Modelling and Estimating Individual and Firm Effects with Count Panel Data

Jean-François Angers[a], Denise Desjardins[b]
Georges Dionne[b] and François Guertin[c]

April 29, 2015

Very preliminary; please do not quote

Abstract

In this article, we propose a new parametric model for the modelling and estimation of accident distributions for drivers working in fleets of vehicles. The analysis uses panel data and takes into account individual and fleet effects in a non-linear model. Our sample contains more than 456,000 observations of vehicles and 87,000 observations of fleets. Non-observable factors are treated as random effects. The distribution of accidents is affected by both observable and non-observable factors from drivers, vehicles and fleets. Past experience of both individual drivers and individual fleets is very significant to explain road accidents. Unobservable factors are also significant, which means that insurance pricing should take into account both observable and unobservable factors in predicting the rate of road accidents under asymmetric information.

Keywords: Accident distributions, drivers in fleet of vehicles, individual effect, firm effect, panel data, Poisson, gamma, Dirichlet, insurance pricing.

JEL codes: C23, C25, C55, G22.

[a] Department of Mathematics and Statistics, University of Montreal, Montreal
[b] Canada Research Chair in Risk Management, HEC Montréal, Montreal
[c] Calcul Québec, University of Montreal, Montreal

1. INTRODUCTION

Since the early 1980s, several researchers have proposed different models to account for correlation resulting from temporal repetitions of observations. Indeed, the use of panel-type individual data has become popular in many economic applications in the fields of labor economics, health economics, firm productivity, patents, transportation, and education (Hausman and Wise, 1979; Gourieroux et al. 1984; Hausman et al. 1984; Cameron and Trivedi, 1996; Hsiao, 1986; Baltagi, 1995; Abowd et al., 1999; Dionne et al., 1997, 1998).

In the domain of count-data applications, the ground-breaking contributions are the articles of Gourieroux et al. (1984) and Hausman et al. (1984). The second contribution proposes a Maximum Likelihood Method (MLE) for estimating the parameters, whereas the first work offers a pseudo-MLE treatment of the data. Both approaches yield the same results under the usual conditions.

Abowd et al. (1999) added another dimension to the estimation of models with panel data by estimating both the worker and the firm effect, allowing for observable and unobservable factors in a wage equation. If, in each period, we have observations on workers from the same firm, these observations may show common observable and non-observable characteristics. These characteristics must be suitably modelled. Abowd et al. (1999) proposes a linear model to estimate the wage equation, which is not suitable for accident distributions because accident distributions are usually non-linear. In this article we extend Hausman et al.'s (1984) parametric model to add a firm effect to the individual effect in the estimation of event distributions, and we apply the model to the accident distributions of trucks belonging to fleets of vehicles.[1]

To our knowledge there is no non-linear econometric model in the literature that estimates individual and firm effects with panel data. The matching of longitudinal individual and firm data is very important in environments where the observed results (here accidents) are a function of both parties' characteristics and unobserved actions. For insurance companies, knowing all the

[1] On insurance application with non-parametric models, see Pinquet (2013). On accident distribution estimation and insurance pricing see also Purcaru and Denuit (2003), Boucher, Denuit, and Montserrat (2008), and Frangos and Vrontos (2001).

sources of accidents involving vehicles belonging to a fleet is essential to develop a fair pricing scheme that takes into account the negligence of each actor. This is also important for the regulator, which has to compute the optimal fines of different infractions that affect accident distributions (Fluet, 1999).

In our application, we estimate the distribution of vehicle accidents for different fleets over time, by first decomposing the explanatory factors into heterogeneous factors linked to vehicles and their drivers, then into heterogeneous factors linked to fleets, and finally into residual factors. Factors linked to vehicles and drivers and those linked to fleets can be correlated. For example, a negligent manager may not spend enough money on mechanical repair of his trucks and might ask his employees to drive too fast. However, the employees may also exceed the speed limit without informing the manager.

As mentioned elsewhere, the modelling of Hausman et al. (1984) is not directly applicable to the ex-post calculation of insurance premiums using a Bayesian model (Angers et al., 2006). However the extended model we propose can be used for the insurance pricing of vehicles that includes individual and firm effects. Our model can also be applied to any non-linear count modelling with random individual and common effects. We may think of different principal-agent output such as operational risk events in banks, innovations in teams, deaths in hospitals, airline accidents, or any other event involving many agents working for a principal under asymmetric information (Holmstrom, 1982).

In section 2, we propose a short literature review of count data models, and section 3 develops our econometric model. Sections 4 and 5 present the data and the results of our estimations. We also analyze our results based on various statistical performance criteria. Section 6 concludes the paper.

2. LITERATURE

2.1 Basic count data models

Most of the econometric models applied to count variables that takes nonnegative values start from the Poisson distribution, where the probability of truck *i* of fleet *f* being involved in y_{fit} observable accidents (or claims) in period *t* can be represented by the following equation (Hausman et al., 1984; Gouriéroux et al. 1984; Gouriéroux, 1999):

$$P\left(Y_{fit} \mid \lambda_{fit}\right) = \frac{e^{-\lambda_{fit}}\left(\lambda_{fit}\right)^{y_{fit}}}{\Gamma\left(y_{fit}+1\right)}.$$

where $\Gamma(.)$ is the gamma function.

By definition of the Poisson law, the mathematical expectation of the number of accidents is equal to the variance, $E\left(Y_{fit}\right) = Var\left(Y_{fit}\right) = \lambda_{fit}$ where Y_{fit} is the random variable representing the number of accidents of truck *i*, fleet *f* in period *t* and $\lambda_{fit}\left(>0\right)$ is the Poisson parameter equal to the mean and the variance of the distribution. In fact, the parameter $\lambda_{fit} = \exp\left(X_{fit}\beta\right)$, where the vector $X_{fit} = \left(x_{fit1},\cdots,x_{fitp}\right)$ represents the *p* characteristics of truck *i* of fleet *f* observed in period *t* and β is a vector of parameters to be estimated. This is the equidispersion model. This modelling implicitly supposes that the distribution of accidents can be explained entirely by observable heterogeneity. In many applications, restricting the equality of the mean and the variance is not always compatible with the data. For car accidents, it often happens that the variance is higher than the average (Dionne and Vanasse, 1989, 1992), meaning that some heterogeneity is not observable by the econometrician and the insurance company.

To take into account the overdispersion property in the data, we can suppose that the parameter λ_{fit} has a random term such that $\lambda_{fit} = e^{X_{fit}\beta+\varepsilon_i} = \alpha_i\gamma_{fit}$ with $\alpha_i = e^{\varepsilon_i}$ and $\gamma_{fit} = e^{X_{fit}\beta}$ and where α_i is the random individual specific effect. Suppose that α_i follows a gamma distribution of parameter $\left(\delta^{-1},\delta^{-1}\right)$, we obtain the negative binomial distribution[2] (NB2 (Cameron and Trivedi, 1986; Boyer, Dionne, and Vanasse, 1992)):

$$P\left(Y_{fit}\mid\gamma_{fit},\delta\right)=\frac{\Gamma\left(\delta^{-1}+y_{fit}\right)}{\Gamma\left(\delta^{-1}\right)\Gamma\left(y_{fit}+1\right)}\left(\frac{\delta^{-1}}{\delta^{-1}+\gamma_{fit}}\right)^{\delta^{-1}}\left(\frac{\gamma_{fit}}{\delta^{-1}+\gamma_{fit}}\right)^{y_{fit}}, \qquad (1)$$

where the mean remains equal to $\exp\left(X_{fit}\beta\right)$ and the variance equals $\lambda_{fit}\left(1+\alpha_i\right)$.

This modelling does not simply allow for overdispersion. It also lets us consider unobserved or latent heterogeneity that is absent from the Poisson model. Unobserved heterogeneity is very important for pricing insurance premiums under asymmetric information (Dionne and Vanasse, 1989, 1992). The above modelling is appropriate for independent observations, meaning those without individual and time effects, and cannot be appropriate for panel data.

2.2 Taking time into account

Let us now consider data that contain observations where the same unit (individual or truck, for example) is observed over several successive periods but without firm or group effects. There are two treatments for panel data in the literature, the fixed effects and the random effects model. In this section we limit our discussion to the random effects NB model applied to short periods of time where the number of periods is fixed and the number of individuals is large. Hausman, Hall, and Griliches (1984) propose an extension of the model expressed by equation (1), which is not designed to take into account repetitions of observations over time. The new model is a hierarchical model that comes directly from the Poisson model.[3] Thus Y_{fit} would be distributed according to the NB2 model of parameters $\alpha_i\gamma_{fit}$ and ϕ_i, where α_i and ϕ_i vary across individuals. Suppose that $\left(1+\alpha_i/\phi_i\right)^{-1}$ follows a beta distribution of parameters (a,b), we obtain a closed form solution for the random effect negative binomial model:

[2] For an analysis of the Poisson lognormal mixture see Greene (2005).
[3] Note that the Poisson model can also be estimated with panel data. We do not consider this possibility here. See Cameron and Trivedi (2013) for a detailed analysis.

$$P\left(Y_{fi1},\cdots,Y_{fiT_i}\right)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma\left(a+\sum_t\gamma_{fit}\right)\Gamma\left(b+\sum_t y_{fit}\right)}{\Gamma\left(a+b+\sum_t\gamma_{fit}+\sum_t y_{fit}\right)}\prod_t\frac{\Gamma\left(y_{fit}+\gamma_{fit}\right)}{\Gamma\left(\gamma_{fit}\right)\Gamma\left(y_{fit}+1\right)}. \tag{2}$$

The NB2 model can also be estimated with individual dummies (or other methods) in a fixed effect version. The β parameters can be inconsistent however because of the incidental parameters problem, but some contributions have shown that the inconsistency may be not important (Allison and Waterman, 2002; Green, 2004).

