Assurances et gestion des risques
Insurance and Risk Management

Modélisation et estimation des effets individuels et d'entreprise avec des données de panel : une application aux parcs de véhicules

Jean-François Angers, Denise Desjardins, Georges Dionne and François Guertin

Volume 73, Number 4, 2006

EN L'HONNEUR DE / IN HONOR OF : CLAIRE LABERGE-NADEAU

URI: https://id.erudit.org/iderudit/1106607ar
DOI: https://doi.org/10.7202/1106607ar

Publisher(s)

Faculté des sciences de l'administration, Université Laval

ISSN

1705-7299 (print)
2371-4913 (digital)

Cite this article

Angers, J.-F., Desjardins, D., Dionne, G. & Guertin, F. (2006). Modélisation et estimation des effets individuels et d'entreprise avec des données de panel : une application aux parcs de véhicules. *Assurances et gestion des risques / Insurance and Risk Management*, *73*(4), 457–497. https://doi.org/10.7202/1106607ar

Article abstract

In this article, we propose a detailed analysis of the modeling and estimation of accident distributions of vehicles belonging to a fleet. This analysis uses panel data to account simultaneously for vehicle as well as fleet effects. The distribution of accidents can be affected by both observable and non-observable factors. Nonobservable factors are modeled as random effects.

Assurances et gestion des risques, vol. 73(4), janvier 2006, 457-497
Insurance and Risk Management, vol. 73(4), January 2006, 457-497

Modélisation et estimation des effets individuels et d'entreprise avec des données de panel : une application aux parcs de véhicules

par Jean-François Angers, Denise Desjardins, Georges Dionne et François Guertin

RÉSUMÉ

Dans cet article, nous proposons une analyse détaillée de la modélisation et de l'estimation des distributions d'accidents de véhicules appartenant à un parc de véhicules. L'analyse tient compte simultanément des effets véhicules et de parcs de véhicules avec des données de panel. La distribution des accidents peut être affectée par des facteurs observables et non observables. Les facteurs non observables sont modélisés comme des effets aléatoires.

Mots clés :Assurance des parcs de véhicules, effet véhicule, effet parc de véhicules, données de panel, modélisation économétrique, estimation.

ABSTRACT

In this article, we propose a detailed analysis of the modeling and estimation of accident distributions of vehicles belonging to a fleet. This analysis uses panel data to account simultaneously for vehicle as well as fleet effects. The distribution of accidents can be affected by both observable and non-observable factors. Non-observable factors are modeled as random effects.

Keywords: Insurance for fleet of vehicles, vehicle effect, fleet effect, panel data, econometric modeling, estimation.

Les auteurs :

Jean-François Angers, Denise Desjardins, Georges Dionne et François Guertin sont des chercheurs au Centre de recherche sur les transports. Georges Dionne est titulaire de la Chaire de recherche du Canada en gestion des risques.

Texte présenté au Congrès de la Société canadienne de science économique, le 13 mai 2005, et au Colloque en sécurité routière en l'honneur de Claire Laberge-Nadeau, le 19 mai 2005. Nous remercions A. Taamouti et deux arbitres anonymes pour leurs commentaires pertinents. Cette recherche a bénéficié du support financier du programme d'action concertée en sécurité routière MTQ-SAAQ-FCAR.

I. INTRODUCTION

Depuis le début des années 1980, plusieurs chercheurs ont proposé différents modèles pour tenir compte des corrélations résultant de répétitions d'observations dans le temps. En effet, l'utilisation de données individuelles de type panel est devenue populaire dans beaucoup d'applications économiques, que ce soit en économie du travail, en économie des transports ou en économie de l'éducation (Abowd et al., 1999; Dionne et al., 1998; Hausman et Wise, 1979; Hsiao, 1986; Baltagi, 1995).

Dans les applications avec données de comptage, c'est l'article de Hausman et al. (1984) qui semble être la première contribution. Comme nous l'avons toutefois déjà discuté, sa modélisation n'est pas applicable pour le calcul ex post des primes d'assurance (Angers et al., 2004). Ces deux contributions proposent des modèles d'estimation paramétriques, une classe de modèles que nous utiliserons dans cette extension. (Pour des applications en assurance avec des modèles non paramétriques ou semi-paramétriques, voir Pinquet, 2000, et Dionne et al., 2001.)

Abowd et al. (1999) ont ajouté une autre dimension au problème d'estimation en considérant également l'effet entreprise, qui est aussi une source de corrélation entre les observations individuelles. En effet, si nous avons, à chaque période, des observations de travailleurs ou de véhicules provenant d'une même entreprise, ces observations peuvent présenter des caractéristiques (observables et non observables) communes. Il est important que ces caractéristiques soient également bien modélisées. Abowd et al. (1999) ont estimé un modèle linéaire avec des effets fixes.

Si nous nous intéressons aux distributions des accidents de véhicules appartenant à différents parcs de véhicules dans le temps, nous pouvons décomposer les facteurs explicatifs en facteurs hétérogènes reliés aux véhicules et à leurs conducteurs, puis en facteurs hétérogènes reliés aux parcs de véhicules et en facteurs résiduels. Les facteurs reliés aux véhicules et ceux reliés aux parcs de véhicules peuvent être corrélés.

Dans cet article, nous proposons une analyse détaillée de la modélisation et de l'estimation des distributions d'accidents de véhicules tenant compte simultanément des effets individuels et de parcs de véhicules avec des données de panel. La distribution des accidents peut être affectée par des facteurs observables et non observables. Les facteurs non observables sont modélisés comme des effets aléatoires. L'expression « flotte de véhicules » est également utilisée dans cet article pour désigner « parcs de véhicules »

Dans la section 2 qui suit, nous proposons notre modélisation théorique. Dans la section 3, nous présentons les résultats des estimations et nous analysons les résultats en fonction de différents critères de performance statistique. La section 4 propose des extensions à l'analyse en guise de conclusion.

2. MÉTHODOLOGIE

2.1 Modèle économétrique d'estimation des distributions d'accidents de véhicules

La plupart des modèles économétriques appliqués à des variables discrètes (ou de comptage) ont pour point de départ la distribution de Poisson, où la probabilité d'être impliqué dans y_{fij} accidents peut être représentée par l'expression suivante:

$$P(Y_{fij} = y_{fij}) = \frac{e^{-\lambda_{fij}}(\lambda_{fij})^{y_{fij}}}{\Gamma(y_{fij}+1)}.$$

Par définition de la loi de Poisson, nous avons que l'espérance mathématique du nombre d'accidents est égale à la variance, $E(Y_{fij}) = Var(Y_{fij}) = \lambda_{fij}$ où Y_{fij} est le nombre d'accidents du camion *i* à la période *j* de la flotte *f* et $\lambda_{fij} = d_{fij}\, e^{X_{fij}\beta}$ (> 0) est le paramètre de la loi de Poisson. Le paramètre d_{fij} mesure l'exposition au risque, le vecteur X_{fij} mesure les caractéristiques observables du camion *i* de la flotte *f* au temps *j* et β est un vecteur de paramètre à estimer. Cette modélisation suppose implicitement que la distribution d'accidents peut être expliquée entièrement par l'hétérogénéité observable. Il est à noter que la restriction de l'égalité de la moyenne et de la variance n'est pas toujours compatible avec les données. Pour les accidents de la route, il arrive souvent que la variance soit supérieure à la moyenne. Pour tenir compte de cette première propriété, nous pouvons supposer que le paramètre λ_{fij} a un terme aléatoire tel que $\lambda_{fij} = d_{fij}e^{X_{fij}\beta+\varepsilon_i} = \alpha_i\, \gamma_{fij}$ avec $\alpha_i = e^{\varepsilon_i}$ et $\gamma_{fij} = d_{fij}e^{X_{fij}\beta}$. En supposant que α_i suit une distribution gamma de paramètre $(\delta^{-1}, \delta^{-1})$, nous obtenons la distribution binomiale négative de paramètres $(\delta^{-1}, \gamma_{fij})$:

$$P(Y_{fij} = y_{fij}) = \frac{\Gamma(\delta^{-1}+y_{fij})}{\Gamma(\delta^{-1})\,\Gamma(y_{fij}+1)}\left(\frac{\delta^{-1}}{\delta^{-1}+\gamma_{fij}}\right)^{\delta^{-1}}\left(\frac{\gamma_{fij}}{\delta^{-1}+\gamma_{fij}}\right)^{y_{fij}}. \quad (1)$$

Cette modélisation est appropriée pour des observations individuelles indépendantes, c'est-à-dire sans effet entreprise ou effet temps.

2.1.1 Prise en compte du temps

Considérons maintenant le fait que les données sont des observations dans lesquelles les mêmes unités (individus, camions, par exemple) sont observées pendant plusieurs périodes successives mais sans effet entreprise. Hausman, Hall and Griliches (1984) ont proposé une extension du modèle de l'équation (1) pour tenir compte des répétitions des observations dans le temps. Supposons que Y_{fij} est iid selon la binomiale négative de paramètres $\alpha_i\ \gamma_{fij}$ et ϕ_i. En supposant maintenant que $\left(1+\frac{\alpha_i}{\phi_i}\right)^{-1}$ suit une distribution bêta de paramètres (a,b), nous obtenons le modèle binomial négatif à effet aléatoire :

$$P\left(Y_{fi1},\ldots,Y_{fiT_{if}}\right)$$

$$=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\left(a+\sum_j\gamma_{fij}\right)\left(b+\sum_j y_{fij}\right)}{\left(a+b+\sum_j\gamma_{fij}\sum_j y_{fij}\right)}\prod_j\frac{\Gamma(y_{fij}+\gamma_{fij})}{\Gamma(\gamma_{fij})\Gamma(y_{fij}+1)}. \quad (2)$$

Une autre façon d'introduire l'aspect panel des données est de supposer que $\lambda_{fij}=\gamma_{fij}(\alpha_i\theta_{ij})$. Le paramètre α_i est un effet aléatoire associé au camion *i*, c'est-à-dire le risque attribuable au camion, tandis que θ_{ij} ajoute la dimension temporelle et doit être interprété comme étant l'effet aléatoire du camion-temps tel que $\sum_{j=1}^{T_i}\theta_{ij}=1$, où T_i est le nombre de périodes où le camion *i* est observé. θ_{ij} est ainsi la proportion du risque non observable attribuable à la période *j*. Nous faisons l'hypothèse *a priori* que les θ_{ij} suivent une Dirichlet $(\nu_1, \nu_2, \ldots, \nu_{T_i})$ et que α_i suit une densité gamma $(T_i\kappa^{-1}, \kappa^{-1})$.

Avec cette notation, la distribution du nombre d'accidents du véhicule, lorsqu'il n'a qu'une seule période, est donnée par (Angers et al., 2004) :

$$P(Y_{fi1})=\frac{\Gamma(y_{fi1}+\kappa^{-1})}{\Gamma(y_{fi1}+1)\,\Gamma(\kappa^{-1})}\left(\frac{\kappa^{-1}}{\kappa^{-1}+\gamma_{fi1}}\right)^{\kappa^{-1}}\left(\frac{\gamma_{fi1}}{\kappa^{-1}+\gamma_{fi1}}\right)^{y_{fi1}}, \quad (3)$$

ce qui est l'équivalent de l'équation (1).

