l'âge 0 ($\Lambda[0]$) est nul, nous pouvons donc exprimer la fonction de probabilité du nombre d'enfants de l'étudiant i comme :

$$\Pr(N_i(T) = n) = \begin{cases} e^{-\Lambda_i[T]} & n = 0 \\ \sum_{k=1}^{n} \frac{(\gamma_i k)^{n-k} e^{-\gamma_i k}}{(n-k)!} \left(G\left(n+1; \Lambda_i[T-k\,a]\right) - G\left(n; \Lambda_i[T-(k-1)a]\right)\right) & n > 0 \end{cases}. \qquad (8)$$

3.4 Hétérogénéité

Tout comme l'analyse de la sinistralité en assurance dommage, plusieurs caractéristiques d'un individu analysé dans le modèle (8) ne peuvent pas être utilisées dans la modélisation. En effet, certains éléments ne sont pas disponibles dans l'analyse (mortalité infantile, adoption, intérêt d'avoir des enfants, problème de fertilité, activité sexuelle importante, etc.). Néanmoins, certaines de ces variables ont un impact sur le taux de natalité d'un individu. Afin de modéliser ces caractéristiques inconnues de chaque étudiant, nous introduisons un paramètre d'hétérogénéité θ dans le modèle. Cette modélisation Bayésienne empirique est une alternative intéressante (voir Boucher et Denuit (2008) pour une utilisation en assurance automobile) car elle permet de modéliser statistiquement plusieurs éléments inconnus de chaque étudiant.

Mathématiquement, on suppose que le facteur d'hétérogénéité est ajouté au paramètre de moyenne de la distribution de comptage. On dénote par h(·) la distribution de l'hétérogénéité θ, et nous supposons que celle-ci est distribuée selon une loi gamma de moyenne 1. Ainsi, pour n > 0 :

$$\begin{aligned} P(N_i(T)=n) &= \int_0^\infty P(N_i(T)=n \mid \theta)h(\theta)d\theta \\ &= \int_0^\infty \sum_{k=1}^{n} \frac{(\gamma_i k)^{n-k} e^{-\gamma_i k}}{(n-k)!} \\ &\quad \begin{pmatrix} G\left(n+1; \Lambda_i[T-k\,a]\theta\right) \\ -G\left(n; \Lambda_i[T-(k-1)a]\theta\right) \end{pmatrix} h(\theta)d\theta \end{aligned} \tag{9}$$

Cette dernière intégrale peut se calculer numériquement, à l'aide de certain logiciel statistique par exemple. Toutefois, le lecteur intéressé peut consulter Boucher et Denuit (2007) pour connaître une forme de la formule fermée de $\int_0^\infty G(.;\theta)h(\theta)d\theta$ lorsque θ suit une loi gamma.

4. APPLICATION À LA POPULATION ÉTUDIANTE

Le choix d'une fonction pour le taux de natalité $\lambda(t)$ se doit d'être choisie. La forme Gaussienne est la forme la plus intuitive, et apparaissant à première vue la plus intéressante. En effet, les figures

1 à 2 nous montrent clairement des distributions à cloche typiquement Gaussiennes. On suppose donc :

$$\lambda(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{(t-\mu)^2}{\sigma^2} \right)\right) \tag{10}$$
$$\Lambda(T) = \Phi(T;\mu,\sigma^2)$$

où $\Phi(T|\mu,\sigma^2)$ est la fonction cumulative de la loi normale. Même si l'utilisation d'une intensité gaussienne est proposée, nous ne sommes pas sans savoir que son utilisation peut générer quelques inconvénients. En effet, un problème potentiel à cette approche est le domaine de la distribution qui couvre l'ensemble des nombres réels (négatifs et positifs), impliquant donc des taux de natalité non-nuls pour des âges négatifs. Un autre problème de cette fonction est sa difficulté a bien modéliser le taux de natalité des étudiants. En effet, il convient de supposer que l'âge moyen pour une première naissance est plus élevé pour des parents ayant effectués des études avancées que pour le reste de la population. L'effet de symétrie de la forme Gaussienne pourrait donc ne pas être approprié puisqu'il impliquerait des taux de natalité positif pour des âges trop avancés. Similairement, dans un but d'analyser le taux de natalité de groupes un peu plus précoce (étudiants du secondaire professionnel), l'effet de symétrie impliquerait des taux de natalité positif pour des enfants de moins de 10 ans.