Estimating the random effect model in (2) can also yield inconsistent random effects estimators because α_i and the vector of observable individual characteristics may be correlated. We can apply the Hausman test statistic to determine whether or not we should reject the null hypothesis that the individual effects are not correlated with the variables in the regression component.

The model in (2) is suitable for estimating parameters with individual effects but cannot take into account the firm or the fleet effect when individual observations belong to different firms with common characteristics that can affect accident distributions.

3. METHODOLOGY: Taking into account simultaneously firm and individual effects in a panel model

We now move on to the generalization of the model, which will allow us to account, simultaneously, for the individual effect, the firm effect and the time effect. We are interested in observations that have common characteristics because they belong to the same firm, for example: workers in a firm, vehicles in a fleet, patients in a hospital or children attending the same school.

Let us suppose that $\lambda_{fit}=\gamma_{fit}\left(\alpha_f\theta_{(f)i}\eta_{(fi)t}\right)$ with $\gamma_{fit}=e^{X_{fit}\beta}$. The vector $X_{fit}=\left(x_{fit1},\cdots,x_{fitp}\right)$ still represents the *p* characteristics of vehicle *i* from fleet *f* observed in period *t*. Here we can have many different fleets over a given number of periods. This vector contains specific information about the vehicle or the driver and other specific information about the fleet. β is a vector of *p* parameters to be estimated. Let α_f be the random effect associated with fleet *f* (i.e. the risk or non-observable characteristics attributable to the fleet), whereas $\theta_{(f)i}$ is the random effect of truck *i* of fleet *f* where $\sum_{i=1}^{I_f}\theta_{(f)i}=1$, I_f being the number of vehicles in fleet *f*. Finally $\eta_{(fi)t}$ is the time random effect of period *t* of truck *i* of fleet *f* such that $\sum_{t=1}^{T_i}\eta_{(fi)t}=1$ where T_i is the number of periods for truck *i*.

We posit that α_f follows a gamma distribution of parameters $\left(\sum_{i=1}^{I_f}T_i\ \kappa^{-1},\kappa^{-1}\right)$. The vector $\theta_{(f)}=(\theta_{(f)1},\theta_{(f)2},\cdots,\theta_{(f)I_f})$ follows a Dirichlet distribution of parameters $(\nu_{(f)1},\nu_{(f)2},\cdots,\nu_{(f)I_f})$ and the vector $\eta_{(fi)}=(\eta_{(fi)1},\eta_{(fi)2},\cdots,\eta_{(fi)T_i})$ follows a Dirichlet distribution of parameters $(\delta_{(fi)1},\delta_{(fi)2},\cdots,\delta_{(fi)T_i})$ where T_i is the number of periods of vehicle *i*.

The conditional distribution of the number of accidents for all the vehicles in fleet *f* is given by:

$$\begin{aligned}P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\mid\alpha_f,\theta_{(f)},\eta_{(fi)}\right)&=\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}P\left(Y_{fit}\mid\lambda_{fit}\right)\\&=\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\frac{e^{-\lambda_{fit}}\left(\lambda_{fit}\right)^{y_{fit}}}{\Gamma\left(y_{fit}+1\right)}\\&=\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\left(\frac{1}{\Gamma\left(y_{fit}+1\right)}\right)\right]\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\left(\lambda_{fit}\right)^{y_{fit}}\right]e^{-\sum_{i=1}^{I_f}\sum_{t=1}^{T_i}\lambda_{fit}}.\end{aligned}\tag{3}$$

Since $\lambda_{fit}=\gamma_{fit}\left(\alpha_f\theta_{(f)i}\eta_{(fi)t}\right)$ then :

$$\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\left(\lambda_{fit}\right)^{y_{fit}}=\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\left(\gamma_{fit}\right)^{y_{fit}}\right]\left[\left(\alpha_f\right)^{\sum_{i=1}^{I_f}\sum_{t=1}^{T_i}y_{fit}}\right]\left[\prod_{i=1}^{I_f}\left(\theta_{(f)i}\right)^{\sum_{t=1}^{T_i}y_{fit}}\right]\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\left(\eta_{(fi)t}\right)^{y_{fit}}\right]\tag{4}$$

Let $S_i=\sum_{t=1}^{T_i}y_{fit}$ and $S_0=\sum_{i=1}^{I_f}\sum_{t=1}^{T_i}y_{fit}=\sum_{i=1}^{I_f}S_i$, equation (4) can be rewritten as follows

$$\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}(\lambda_{fit})^{y_{fit}}=\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}(\gamma_{fit})^{y_{fit}}\right]\left[(\alpha_f)^{S_0}\right]\left[\prod_{i=1}^{I_f}\left(\theta_{(f)i}\right)^{S_i}\right]\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\left(\eta_{(fi)t}\right)^{y_{fit}}\right].$$

Moreover, the summation $\sum_{i=1}^{I_f}\sum_{t=1}^{T_i}\lambda_{fit}$ in equation (3) can be written as

$$\alpha_f\sum_{i=1}^{I_f}\theta_{(f)i}\sum_{t=1}^{T_i}\gamma_{fit}\eta_{(fi)t}\,.$$

The joint distribution of the number of accidents for all the vehicles in fleet *f* is given by:

$$P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\right)=\int\cdots\int P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\mid\eta_{(fi)}\right)f\left(\underset{\sim}{\eta}_{(fi)}\right)d\eta_{(fi)} \tag{5}$$

with
$$P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\mid\eta_{(fi)}\right)=\int\cdots\int P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\mid\theta_{(f)},\eta_{(fi)}\right)f\left(\theta_{(f)}\right)d\theta_{(f)} \tag{6}$$

and
$$P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\mid\theta_{(f)},\eta_{(fi)}\right)=\int_0^\infty P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\mid\alpha_f,\theta_{(f)},\eta_{(fi)}\right)f\left(\alpha_f\right)d\alpha_f\,. \tag{7}$$

By integrating equation (7) with respect to α_f, we obtain

$$\frac{\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\frac{(\gamma_{fit})^{y_{fit}}}{\Gamma(y_{fit}+1)}\right]\left[\prod_{i=1}^{I_f}\left(\theta_{(f)i}\right)^{S_i}\right]\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\left(\eta_{(fi)t}\right)^{y_{fit}}\right]\left[\left(\kappa^{-1}\right)^{\sum_{i=1}^{I_f}T_i\kappa^{-1}}\right]\left[\Gamma\left(S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)\right]}{\Gamma\left(\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)\left[\left(\kappa^{-1}+\sum_{i=1}^{I_f}\theta_{(f)i}\sum_{t=1}^{T_i}\gamma_{fit}\eta_{(fi)t}\right)^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}\right]}. \tag{8}$$

By replacing $P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\mid\theta_{(f)},\eta_{(fi)}\right)$ in equation (6) by its value given in (8) and by replacing the density function $f\left(\theta_{(f)}\right)$by the density of a parametric Dirichlet distribution of parameters ($\nu_{(f)1},\nu_{(f)2},\cdots,\nu_{(f)I_f}$), we obtain the following expression:

$$P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}} \mid \eta_{(fi)}\right)$$

$$= \frac{\left[\Gamma\left(S_0+\sum_{i=1}^{I_f} T_i\kappa^{-1}\right)\right]\left[\left(\kappa^{-1}\right)^{\sum_{i=1}^{I_f} T_i\kappa^{-1}}\right]\left[\Gamma\left(\sum_{i=1}^{I_f}\nu_{(f)i}\right)\right]\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\left(\gamma_{fit}\eta_{(fi)t}\right)^{y_{fit}}\right]}{\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\Gamma\left(y_{fit}+1\right)\right]\left[\Gamma\left(\sum_{i=1}^{I_f} T_i\kappa^{-1}\right)\right]\left[\prod_{i=1}^{I_f}\Gamma\left(\nu_{(f)i}\right)\right]} \int\cdots\int \frac{\prod_{i=1}^{I_f}\left(\theta_{(f)i}\right)^{S_i+\nu_{(f)i}-1}}{\left(\kappa^{-1}+\sum_{i=1}^{I_f}\theta_{(f)i}\sum_{t=1}^{T_i}\gamma_{fit}\eta_{(fi)t}\right)^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}} d\theta_{(f)} . \quad (9)$$

We must estimate the multidimensional integral:

$$\int\cdots\int \frac{\prod_{i=1}^{I_f}\left(\theta_{(f)i}\right)^{S_i+\nu_{(f)i}-1}}{\left(\kappa^{-1}+\sum_{i=1}^{I_f}\theta_{(f)i}\sum_{t=1}^{T_i}\gamma_{fit}\eta_{(fi)t}\right)^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}} d\theta_{(f)}$$

of equation (9) to obtain the model’s parameters. We analyze two possibilities:

3.1. All trucks have identical *a priori* risk

This first possibility, which greatly simplifies the estimations, is to suppose that all the γ_{fit} of the I_f vehicles are identical and equal to γ, for all periods. Under this hypothesis, the multidimensional integral of equation (9) is reduced to:

$$\int\cdots\int \frac{\prod_{i=1}^{I_f}\left(\theta_{(f)i}\right)^{S_i+\nu_{(f)i}-1}}{\left(\kappa^{-1}+\gamma\sum_{i=1}^{I_f}\theta_{(f)i}\sum_{t=1}^{T_i}\eta_{(fi)t}\right)^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}} d\theta_{(f)} = \frac{\prod_{i=1}^{I_f}\Gamma\left(S_i+\nu_{(f)i}\right)}{\left(\left(\kappa^{-1}+\gamma\right)^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}\right)\Gamma\left(\sum_{i=1}^{I_f}\left(S_i+\nu_{(f)i}\right)\right)} \quad (10)$$

and the joint distribution of the number of accidents at period *t* for the I_f vehicles in fleet *f* is given by the following expression:

$$P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}} \mid \eta_{(fi)}\right)$$

$$= \frac{\left[\Gamma\left(S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)\right]\left[\left(\kappa^{-1}\right)^{\sum_{i=1}^{I_f}T_i\kappa^{-1}}\right]\left[\Gamma\left(\sum_{i=1}^{I_f}\nu_{(f)i}\right)\right]\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\left(\gamma_{fit}\eta_{(fi)t}\right)^{y_{fit}}\right]}{\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\Gamma\left(y_{fit}+1\right)\right]\left[\Gamma\left(\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)\right]\left[\prod_{i=1}^{I_f}\Gamma\left(\nu_{(f)i}\right)\right]} \frac{\prod_{i=1}^{I_f}\Gamma\left(S_i+\nu_{(f)i}\right)}{\left(\left(\kappa^{-1}+\gamma\right)^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}\right)\Gamma\left(\sum_{i=1}^{I_f}\left(S_i+\nu_{(f)i}\right)\right)} \tag{11}$$

Further, by replacing $P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}} \mid \eta_{(fi)}\right)$ in equation (5) by the expression in (11) and by replacing the density function $f\left(\eta_{(fi)}\right)$by the density of a parametric Dirichlet distribution of parameters ($\delta_{(fi)1},\delta_{(fi)2},\cdots,\delta_{(fi)T_i}$), we obtain the following approximation for (5), the joint distribution of the number of accidents for all the vehicles in all fleets:

$$P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\right)=$$

$$\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\frac{\left(\gamma_{fit}\right)^{y_{fit}}}{\Gamma\left(y_{fit}+1\right)}\right]\frac{\left[\Gamma\left(S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)\right]\left[\left(\kappa^{-1}\right)^{\sum_{i=1}^{I_f}T_i\kappa^{-1}}\right]\left[\Gamma\left(\sum_{i=1}^{I_f}\nu_{(f)i}\right)\right]}{\left[\Gamma\left(\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)\right]\left[\prod_{i=1}^{I_f}\Gamma\left(\nu_{(f)i}\right)\right]}\frac{\prod_{i=1}^{I_f}\Gamma\left(S_i+\nu_{(f)i}\right)}{\left(\left(\kappa^{-1}+\gamma\right)^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}\right)\Gamma\left(\sum_{i=1}^{I_f}\left(S_i+\nu_{(f)i}\right)\right)} \tag{12}$$

$$\times\left[\frac{\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\Gamma\left(y_{fit}+\delta_{(fi)t}\right)}{\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\Gamma\left(\delta_{(fi)t}\right)}\right]\left[\frac{\prod_{i=1}^{I_f}\Gamma\left(\sum_{t=1}^{T_i}\delta_{(fi)t}\right)}{\prod_{i=1}^{I_f}\Gamma\left(S_i+\sum_{t=1}^{T_i}\delta_{(fi)t}\right)}\right]$$

The main working hypothesis for this first scenario supposes implicitly that all the vehicles in the fleet represent identical *a priori* risks, which is probably a very strong hypothesis because, as we shall see, several variables distinguishing vehicles and driver behavior are significant in estimating the probabilities of accidents of different vehicles. Another possibility is to divide the vehicles into homogeneous risk groups, as insurers do when classifying risks.

3.2 Trucks belong to different groups

Under this second possibility, we suppose that $\gamma_{fit} = \overline{\gamma}_{fi} \forall t = 1, \cdots, T_i$ where $\overline{\gamma}_{fi} = \frac{1}{T_i} \sum_{i=1}^{T_i} \gamma_{fit}$. The integral of equation (9) becomes:

$$\int \cdots \int \frac{\prod_{i=1}^{I_f} \left(\theta_{(f)i}\right)^{S_i + \nu_{(f)i} - 1}}{\left(\kappa^{-1} + \sum_{i=1}^{I_f} \theta_{(f)i} \overline{\gamma}_{fi}\right)^{S_0 + \sum_{i=1}^{I_f} T_i \kappa^{-1}}} d\theta_{(f)} . \tag{13}$$

We can separate the vehicles into two groups (high risk and low risk) and define $G_1 = 1, \cdots, g_1$ as the set of all vehicles of the first group with $\overline{\gamma}_{g_1} = \frac{\sum_{i=1}^{g_1} \overline{\gamma}_{fi}}{g_1}$, and $G_2 = g_1 + 1, \cdots, I_f$, as the set of all vehicles of the second group with $\overline{\gamma}_{g_2} = \frac{\sum_{i=g_1+1}^{I_f} \overline{\gamma}_{fi}}{g_2}$.

The integral of equation (13) thus becomes:

$$\int \cdots \int \frac{\left[\prod_{i=1}^{g_1} \left(\theta_{(f)i}\right)^{S_i + \nu_{(f)i} - 1}\right]\left[\prod_{i=g_1+1}^{I_f} \left(\theta_{(f)i}\right)^{S_i + \nu_{(f)i} - 1}\right]}{\left(\kappa^{-1} + \overline{\gamma}_{g_1} \sum_{i=1}^{g_1} \theta_{(f)i} + \overline{\gamma}_{g_2} \sum_{i=g_1+1}^{I_f} \theta_{(f)i}\right)^{S_0 + \sum_{i=1}^{I_f} T_i \kappa^{-1}}} d\theta_{(f)} . \tag{14}$$

Let $v = \sum_{i=1}^{I_f} \theta_{(f)i}$, $u_i = \frac{\left(\theta_{(f)i}\right)^{\chi_{1i}}}{v}$ and $w_i = \frac{\left(\theta_{(f)i}\right)^{\chi_2}}{1 - v}$. The integral of equation (14) can be rewritten as follows:

$$\int\cdots\int \frac{\left[\left[\prod_{i\neq g_1}(vu_i)^{S_i+\nu_{(f)i}-1}\right]\left[\prod_{i\neq I_f}\left((1-v)w_{ij}\right)^{S_i+\nu_{(f)i}-1}\right]\left[v\left(1-\sum_{i\neq g_1}u_i\right)\right]^{S_{g_1}+\nu_{(f)g_1}-1}\left[(1-v)\left(1-\sum_{i\neq I_f}w_i\right)\right]^{S_{I_f}+\nu_{(f)I_f}-1}\right]}{\left(\left(\kappa_f^{-1}+\overline{\gamma}_{g1}\right)v+\left(\kappa_f^{-1}+\overline{\gamma}_{g2}\right)(1-v)\right)^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}} v^{g_1-1}(1-v)^{I_f-g_1-1}\,dudwdv$$

By integrating we obtain

$$=\frac{\prod_{i=1}^{I_f}\Gamma\left(S_i+\nu_{(f)i}\right)}{\left[\Gamma\left(\sum_{i=1}^{I_f}S_i+\nu_{(f)i}\right)\right]\left[\left(\kappa^{-1}+\overline{\gamma}_{g_2}\right)^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}\right]}\,{}_2F_1\left(\sum_{i=1}^{g_1}S_i+\nu_{(f)i},S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1},\sum_{i=1}^{I_f}S_i+\nu_{(f)i},\left(\frac{\overline{\gamma}_{g_2}-\overline{\gamma}_{g_1}}{\kappa^{-1}+\overline{\gamma}_{g_2}}\right)\right). \tag{15}$$