Nous abordons maintenant la présentation détaillée du modèle pour deux périodes. Ce modèle est une réinterprétation du modèle proposé par Angers et al. (2004) pour tenir compte de l'effet entreprise dans un environnement à une seule période. Nous l'utilisons ici pour introduire uniquement l'effet temps dans une première étape. Nous ajouterons ensuite l'effet flotte dans la section 2.1.2. Commençons par un modèle à deux périodes.

La probabilité du nombre d'accidents observés du camion *i* ayant deux périodes conditionnellement à θ_{i1}, est égale à :

$$P(Y_{fi1}, Y_{fi2} \mid \theta_{i1})$$

$$= \left[\frac{(\gamma_{fi1})^{y_{fi1}} (\gamma_{fi2})^{y_{fi2}} (\theta_{i1})^{y_{fi1}} (1-\theta_{i1})^{y_{fi2}} \kappa^{-2\kappa^{-1}}}{\Gamma(y_{fi1}+1)\,\Gamma(y_{fi2}+1)\,\Gamma(2\kappa^{-1})} \right]$$

$$\times \int_0^{\infty} \left[\alpha_i^{\sum_{j=1}^{2} y_{fij} + 2\kappa^{-1} - 1} \right] \left[e^{-\alpha_i (\theta_{i1}\gamma_{fi1} + (1-\theta_{i1})\gamma_{fi2} + \kappa^{-1})} \right] d\alpha_i .$$

Ainsi, en intégrant par rapport à α_i, le côté droit devient :

$$\left[\frac{(\gamma_{fi1})^{y_{fi1}} (\gamma_{fi2})^{y_{fi2}} (\theta_{i1})^{y_{fi1}} (1-\theta_{i1})^{y_{fi2}}}{\Gamma(y_{fi1}+1)\,\Gamma(y_{fi2}+1)} \right] \left[\frac{\kappa^{-2\kappa^{-1}}}{\Gamma(2\kappa^{-1})} \right]$$
$$\times \frac{\Gamma\left(2\kappa^{-1} + \sum_{j=1}^{2} y_{fij} \right)}{\left(\kappa^{-1} + \theta_{i1}\gamma_{fi1} + (1-\theta_{i1})\gamma_{fi2}\right)^{2\kappa^{-1} + \sum_{j=1}^{2} y_{fij}}} . \tag{4}$$

En remplaçant ensuite l'expression (4) dans la distribution d'accidents $P(Y_{fi1}, Y_{fi2})$, nous obtenons :

$$P(Y_{fi1}, Y_{fi2})$$
$$= \int_0^1 \left[\frac{(\gamma_{fi1})^{y_{fi1}} (\gamma_{fi2})^{y_{fi2}} (\theta_{i1})^{y_{fi1}} (1-\theta_{i1})^{y_{fi2}}}{\Gamma(y_{fi1}+1)\,\Gamma(y_{fi2}+1)} \right] \left[\frac{\kappa^{-2\kappa^{-1}}}{\Gamma(2\kappa^{-1})} \right]$$
$$\times \frac{\Gamma\left(2\kappa^{-1} + \sum_{j=1}^{2} y_{fij} \right)}{\left(\kappa^{-1} + \theta_{i1}\gamma_{fij} + (1-\theta_{i1})\gamma_{fi2}\right)^{2\kappa^{-1} + \sum_{j=1}^{2} y_{fij}}} f(\theta_{i1}) d\theta_{i1} . \tag{5}$$

En remplaçant $f(\theta_{i1})$ dans l'équation (5) par la densité Dirichlet (ou bêta lorsqu'il y a deux paramètres) de paramètres ν_1, ν_2, soit :

$$f(\theta_{i1}) = \frac{\Gamma\left(\sum_{i=1}^{2} \nu_i\right)}{\prod_{i=1}^{2} \Gamma(\nu_i)} (\theta_{i1})^{\nu_1 - 1}(1-\theta_{i1})^{\nu_2 - 1},$$

nous obtenons :

$$P(Y_{i1}, Y_{i2})$$

$$= \prod_{j=1}^{2} \left[\frac{(\gamma_{fi1})^{y_{fij}}}{\Gamma(y_{fij}+1)} \right] \left[\frac{\Gamma\left(2\kappa^{-1} + \sum_{j=1}^{2} y_{fij}\right) \kappa^{-2\kappa^{-1}}}{\Gamma(2\kappa^{-1})} \right] \left(\frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\,\Gamma(\nu_2)} \right) \quad (6)$$

$$\times \int_0^1 \frac{(\theta_{i1})^{y_{fi1}+\nu_1-1}(1-\theta_{i1})^{y_{fi2}+\nu_2-1}}{(\kappa^{-1} + \theta_{i1}\gamma_{fi1} + (1-\theta_{i1})\gamma_{fi2})^{2\kappa^{-1}+\sum_{j=1}^{2} y_{fij}}} d\theta_{i1}.$$

Il nous reste à évaluer l'intégrale

$$\int_0^1 \frac{(\theta_{i1})^{y_{fi1}+\nu_1-1}(1-\theta_{i1})^{y_{fi2}+\nu_2-1}}{(\kappa^{-1} + \theta_{i1}\gamma_{fi1} + (1-\theta_{i1})\gamma_{fi2})^{2\kappa^{-1}+\sum_{j=1}^{2} y_{fij}}} d\theta_{i1}$$

de l'équation (6).

Pour ce faire, écrivons l'expression $\kappa^{-1} + \theta_{i1}\,\gamma_{fi1} + (1-\theta_{i1})\,\gamma_{fi2}$ de la façon suivante :

$$(\kappa^{-1} + \gamma_{fi2}) \left[1 - \left(\frac{\gamma_{fi2} - \gamma_{fi1}}{\kappa^{-1} + \gamma_{fi2}} \right) \theta_{i1} \right]$$

et nous obtenons l'intégrale égale à (Angers et al., 2004) :

$$\int_0^1 \frac{(\theta_{i1})^{y_{fi1}+\nu_1-1}(1-\theta_{i1})^{y_{fi2}+\nu_2-1}}{\left(\left[1-\left(\frac{\gamma_{fi2}-\gamma_{fi1}}{\kappa^{-1}+\gamma_{fi2}}\right)\theta_{i1}\right]\right)^{2\kappa^{-1}+\sum_{j=1}^{2} y_{fij}}} d\theta_{i1}$$

$$= \left[\frac{\prod_{j=1}^{2}\Gamma(y_{fij}+\nu_j)}{\Gamma\left(\sum_{j=1}^{2}(y_{fij}+\nu_j)\right)}\right] {}_2F_1\left(y_{fi1}+\nu_1; 2\kappa^{-1}+\sum_{j=1}^{2} y_{fij}; \sum_{j=1}^{2}(y_{fij}+\nu_j); \left(\frac{\gamma_{fi2}-\gamma_{fi1}}{\kappa^{-1}+\gamma_{fi2}}\right)\right),$$

où ${}_2F_1$ est la fonction hypergéométrique qui s'écrit comme suit :

$$1+\sum_{\ell=1}^{\infty}\left[\frac{(y_{fi1}+\nu_1)^{[\ell]}\left(2\kappa^{-2}+\sum_{j=1}^{2} y_{fij}\right)^{[\ell]}}{\left(\sum_{j=1}^{2}(y_{fij}+\nu_j)\right)^{[\ell]}} \times \frac{\left(\frac{\gamma_{fi2}-\gamma_{fi1}}{\kappa^{-1}+\gamma_{fi2}}\right)^{\ell}}{\ell!}\right],$$

avec $h^{[\ell]} = h(h+1) \cdots (h+\ell+1)$, une fonction factorielle croissante (Gradshteyn and Ryzhik, 1980, section 9.1).

La probabilité du nombre d'accidents observés du camion *i* ayant deux périodes est donnée par :

$$P(Y_{fi1}, Y_{fi2})$$

$$= \prod_{j=1}^{2}\left[\frac{(\gamma_{fij})^{y_{fij}}}{\Gamma(y_{fij}+1)}\right]\left[\frac{\Gamma\left(2\kappa^{-1}+\sum_{j=1}^{2} y_{fij}\right)\kappa^{-2\kappa^{-1}}}{\Gamma(2\kappa^{-1})}\right]\left(\frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\,\Gamma(\nu_2)}\right)$$

$$\times \frac{1}{(\kappa^{-1}+\gamma_{fi2})^{2\kappa^{-1}+\sum_{j=1}^{2} y_{fij}}}\left[\frac{\prod_{j=1}^{2}\Gamma(y_{fij}+\nu_j)}{\Gamma\left(\sum_{j=1}^{2}(y_{fij}+\nu_j)\right)}\right]$$

$$\times {}_2F_1\left(y_{fi1}+\nu_1; 2\kappa^{-1}+\sum_{j=1}^{2} y_{fij}; \sum_{j=1}^{2}(y_{fij}+\nu_j); \left(\frac{\gamma_{fi2}-\gamma_{fi1}}{\kappa^{-1}+\gamma_{fi2}}\right)\right).$$

Généralisons maintenant le modèle où le camion i à T_i périodes. La probabilité du nombre d'accidents observés du camion i est donnée par :

$$P(Y_{fi1}, \ldots, Y_{fiT_i}) = \int_{\sum_{j=1}^{T_i} \theta_{ij} = 1} \cdots \int P(Y_{fi1}, \ldots, Y_{fiT_i} \mid \theta_{i1}, \ldots, \theta_{iT_i}) f(\theta_{i1}, \ldots, \theta_{iT_i}) \, d\theta_{i1}, \ldots, d\theta_{iT_i - 1} \quad (7)$$

En utilisant la probabilité conditionnelle

$$P(Y_{fi1}, \ldots, Y_{fiT_i} \mid \theta_{i1}, \ldots, \theta_{iT_i}) = \int_0^{\infty} P(Y_{fi1}, \ldots, Y_{fiT_i} \mid \alpha_i, \theta_{i1}, \ldots, \theta_{iT_i}) f(\alpha_i) \, d\alpha_i$$

et en intégrant, nous obtenons :

$$P(Y_{fi1}, \ldots, Y_{fiT_i} \mid \theta_{i1}, \ldots, \theta_{iT_i}) = \left[\prod_{j=1}^{T_i} \left(\frac{(\gamma_{fij})^{y_{fij}} (\theta_{ij})^{y_{fij}}}{\Gamma(y_{fij} + 1)} \right) \right] \left[\frac{\kappa^{-T_i \kappa^{-1}}}{\Gamma(T_i \kappa^{-1})} \right] \frac{\Gamma\left(T_i \kappa^{-1} + \sum_{j=1}^{T_i} y_{fij} \right)}{\left(\kappa^{-1} + \sum_{j=1}^{T_i} \theta_{ij} \gamma_{fij} \right)^{T_i \kappa^{-1} + \sum_{j=1}^{T_i} y_{fij}}} \quad (8)$$