Pour obtenir la courbe de natalité des étudiants de cycle supérieurs, il n'est pas possible de modéliser directement les données du sondage à partir des équations (8) et (9). En effet, le problème d'une telle modélisation est qu'elle suppose que les étudiants du sondage ont toujours été aux études. Cette supposition n'a pas de sens pour une partie importante des étudiants sondés. En effet, alors qu'il semble logique qu'un doctorant de 25 ans, ayant débuté ses études de troisième cycle à 24 ans, ait toujours été aux études (depuis l'âge de 5 ans jusqu'au moment du sondage), cette situation ne tient pas pour toute la population étudiante. En effet, à l'opposé, il semble hautement improbable qu'un étudiant de 36 ans, indiquant avoir débuté ses études de maitrise à 35 ans, soit resté dans le milieu académique toute sa vie. Cet étudiant a certainement quitté le monde universitaire pendant quelques années à un moment ou l'autre de sa vie. Il arrive fréquemment que des étudiants quittent le monde universitaire (pour plusieurs raisons) afin d'aller dans le marché du travail, pour ensuite revenir aux études. Ainsi, il est important de comprendre que le nombre d'enfants à charge indiqué dans le sondage du CNCS ne correspond pas au nombre d'enfants nés pendant les études. Il n'est donc pas possible d'utiliser directement cette information pour modéliser le taux de natalité des étudiants.

Nous proposons ainsi de découper la fonction d'intensité $\lambda(t)$ en divers morceaux afin qu'elle puisse mieux correspondre à la réalité des étudiants. Ainsi, nous proposons 3 composantes pour le taux de natalité :

1. Le taux de natalité lors de leurs études secondaire, collégiale et universitaire (typiquement jusqu'à l'âge 27 ans). Nous utilisons un $\lambda(t)$ de forme gaussienne, puisque similaire à la forme observée du taux de natalité, pour modéliser le taux de natalité de ces étudiants;

2. Un taux de natalité pour les étudiants ayant quitté le monde universitaire. Dans la base de données utilisée, nous avons l'âge auquel ils ont commencé leurs études de cycles supérieurs. Ainsi, nous supposons que les étudiants ayant commencé leurs études de maitrise après l'âge de 27 ans et que les doctorants ayant débuté après 29 ans n'ont pas toujours été étudiants (ces valeurs ont été déterminé par diverses comparaisons statistiques). Pour ces étudiants, nous modélisons ainsi différemment l'intensité du modèle entre l'âge de 27 ans et leur âge déclaré de retour aux études. On suppose que le taux de natalité de cette période de temps se comporte comme celui de la population générale, à un facteur de proportionnalité près dépendant du sexe de l'individu.

3. Pour les étudiants n'ayant pas toujours été étudiant, nous utilisons ensuite un autre taux de natalité pour la période de leur vie correspondant à leur âge déclaré de retour aux études et leur âge actuel. Dans ce cas précis, nous supposons aussi que le taux de natalité se comporte comme celui de la population générale, à un facteur de proportionnalité près.

4.1 Résultats

Nous utilisons les données du sondage de la CNCS, à partir du modèle décrit à la sous-section précédente, et en utilisant l'équation (9) avec une intensité $\lambda(t)$ de forme gaussienne telle qu'exprimée à l'équation 10. Une hétérogénéité log-normale de paramètres $\mu_{\text{Lognormale}} = -\eta^2/2$ et $\sigma^2_{\text{Lognormale}} = \eta^2$ (différent du σ^2 de l'intensité) est utilisée. En utilisant les régresseurs dans le modèle de comptage, nous pouvons ainsi exprimer la probabilité d'avoir $N = n$ enfants à charge comme :

$$P(N_i(T)=n_i)=\int_0^{\infty}\sum_{k=1}^{n}\frac{(\gamma_i k)^{n-k}e^{-\gamma_i k}}{(n-k)!}\begin{pmatrix}G\left(n+1;\Lambda_i[T-k\,a]\theta\right)\\-G\left(n;\Lambda_i[T-(k-1)a]\theta\right)\end{pmatrix}h(\theta)d\theta$$