Thus, by replacing the integral in equation (14) by its value given in (15) we obtain the following approximation for $P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\mid\eta_{(fi)}\right)$ in (9):

$$\frac{\Gamma\left(S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)\left(\kappa^{-1}\right)^{\sum_{i=1}^{I_f}T_i\kappa^{-1}}\Gamma\left(\sum_{i=1}^{I_f}\nu_{(f)i}\right)\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\left(\gamma_{fit}\eta_{(fi)t}\right)^{y_{fit}}}{\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\Gamma\left(y_{fit}+1\right)\Gamma\left(\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)\prod_{i=1}^{I_f}\Gamma\left(\nu_{(f)i}\right)}\left(\frac{\prod_{i=1}^{I_f}\Gamma\left(S_i+\nu_{(f)i}\right)}{\Gamma\left(\sum_{i=1}^{I_f}S_i+\nu_{(f)i}\right)}\right)\frac{1}{\left(\kappa^{-1}+\overline{\gamma}_{g_2}\right)^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}} \tag{16}$$
$$\times\,{}_2F_1\left(\sum_{i=1}^{g_1}S_i+\nu_{(f)i},S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1},\sum_{i=1}^{I_f}S_i+\nu_{(f)i},\left(\frac{\overline{\gamma}_{g_2}-\overline{\gamma}_{g_1}}{\kappa^{-1}+\overline{\gamma}_{g_2}}\right)\right)$$

where ${}_2F_1$ is a hypergeometric function whose value is equal to:

$$1+\sum_{\ell=1}^{\infty}\left[\frac{\left(\sum_{i=1}^{g_1}S_i+\nu_{(f)i}\right)^{[\ell]}\left(S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)^{[\ell]}}{\left(\sum_{i=1}^{I_f}S_i+\nu_{(f)i}\right)^{[\ell]}}\frac{\left(\frac{\overline{\gamma}_{g_2}-\overline{\gamma}_{g_1}}{\kappa^{-1}+\overline{\gamma}_{g_2}}\right)^{\ell}}{\ell!}\right],$$

with $h^{[\ell]}=h(h+1)\cdots(h+\ell+1)$, being an increasing factorial function.

Further, by replacing $P\left(Y_{f11},\cdots,Y_{fI_fT_{I_f}}\mid\eta_{(fi)}\right)$ in equation (5) by the expression in (16) and by replacing the density function $f\left(\eta_{(fi)}\right)$ by the density of a parametric Dirichlet distribution of parameters ($\delta_{(fi)1}, \delta_{(fi)2}, \cdots, \delta_{(fi)T_i}$), we obtain the following approximation for (5), the joint distribution of the number of accidents for all the vehicles in all fleets:

$$\left[\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\frac{(\gamma_{fit})^{y_{fit}}}{\Gamma(y_{fit}+1)}\right]\left[\frac{\Gamma\left(S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)}{\Gamma\left(\sum_{i=1}^{I_f}T_i\kappa^{-1}\right)}\right]\left[\frac{(\kappa^{-1})^{\sum_{i=1}^{I_f}T_i\kappa^{-1}}}{(\kappa^{-1}+\overline{\gamma}_{g_2})^{S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}}}\right]\left[\frac{\Gamma\left(\sum_{i=1}^{I_f}\nu_{(f)i}\right)}{\Gamma\left(\sum_{i=1}^{I_f}S_i+\nu_{(f)i}\right)}\right]\left[\frac{\prod_{i=1}^{I_f}\Gamma(S_i+\nu_{(f)i})}{\prod_{i=1}^{I_f}\Gamma(\nu_{(f)i})}\right]$$
$$\times\left[\frac{\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\Gamma(y_{fit}+\delta_{(fi)t})}{\prod_{i=1}^{I_f}\prod_{t=1}^{T_i}\Gamma(\delta_{(fi)t})}\right]\left[\frac{\prod_{i=1}^{I_f}\Gamma\left(\sum_{t=1}^{T_i}\delta_{(fi)t}\right)}{\prod_{i=1}^{I_f}\Gamma\left(S_i+\sum_{t=1}^{T_i}\delta_{(fi)t}\right)}\right]\times {}_2F_1\left(\sum_{i=1}^{g_1}(S_i+\nu_{(f)i}), S_0+\sum_{i=1}^{I_f}T_i\kappa^{-1}, \sum_{i=1}^{I_f}(S_i+\nu_{(f)i}), \left(\frac{\overline{\gamma}_{g_2}-\overline{\gamma}_{g_1}}{\kappa^{-1}+\overline{\gamma}_{g_2}}\right)\right). \tag{17}$$

It should be noted that equation (12) is a particular case of equation (17). This procedure in estimating the integral can be generalized to several homogeneous groups, but it is not obvious that the precision gained would be greater than that corresponding to a Monte Carlo approximation of the multivariate integral of equation (9), which is not presented here. In Section 5, we present the econometric results obtained from equation (17).

3.3 Parameters estimation

Let $\nu_{(f)i} = \nu \ \forall i$ and $\delta_{(fi)t} = \delta \ \forall t$, we can apply the maximum likelihood method to estimate the unknown parameters, ν, κ^{-1}, δ and β of the log likelihood function of equation (17). In the application, presented in section 5, we will apply the Newton-Raphson method of estimation. This is a nonlinear optimization subroutine available in the IML procedure in SAS. The initial values of the vector β are the maximum likelihood estimates of the NB2 model, and the initial values for ν, κ^{-1}, δ parameters are set to one.

To determine the variance-covariance matrix of the asymptotic distribution, we apply the subroutine NLPFDD in the IML procedure of SAS to compute finite difference approximations for first and second order derivatives to obtain the Hessian matrix at $\hat{\nu}, \hat{\kappa}^{-1}, \hat{\delta}$ and $\hat{\beta}$.

The size of the data is quite large; to reduce the computation time drastically, we compute the log likelihood function with a homemade C program outside of the SAS system. We created an interface to communicate between the SAS program and the C program. We invoked a method for passing data to and from the SAS system called "named pipe". The execution speed in SAS/IML is very slow compared with the C program. We obtained a speed of 20 to 30 times faster. We can further gain in execution speed by adding many machines in parallel because each machine can make the calculations independently.

To divide the trucks of a fleet into two homogeneous groups as shown in equation (14), we take the maximum likelihood estimates $\left(\hat{\beta}\right)$ of the NB2 model to estimate $\hat{\gamma}_{fit} = e^{X_{fit}\hat{\beta}}$ for all the vehicles. Given that a truck has an estimate by year of follow-up, we calculated its mean $\hat{\bar{\gamma}}_{fi} = \frac{1}{T_i}\sum_{i=1}^{T_i}\hat{\gamma}_{fit}$. We sorted $\hat{\bar{\gamma}}_{fi}$ for $i = 1, \cdots, I_f$ and calculated the difference $\left(\hat{\bar{\gamma}}_{fi+1} - \hat{\bar{\gamma}}_{fi}\right)$ for $i = 1, \cdots, I_f - 1$. After, we choose a cut-off point γ_{fc} where c is such that $\arg\max\left(\hat{\bar{\gamma}}_{fi+1} - \hat{\bar{\gamma}}_{fi}\right)$. The truck *i* is in group 1 if $\hat{\bar{\gamma}}_{fi} < \gamma_{fc}$ or is in group 2 if $\hat{\bar{\gamma}}_{fi} \geq \gamma_{fc}$ for all the observations of the truck *i*.

For example, for a fleet of 8 trucks and 20 observations (truck-years) as shown in Table 1, $c = \arg\max\left(0.08732\right) = 6$. Then the cutoff point $\gamma_{fc} = \hat{\bar{\gamma}}_{f6} = 0.23357$. All observations of truck 1 to truck 5 will therefore be in group 1 (low risk group) and all the others will be in group 2 (high risk group) . If we use the median or the mean instead of the maximum difference then the cut-off point will be respectively 0.12067 ((0.09509+0.14625)/2) and 0.14605, and truck 5 will change to group 2. However, it is more appropriate to be in group 1 because $\hat{\bar{\gamma}}_{f5}$ is nearer to those in group 1.

Table 1
Example of group division, fleet of 8 trucks and 20 truck-years

Truck	Year	$\hat{\gamma}_{fit}$	$\hat{\bar{\gamma}}_{fi}$	Difference	Group
1	94	0.02527			1
	95	0.06524	0.04526		1
2	91	0.02417			1
	92	0.07178			1
	93	0.06422			1
	94	0.07340			1
	95	0.06423	0.05956	0.01430	1
3	91	0.09947			1
	92	0.09067	0.09067	0.03111	1
4	91	0.09677			1
	92	0.09817			1
	93	0.09033	0.09509	0.00442	1
5	91	0.15184			1
	92	0.14065	0.14625	0.05116	1
6	91	0.22807			2
	92	0.23906	0.23357	0.08732	2
7	91	0.25807			2
	92	0.23906	0.24857	0.01500	2
8	91	0.25989			2
	92	0.23906	0.24948	0.00091	2
Mean			0.14605		

4. DATA

The *Société de l'assurance automobile du Québec* (SAAQ) provided the files for our data set. The SAAQ (1998, 1999) is in charge of monitoring whether vehicles engaged in road transportation of people or merchandise comply with applicable laws and regulations. The SAAQ is also the insurer for bodily injuries linked to traffic accidents for individuals and fleets of vehicles.

Our starting point is the whole population of carriers registered in a SAAQ file on July 1997. To be in that file of carriers, a carrier must be the owner of a vehicle that meets one of the conditions below:

- The vehicle belongs to one of the following plate categories: commercial; general or bulk transportation of merchandise;
- The plate status is “H” for hold;
- The vehicle is a truck;
- The vehicle’s authorization status is active or reconstructed;
- For trucks, the net mass exceeds 3,000 kg and the vehicle is not used for emergencies.