Ainsi, en remplaçant $P(Y_{fi1}, \ldots, Y_{fiT_i} \mid \theta_{i1}, \ldots, \theta_{iT_i})$ dans l'équation (7) par l'expression de l'équation (8) et la fonction de densité $f(\theta_{i1}, \ldots, \theta_{iT_i})$ par la densité Dirichlet de paramètres $(\nu_1, \nu_2, \ldots, \nu_{T_i})$ nous obtenons :

$$P(Y_{fij}, \ldots, Y_{fij}) = \left[\prod_{j=1}^{T_i} \left(\frac{(\gamma_{fij})^{y_{fij}}}{\Gamma(y_{fij} + 1)} \right) \right] \frac{\kappa^{-T_i \kappa^{-1}} \Gamma\left(T_i \kappa^{-1} + \sum_{j=1}^{T_i} y_{fij} \right) \Gamma\left(\sum_{j=1}^{T_i} \nu_j \right)}{\Gamma(T_i \kappa^{-1}) \prod_{j=1}^{T_i} \Gamma(\nu_j)}$$

$$\times \int_{\sum_{j=1}^{T_i} \theta_{ij} = 1} \cdots \int \frac{\prod_{j=1}^{T_i} (\theta_{fij})^{y_{fij} + \nu_j - 1}}{\left(\kappa^{-1} + \sum_{j=1}^{T_i} \theta_{ij} \gamma_{fij} \right)^{T_i \kappa^{-1} + \sum_{j=1}^{T_i} y_{fij}}} \, d\theta_{i1} \ldots d\theta_{iT_i - 1}. \quad (9)$$

Il nous reste à évaluer l'intégrale

$$\int_{\sum_{j=1}^{T_i}\theta_{ij}=1} \cdots \int \frac{\prod_{j=1}^{T_i} (\theta_{ij})^{y_{fij}+v_j-1}}{\left(\kappa^{-1}+\sum_{j=1}^{T_i}\theta_{ij}\gamma_{fij}\right)^{T_i\kappa^{-1}+\sum_{j=1}^{T_i} y_{fij}}} d\theta_{i1}\cdots d\theta_{iT_i-1}$$

de l'équation (9).

Comme dans Angers et al. (2004), il y a trois façons de le faire sur lesquelles nous allons revenir dans le cadre du modèle général de la section suivante.

2.1.2 Prise en compte simultanée de l'effet temps et de l'effet entreprise

Nous procédons maintenant à la généralisation du modèle pour tenir compte simultanément de l'effet entreprise et de l'effet temps. En effet, nous nous intéressons à des observations qui ont des caractéristiques communes parce qu'elles font partie d'une même entreprise. On peut penser à des travailleurs d'une firme, à des véhicules d'une flotte ou à des enfants d'une même école.

Posons $\lambda_{fij} = \gamma_{fij}\alpha_f\theta_{fij}$ avec $\gamma_{fij} = d_{fij}e^{X_{fij}\beta}$, où d_{fij} mesure le nombre de jours où le véhicule *i* de la flotte *f* est autorisé à circuler durant la période *j*, divisé par le nombre de jours total de la période *j*. C'est une mesure d'exposition au risque d'accident. Le vecteur $X_{fij} = (x_{fij1}, ..., x_{fijp})$ contient les *p* caractéristiques du camion *i* de la flotte *f* observées à la période *j* ; ce vecteur contient des informations spécifiques au véhicule et d'autres spécifiques à la flotte. Le paramètre α_f est l'effet aléatoire associé à la flotte *f*, c'est-à-dire le risque ou les caractéristiques non observables attribuables à la flotte, tandis que le paramètre θ_{fij} est l'effet aléatoire du camion *i* de la flotte *f* à la période *j*.

$$\text{Posons } \delta_{fij} = \begin{cases} 1 & \text{si le camion } i \text{ de la flotte } f \text{ est présent à la période } j \\ 0 & \text{sinon.} \end{cases}$$

Nous supposons que $\sum_{i=1}^{I_f}\sum_{j=1}^{T} \delta_{fij}\theta_{fij} = 1$, où I_f est le nombre total de véhicules dans la flotte *f* et *T* est le nombre maximum de périodes disponible dans la base de données. En d'autres termes, θ_{fij} est la proportion du risque de la flotte f attribuable au véhicule i à la période j; ainsi, le risque total non observable du véhicule i de la flotte f à la période j est défini par $\alpha_f\theta_{fij}$. Pour alléger la notation lors du développement des équations, nous faisons l'hypothèse que $\delta_{fij} \equiv 1$ pour tous les camions et toutes les périodes. (Note : les résultats finaux seront présentés sans cette hypothèse).

Nous faisons l'hypothèse que $\theta_{f11}, ..., \theta_{f1T}, ..., \theta_{fI_f1}, ..., \theta_{fI_fT}$ suivent une distribution Dirichlet de paramètres $(\nu_{11}, ..., \nu_{1T}, ..., \nu_{I_f1}, ..., \nu_{I_fT})$ et que α_f suit une distribution gamma de paramètres $(\bar{I}_f\kappa_f^{-1}, \kappa_f^{-1})$, $\bar{I}_f = \frac{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}}{n_{fp}}$ étant le nombre moyen annuel de camions de la flotte f et n_{fp} le nombre de périodes ayant au moins un véhicule de la flotte f.

La distribution du nombre d'accidents de tous les véhicules de la flotte f est donnée par :

$$P(Y_{f11}, \ldots, Y_{fI_fT})$$

$$= \int_{\sum\sum \theta_{fij}=1} \cdots \int P(Y_{f11}, \ldots, Y_{fI_fT} \mid \theta_{f11}, \ldots, \theta_{fI_fT})\, f(\theta_{f11}, \ldots, \theta_{fI_fT})\, d\theta_{f11} \ldots d\theta_{fI_fT-1} \tag{10}$$

où

$$\theta_{fI_fT} = 1 - \sum\sum_{(i,j)\neq(I_f,T)} \theta_{fij}\,.$$

Nous pouvons réécrire la probabilité conditionnelle :

$$P(Y_{f11}, \ldots, Y_{fI_fT} \mid \theta_{f11}, \ldots, \theta_{fI_fT}) = \int_0^{\infty} P(Y_{f11}, \ldots, Y_{fI_fT} \mid \alpha_f, \theta_{f11}, \ldots, \theta_{fI_fT})\, f(\alpha_f)\, d\alpha_f$$

et, en intégrant par rapport à α_f, nous obtenons la probabilité jointe conditionnelle d'accident :

$$P(Y_{f11}, \ldots, Y_{fI_fT} \mid \theta_{f11}, \ldots, \theta_{fI_fT})$$

$$= \left[\prod_{i=1}^{I_f}\prod_{j=1}^{T} \frac{(\gamma_{fij})^{y_{fij}}(\theta_{fij})^{y_{fij}}}{\Gamma(y_{fij}+1)}\right]\left[\frac{\kappa_f^{-\bar{I}_f\kappa_f^{-1}}}{\Gamma(\bar{I}_f\kappa_f^{-1})}\right] \frac{\Gamma\left(\bar{I}_f\kappa_f^{-1} + \sum_{i=1}^{I_f}\sum_{j=1}^{T} y_{fij}\right)}{\left(\kappa_f^{-1} + \sum_{i=1}^{I_f}\sum_{j=1}^{T}\theta_{fij}\gamma_{fij}\right)^{\bar{I}_f\kappa_f^{-1} + \sum_{i=1}^{I_f}\sum_{j=1}^{T} y_{fij}}} \tag{11}$$

Ainsi, en remplaçant $P(y_{f11}, ..., y_{fI_fT} \mid \theta_{f11}, ..., \theta_{fI_fT})$ dans l'équation (10) par sa valeur donnée en (11) et en remplaçant la fonction de densité $f(\theta_{f11}, ..., \theta_{fI_fT})$ par la densité d'une Dirichlet de paramètres $(\nu_{11}, ..., \nu_{1T_1}, ..., \nu_{I_f1}, ..., \nu_{I_fT})$, nous obtenons l'expression suivante :

$$P(Y_{f11}, \ldots, Y_{fI_fT})$$

$$= \left[\prod_{i=1}^{I_f}\prod_{j=1}^{T}\frac{(\gamma_{fij})^{y_{fij}}}{\Gamma(y_{fij}+1)}\right]\frac{\kappa_f^{-\bar{I}_f\kappa_f^{-1}}\Gamma\left(\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}y_{fij}\right)\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\nu_{ij}\right)}{\Gamma(\bar{I}_f\kappa_f^{-1})\prod_{i=1}^{I_f}\prod_{j=1}^{T}\Gamma(\nu_{ij})} \tag{12}$$

$$\times \int \underset{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\theta_{fij}=1}{\cdots} \int \frac{\prod_{i=1}^{I_f}\prod_{j=1}^{T}(\theta_{fij})^{y_{fij}+\nu_{ij}-1}}{\left(\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}\theta_{fij}\gamma_{fij}\right)^{\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}y_{fij}}}\, d\theta_{f11}\ldots d\theta_{fI_fT-1}\,.$$

Nous devons évaluer l'intégrale à plusieurs dimensions :

$$\int \underset{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\theta_{fij}=1}{\cdots} \int \frac{\prod_{i=1}^{I_f}\prod_{j=1}^{T}(\theta_{fij})^{y_{fij}+\nu_{ij}-1}}{\left(\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}\theta_{fij}\gamma_{fij}\right)^{\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}y_{fij}}}\, d\theta_{f11}\ldots d\theta_{fI_fT-1}$$

de l'équation (12) pour estimer les paramètres du modèle. Trois possibilités sont maintenant envisagées :

1. Une première possibilité, qui simplifie beaucoup les calculs, est de supposer que tous les γ_{fij} des I_f véhicules sont identiques et égaux à γ_f et ce, pour le nombre T de périodes. Sous cette hypothèse, l'intégrale multidimensionnelle de l'équation (12) est réduite à :

$$\frac{1}{(\kappa_f^{-1}+\gamma_f)^{\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T} y_{fij}}} \int \underset{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\theta_{fij}=1}{\cdots} \int \prod_{i=1}^{I_f}\prod_{j=1}^{T}(\theta_{fij})^{y_{fij}+\nu_{ij}-1} d\theta_{f11}\cdots d\theta_{fI_fT-1}$$

$$= \frac{\prod_{i=1}^{I_f}\prod_{j=1}^{T}\Gamma(y_{fij}+\nu_{ij})}{\left((\kappa_f^{-1}+\gamma_f)^{\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T} y_{fij}}\right)\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}(y_{fij}+\nu_{ij})\right)}.$$

L'hypothèse de travail de ce premier scénario suppose implicitement que tous les véhicules de la flotte représentent des risques *a priori* identiques, ce qui est probablement une hypothèse très forte car, comme nous le verrons, plusieurs variables distinguant les véhicules, les habitudes de conduite des conducteurs et les caractéristiques observables des flottes sont significatives dans l'estimation des probabilités d'accident. Une autre possibilité est de diviser les véhicules en différents groupes homogènes de risque, comme le font les assureurs en classifiant les risques.

2. Sous cette deuxième possibilité, nous pouvons séparer les véhicules en deux groupes et définir $G_1 = 1, ..., g_1$ comme l'ensemble des véhicules-périodes du premier groupe avec

$$\gamma_{fg_1} = \frac{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\chi_{1ij}\gamma_{fij}}{g_1}, \text{ où}$$

$$\chi_{1ij} = \begin{cases} 1 & \text{si le camion } i \text{ appartient au groupe 1} \\ & \text{à la période } j \\ 0 & \text{sinon.} \end{cases}$$

et $G_2 = g_1 + 1, ..., I_fT$, comme l'ensemble des véhicules-périodes du deuxième groupe avec $\gamma_{fg_2} = \dfrac{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\chi_{2ij}\gamma_{fij}}{g_2}$, où

$$\chi_{2ij} = \begin{cases} 1 & \text{si le camion } i \text{ appartient au groupe 2} \\ & \text{à la période } j \\ 0 & \text{sinon.} \end{cases}$$

et $g_2 = I_fT - g_1$.