avec :

$$\Lambda_i(T)=\exp(\beta_1)\Phi(T;\mu,\sigma^2)+\exp(\beta_2)*c_2+\exp(\beta_3)*c_3 \qquad (11)$$

où μ, σ, β_1, β_2 et β_3 sont des paramètres à évaluer, en plus du paramètre η^2 provenant de la distribution de l'hétérogénéité. En utilisant les données de l'Institut de la Statistique du Québec sur les naissances multiples, nous supposons γ_i = exp(–3,4) = 0.0333 (signifiant 1.0333 bébé par naissance en moyenne). Les variables c_2 et c_3 représentent l'intensité cumulée des étudiants pour les périodes de temps pendant lesquelles ils n'étaient pas aux études, et pendant lesquelles ils sont retournées aux études (points 2 et 3 de la sous-section précédente).

Avec ce modèle, nous obtenons les paramètres exprimés dans le tableau 3.

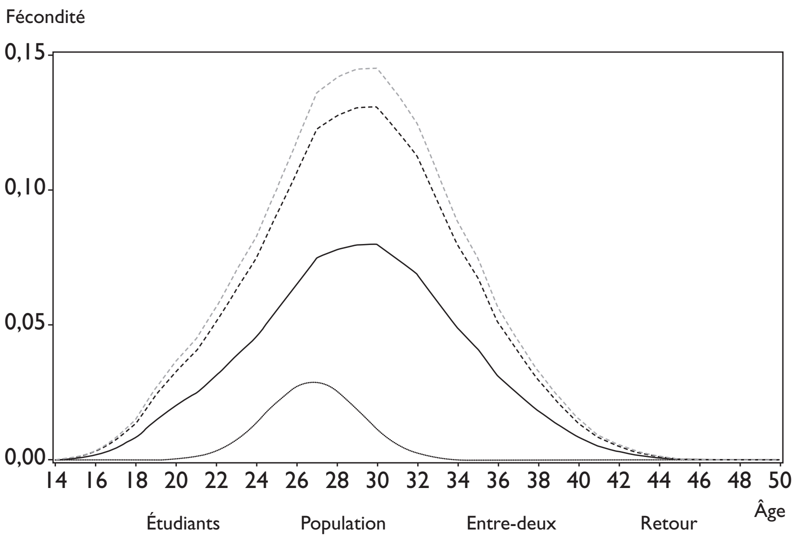
TABLEAU 3
ESTIMATEURS DES PARAMÈTRES POUR LE MODÈLE DE COMPTAGE

Paramètre	Estimateur	Écart-type
β_1	-1.7902	0.3711
β_2	0.1045	0.0889
β_3	-0.4922	0.4196
μ	26.3172	1.3853
σ^2	2.2872	0.9143
η^2	0.5579	0.1673

Le graphique 3 expose la courbe du taux de natalité selon l'âge pour divers types de population. L'analyse statistique du taux de natalité de ces différents groupes nous permet de constater que le taux de natalité des étudiants (courbe « Étudiants »), qui n'ont jamais quitté le milieu scolaire, et qui sont actuellement aux études supérieures, est bien inférieure au taux de la population (courbe « Population »). La courbe « Entre-deux » représente le taux de natalité des individus qui ne sont pas aux études, mais qui y reviendront pour des études supérieures. Ce taux de natalité est supérieur au taux de natalité de la population.

Lors d'un retour aux études, l'analyse statistique basée sur le sondage de la CNCS indique que le taux de natalité de la population étudiante est encore une fois inférieure à celui de la population. Ces observations nous permettent de voir que la poursuite d'études universitaires diminuent fortement le taux da natalité, plus particulièrement dans le cas des étudiants poursuivant leur formation dans des programmes d'études de cycles supérieurs.