Linked to each carrier, the data contain: (1) information on violations (with convictions) committed by the carrier during the 1989-1998 period, either for non-compliance with the Highway Safety Code’s provisions on mechanical inspection; with rules on vehicles and their equipment; with codes on driving and hours of service or for oversize or poor load securement, etc., and (2) information identifying the carrier.

We also have access to information on vehicles registered in Quebec for the period of January 1, 1990 to December 31, 1998. We can link vehicles to carriers. From the authorization status, we obtain information describing the vehicle and plates. For each plate number, we have data covering the 1990-1998 period drawn from the files on mechanical inspection of vehicles and from the record of drivers’ violations with conviction and demerit points for speeding, failure to stop at a red light or stop sign, and illegal passing, and for accidents.

The choice and the description of the variables used in this study can be found in Appendix A.

5. RESULTS

5.1 Descriptive statistics

5.1.1 By fleet

We have 17,542 fleets with at least two trucks with a follow-up of at least two periods. In December 31, 1998, the average number of years each carrier is in the sample is seven, the minimum is one year and three months, and the maximum is 20 years and 9 months.

Table 2 shows the distribution of fleets according to their main economic sector: 76.75% of 17,542 carriers are independent trucking firms, 13.09% are bulk public trucking firms and 8.70% are general public trucking firms. The sector is unknown for only a few firms. In addition, a few fleets also transport passengers or are short-term leasing firms.

Table 2
Distribution of firm's main activity

Firm's main activity	N	%
Unknown (sect_00)	63	0.36
Transporting passengers (sect_14)	71	0.40
General public trucking (sect_05)	1,526	8.70
Independent trucking (sect_07)	13,464	76.75
Short-term leasing firm (sect_08)	121	0.69
Bulk public trucking (sect_06)	2,297	13.09
Total	17,542	100.00

We note in Table 3 that approximately 4% of 17,542 fleets have over 20 trucks. On average, a truck has 3.87 observation periods ranging from 3.38 (a truck from a fleet of size 2) to 4.30 (a truck from a fleet of 10 to 20 trucks).

Table 3
Size of fleet distribution

Size of fleet	N	%	Average observation period per truck
2	6,888	39.27	3.38
3	3,203	18.26	4.07
4 to 5	3,285	18.73	4.18
6 to 9	2,171	12.38	4.29
10 to 20	1,298	7.40	4.30
21 to 50	496	2.83	4.24
More than 50	201	1.15	4.03
Total	17,542	100.00	3.87

We note in Table 4 that a quarter of the 17,542 carriers have eight years of follow-up.

Table 4
Number of years of follow-up of the firm

Number of years of follow-up	N	%
2	3,649	20.80
3	2,512	14.32
4	2,075	11.83
5	1,654	9.43
6	1,645	9.38
7	1,567	8.93
8	4,440	25.31
Total	17,542	100.00

We note in Table 5 that there are 3,629 fleets for which we have two consecutive years of follow-up, which is 99.5% (3,629/3,649) of fleets with two observation periods (Table 4). This percentage varies from 98.96 (3 periods) to 87.17% (7 periods). The higher the number of years of follow-up, the higher the percentage of carriers with absences during the reporting period.

Table 5
Number of consecutive years of follow-up of the fleets by year of follow-up start, Quebec 1991 to 1997

Number of years of follow-up	Year of follow-up start							Total
	1991	1992	1993	1994	1995	1996	1997	
2	949	296	263	272	332	325	1,192	3,629
3	708	251	178	180	231	938		2,486
4	619	174	166	169	807			1,935
5	448	181	131	739				1,499
6	565	144	783					1,492
7	483	883						1,366
8	4,440							4,440
Total	8,212	1,929	1,521	1,360	1,370	1,263	1,192	16,847

Table 6 shows the distribution of the size of the fleet by year and total of 8 years. In 1991, we have 8,650 fleets, this number increases to 11,965 fleets in 1996 and decreases to 10,321 in 1998 for a total of 87,771 fleet-years. Among the 87,771 fleet-years, 46.51% have two vehicles and about 3% have over 20 vehicles.

Table 6
Size of fleet distribution (in %) by year

Size of fleet	% by year								% total
	1991	1992	1993	1994	1995	1996	1997	1998	
2	47.86	46.31	46.60	46.82	45.98	46.08	45.83	47.03	46.51
3	19.63	19.75	19.92	19.71	19.45	19.30	19.28	19.10	19.51
4 to 5	15.26	15.99	15.75	15.54	16.28	16.02	16.40	16.26	15.96
6 to 9	9.16	9.40	9.19	9.46	9.62	9.88	9.62	9.27	9.47
10 to 20	5.45	5.72	5.76	5.74	5.76	5.68	5.95	5.64	5.72
21 to 50	1.97	2.06	2.00	1.89	2.05	2.14	2.08	2.01	2.03
More than 50	0.68	0.78	0.78	0.83	0.86	0.89	0.85	0.70	0.80
Number of fleets	8,650	10,691	11,132	11,445	11,733	11,965	11,834	10,321	87,771

From Table 7, we observe that 9,963 fleets remain in the same class of fleet size during the eight years of observation, which is 56.80% of 17,542 fleets (sum of the diagonal of Table 7). There are 2,722 fleets whose size varies between 2 and 3 trucks, and 1,423 fleets whose size varies between 2 trucks to 4-5 trucks.

Table 7
Minimum and maximum fleet size distribution during the follow-up, Québec 1991-1998

Minimum size of fleet		Maximum size of fleet							Total	
		2	3	4 to 5	6 to 9	10 to 20	21 to 50	+ 50	N	%
2		7,884	2,722	1,423	365	101	19	2	12,561	71.35
3			790	946	368	85	14	2	2,205	12.57
4 to 5				551	644	172	22	1	1,390	7.92
6 to 9					320	394	57	12	783	4.46
10 to 20						268	152	22	442	2.52
21 to 50							93	56	149	0.85
More than 50								57	57	0.32
Total	N	7,884	3,512	2,920	1,697	1,020	357	152	17,542	100.00
	%	44.94	20.20	16.65	9.67	5.81	2.04	0.84	100.00	

We observe from Table 8 that the average accident rate of trucks per fleet is lowest for the year 1997, followed by 1993, 1998, 1996 and 1994. In the years 1991, 1992 and 1995, the highest average rates of truck accidents per fleet were recorded. These observations are almost stable for different fleet sizes.

Table 8
Average truck accidents per fleet according to size of fleet and year.

Size of fleet	Average truck accident per fleet by year 1991	1992	1993	1994	1995	1996	1997	1998	Total
2	0.2626	0.2480	0.2219	0.2215	0.2219	0.2155	0.1809	0.2186	0.2224
3	0.4370	0.4154	0.3811	0.4007	0.4194	0.3712	0.3129	0.4049	0.3909
4 to 5	0.6689	0.6864	0.6030	0.6296	0.6408	0.5863	0.5507	0.6490	0.6239
6 à 9	1.3914	1.2259	1.0909	1.1311	1.1833	1.0981	1.0018	1.1996	1.1550
10 to 20	2.6730	2.6127	2.3744	2.5099	2.4527	2.4824	2.0767	2.5223	2.4497
21 to 50	5.9176	5.3818	5.0448	5.3565	5.8875	5.2461	4.8618	5.7681	5.4094
More than 50	22.6780	22.4096	21.7701	22.0421	22.0198	21.0935	18.4700	22.5417	21.5014
Average truck accidents per fleet	0.8575	0.8561	0.7824	0.8157	0.8531	0.8153	0.7106	0.8120	0.8109

5.1.2 By truck

We have 111,106 different trucks in the database, nearly three-quarters of which are for commercial use, including transportation of goods without a permit from the Commission de Transports du Québec (CTQ). As indicated in Table 9, 17.52% of the trucks are used for transportation of goods other than bulk.

Table 9
Vehicle use distribution.

Vehicle use	N	%
Commercial use, including transport of goods without CTQ permit. (combr)	82,798	74.52
Transport of goods other than in bulk (tbrgn)	19,470	17.52
Transport of goods in bulk (tbrvr)	8,838	7.95
Total	111,106	100.00

We note in Table 10 that 78.80% of the 111,106 trucks use diesel as fuel.

Table 10
Type of fuel distribution

Type of fuel	N	%
Diesel	87,546	78.80
Gas	22,999	20.70
Other	561	0.50
Total	111,106	100.00

Table 11 illustrates that 21.15% of the 111,106 trucks have six axles or more and 28.57% have two axles and weigh more than 4,000 kg, and Table 12 shows that 64.95% of the 111,106 trucks have 6 to 7 cylinders. Only 1.15% has 5 cylinders or fewer.

Table 11
Number of axles distribution

Number of axles	N	%
2 axles (3,000 to 4,000 kg)	15,960	14.36
2 axles (More than 4,000 kg)	31,747	28.57
3 axles	21,856	19.67
4 axles	7,377	6.64
5 axles	10,666	9.60
6 axles and more	23,500	21.15
Total	111,106	100.00

Table 12
Number of cylinders distribution

Number of cylinders	N	%
Unknown	501	0.45
1 to 5 cylinders	1,283	1.15
6 to 7 cylinders	71,159	64.05
8 or more than 10 cylinders	38,163	34.35
Total	111,106	100.00

Table 13 indicates that 10.64% of the 111,106 trucks have 8 years of follow-up, which represents 10.64% of the population.