Dans cette section, l'indice f n'apparaît pas aux différents paramètres ν, g_1, g_2, car les estimations sont effectuées pour une flotte en particulier. L'intégrale de l'équation (12) devient ainsi :

$$\int_{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\theta_{fij}=1} \cdots \int \frac{\left[\prod_{i=1}^{I_f}\prod_{j=1}^{T}(\theta_{fij})^{\chi_{1ij}(c_{ij}-1)}\prod_{i=1}^{I_f}\prod_{j=1}^{T}(\theta_{fij})^{\chi_{2ij}(c_{ij}-1)}\right]}{\left(\kappa_f^{-1}+\gamma_{fg_1}\sum_{i=1}^{I_f}\sum_{j=1}^{T}\chi_{1ij}\theta_{fij}+\gamma_{fg_2}\sum_{i=1}^{I_f}\sum_{j=1}^{T}\chi_{2ij}\theta_{fij}\right)^d} d\theta_{f11}\cdots d\theta_{fI_fT-1}$$

avec $c_{ij} = y_{fij} + \nu_{ij}$ et $d = \bar{I}_f\kappa^{-1} + \sum_{i=1}^{I_f}\sum_{j=1}^{T} y_{fij}$.

En posant $v = \sum_{i=1}^{I_f}\sum_{j=1}^{T}\chi_{1ij}\theta_{fij}$, $u_{ij} = \dfrac{\chi_{1ij}\theta_{fij}}{v}$, et $w_{ij} = \dfrac{\chi_{2ij}\theta_{fij}}{1-v}$. Notons que $\sum_{i=1}\sum_{j=1} u_{ij} = 1$ et $\sum_{i=1}\sum_{j=1} w_{ij} = 1$. Cette intégrale peut se réécrire de la façon suivante, où $\underset{\sim}{u}$ est le vecteur des valeurs du groupe 1 et $\underset{\sim}{w}$ le vecteur des valeurs du groupe 2 :

$$\int\cdots\int \frac{\left[\prod_{i=1}^{I_f}\prod_{j=1}^{T}(vu_{ij})^{\chi_{1ij}(c_{ij}-1)}\right]\left[\prod_{i=1}^{I_f}\prod_{j=1}^{T}((1-v)w_{ij})^{\chi_{2ij}(c_{ij}-1)}\right]}{((\kappa_f^{-1}+\gamma_{fg_1})v+(\kappa_f^{-1}+\gamma_{fg_2})(1-v))^d} v^{g_1-1}(1-v)^{g_2-1} d\underset{\sim}{u}d\underset{\sim}{w}dv$$

$$= \left(\frac{\prod_{i=1}^{I_f}\prod_{j=1}^{T}\Gamma\left(y_{fij}+\nu_{ij}\right)}{\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\left(y_{fij}+\nu_{ij}\right)\right)\left(\kappa_f^{-1}+\gamma_{fg_2}\right)^{\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}y_{fij}}}\right) \tag{13}$$

$$\times {}_2F_1\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\chi_{1ij}\left(y_{fij}+\nu_{ij}\right),\left(\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}y_{fij}\right),\sum_{i=1}^{I_f}\sum_{j=1}^{T}\left(y_{fij}+\nu_{ij}\right),\left(\frac{\gamma_{fg_2}-\gamma_{fg_1}}{\kappa_f^{-1}+\gamma_{fg_2}}\right)\right).$$

Cette façon de procéder pour évaluer l'intégrale peut être généralisée à plusieurs groupes homogènes. Par exemple, une généralisation à trois groupes impliquerait une série infinie dont les coefficients seraient des fonctions hypergéométriques ${}_2F_1$, comme celle exprimée en (13). Il n'est cependant pas évident que le gain de précision obtenu serait beaucoup supérieur. Une approximation Monte Carlo de l'intégrale multivariée de l'équation (12) nous permettrait d'évaluer le gain potentiel. Nous abordons maintenant l'approximation Monte Carlo de cette intégrale.

3. Si nous voulons évaluer l'intégrale

$$\int_{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\theta_{fij}=1} \cdots \int \frac{\prod_{i=1}^{I_f}\prod_{j=1}^{T}(\theta_{fij})^{y_{fij}+\nu_{ij}-1}}{\left(\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}\theta_{fij}\gamma_{fij}\right)^{\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}y_{fij}}} d\theta_{f11}\ldots d\theta_{fI_fT-1}$$

par la méthode de Monte Carlo, nous pouvons utiliser la fonction d'importance $h(\underset{\sim}{\theta})$ (Lange, 1999) où ($\underset{\sim}{\theta} = \theta_{f11}, \ldots, \theta_{fI_fT}$) tel que :

$$\int_{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}\theta_{fij}=1} \cdots \int g(\underset{\sim}{\theta})\, d\underset{\sim}{\theta} = \int_{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}\theta_{fij}=1} \cdots \int \frac{g(\theta)}{h(\theta)} h(\underset{\sim}{\theta})\, d\underset{\sim}{\theta}$$

$$= \int_{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}\theta_{fij}=1} \cdots \int w(\underset{\sim}{\theta})\, h(\underset{\sim}{\theta})\, d\underset{\sim}{\theta} \approx \frac{1}{N}\sum_{l=1}^{N} w(\underset{\sim}{\theta}^l)$$

avec

$$w(\underset{\sim}{\theta}) = \frac{g(\underset{\sim}{\theta})}{h(\underset{\sim}{\theta})}.$$

En posant :

$$h(\underset{\sim}{\theta}) = \frac{\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}(y_{fij}+\nu_i)\right)}{\prod_{i=1}^{I_f}\prod_{j=1}^{T}\Gamma(y_{fij}+\nu_i)} \prod_{i=1}^{I_f}\prod_{j=1}^{T}(\theta_{fi})^{y_{fij}+\nu_i-1},$$

nous pouvons réécrire l'expression de l'intégrale multivariée en multipliant son numérateur et son dénominateur par la fonction $h(\underset{\sim}{\theta})$ telle que définie plus haut et obtenir, après simplifications, l'expression suivante :

$$\int_{\sum_{i=1}^{I_f}\sum_{j=1}^{T}\theta_{fij}=1} \cdots \int \frac{1}{\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}(\kappa_f^{-1}+\gamma_{fij})\,\theta_{fij}\right)^{\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T} y_{fij}} \dfrac{\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}(y_{fij}+\nu_i)\right)}{\prod_{i=1}^{I_f}\prod_{j=1}^{T}\Gamma(y_{fij}+\nu_i)}}$$

$$\times \frac{\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}(y_{fij}+\nu_{ij})\right)}{\prod_{i=1}^{I_f}\prod_{j=1}^{T}\Gamma(y_{fij}+\nu_{ij})} \prod_{i=1}^{I_f}\prod_{j=1}^{T}(\theta_{fij})^{y_{fij}+\nu_i-1}\, d\theta_{f11}\cdots d\theta_{fI_fT-1}$$

qui peut être évaluée par

$$\left(\frac{\prod_{i=1}^{I_f}\prod_{j=1}^{T}(\Gamma(y_{fij}+\nu_{ij}))^{\delta_{fij}}}{\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}(y_{fij}+\nu_{ij})\right)}\right)\frac{1}{N}\sum_{l=1}^{N}\left[\frac{1}{\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}(\kappa_f^{-1}+\gamma_{fij})\,\theta_{fi}^{l}\right)^{\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}y_{fij}}}\right].$$

En fait, nous pouvons, par exemple, générer des nombres aléatoires a_{fij}^{ℓ} qui sont des valeurs de la densité gamma de paramètres $y_{fij}+\nu_{ij}$ et 1 pour $i = 1, \ldots, I_f$, $j = 1, \ldots, T$ et $\ell = 1, \ldots, N$, où N est le nombre d'itérations de l'approximation de Monte Carlo. Ainsi, en posant $\theta_{fij}^{l} = \dfrac{a_{fij}^{l}}{\sum_{i=1}^{I_f}\sum_{j=1}^{T_{I_f}} a_{fij}^{l}}$, nous obtenons des valeurs d'une Dirichlet $(y_{f11}+\nu_{11}, \ldots, y_{fI_fT_{I_f}}+\nu_{I_fT})$.

Si nous réintroduisons les δ_{fij}, $P(y_{f11}, ..., y_{fI_fT})$ est :

1) égale à

$$\left[\prod_{i=1}^{I_f}\prod_{j=1}^{T}\left(\frac{(\gamma_{fij})^{y_{fij}}}{\Gamma(y_{fij}+1)}\right)^{\delta_{fij}}\right]\frac{\kappa_f^{-\bar{I}_f\kappa_f^{-1}}\Gamma\left(\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}y_{fij}\right)\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}\nu_{ij}\right)}{\Gamma(\bar{I}_f\kappa_f^{-1})\prod_{i=1}^{I_f}\prod_{j=1}^{T}(\Gamma(\nu_{ij}))^{\delta_{fij}}}$$

$$\times\frac{\prod_{i=1}^{I_f}\prod_{j=1}^{T}(\Gamma(y_{fij}+\nu_{ij}))^{\delta_{fij}}}{\left((\kappa_f^{-1}+\gamma_f)^{\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}y_{fij}}\right)\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}(y_{fij}+\nu_{ij})\right)}$$

2) approximativement égale à

$$\frac{\Gamma\left(\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}y_{fij}\right)}{\Gamma(\bar{I}_f\kappa_f^{-1})}(\kappa_f^{-1})^{\bar{I}_f\kappa^{-1}}\left(\frac{1}{\kappa_f^{-1}+\gamma_{fg_2}}\right)^{\bar{I}_f\kappa_f^{-1}+\sum\sum\delta_{fij}y_{fij}}\prod_{i=1}^{I_f}\prod_{j=1}^{T}\left(\frac{\Gamma(y_{fij}+\nu_{ij})(\gamma_{fij})^{y_{fij}}}{\Gamma(y_{fij}+1)\Gamma(\nu_{ij})}\right)^{\delta_{fij}}$$

$$\times\frac{\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}\nu_{ij}\right)}{\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}(y_{fij}+\nu_{ij})\right)} \tag{14}$$

$$\times{}_2F_1\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\chi_{1ij}(y_{fij}+\nu_{ij}),\left(\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}y_{fij}\right),\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}(y_{fij}+\nu_{ij}),\left(\frac{\gamma_{fg_2}-\gamma_{fg_1}}{\kappa_f^{-1}+\gamma_{fg_2}}\right)\right),$$

où ${}_2F_1$ est la fonction hypergéométrique définie à la section 2.1.1.