FIGURE 3
TAUX DE NATALITÉ DES ÉTUDIANTS SELON DIVERSES CLASSES



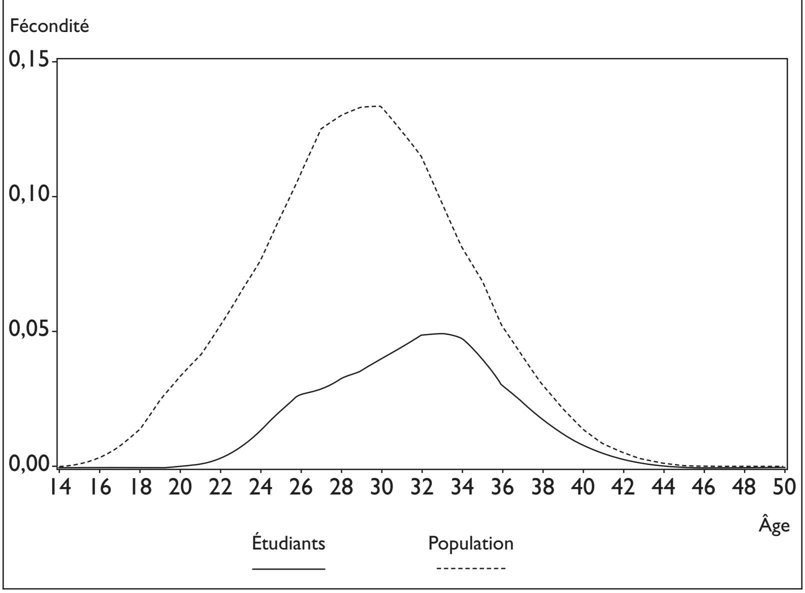
4.1.1 Taux de natalité de tous les étudiants

En ayant obtenu la courbe du taux de natalité de divers profils d'étudiants, il convient maintenant de modéliser le taux de natalité de tous les étudiants, peu importe s'ils sont de retour aux études ou non. En effet, l'intérêt est de déterminer le taux de natalité d'individus étant aux études. Ainsi, les courbes modélisant le taux de natalité des hommes et des femmes ayant quitté l'université n'a pas d'intérêt dans notre analyse (courbes « entre-deux »). La figure 4 illustre le résultat de la modélisation du taux de natalité des étudiants, où le taux de natalité a été construit en pondérant adéquatement le nombre d'étudiants provenant d'un parcours entièrement académique et les étudiants revenant aux études après un exil sur le marché du travail, par exemple. Cette pondération a été effectuée en se basant sur le sondage du CNCS, où l'on peut constater que la proportion par âge d'étudiants qui n'ont jamais quitté le milieu scolaire peut s'approximer par l'équation :

$$Prop(\text{âge}) = \max\left[0, \min\left[\frac{\text{âge} - 27}{7}, 1\right]\right] \tag{12}$$

En utilisant cette proportion d'étudiants ayant quitté ou non le milieu académique, nous pouvons ainsi calculer le taux de natalité des étudiants de cycles d'études supérieures, tel qu'illustrer dans le graphique 4.

FIGURE 4
TAUX DE NATALITÉ DE LA POPULATION ÉTUDIANTE GÉNÉRALE



Comparativement au taux de fécondité des femmes au Québec, nous pouvons constater que le taux de natalité des étudiants de cycle d'études supérieures au Québec est beaucoup plus faible. Il pourrait être intéressant de se questionner à savoir si un tel taux est profitable pour le Québec moderne, ou s'il serait préférable que le Québec encourage davantage les étudiants de cycles supérieurs à avoir des enfants.

4.1.2 Nombre de naissances et coût de l'assurance parentale

Les taux de natalité calculés nous permettent d'estimer le nombre total de naissances dont les parents sont étudiants. Avec les données collectées, il est maintenant possible d'estimer les coûts d'un régime d'assurance parentale pour les étudiants. En effet, en utilisant la réparition des étudiants selon leur groupe d'âge, tel qu'illustré plus tôt dans le tableau 2, il pourra être possible de prédire le nombre de naissances provenant de parents aux études (niveau maîtrise et doctorat). Nous utilisons des techniques de lissage linéaire, à travers la procédure PROC EXPAND du système SAS, pour étaler le nombre d'étudiants par âge. Le nombre de naissances projetées par âge de l'étudiant sont exprimés dans le tableau 4. Le tableau indique aussi les résultats obtenus si le taux de natalité des étudiants de cycles supérieurs étaient équivalent à celui de la population globale.