Table 13
Number of years of follow-up of the truck

Number of years of follow-up	N	%
2	30,716	27.65
3	23,270	20.94
4	17,831	16.05
5	11,998	10.80
6	9,241	8.32
7	6,225	5.60
8	11,825	10.64
Total	111,106	100.00

We note in Table 14, that there are 30,432 trucks for which we have two consecutive years of follow-up, which corresponds to 99.07% (30,432/30,716) of trucks with two observation periods (Table 13). This percentage varies from 98.43 (3 periods) to 97.65 (7 periods).

Table 14
Number of consecutive years of follow-up of the trucks by year of follow-up start, Quebec 1991 to 1997.

Number of year of follow-up	Year of follow-up start							Total
	1991	1992	1993	1994	1995	1996	1997	
2	8,326	2,581	2,193	2,081	2,844	2,351	10,056	30,432
3	6,421	2,291	1,624	1,855	1,947	8,766		22,904
4	5,273	1,535	1,711	1,524	7,304			17,347
5	3,967	1,289	1,067	5,226				11,549
6	3,680	818	4,441					8,939
7	2,630	3,449						6,079
8	11,825							11,825
Total	42,122	11,963	11,036	10,686	12,095	11,117	10,056	109,075

5.1.3 By truck-years

There are 43,037 trucks in 1991. This number increased to 63,749 in 1996. It decreases to 52 392 in 1998 for a total of 456,177 truck-years, 15% of which had an accident during one year (Table 15). In 1991, nearly 86 out of 100 vehicles had no accident; this percentage rises to 88 out of 100 in 1997.

Table 15
Number of truck accidents distribution according to year of observation

Number of truck accidents	% (by year of observation)								Total
	1991	1992	1993	1994	1995	1996	1997	1998	
0	85.61	86.07	87.19	86.66	86.47	87.04	88.44	86.52	86.78
1	12.22	11.93	11.05	11.47	11.53	11.14	10.11	11.56	11.36
2	1.82	1.67	1.47	1.58	1.65	1.49	1.24	1.60	1.56
3	0.28	0.27	0.23	0.23	0.28	0.26	0.17	0.26	0.24
4 and more	0.07	0.06	0.05	0.06	0.07	0.06	0.04	0.06	0.06
Number of trucks	43,037	55,388	57,795	59,347	61,917	63,749	62,552	52,392	456,177
Means truck crash	0.1696	0.1632	0.1489	0.1556	0.1596	0.1515	0.1327	0.1578	0.1541

Table 16 presents the variation of average annual accidents per truck relative to the number of driver's violations of the Highway Safety Code during the year preceding the accidents. Violations committed by drivers are very powerful in explaining truck accidents during the next year. Indeed, we observe that the year *t* accident rate is an increasing function of previous year violations committed by drivers.

Table 16
Average truck accidents according to the driver's violations committed the previous year.

Violations committed by the driver the previous year	Year								Total
	1991	1992	1993	1994	1995	1996	1997	1998	
For speeding									
0	0.1642	0.1586	0.1432	0.1486	0.1516	0.1435	0.1240	0.1498	0.1472
1	0.2974	0.2592	0.2640	0.2723	0.2631	0.2523	0.2161	0.2609	0.2556
2	0.2701	0.3410	0.3045	0.4000	0.3566	0.3249	0.3207	0.3281	0.3337
3 and more	0.4194	0.5000	0.2424	0.4651	0.5506	0.4821	0.3973	0.4600	0.4505
For driving with a suspended license									
0	0.1696	0.1629	0.1485	0.1547	0.1584	0.1507	0.1321	0.1574	0.1535
1 and more	0.7500	0.5217	0.3750	0.4076	0.3549	0.3426	0.3017	0.3265	0.3566
For running a red light									
0	0.1679	0.1617	0.1473	0.1538	0.1571	0.1491	0.1308	0.1555	0.1521
1	0.2726	0.2846	0.2764	0.2999	0.3350	0.3135	0.2981	0.3413	0.3036
2 and more	0.5294	0.6667	0.3846	0.2727	0.6000	0.7272	0.2308	0.2727	0.5040

Violations committed by the driver the previous year	Year								Total
	1991	1992	1993	1994	1995	1996	1997	1998	
For disobeying stop signs or police signals									
0	0.1677	0.1618	0.1474	0.1541	0.1572	0.1498	0.1315	0.1561	0.1524
1	0.3204	0.3140	0.2797	0.2823	0.3570	0.2931	0.2411	0.3100	0.2993
2 and more	0.5000	0.2857	0.2500	0.5833	0.2941	0.5263	0.3125	0.5000	0.4016
For failing to wear a seat belt									
0	0.1689	0.1626	0.1481	0.1554	0.1588	0.1508	0.1316	0.1576	0.1534
1	0.2304	0.2246	0.2293	0.1770	0.2376	0.2100	0.2096	0.2124	0.2164
2 and more	0.4138	0.4333	0.2571	0.2750	0.1774	0.2653	0.3137	0.1200	0.2741
For overweight									
0	0.1649	0.1583	0.1448	0.1517	0.1544	0.1461	0.1293	0.1540	0.1497
1	0.2430	0.2764	0.2410	0.2501	0.2432	0.2383	0.1889	0.2631	0.2394
2 and more	0.3387	0.2956	0.3364	0.2926	0.3552	0.3026	0.2155	0.3874	0.3065
For oversize									
0	0.1695	0.1632	0.1488	0.1554	0.1596	0.1515	0.1326	0.1577	0.1540
1 and more	0.2836	0.1000	0.2917	0.2603	0.1574	0.1545	0.2269	0.2821	0.2119
For poorly secured loads									
0	0.1688	0.1625	0.1482	0.1550	0.1587	0.1509	0.1323	0.1570	0.1534
1 and more	0.3185	0.3198	0.2667	0.2665	0.2656	0.2621	0.2214	0.3778	0.2791
For exceeding hours of service									
0	0.1696	0.1632	0.1486	0.1556	0.1592	0.1513	0.1325	0.1575	0.1539
1 and more	0.5714	0.3000	0.6333	0.1951	0.3529	0.2743	0.3881	0.3571	0.3496
For failure to undergo mechanical inspection									
0	0.1691	0.1626	0.1474	0.1546	0.1578	0.1509	0.1321	0.1572	0.1532
1 and more	0.2890	0.3180	0.2388	0.2534	0.3024	0.2251	0.2168	0.2768	0.2591

5.2 Estimation of the models

For comparison we first estimate the basic Poisson and Negative Binomial models. The results for the Poisson model are presented in column 2 in Table 17. We fit our initial Poisson specification model with the number of truck accidents per year as the dependent variable and 42 independent variables for the different characteristics of the trucks, drivers and fleets. Several variables measure observable heterogeneity. Some of these variables (type of fuel, number of cylinders, number of axles, type of vehicle used) are characteristics concerning vehicles, whereas others (sector, fleet size, etc.) have to do with the fleet. We also include the number of violations of the trucking standards the year before the accidents and the number of violations of the road safety code leading to demerit points the year before the accidents**.** The first group of violations is more related to fleet behavior, and the second group is more related to drivers' behavior**.** Almost all coefficients are significant at 1%.

In this table, we also present the corresponding estimate of the NB2 model in column 3. The estimate of δ is equal to 0.8135 with the standard deviation of 0.0187. The implied variance to mean ratio $(1+\delta)/\delta$ is 2.23, which is greater than 1. Thus, the NB2 model specification allows for overdispersion in accident distribution. Otherwise, the coefficients of the observable characteristics are very stable between the two models.

These results do not control for truck-specific effect so the serial correlation of residuals may be a problem having panel data. To add specific truck effect to the NB2 model, we also consider a random effect specification. The result is presented in column 2 in Table 18. The third column presents the estimates of our Gamma/Dirichlet model when we add random firm-specific effect.

Table 17

Estimation of the parameters of the distribution of the number of annual truck accidents for the 1991-1998 period (fleet of two trucks or more and trucks with two periods or more) and the Poisson and NB2 models.