3) approximativement égale à

$$\left[\prod_{i=1}^{I_f}\prod_{j=1}^{T}\left(\frac{(\gamma_{fij})^{y_{fij}}}{\Gamma(y_{fij}+1)}\right)^{\delta_{fij}}\right]\frac{\kappa_f^{-\bar{I}_f\kappa_f^{-1}}\Gamma\left(\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}y_{fij}\right)\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}\nu_{ij}\right)}{\Gamma(\bar{I}_f\kappa_f^{-1})\prod_{i=1}^{I_f}\prod_{j=1}^{T}(\Gamma(\nu_{ij}))^{\delta_{fij}}}$$

$$\times\left(\frac{\prod_{i=1}^{I_f}\prod_{j=1}^{T}(\Gamma(y_{fij}+\nu_{ij}))^{\delta_{fij}}}{\Gamma\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}(y_{fij}+\nu_{ij})\right)}\right)\frac{1}{N}\sum_{l=1}^{N}\left[\frac{1}{\left(\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}(\kappa_f^{-1}+\gamma_{fij})\,\theta_{fi}^{l}\right)^{\bar{I}_f\kappa_f^{-1}+\sum_{i=1}^{I_f}\sum_{j=1}^{T}\delta_{fij}y_{fij}}}\right].$$

Dans la section suivante, nous utiliserons la seconde méthode d'estimation. La première est très restrictive et la troisième très difficile et très longue à réaliser. Cette dernière pourrait par contre être utile pour vérifier la précision de la seconde (voir Angers et al., 2004, pour une application de ce type de vérification pour l'effet flotte).

2.2 Base de données

Nous avons créé une base de données à partir de fichiers en provenance de la Société d'assurance automobile du Québec (SAAQ). La SAAQ est chargée de vérifier si les véhicules faisant du transport routier de personnes et de marchandises sont conformes aux lois et règlements en vigueur. Elle est aussi l'assureur pour les dommages corporels reliés aux accidents de la route.

Le point de départ de la base de données est l'ensemble des transporteurs enregistrés en date du 30 janvier 1999. À un transporteur sont liées les données contenant : 1) les informations sur les infractions relatives à la politique de conformité faisant l'objet de condamnation et que le transporteur a commises entre 1989 et 1998, soit pour le non-respect des dispositions du Code de la sécurité routière à l'égard de la vérification mécanique, pour le non-respect des règles aux véhicules et à leur équipement, pour le non-respect des heures de conduite et de travail, pour la dimension excédentaire ou pour l'arrimage inadéquat, etc., 2) les informations permettant d'identifier le transporteur en date du 30 janvier 1999.

Nous accédons à l'information sur les véhicules immatriculés au Québec et ce, pour la période du 1er janvier 1990 au 31 décembre 1998. Nous pouvons relier les véhicules au transporteur. Un véhicule est éligible, si :

- le véhicule appartient à une des catégories de plaque suivantes : autobus public, privé ou scolaire, commercial, transport de marchandises en général et en vrac;
- le statut de la plaque est « D » pour détenir;
- le type d'usage du véhicule est autobus, camion ou automobile;
- le statut du véhicule relié à l'autorisation est actif ou reconstruit;
- pour le type véhicule : usage camion et automobile, la masse nette du véhicule est plus grande que 3 000 kg et le type d'utilisation catégorie usage est différent d'urgence.

De l'autorisation, nous obtenons l'information décrivant le véhicule et les plaques. Pour chaque numéro de plaque, nous avons les données correspondant au dossier vérification mécanique des véhicules pour le années 1990 à 1998, les infractions entraînant des points d'inaptitude pour les années 1990 et 1998 et faisant l'objet de condamnation soit pour excès de vitesse, omission de se conformer à un feu rouge ou à un panneau d'arrêt, dépassement prohibé, etc., et les accidents pour les années de 1990 à 1998 (voir Dionne et al., 1999, pour une description détaillée de la base de données).

2.3 Choix et description des variables

L'unité d'observation est un véhicule éligible ayant au moins un jour d'autorisation de circuler au cours d'une année *j*. Nous analysons les accidents totaux présents dans les fichiers de la SAAQ, c'est-à-dire tous les accidents routiers au Québec ayant fait l'objet d'un rapport de police, ce qui comprend des accidents avec des dommages matériels seulement.

Variable dépendante

Y_{fij} = le nombre d'accidents dans lequel le véhicule *i* de la flotte *f* a été impliqué au cours de l'année *j*. Y_{fij} peut prendre les valeurs 0, 1, 2, 3, 4 et plus. Seulement 81 camions sur 515 851 ont plus de quatre accidents, soit 0,016 %.

Variable d'exposition

d_{fij} = le nombre de jours où l'autorisation de circuler est active pour l'année *j* divisé par le nombre total de jours.

Variables explicatives

Nous avons deux types de variables explicatives : celles concernant le transporteur et celles concernant le véhicule.

Variables concernant le transporteur

– *Taille de la flotte pour l'année j* : 7 variables dichotomiques ont été créées.

  La taille de deux véhicules est utilisée comme catégorie de référence. Des coefficients estimés positifs et significatifs indiqueront donc que les véhicules sont davantage à risque d'accident que ceux dans la classe de deux véhicules.

– *Secteur d'activité économique du transporteur routier* : 5 variables dichotomiques ont été créées pour les véhicules transportant des biens :

sect_14 = 1 si le principal secteur d'activité est le transport de personnes;

sect_05 = 1 si le secteur d'activité est le camionnage public général;

sect_06 = 1 si le secteur d'activité est le camionnage public en vrac;

sect_07 = 1 si le secteur d'activité est le camionnage pour compte propre;

sect_08 = 1 si le secteur d'activité est une entreprise de location à court terme.

Le secteur d'activité « camionnage public en vrac » est utilisé comme catégorie de référence. Les coefficients estimés négatifs et significatifs pour les autres secteurs d'activité économiques du transporteur indiqueront donc que les véhicules de ces secteurs d'activité représentent des risques moins élevés que ceux du groupe de référence (et inversement pour les coefficients positifs et significatifs).

- Sept (7) variables ont été créées pour les véhicules effectuant *le transport de biens* afin de mesurer le nombre d'infractions faisant l'objet de condamnation par véhicule commises l'année précédente à l'année *j* par un transporteur routier :
  - *Nombre d'infractions de surcharge par véhicule commises l'année précédente à l'année j par un transporteur routier.*
  - *Nombre d'infractions pour dimension excédentaire par véhicule commises l'année précédente à l'année j par un transporteur routier.*
  - *Nombre d'infractions pour arrimage inadéquat par véhicule commises l'année précédente à l'année j par un transporteur routier.*
  - *Nombre d'infractions pour non-respect des dispositions du Code de la sécurité routière à l'égard du transport de matières dangereuses par véhicule commises l'année précédente à l'année j par un transporteur routier.*
  - *Nombre d'infractions pour non-respect des heures de conduites et de travail par véhicule commises l'année précédente à l'année j par un transporteur routier.*
  - *Nombre d'infractions pour non-respect des dispositions du Code de la sécurité routière à l'égard de la vérification mécanique par véhicule commises l'année précédente à l'année j par un transporteur routier.*

- *Nombre d'infractions autres que celles déjà mentionnées par véhicule commises l'année précédente à l'année j par un transporteur routier.*

Un signe positif est prédit pour toutes ces variables, car plus d'infractions devrait générer, en moyenne, plus d'accidents.

Variables concernant le véhicule

– *Nombre de cylindres du véhicule :* 4 variables dichotomiques ont été créées :

cyl_0 = 1	si le nombre de cylindres du véhicule est inconnu;
cyl1_5 = 1	si le nombre de cylindres du véhicule est de 1 à 5;
cyl6_7 = 1	si le nombre de cylindres du véhicule est de 6 ou 7;
cyl_8p = 1	si le nombre de cylindres du véhicule est de 8 ou plus de 10.

Le groupe de véhicules ayant 8 cylindres ou plus de 10 cylindres est utilisé comme catégorie de référence. Les coefficients estimés négatifs et significatifs pour les autres groupes du nombre de cylindres du véhicule indiqueront donc que ces groupes de véhicules représentent des risques moins élevés que ceux du groupe de référence.

– *Type de carburant du véhicule :* 3 variables dichotomiques ont été créées :

diesel = 1	si le véhicule utilise du diesel comme carburant;
essence = 1	si le véhicule utilise de l'essence comme carburant;
carb_aut = 1	si le véhicule utilise un autre type de carburant.

Le groupe de véhicules utilisant du diesel comme carburant est considéré comme catégorie de référence. Les coefficients estimés négatifs et significatifs pour les autres groupes de carburant du véhicule indiqueront donc que ces groupes de véhicules représentent des risques moins élevés que ceux du groupe de référence.

– *Nombre d'essieux maximal du véhicule :* 6 variables dichotomiques ont été créées :

ess_2 = 1	si le véhicule a deux essieux et la masse est comprise entre 3 000 et 4 000 kg;
ess_2p = 1	si le véhicule a deux essieux et la masse est supérieure à 4 000 kg;

ess_3 = 1 si le véhicule a un maximum de trois essieux pouvant le supporter;

ess_4 = 1 si le véhicule a un maximum de quatre essieux pouvant le supporter;

ess_5 = 1 si le véhicule a un maximum de cinq essieux pouvant le supporter;

ess_6p = 1 si le véhicule a six essieux et plus pouvant le supporter.

Le groupe de véhicules ayant deux essieux et dont la masse est comprise entre 3 000 et 4 000 kg est utilisé comme catégorie de référence. Les coefficients estimés positifs et significatifs pour les autres groupes du nombre maximal d'essieux pouvant supporter le véhicule indiqueront donc que ces groupes de véhicules représentent des risques plus élevés que ceux du groupe de référence.

– *Type d'utilisation du véhicule* : 3 variables dichotomiques pour les véhicules transportant des biens :

compr = 1 si l'usage du véhicule est destiné à l'utilisation commerciale, incluant le transport de biens sans permis de la C.T.Q.;

tbrgn = 1 si l'usage du véhicule est destiné au transport de biens autre que « vrac », exigeant un permis de la C.T.Q.;

tbrvr = 1 si l'usage du véhicule est destiné au transport de matières en « vrac ».

Le groupe de véhicules transportant des matières en vrac est utilisé comme catégorie de référence. Les coefficients estimés négatifs et significatifs pour les autres groupes d'utilisation du véhicule indiqueront donc que ces groupes de véhicules représentent des risques moins élevés que ceux du groupe de référence.

– Six (6) variables ont été créées afin de mesurer le nombre d'infractions faisant l'objet de condamnation par véhicule commises l'année précédente à l'année *j* par un ou des conducteurs :

- *Nombre d'infractions pour excès de vitesse par véhicule commises l'année précédente à l'année j.*
- *Nombre d'infractions pour conduite durant une sanction par véhicule commises l'année précédente à l'année j.*

- *Nombre d'infractions pour omission de se conformer à un feu rouge par véhicule commises l'année précédente à l'année j.*
- *Nombre d'infractions pour omission de se conformer à un panneau d'arrêt ou aux signaux d'agent par véhicule commises l'année précédente à l'année j.*
- *Nombre d'infractions pour omission de porter la ceinture par véhicule commises l'année précédente à l'année j.*
- *Nombre d'infractions autres que celles mentionnées par véhicule commises l'année précédente à l'année j.*

Un signe positif est prédit pour toutes ces variables, car plus d'infractions devrait générer, en moyenne, plus d'accidents.

3. RÉSULTATS

3.1. Statistiques descriptives

Le Tableau 1 donne la répartition de la taille de la flotte. On retrouve 55,26 % des 24 282 entreprises ayant deux véhicules et 2 % ayant plus de 20 véhicules.