Afin de calculer le coût de l'assurance parentale, nous supposons que le montant de prestation du congé de maternité ou du congé parental correspond à une bourse de 5000 $ par naissance[1]. Nous supposons encore une fois que tous les étudiants ayant des enfants se prévaudront de leurs congés parentaux. De plus, il est à noter que lors de la naissance de l'enfant, l'un ou l'autre des conjoints peut utiliser le RQAP afin de bénéficier d'un congé parental. Selon les statistiques officielles du RQAP, compilées par le Conseil de gestion de l'assurance parentale, la répartition des prestataires servis selon le sexe (du 1er janvier au 30 avril 2009) était de 70,2 % pour les femmes et 29,8 % pour les hommes. Bien entendu, puisque le système d'assurance pour étudiants n'est pas en vigueur en ce moment, il n'est pas possible de connaître ce que sera la répartition des prestations pour le régime d'assurance parentale des étudiants.

On peut distinguer 4 types de parents :

1. Père et mère non-étudiants;

2. Père étudiant, mère non-étudiante;

3. Mère étudiante, père non-étudiant;

4. Père et mère étudiants.

Le premier cas est la seule situation n'ayant pas d'influence sur notre analyse. Toutefois, la statistique officielle de la répartition des prestataires du RQAP selon le sexe est calculée uniquement pour cette situation. Un questionnement par rapport aux trois autres situations est donc pertinent. Aucune statistique, ou exemple comparable n'est disponible pour l'évaluation des ces statistiques. Ainsi, par conservatisme, nous supposons que toutes les mères utiliseront leur congé de maternité et que tous les pères utiliseront leur congé de

TABLEAU 4
PRÉDICTION DES NAISSANCES

Âge	Nombre d'étudiant	Taux de nat. étudiant (%)	Taux de fert. population (%)	Nb. prédits de bébés (étudiant)	Nb. prédits de bébés (population)
20	1183	0.04	3.25	0	40
21	1677	0.12	4.00	2	69
22	2174	0.33	5.04	7	113
23	2670	0.73	6.23	20	172
24	3176	1.36	7.36	44	242
25	3665	2.09	8.92	79	338
26	4161	2.69	10.5	116	450
27	4527	2.87	12.1	134	565
28	4114	3.25	12.6	138	534
29	3547	3.53	12.9	129	471
30	2992	3.95	12.9	122	398
31	2436	4.34	12.0	109	303
32	1940	4.78	11.1	96	222
33	1759	4.83	9.45	88	172
34	1637	4.68	7.87	79	133
35	1516	3.95	6.65	62	104
36	1399	2.99	5.03	43	73
37	1280	2.34	3.94	31	52
38	1206	1.75	2.94	22	37
39	1139	1.25	2.11	15	25
40	1074	0.82	1.37	9	15
41	1004	0.49	0.82	5	8
42	939	0.30	0.51	3	5
43	885	0.16	0.28	2	3
44	835	0.08	0.13	1	1
Total	58928			1358	4546

paternité, signifiant qu'il est donc possible qu'une prestation de 5000 $ soit donnée à la mère et au père pour une seule et même naissance si les deux parents sont étudiants aux cycles supérieurs. Par conservatisme encore une fois, nous supposerons que 100 % des congés sont utilisés par les parents.

Selon ce modèle, en dollars de 2006, l'assurance parentale au Québec aurait donc coûté $5000 \times 1358/(1 + \exp(-3,4)) = 6.57$ millions de dollars. Par étudiant, la valeur de l'assurance est ainsi d'environ 111,50 $.

L'introduction d'un système d'assurance parentale pour étudiants entraînera nécessairement une augmentation du taux de natalité chez les étudiants. Le cas où le taux de natalité se comporterait comme celui de la population générale, correspondant à un coût d'assurance de 22 millions de dollars ou 373 $ par étudiant, nous apparaît ainsi comme la valeur maximale possible du coût de l'assurance parentale pour les étudiants. Cette situation nous semble toutefois extrême et se veut plutôt une borne maximale de l'estimation des coûts.

5. CONCLUSION

Les étudiants aux cycles supérieurs réussissent souvent à subvenir à leurs besoins de manière autonome en ayant recours à l'aide financière aux études, en travaillant à temps partiel, en bénéficiant de bourses d'excellence ou en recevant un support financier de leur directeur de recherche. Toutefois, lorsque ces étudiants deviennent parents, ceux-ci n'ont pas droit à une sécurité financière pour une grande partie de leur revenu s'ils décident de cesser momentanément leurs études. Il est ainsi pertinent de se questionner sur la nécessité d'implanter une assurance parentale pour les étudiants afin de protéger de la précarité une partie de la population.