Explanatory variables	Poisson model		NB2model	
	Coefficient	Standard deviation	Coefficient	Standard deviation
Constant	-3.5846*	0.0415	-3.5895*	0.0438
Number of years as carrier at 31 December	-0.0424*	0.0026	-0.0432*	0.0028
Sector of activity in 1998				
Other sector	-0.2766*	0.0804	-0.2694*	0.0840
General public trucking	0.0933*	0.0210	0.0977*	0.0227
Bulk public trucking	Reference group		Reference group	
Private trucking	0.1548*	0.0177	0.1595*	0.0191
Short-term rental firm	0.4055*	0.0275	0.4185*	0.0300
Size of fleet				
2	Reference group		Reference group	
3	0.1245*	0.0161	0.1246*	0.0171
4 to 5	0.1900*	0.0151	0.1926*	0.0160
6 to 9	0.2764*	0.0148	0.2797*	0.0158
10 to 20	0.3704*	0.0142	0.3761*	0.0152
21 to 50	0.3698*	0.0151	0.3782*	0.0161
More than 50	0.3837*	0.0142	0.3892*	0.0151
Number of days authorized to circulate in previous year	1.6703*	0.0290	1.6765*	0.0300
Number of violations of trucking standards in previous year				
For overload	0.1456*	0.0104	0.1502*	0.0119
For excessive size	0.1607	0.0825	0.1615	0.0916
For poorly secured cargo	0.2927*	0.0329	0.2991*	0.0383
For failure to respect service hours	0.2771*	0.0598	0.2880*	0.0718
For failure to pass mechanical inspection	0.2819*	0.0280	0.2977	0.0321
For other reasons	0.2812*	0.0699	0.2602*	0.0807
Type of vehicle use				
Commercial use including transport of goods without C.T.Q. permit	-0.1167*	0.0177	-0.1249*	0.0192
Transport of other than "bulk" goods	-0.0325	0.0203	-0.0387	0.0221
Transport of "bulk" goods	Reference group		Reference group	
Type of fuel				
Diesel	Reference group		Reference group	
Gas	-0.3922*	0.0124	-0.3939*	0.0130
Other	-0.3169*	0.0684	-0.3161*	0.0713
Number of cylinders				
1 to 5 cylinders	0.3536*	0.0360	0.3527*	0.0385
6 to 7 cylinders	0.3752*	0.0114	0.3763*	0.0120
8 or more than 10 cylinders	Reference group		Reference group	
Number of axles				
2 axles (3,000 to 4,000 kg)	-0.1603*	0.0177	-0.1616*	0.0188
2 axles (more than 4,000 kg)	-0.1505*	0.0122	-0.1541*	0.0132
3 axles	-0.1156*	0.0124	-0.1203*	0.0133
4 axles	-0.1818*	0.0163	-0.1817*	0.0175

Explanatory variables	Poisson model		NB2model	
	Coefficient	Standard deviation	Coefficient	Standard deviation
5 axles	-0.2040*	0.0145	-0.2056*	0.0156
6 axles or more	Reference group		Reference group	
Number of violations with demerit points year before				
For speeding	0.2961*	0.0092	0.3098*	0.0108
For driving with suspended license	0.4895*	0.0350	0.5590*	0.0465
For running a red light	0.4549*	0.0226	0.4723*	0.0259
For ignoring stop sign or traffic officer	0.4953*	0.0244	0.5107*	0.0279
For not wearing a seat belt	0.2295*	0.0281	0.2386*	0.0313
Observation period				
1991	0.0099	0.0222	0.0142	0.0240
1992	-0.0225	0.0202	-0.0195	0.0217
1993	-0.0881*	0.0189	-0.0876*	0.0203
1994	-0.0228	0.0174	-0.0218	0.0187
1995	-0.0012	0.0163	-0.0011	0.0175
1996	-0.0463*	0.0157	-0.0453*	0.0168
1997	-0.1605*	0.0158	-0.1597*	0.0168
1998	Reference group		Reference group	
$\hat{\delta}$			0.8135*	0.0187
Number of observations:	456, 117		456, 117	

* Significant at 1%.

Table 18

Estimation of the parameters of the distribution of the number of annual truck accidents for the 1991-1998 period (fleet of two trucks or more and trucks with two periods or more), and the Hausman's and Gamma/Dirichlet models.

Explanatory variables	Hausman's model		Gamma/Dirichlet model	
	Coefficient	Standard deviation	Coefficient	Standard deviation
Constant	-0.1034	0.0830	-3.9039*	0.0570
Number of years as carrier at 31 December	-0.0437*	0.0031	-0.0464*	0.0043
Sector of activity in 1998				
Other sector	-0.2574*	0.0932	-0.1378	0.1161
General public trucking	0.1002*	0.0252	0.1672*	0.0304
Bulk public trucking	Reference group		Reference group	
Private trucking	0.1571*	0.0213	0.2262*	0.0257
Short-term rental firm	0.4464*	0.0336	0.5591*	0.0484
Size of fleet				
2	Reference group		Reference group	
3	0.1255*	0.0180	0.0806*	0.0205
4 to 5	0.1937*	0.0172	0.1394*	0.0205
6 to 9	0.2795*	0.0171	0.2152*	0.0210
10 to 20	0.3613*	0.0166	0.2939*	0.0209
21 to 50	0.3573*	0.0177	0.3014*	0.0223
More than 50	0.3588*	0.0167	0.3087*	0.0217
Number of days authorized to drive in previous year	1.6859*	0.0299	2.0517*	0.0299
Number of violations of trucking standards in year before				
For overload	0.1217*	0.0117	0.0978*	0.0114

Explanatory variables	Hausman's model		Gamma/Dirichlet model	
	Coefficient	Standard deviation	Coefficient	Standard deviation
For excessive size	0.1418	0.0884	0.1566	0.0861
For poorly secured cargo	0.2520*	0.0363	0.2005*	0.0354
For failure to respect service hours	0.2502*	0.0666	0.1969*	0.0659
For failure to pass mechanical inspection	0.2390*	0.0308	0.1770*	0.0297
For other reasons	0.2628*	0.0780	0.1658	0.0741
Type of vehicle use				
Commercial use including transport of goods without C.T.Q. permit	-0.1410*	0.0213	-0.1926*	0.0213
Transport of other than "bulk" goods	-0.0515	0.0244	-0.1132*	0.0243
Transport of "bulk" goods	Reference group		Reference group	
Type of fuel				
Diesel	Reference group		Reference group	
Gas	-0.4090*	0.0145	-0.3959*	0.0136
Other	-0.3241*	0.0778	-0.3107*	0.0734
Number of cylinders				
1 to 5 cylinders	0.3587*	0.0440	0.2186*	0.0401
6 to 7 cylinders	0.3777*	0.0136	0.3761*	0.0126
8 or more than 10 cylinders	Reference group		Reference group	
Number of axles				
2 axles (3,000 to 4,000 kg)	-0.1614*	0.0210	-0.2885*	0.0208
2 axles (more than 4,000 kg)	-0.1713*	0.0150	-0.2810*	0.0150
3 axles	-0.1557*	0.0151	-0.1260*	0.0149
4 axles	-0.1897*	0.0199	-0.1331*	0.0191
5 axles	-0.2184*	0.0173	-0.1934*	0.0174
6 axles or more	Reference group		Reference group	
Number of violations with demerit points year before				
For speeding	0.2582*	0.0105	0.1920*	0.0102
For driving with suspended license	0.4489*	0.0426	0.3834*	0.0419
For running a red light	0.3838*	0.0247	0.3054*	0.0238
For ignoring stop sign or traffic officer	0.4268*	0.0267	0.3514*	0.0256
For not wearing a seat belt	0.2048*	0.0304	0.1517*	0.0294
Observation period				
1991	0.0176	0.0251	0.0732	0.0330
1992	-0.0196	0.0225	0.0523	0.0292
1993	-0.0849*	0.0208	0.0789*	0.0259
1994	-0.0213	0.0190	0.1825*	0.0226
1995	0.0004	0.0175	0.2057*	0.0196
1996	-0.0434*	0.0165	0.1185*	0.0175
1997	-0.1588*	0.0163	-0.0798*	0.0163
1998	Reference group		Reference group	
$\hat{a}$	57.8437*	3.5792		
$\hat{b}$	1.8247*	0.0384		
$\hat{\nu}$			2.0619*	0.0439
$\hat{\kappa}$			12.6466*	0.2503
$\hat{\delta}$			4.6684*	0.3260
Number of observations:	456, 117		456, 117	

* Significant at 1%.

The random effect parameters are significant in both models. Given that the (a,b) parameters of the Hausman model are very difficult to interpret (Greene, 2005), let us concentrate on the Gamma/Dirichlet model proposed in this article. The significance of the three random effect parameters means that the random effect associated with the fleets (or the non-observable risk of the fleets) ($\hat{\kappa}$), as well the random effect of the trucks or the drivers ($\hat{\nu}$) and the random time effect ($\hat{\delta}$) significantly affect the truck distribution of accidents even when we control for many observable characteristics.

Suppose we are proposing a parametric model to rate insurance for vehicles belonging to a fleet. According to the results in Table 18, this premium will be a function of observable characteristics of the vehicle and fleet of the vehicle, as well a function of violations of the road-safety code committed by drivers and carriers. This will not be enough because many unobservable characteristics of trucks, drivers and carriers also affect the trucks' distribution of accidents. The premiums will also have to be adjusted using the parameters of the random effects so as to account for the impact that the unobservable characteristics or actions of carrier, truck and drivers and even time can have on the truck accident rate. This form of rating makes it possible to visualize the impact (observable and non-observable) of behaviors of owners and drivers on the predicted rate of accidents, and consequently on premiums.