TABLEAU 1
TAILLE DE LA FLOTTE

Taille de la flotte	Nombre d'entreprises	%
2 véhicules	13 418	55,26
3 véhicules	4 310	17,75
4 à 5 véhicules	3 196	13,16
6 à 9 véhicules	1 820	7,20
10 à 20 véhicules	1 009	4,16
21 à 50 véhicules	376	1,55
51 véhicules et plus	153	0,63
Total	24 282	100,00

On observe, au Tableau 2, que les moyennes d'accidents varient sensiblement d'une période à l'autre, de même que les écarts types. Il est à noter que le mot annuel signifie « nombre de jours avec autorisation de conduire durant une année ». Cette notification s'applique également à tous les tableaux suivants. Les variations des moyennes selon la taille des flottes (Tableau 3) sont beaucoup plus importantes. Les taux moyens d'accident augmentent avec la taille à l'exception de la classe 51 véhicules et plus, qui présente une moyenne inférieure à celles des classes 10 à 20 et 21 à 50 véhicules, mais toujours supérieure à la moyenne des tailles inférieures à 20 véhicules. Ces résultats peuvent être expliqués par une plus grande difficulté à surveiller les conducteurs de camions ou par une plus grande exposition au risque. En effet, les camions de grandes flottes peuvent circuler plus d'heures par année que les autres véhicules avec des conducteurs travaillant sur plusieurs horaires durant une journée.

TABLEAU 2
NOMBRE MOYEN D'ACCIDENTS ANNUELS PAR CAMION SELON LA PÉRIODE D'OBSERVATION

Période d'observation	Nombre de camions	Nombre de camions-années	Moyenne d'accident	Écart type
1991	55 226	48 709,72	0,1603	0,4344
1992	60 697	53 245,90	0,1445	0,4125
1993	61 613	54 778,52	0,1442	0,4094
1994	63 323	56 393,68	0,1536	0,4202
1995	66 346	58 956,26	0,1573	0,4366
1996	68 478	60 818,63	0,1485	0,4212
1997	66 840	59 801,32	0,1505	0,4201
1998	73 328	65 407,77	0,1605	0,4405
Total	515 851	458 112,09	0,1523	0,4246

TABLEAU 3
NOMBRE MOYEN D'ACCIDENTS ANNUELS PAR CAMION SELON LA TAILLE DE LA FLOTTE

Taille de la flotte	Nombre de camions	Nombre de camions-années	Moyenne d'accident	Écart type
2 véhicules	93 004	81 275,92	0,1120	0,3596
3 véhicules	57 165	50 490,06	0,1306	0,3861
4 à 5 véhicules	66 689	59 337,28	0,1421	0,4089
6 à 9 véhicules	64 598	57 861,31	0,1662	0,4431
10 à 20 véhicules	73 261	65 689,77	0,1781	0,4582
21 à 50 véhicules	59 816	53 499,64	0,1797	0,4598
51 véhicules et plus	101 318	89 958,11	0,1678	0,4515

Le Tableau 4 montre que le nombre d'accidents moyen varie également beaucoup entre les secteurs d'activité. Le secteur transport par autobus a moins d'accidents, alors que celui de l'entreprise de location à court terme a une moyenne très élevée. On peut expliquer le premier résultat par une plus faible exposition au risque, étant donné que le transport par camion n'est pas l'activité principale de l'entreprise, dont la principale activité est le transport par autobus. Les entreprises de location à court terme, par contre, devraient avoir des niveaux de kilométrage élevés et moins de surveillance des conducteurs.

TABLEAU 4
NOMBRE MOYEN D'ACCIDENTS ANNUELS PAR CAMION SELON LE SECTEUR D'ACTIVITÉ DE L'ENTREPRISE

Secteur d'activité principal de l'entreprise	Nombre de camions	Nombre de camions-années	Moyenne d'accident	Écart type
Transport par autobus	1 992	1 785,77	0,0730	0,2836
Camionnage public général	84 328	75 587,43	0,1944	0,4674
Camionnage public en vrac	48 434	43 300,14	0,1500	0,4303
Camionnage pour compte propre	366 605	325 749,88	0,1406	0,4091
Entreprise de location à court terme	13 486	10 884,40	0,2005	0,4853

Le Tableau 5 indique clairement que les transporteurs qui ont des véhicules ayant eu des infractions commises ont plus d'accidents. L'écart type est également plus grand.

TABLEAU 5
NOMBRE MOYEN D'ACCIDENTS ANNUELS PAR CAMION SELON LE NOMBRE D'INFRACTIONS RELATIVES À LA POLITIQUE DE CONFORMITÉ QUE L'ENTREPRISE A COMMISES L'ANNÉE PRÉCÉDENTE

Nombre d'infractions que l'entreprise a commises l'année précédente	Nombre de camions	Nombre de camions-années	Moyenne d'accident	Écart type
Pour surcharge				
0	493 971	438 596,88	0,1470	0,4156
1	18 329	16 358,51	0,2652	0,5715
2 et plus	3 551	3 156,70	0,3133	0,6238
Pour dimension excédentaire				
0	515 181	457 524,00	0,1521	0,4243
1 et plus	670	588,09	0,3112	0,6190
Pour arrimage inadéquat				
0	513 140	455 693,26	0,1518	0,4239
1 et plus	2 711	2 418,83	0,2567	0,5374
Pour non-respect des heures de conduite				
0	515 309	457 636,45	0,1522	0,4243
1 et plus	542	475,64	0,3361	0,6409
Pour non-respect de la vérification mécanique				
0	511 427	454 343,29	0,1510	0,4224
1 et plus	4 424	3 768,80	0,2799	0,5863
Pour d'autres raisons				
0	515 166	457 533,75	0,1522	0,4241
1 et plus	685	578,34	0,2749	0,6785

Le Tableau 6 indique que l'utilisation commerciale des camions explique moins d'accidents.

TABLEAU 6
NOMBRE MOYEN D'ACCIDENTS ANNUELS PAR CAMION SELON LE TYPE D'UTILISATION DU VÉHICULE

Type d'utilisation du véhicule	Nombre de camions	Nombre de camions-années	Moyenne d'accident	Écart type
Utilisation commerciale incluant le transport des biens sans permis C.T.Q.	380 873	337 612,16	0,1371	0,4030
Transport de biens autre que vrac	91 587	80 543,83	0,1952	0,4751
Transport de biens en vrac	43 391	39 956,10	0,1912	0,4792

On remarque au Tableau 7 que l'utilisation du diesel comme carburant correspond à plus d'accidents par camion. Ce résultat peut aussi refléter un plus grand kilométrage moyen.

TABLEAU 7
NOMBRE MOYEN D'ACCIDENTS ANNUELS PAR CAMION SELON LE TYPE DE CARBURANT

Type de carburant	Nombre de camions	Nombre de camions-années	Moyenne d'accident	Écart type
Diesel	398 708	356 153,75	0,1729	0,4510
Essence	114 626	99 656,10	0,0828	0,3097
Autre	2 517	2 302,23	0,0819	0,3015

Les camions avec 8 ou plus de 10 cylindres ont moins d'accidents (Tableau 8) tandis que ceux qui ont 6 essieux et plus en ont plus (Tableau 9).

TABLEAU 8
NOMBRE MOYEN D'ACCIDENTS ANNUELS PAR CAMION SELON LE NOMBRE DE CYLINDRES DU VÉHICULE

Nombre de cylindres	Nombre de camions	Nombre de camions-années	Moyenne d'accident	Écart type
1 a 5 cylindres	5 817	5 185,35	0,1750	0,4589
6 à 7 cylindres	326 418	292 170,77	0,1810	0,4634
8 ou plus de 10 cylindres	180 360	157 902,68	0,1012	0,3385

TABLEAU 9
NOMBRE MOYEN D'ACCIDENTS ANNUELS PAR CAMION SELON LE NOMBRE D'ESSIEUX DU VÉHICULE

Nombre d'essieux	Nombre de camions	Nombre de camions-années	Moyenne d'accident	Écart type
2 essieux (3 000 à 4 000 kg)	78 514	69 819,60	0,1173	0,3572
2 essieux (plus de 4 000 kg)	147 951	130 761,88	0,1246	0,3849
3 essieux	100 393	89 774,22	0,1463	0,4317
4 essieux	35 802	32 114,63	0,1420	0,4061
5 essieux	47 410	41 728,74	0,1623	0,4226
6 essieux et plus	105 781	93 913,00	0,2167	0,5026

Les infractions entraînant des points d'inaptitude commises par les conducteurs ont également un très grand pouvoir explicatif des accidents.

TABLEAU 10
NOMBRE MOYEN D'ACCIDENTS ANNUELS PAR CAMION SELON LE NOMBRE D'INFRACTIONS ENTRAÎNANT DES POINTS D'INAPTITUDE QUE LE CONDUCTEUR A COMMISES L'ANNÉE PRÉCÉDENTE

Infractions que le conducteur a commises l'année précédente	Nombre de camions	Nombre de camions-années	Moyenne d'accident	Écart type
Pour excès de vitesse				
0	486 217	431 915,58	0,1432	0,4120
1	25 873	22 918,95	0,2805	0,5562
2	3 088	2 689,98	0,3280	0,5788
3 et plus	673	587,58	0,5529	0,8164
Pour conduite durant une sanction				
0	514 255	456 753,18	0,1516	0,4236
1 et plus	1 596	1 358,91	0,3544	0,6131
Pour omission de se conformer à un feu rouge				
0	509 331	452 382,42	0,1503	0,4214
1	6 375	5 606,02	0,3064	0,6008
2 et plus	145	123,65	0,2825	0,5907
Pour panneau d'arrêt ou signaux d'agent				
0	510 148	453 033,51	0,1510	0,4226
1	5 567	4 956,25	0,2667	0,5490
2 et plus	136	122,33	0,7123	1,0510
Pour omission de porter la ceinture				
0	510 871	453 664,64	0,1512	0,4226
1	4 622	4 133,16	0,2678	0,5871
2 et plus	358	314,29	0,2617	0,5466
Autre				
0	513 805	456 297,76	0,1505	0,4216
1 et plus	2 046	1 814,33	0,6266	0,7823

3.2 Estimation du modèle

Les résultats du Tableau 11 correspondent à l'approximation hypergéométrique présentée à la section 2.1.2 (deuxième méthode). Nous avons divisé les données en deux groupes de la façon suivante. Premièrement, nous avons estimé les coefficients à l'aide du modèle de la binomiale négative pour pouvoir calculer $\hat{\gamma} = de^{X\hat{\beta}}$ pour tous les véhicules. Nous avons calculé la moyenne des $\hat{\gamma}$. Le véhicule se trouve dans le groupe 1 si $\hat{\gamma}$ est plus petit ou égal à la moyenne ; sinon, il est dans le groupe 2.