Le présent papier propose un nouveau modèle de comptage, basée sur une moyenne non-homogène, afin de modéliser le taux de natalité d'un groupe de personnes. Le modèle proposé tient compte des naissances multiples, du temps d'attente minimal entre deux accouchements et de l'hétérogénéité de la population. Le modèle a été appliqué afin de modéliser le taux de natalité des étudiants de maîtrise et de doctorat au Québec. À partir de diverses données publiques, il a été possible de prédire un nombre de naissances pour ce type d'étudiants. Il a aussi été possible d'estimer le coût d'assurance d'un éventuel régime d'assurance parentale pour les étudiants aux cycles supérieurs du Québec. L'estimation est basée sur un sondage du CNCS effectué en 2006 auprès des étudiants aux cycles supérieurs du Québec. Les résultats sont donc des estimations du coût d'un système d'assurance parentale pour étudiants qui aurait été en force en 2006. Les montants indiqués sont en dollars de 2006 et les statistiques des étudiants (tel que le taux de natalité) se doivent d'être comparées avec celles de la population en 2006.

En estimant certaines informations et en posant certaines hypothèses, nous avons pu calculer un coût d'assurance hypothétique pour ce régime. Le coût d'assurance d'un tel régime serait d'environ 111,50 dollars par étudiant. En plus de donner une meilleure sécurité financière aux parents, l'un des objectifs du RQAP était d'augmenter le taux de natalité au Québec. Sachant ainsi qu'il est fort probable que l'introduction d'un tel système d'assurance parentale pourrait augmenter le taux de natalité et ainsi le coût de l'assurance parentale pour étudiant. Dans une situation pouvant être qualifiée d'extrême, où le taux de natalité des étudiants de cycles supérieurs pourrait être équivalent à celui de la population, nous évaluons le coût du régime à 373 dollars par étudiant.

Toutefois, cette augmentation du taux de natalité chez les étudiants pourrait aussi signifier une diminution du taux de natalité des finissants. En effet, en ce moment, les étudiants actuels ont un incitatif financier à attendre la fin de leurs études pour avoir un enfant. Puisque la prestation d'assurance parentale d'un étudiant (5000 $ dans notre modèle) est beaucoup plus faible que celle en vigueur au RQAP (tableau 1), le coût réel de l'introduction d'une assurance parentale pour étudiants pourrait finalement être beaucoup moins importante que ce qui a été calculée.

Références

Alho, J. and Spencer, B. (2005). *Statistical Demography and ForecastingSeries: Springer Series in Statistics*. Springer Series in Statistics.

Boucher, J.-P. and Denuit, M. (2007). Duration Dependence Models for Claim Counts. *Deutsche Gesellschaft fur Versicherungsmathematik (German Actuarial Bulletin)*, 28: 29-45.

Boucher, J.-P. and Denuit, M. (2008). Crédibilité linéaire multivariée utilisant le nombre de périodes avec réclamations : modèles de Poisson, modèles à barrière et modèles gonflés à zéro. *Assurances et gestion des risques*, 75(4).

Boucher, Denuit, and Guillén, Boucher, J.-P., Denuit, M., and Guillén, M. (2007). Risk Classification for Claim Counts: A Comparative Analysis of Various Zero-Inflated Mixed Poisson and Hurdle Models. *North American Actuarial Journal*, 11-4: 110-131.

CNCS (2007). Les sources et les modes de financement des étudiants aux cycles supérieurs. Tech. rep., Conseil national des cycles supérieurs.

Delwarde, A. and Denuit, M. (2005). *Construction de tables de mortalité, Périodiques et prospectives*. Economica.

Ross, S. M. (1996). *Stochastic Processes, 2nd ed.*, New York : Wiley.

Santos Silva, J. and Windmeijer, F. (2001). Two-part Multiple Spell Models for Health Care Demand. *Journal of Econometrics*, 104: 67-89.

Notes

1. Les naissances multiples ne changent pas l'indemnité de 5000 $.