TABLEAU 11
ESTIMATION DES PARAMÈTRES DE LA DISTRIBUTION DU NOMBRE D'ACCIDENTS ANNUELS DES CAMIONS POUR LES ANNÉES 1991-1998 (flotte de deux camions et plus) (Approximation hypergéométrique)

Variables explicatives	Coefficient	Écart type	Statistique t	P-value
ν (1991)	2,1992	0,0728	30,2221	< 0,0001
ν (1992)	2,2243	0,0729	30,5278	< 0,0001
ν (1993)	2,0520	0,0672	30,5241	< 0,0001
ν (1994)	2,1660	0,0706	30,6884	< 0,0001
ν (1995)	2,2017	0,0716	30,7594	< 0,0001
ν (1996)	2,0674	0,0672	30,7513	< 0,0001
ν (1997)	1,8283	0,0599	30,5456	< 0,0001
ν (1998)	2,0085	0,0658	30,5420	< 0,0001
κ (taille = 2)	1,9655	0,0776	25,3168	< 0,0001
κ (taille = 3)	1,8303	0,0851	21,4968	< 0,0001
κ (taille = 4-5)	2,4055	0,1075	22,3761	< 0,0001
κ (taille = 6-9)	3,3330	0,1679	19,8512	< 0,0001
κ (taille = 10-20)	6,6952	0,3692	18,1362	< 0,0001
κ (taille = 21-50)	16,5945	1,3546	12,2502	< 0,0001
κ (taille = 51 et plus)	110,5875	11,9894	9,2238	< 0,0001
Constante	-2,2529	0,0363	-62,1475	< 0,0001
Période d'observation	-0,0406	0,0018	-22,3185	< 0,0001

FIGURE 11 (suite)

Variables explicatives	Coefficient	Écart type	Statistique t	P-value
Secteur d'activité principal de l'entreprise				
Transport par autobus	-0,2482	0,1409	-1,7618	0,0781
Camionnage public général	0,2266	0,0322	7,0456	< 0,0001
Camionnage de compte propre	0,1065	0,0252	4,2308	< 0,0001
Entreprise de location de court terme	0,5555	0,0791	7,0259	< 0,0001
Camionnage public en vrac*	–	–	–	–
Nombre de jours d'autorisation de circuler l'année précédente	1,9713	0,0256	76,9592	< 0,0001
Nombre d'infractions que le transporteur a commises l'année précédente				
Pour surcharge	0,0816	0,0108	7,5281	< 0,0001
Pour dimension excédentaire	0,2275	0,0797	2,8541	0,0043
Pour arrimage inadéquat	0,2015	0,0334	6,0391	< 0,0001
Pour non-respect des heures de conduite	0,1614	0,0629	2,5648	0,0103
Pour non-respect de la vérification mécanique	0,1901	0,0277	6,8719	< 0,0001
Pour autre raison	0,1875	0,0697	2,6884	0,0072
Type d'utilisation du véhicule				
Commerciale incluant le transport des biens sans permis CTQ	-0,1289	0,0202	-6,3863	< 0,0001
Transport des biens autre que vrac	-0,0917	0,0233	-3,9357	0,0001
Transport des biens en vrac*				
Type de carburant				
Diesel*				
Essence	-0,3034	0,0128	-23,7294	< 0,0001
Autre	-0,3313	0,0703	-4,7152	< 0,0001

FIGURE 11 (suite)

Variables explicatives	Coefficient	Écart type	Statistique t	P-value
Nombre de cylindres				
1 à 5 cylindres	0,1466	0,0377	3,8843	0,0001
6 à 7 cylindres	0,2820	0,0117	24,0569	< 0,0001
8 ou plus de 10 cylindres*	–	–	–	–
Nombre d'essieux				
2 essieux (3 000 à 4 000 kg)	-0,3547	0,0197	-17,9773	< 0,0001
2 essieux (plus de 4 000 kg)	-0,3644	0,0145	-25,1319	< 0,0001
3 essieux	-0,1962	0,0145	-13,5403	< 0,0001
4 essieux	-0,1808	0,0186	-9,7314	< 0,0001
5 essieux	-0,2070	0,0169	-12,2170	< 0,0001
6 essieux et plus*				
Nombre d'infractions que le conducteur a commises l'année précédente				
Pour excès de vitesse	0,1881	0,0095	19,8148	< 0,0001
Pour conduite durant une sanction	0,3730	0,0386	9,6735	< 0,0001
Pour omission de se conformer à un feu rouge	0,3030	0,0223	13,5701	< 0,0001
Pour panneau d'arrêt ou signaux d'agent	0,3561	0,0240	14,8207	< 0,0001
Pour omission de porter la ceinture de sécurité	0,1456	0,0277	5,2516	< 0,0001
Autre	1,0184	0,0282	36,1576	< 0,0001

FIGURE 11 (suite)

Variables explicatives	Coefficient	Écart type	Statistique t	P-value
Taille de la flotte				
2 véhicules*				
3 véhicules	0,3441	0,0186	18,4947	< 0,0001
4 à 5 véhicules	0,4984	0,0188	26,5570	< 0,0001
6 à 9 véhicules	0,6333	0,0199	31,9056	< 0,0001
10 à 20 véhicules	0,7202	0,0217	33,1696	< 0,0001
21 à 50 véhicules	0,7343	0,0275	26,6915	< 0,0001
51 véhicules et plus	0,7101	0,0368	19,3119	< 0,0001
Log de la vraisemblance	-222 811			
Nombre d'entreprises	24 282			
Nombre de véhicules	163 844			
Nombre d'observations	515 851			

Note : * indique le groupe de référence.

Plusieurs variables ont été utilisées pour mesurer l'hétérogénéité observable. Certaines de ces variables (infractions au Code de la sécurité routière, type de carburant, nombre de cylindres, nombre d'essieux, type de véhicule utilisé) sont des caractéristiques concernant les véhicules, alors que d'autres (secteur d'activité, taille de la flotte, infraction flotte au Code de la sécurité routière) concernent la flotte. Elles sont pratiquement toutes significatives et du bon signe, un résultat non surprenant, étant donné le nombre élevé d'observations.

Nous avons aussi estimé un paramètre ν par année et des paramètres κ pour différentes tailles de flotte, afin de tenir compte de l'hétérogénéité non observable. Les coefficients ν et κ sont tous significativement différents de zéro. À première vue, les coefficients paraissent assez semblables, sauf ceux pour les tailles élevées. Afin de vérifier si certains ν et κ diffèrent et pour les identifier, nous avons procédé à un test d'hypothèses que nous présentons maintenant.

3.2.1 Test d'hypothèses

Nous voulons vérifier l'hypothèse suivante : $H_0 : A_\nu \underset{\sim}{\nu} = 0$ vs $H_1 : A_\nu \underset{\sim}{\nu} \neq 0$, c'est-à-dire vérifier s'il y a au moins un ν différent des autres. Sous H_0, nous avons que :

$$\hat{Z}_v'(A_v\hat{\Sigma}_vA_v')^{-1}\hat{Z}_v \sim \chi_7^2, \text{ en posant } \hat{Z}_v = A_v\hat{\underset{\sim}{v}},$$

$$\text{avec } A_v = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \text{ et } \hat{\underset{\sim}{v}} = \begin{pmatrix} 2{,}199191 \\ 2{,}224262 \\ 2{,}051964 \\ 2{,}166003 \\ 2{,}201651 \\ 2{,}067379 \\ 1{,}828278 \\ 2{,}008532 \end{pmatrix}.$$

Or, $\hat{Z}_v'(A_v\hat{\Sigma}_vA_v')^{-1}\hat{Z}_v$ est égal à 188,8857 et $\chi^2_{7;99,5} = 20{,}3$. Nous rejetons donc l'hypothèse H_0 avec un niveau de signification $\alpha = 0{,}5\ \%$.

Scheffé (1999) a montré que les intervalles de confiance simultanées pour tous les contrastes de v ont la forme suivante:

$$\hat{Z}_v - \sqrt{\text{vecdiag}(A_v\hat{\Sigma}_vA_v')' \times 7F_{\alpha;7,\infty}} \leq Z_v$$
$$\leq \hat{Z}_v + \sqrt{\text{vecdiag}(A_v\hat{\Sigma}_vA_v')' \times 7F_{\alpha;7,\infty}}$$

Pour $\alpha = 0{,}05$, nous avons que $F_{\alpha;7,\infty} = 2{,}01$. Le Tableau 12 donne les intervalles de confiance pour les différences des v pour des paires d'années consécutives pour différents niveaux de spécification et le Tableau 13 donne le niveau de signification pour la différence de v entre une paire d'années.

TABLEAU 12
INTERVALLES DE CONFIANCE POUR LA DIFFÉRENCE DES ν POUR DES PAIRES D'ANNÉES CONSÉCUTIVES POUR DIFFÉRENTS NIVEAUX DE SPÉCIFICATION

Intervalle de confiance	$\nu_i - \nu_j$	à 90 %		à 95 %		à 97,5 %		à 99 %		à 99,5 %	
		INF	SUP	INF	SUP	INF	SUP	INF	SUP	INF	SUP
1991-1992	-0,0251	-0,1468	0,0966	-0,1566	0,1065	-0,1655	0,1154	-0,1759	0,1257	-0,1831	0,1330
1992-1993	0,1723	0,0550	0,2896	0,0455	0,2991	0,0370	0,3076	0,0270	0,3176	0,0200	0,3246
1993-1994	-0,1140	-0,2280	-0,0000	-0,2372	0,0092	-0,2455	0,0175	-0,2552	0,0272	-0,2620	0,0339
1994-1995	-0,0356	-0,1491	0,0778	-0,1583	0,0870	-0,1666	0,0953	-0,1762	0,1050	-0,1830	0,1117
1995-1996	0,1342	0,0239	0,2446	0,0150	0,2536	0,0069	0,2616	-0,0024	0,2710	-0,0090	0,2776
1996-1997	0,2391	0,1313	0,3463	0,1232	0,3550	0,1154	0,3628	0,1063	0,3719	0,0999	0,3783
1997-1998	-0,1803	-0,2863	-0,0743	-0,2948	-0,0657	-0,3026	-0,0579	-0,3116	-0,0489	-0,3179	-0,0426

TABLEAU 13
NIVEAU DE SIGNIFICATION POUR LA DIFFÉRENCE DES ν

$\nu_i - \nu_j$	1992	1993	1994	1995	1996	1997	1998
1991	NS	***	NS	NS	*	*****	*****
1992		*****	NS	NS	*****	*****	*****
1993			*	****	NS	*****	NS
1994				NS	NS	*****	*****
1995					***	*****	*****
1996						*****	NS
1997							*****

* significatif à 10 %; ** significatif à 5 %; *** significatif à 2,5 %;
**** significatif à 1 %; ***** significatif à 0,5 %

Les résultats indiquent clairement que l'année 1997, ainsi que l'année 1998, diffèrent des autres, ce qui supporte l'estimation de plusieurs ν.

Nous avons procédé au même exercice pour les κ. Y a-t-il au moins un κ différent des autres ? En d'autres termes, nous voulons vérifier l'hypothèse suivante : $H_0 : A_\kappa \underset{\sim}{\kappa} = 0$ vs $H_1 : A_\kappa \underset{\sim}{\kappa} \neq 0$.

Nous avons que $\hat{Z}'_\kappa (A_\kappa \hat{\Sigma}_\kappa A'_\kappa)^{-1} \hat{Z}_\kappa = 422{,}4979$ et $\chi^2_{6;99,5} = 18{,}5$. Nous rejetons donc l'hypothèse H_0 avec un niveau de signification $\alpha = 0{,}5\ \%$.

Seules les petites tailles semblent avoir un effet flotte commun. Autrement, ils sont tous différents à un degré de confiance de 99,5 %.

TABLEAU 14
INTERVALLES DE CONFIANCE POUR LA DIFFÉRENCE DES K POUR DES PAIRES DE TAILLE DE FLOTTE CONSÉCUTIVES POUR DIFFÉRENTS NIVEAUX DE SPÉCIFICATION

Intervalle de confiance	$\kappa_i - \kappa_j$	à 90 %		à 95 %		à 97,5 %		à 99 %		à 99,5 %	
		INF	SUP	INF	SUP	INF	SUP	INF	SUP	INF	SUP
2-3	0,1352	-0,2671	0,5375	-0,3030	0,5734	-0,3342	0,6047	-0,3708	0,6412	-0,3964	0,6668
3–4-5	-0,5752	-1,0557	-0,0947	-1,0986	-0,0518	-1,1359	-0,0145	-1,1796	0,0292	-1,2101	0,0597
4-5–6-9	-0,9274	-1,6265	-0,2284	-1,6888	-0,1661	-1,7431	-0,1118	-1,8066	-0,0483	-1,8510	-0,0039
6-9–10-20	-3,3622	-4,7849	-1,9400	-4,9114	-1,8130	-5,0218	-1,7026	-5,1511	-1,5734	-5,2414	-1,4830
10-20–21-50	-9,8993	-14,8288	-4,9703	-15,2682	-4,5305	-15,6508	-4,1478	-16,0988	-3,6999	-16,4119	-3,3868
21-50–51 et +	-93,9929	-136,7249	-51,5580	-140,2147	-47,7712	-143,5089	-44,4769	-147,3652	-40,6206	-150,0611	-37,9248

TABLEAU 15
NIVEAU DE SIGNIFICATION POUR LA DIFFÉRENCE DES κ

$\kappa_i - \kappa_j$	3	4-5	6-9	10-20	21-50	51 et +
2	NS	NS	*****	*****	*****	*****
3		***	*****	*****	*****	*****
4-5			*****	*****	*****	*****
6-9				*****	*****	*****
10-20					*****	*****
21-50						*****

* significatif à 10 %; ** significatif à 5 %; *** significatif à 2,5 %; **** significatif à 1 %; ***** significatif à 0,5 %

3.3 Analyse des résidus

Une étape importante de recherche consiste à vérifier jusqu'à quel point la modélisation proposée est plus performante comparativement à d'autres modèles moins généraux. Le Tableau 16 présente différentes statistiques qui comparent quatre différents modèles. Nous avons calculé la médiane, l'étendue interquartile (IQR), la médiane de l'écart absolue (MAD), la moyenne de l'erreur absolue (MAE) et l'écart type des résidus. Si Q_1, Q_2 et Q_3 correspondent au premier, deuxième et troisième quartiles des résidus (e), alors $IQR = Q_3 - Q_1$, $MAD = \text{médiane}_{1 \le i \le n} \{ | e_{fij} - Q_2 | \}$ et $MAE = n^{-1} \sum_{i=1}^{n} |e_{fij}|$.

Si nous utilisons les critères du maximum de vraisemblance et du BIC, les résultats indiquent que le modèle général (Gamma-Dirichlet temps et flotte) est plus performant, et les différences dans les valeurs du BIC et du maximum de vraisemblance sont aussi différentes que celles anticipées. Nous obtenons que le ratio de vraisemblance mesurant l'ajout de l'effet flotte (Gamma-Dirichlet temps et flotte) au modèle Gamma-Dirichlet temps est égal à :

$$-2(\text{Log } L_{44} - \text{Log } L_{51}) = -2(-226\ 188{,}72 + 222\ 811{,}18) = 6\ 755{,}08,$$

ce qui est largement supérieur à une valeur d'une $\chi^2_7(51 - 44)$ à un niveau de signification de 1 % égale à 18,5.

TABLEAU 16
STATISTIQUES DESCRIPTIVES DES RÉSIDUS EFIJ DE QUATRE MODÈLES DIFFÉRENTS*

Statistique	Binomiale négative	Binomiale négative à effet aléatoire	Gamma-Dirichlet temps	Gamma-Dirichlet temps et flotte
MÉDIANE	-0,1203	-0,1021	-0,1000	-0,0821
IQR	0,1364	0,1111	0,1047	0,1218
MAD	0,0681	0,0556	0,0521	0,0567
MAE	0,2542	0,2312	0,2174	0,2150
ÉCART TYPE	0,4215	0,3870	0,3620	0,3639
Log L	-228 086,04	-225 384,30	-226 188,72	-222 811,18
BIC	456 658,76	451 268,44	452 877,28	446 293,19
Nombre de paramètres	37	38	44	51

* $e_{fij} = y_{fij} - E[Y_{fij} \mid \text{données}] = y_{fij} - E[E[Y_{fij} \mid \lambda_{fij}] \mid \text{données}] = y_{fij} - E[\lambda_{fij} \mid \text{données}]$

Il est aussi à noter que les résultats obtenus dans le Tableau 11 proviennent d'une approximation d'une intégrale en n'utilisant que deux groupes. Une analyse utilisant la méthode de Monte Carlo pour l'estimation numérique de l'intégrale nous donnerait une mesure plus exacte du degré d'approximation de notre méthode. Une autre extension serait de tester comment la sélection des deux groupes influence les résultats.

D'autres statistiques du Tableau 16 indiquent que la médiane du modèle général est plus près de zéro mais l'IQR est plus élevé. Les valeurs du MAD et de l'écart type des résidus sont, par contre, assez semblables des autres modèles.

Une autre analyse importante serait de calculer les corrélations entre les effets aléatoires, entre les résidus et les effets aléatoires, et entre $X\beta$ et les effets aléatoires et les résidus. Abowd et al. (2002) ont montré que la méthode d'estimation proposée dans Abowd et al. (1999) permettait d'éliminer les corrélations entre les résidus et la plupart des autres effets fixes. Ils ont aussi montré que l'effet individuel non observable explique beaucoup plus la variation de la variable dépendante que l'effet firme non observable.

Nous avons commencé à réaliser ces études de sensibilité des résultats mais n'avons pas encore de résultats. Il est à noter que notre modèle est non linéaire avec des effets aléatoires au lieu d'effets fixes, ce qui rend beaucoup plus difficile l'analyse des corrélations.

4. CONCLUSION

Dans cet article, nous proposons une importante extension aux modèles d'estimation des distributions d'accident avec données de panel. Nous ajoutons à l'effet temps un effet entreprise ou flotte pour tenir compte des corrélations potentielles entre les véhicules d'une même flotte. Nos résultats indiquent que l'ajout de l'effet flotte améliore significativement le pouvoir explicatif du modèle.

Plusieurs extensions sont possibles. La première consiste à estimer le modèle avec la méthode Monte Carlo pour vérifier la précision de l'approximation hypergéométrique. Il est possible que l'utilisation de seulement deux groupes ne soit pas suffisante. Une autre extension serait de générer une formule de tarification des flottes de véhicules de type bonus malus, en tenant compte à la fois de l'effet temps et de l'effet flotte. Finalement, il serait important de vérifier aussi jusqu'à quel point la méthode proposée réduit les corrélations entre les résidus et les variables explicatives et élimine les biais d'estimation.

Références

Abowd J-M., F. Kramarz, D.N. Margolis (1999) « High Wage Workers and High Wage Firms », *Econometrica*, 251-333.

Abowd J-M., R.H. Creecy, F. Kramarz (2002) « Computing Pearson and Firm Effects Using Linked Longitudinal Employer-Employee Data », Technical paper TP-2002-06, US Census Bureau, Suitland, USA.

Angers J-F., D. Desjardins, G. Dionne, F. Guertin (2004) « Vehicle and Fleet Random Effects in a Model of Insurance Rating for Fleets of Vehicles », Document de recherche, CRT et Chaire de recherche du Canada en gestion des risques, HEC Montréal, 67 pages (révisé en 2005). *ASTIN Bulletin* (à paraître)

Baltagi B.H. (1995) *Econometric Analysis of Panel Data*, Wiley, Chichester, p. 257.

Dionne G., D. Desjardins, J. Pinquet (2001) « Experience Rating Schemes for Fleets of Vehicles », *ASTIN Bulletin* 31, 1, 85-109.

Dionne, G., D. Desjardins, J. Pinquet (1999) « L'évaluation du risque d'accident des transporteurs en fonction de leur secteur d'activité, de la taille de leur flotte et de leur dossier d'infractions », rapport de recherche 99-28, Laboratoire sur la sécurité des transports du Centre de recherche sur les transports, Université de Montréal, 154 p.

Dionne G., R. Gagné, C. Vanasse (1998) « Inferring Technological Parameters from Incomplete Panel Data », *Journal of Econometrics* 87, 303-327.

Dionne G., C. Vanasse (1992) « Automobile Insurance Ratemarking in the Presence of Asymmetrical Information », *Journal of Applied Econometrics* 7, 149-165.

Dionne G., C. Vanasse (1989) « A Generalization of Automobile Insurance Rating Models: The Negative Binomial Distribution with a Regression Component », *ASTIN Bulletin* 19, 199-212.

Frangos N., S.D. Vrontos (2001) « Design of Optimal Bonus-Malus Systems with a Frequency and a Security Component on an Individual Basis in Automobile Insurance », *ASTIN Bulletin* 31, 1-22.

Fluet C. (1999) « Commercial Vehicle Insurance: Should Fleet Policies Differ From Single Vehicule Plans? », dans *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation*, G. Dionne et C. Laberge-Nadeau (Eds.), Kluwer Academic Press, 101-117.

Gouriéroux C. (1999) *Statistiques de l'assurance*, Economica, 297 pages.

Gradshteyn I.S. and Ryzhik I.M. (1980) « Table of Integrals, Series and Products », Academic Press Inc., New York, 1160 pages.

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

Hausman J.A., D.A. Wise (1979) « Attrition Bias in Experimental and Panel Data: the Gary Income Maintenance Experiment », *Econometrica* 47, 455-473.

Hsiao C. (1986) « Analysis of Panel Data », Econometric Society Monographs, no 11, Cambridge University Press, Cambridge.

Lange K. (1999) « Numerical Analysis for Statisticians », Springer, New York, section 21.2, 189-198.

Marie-Jeanne P. (1994) « Problèmes spécifiques des flottes automobiles », Proceedings of the ISUP conference « Cours Avancé sur l'Assurance Automobile ».

Pinquet J. (2000) « Experience Rating Trough Heterogeneous Models », dans *Handbook of Insurance*, G. Dionne (Ed.), Kluwer Academic Publishers, 459-500.

Purcaru O., M. Denuit (2003) « Dependence in Dynamic Claim Frequency Credibility Models », *ASTIN Bulletin* 33, 1, 23-40.

Scheffé H. (1999) « The Analysis of Variance », Wiley, New York, 477 p.

Société de l'assurance automobile du Québec (SAAQ) (1999) « Politique d'évaluation des propriétaires et des exploitants de véhicules lourds », Direction des communications, 107 p.

Société de l'assurance automobile du Québec (SAAQ) (1998) Dossier statistique, « Bilan 1997 des taxis, des autobus et des camions et tracteurs routiers », Service des études et des stratégies en sécurité routière, Direction de la planification et de la statistique, 166 p.