5.3 Fit statistics of different models according to fleet size

We now analyze the performance of the models. Model fits are based on the log likelihood statistics as well on other measures of information criteria such as the Akaike's information criterion (AIC) and the Bayesian information criterion (BIC). One advantage of using these information criterion measures is that they can compare non-nested models. The Bayesian Information Criterion (BIC) is equal to

$$\mathrm{BIC} = -2\ln \mathrm{L} + \mathrm{k}\ln(\mathrm{N})$$

and the Akaike's information criterion (AIC) is equal to

$$\mathrm{AIC} = -2\ln \mathrm{L} + 2\mathrm{k}$$

where k and N are the number of parameters and observations respectively. The BIC and AIC measures indicate a better fit when they are smaller. For two models estimated from the same data set, the model with the smaller BIC and AIC is preferable. We note in Table 19 that the Gamma/Dirichlet model is preferred to the Hausman model as the fleet size increases. In fact, the Hausman model performs better only when we add small fleets with two trucks. This result seems to indicate that the fleet or the enterprise effect is less important for very small fleets. Not surprisingly, both models in Table 18 dominate the two models in Table 17 whatever the fleet size.

Table 19
Fit Statistics of four different models with four different data sets.

Statistics	Poisson	Negative binomial	Hausman model	Gamma/Dirichlet model
17,542 fleets having more than 1 truck				
Log L	-201,123.82	-199,518.13	-197,165.16	-197,210.93
BIC	402,794.93	399,596.58	394,903.67	395,008.24
AIC	402,331.65	399,122.27	394,418.32	394,511.86
Number of trucks	111,106	111,106	111,106	111,106
Number of observations	456,177	456,177	456,177	456,177
Number of parameters	42	43	44	45
10,654 fleets having more than 2 trucks				
Log L	-186,034.41	-184,522.31	-182,285.05	-182,019.42
BIC	372,598.65	369,587.63	365,125.78	364,607.45
AIC	372,150.82	369,128.63	364,656.10	364,126.84
Number of trucks	97,330	97,330	97,330	97,330
Number of observations	409,561	409,561	409,561	409,561
Number of parameters	41	42	43	44
7,451 fleets having more than 3 trucks				
Log L	-171,997.73	-170,571.72	-168,456.10	-167,939.79
BIC	344,508.35	341,669.16	337,450.74	336,430.95
AIC	344,075.47	341,225.45	336,996.20	335,965.58
Number of trucks	87,721	87,721	87,721	87,721
Number of observations	370,442	370,442	370,442	370,442
Number of parameters	40	41	42	43
5,423 fleets having more than 4 trucks				
Log L	-158,974.84	-157,640.19	-155,634.41	-154,880.42
BIC	318,458.77	315,802.20	311,803.36	310,308.11
AIC	318,029.70	315,362.39	311,352.82	309,844.84
Number of trucks	79,609	79,609	79,609	79,609
Number of observations	336,772	336,772	336,772	336,772
Number of parameters	40	41	42	43

Bayesian Information Criterion $(BIC) = -2\ln L + k\ln(N)$; Akaikes Information Criterion $(AIC) = -2\ln L + 2k$ where k and N are the number of parameters and observations respectively, LogL = Log Likelihood ratio.

Because the large majority of trucks belong to fleets that have more than two trucks, it is clear that our model permits better estimation of accident distributions than the Hausman model does. Detailed estimation results of the models with fleets having more than four trucks are presented in Appendix B. We now analyze the consistency of the random effects estimators presented in Table 18 for the Gamma-Dirichlet model.

6 CONCLUSION

In this article, we propose a new parametric model with random effects for the estimation of accident distribution in the presence of individual and firm effects. This type of model is necessary to compute insurance premiums for drivers or vehicles belonging to a fleet because the characteristics and the management behavior of the fleets can affect the accident rate of vehicles and their drivers. For example, the manager of a given fleet may have a high risk appetite and ask their drivers to drive faster or to work more than the regulated number of hours during a week. He may also ask them to transport poorly secured cargo. A pricing rule that includes the observable and non-observable characteristics of all parties that affect accident distributions should consequently be fairer, and introduce the appropriate incentives of all parties under asymmetric information.

The methodology developed in this study can be applied to estimating event distributions in many other domains than insurance pricing. Since 2004, banks are regulated by Basel II for keeping capital for operational risk. The operational risk of different banks is a function of the observable characteristics and the non-observable behavior of the personnel and of management. A similar environment is present for the default risk of different firms or for the accident risk of any public institution or transportation firm including airline accidents. Other domains of applications include the failure or success rate of hospitals, universities, or any institution with principal-agent situations with teams.

APPENDIX

APPENDIX A: CHOICE AND DESCRIPTION OF VARIABLES

The unit of observation is an eligible vehicle with authorization to circulate at least one day in year *t*, and which has been followed up for at least two years. We analyze the accident totals found in SAAQ files. These totals include all the traffic accidents causing bodily injuries and all accidents causing material damage reported by the police in Quebec.

Dependent variable

Y_{fit} = the number of accidents in which vehicle *i* of fleet *f* has been involved during year *t*. Y_{fit} can take the values 0, 1, 2, 3, 4 and over.

Explanatory variables

We have two types of explanatory variables: those concerning the carrier and those concerning vehicles and drivers.

Variables concerning the carrier

- *Size of fleet for year t:* 7 dichotomous variables have been created.
  The two-vehicle size is used as the reference category. Coefficients estimated as positive and significant will thus indicate that vehicles are more at risk of accidents than those in the two-vehicle category.

- *Sector of economic activity*: 5 dichotomous variables have been created for vehicles transporting goods :

  sect_14 = 1 if the main sector of activity is transporting passengers;
  sect_05 = 1 if the sector of activity is general public trucking;
  sect_06 = 1 if the sector of activity is public bulk trucking;
  sect_07 = 1 if the sector of activity is independent trucking;
  sect_08 = 1 if the sector of activity is a short-term leasing firm.

  The "public bulk trucking" sector is used as the reference category. Coefficients estimated as negative and significant for the carrier's other sectors of economic activity will thus indicate that the vehicles of these sectors run lower risks than those in the reference group (and inversely for positive and significant coefficients).

- Seven (7) variables have been created for vehicles engaged in the *transportation of goods,* to measure the number of convictions per vehicle in the year preceding year *t* for each carrier:

  - *Number of violations per vehicle for overweight committed by a carrier in the year preceding year t.* A positive sign is predicted, because more overweight violations should, on average, generate more accidents.

- *Number of violations per vehicle for oversize committed by a carrier in the year preceding year t:* A positive sign is predicted, because more violations for oversize should, on average, generate more accidents.
- *Number of violations per vehicle for poorly secured loads committed by a carrier in the year preceding year t:* A positive sign is predicted, because more violations for poorly secured loads should, on average, generate more accidents.
- *Number of violations per vehicle of Highway Safety Code provisions regarding transportation of hazardous materials committed by a carrier in the year preceding year t:* A positive sign is predicted, because more violations of regulations for the transportation of hazardous materials should, on average, generate more accidents.
- *Number of violations per vehicle of hours-of-service regulations committed by a carrier in the year preceding year t:* A positive sign is predicted because more violations of hours-of-service regulations should, on average, generate more accidents.
- *Number of violations per vehicle of Highway Safety Code provisions regarding mechanical inspection committed by a carrier in the year preceding year t:* A positive sign is predicted, because more violations of regulations regarding mechanical inspection should, on average, generate more accidents.
- *Number of violations per vehicle, other than those already mentioned, committed by a carrier in the year preceding year t:* A positive sign is predicted, because more violations other than those already mentioned should, on average, generate more accidents.

Variables concerning vehicles and drivers (a vehicle may have more than one driver)

➢ *Vehicle's number of cylinders***:** 4 dichotomous variables have been created:

cyl_0 = 1 if the vehicle's number of cylinders is not known;
cyl1_5 = 1 if the vehicle has 1 to 5 cylinders;
cyl6_7 = 1 if the vehicle has 6 to 7 cylinders;
cyl_8p = 1 if the vehicle has 8 or more than 10 cylinders.

The group of vehicles with 8 or more than 10 cylinders is used as the reference category. Coefficients estimated as negative and significant for the other groups of vehicle/number of cylinders indicate that these groups run lower risks than those in the reference group.

➢ *Vehicle's type of fuel:* 3 dichotomous variables have been created:

diesel = 1 if the vehicle uses diesel as fuel;
fuel = 1 if the vehicle uses gas as fuel;
other = 1 if the vehicle uses another type of fuel.

The group of vehicles using diesel as fuel is considered the reference category. Coefficients estimated as negative and significant for the other groups of vehicle/fuel will thus indicate that these groups of vehicles run lower risks than those in the reference group.

➢ *maximum number of axles:* 7 dichotomous variables have been created:

ess_0 =	1	if the maximum number of axles does not apply to this type of vehicle;
ess_2 =	1	if the vehicle has two axles and a mass of between 3,000 and 4,000 kg;
ess_2p =	1	if the vehicle has two axles and a mass higher than 4,000 kg;
ess_3 =	1	if the vehicle is supported by a maximum of three axles;
ess_4 =	1	if the vehicle is supported by a maximum of four axles;
ess_5 =	1	if the vehicle is supported by a maximum of five axles;
ess_6p =	1	if the vehicle is supported by six or more axles.

The group of vehicles with two axles and a mass of between 3 000 and 4 000 kg is used as the reference category. Coefficients estimated as positive and significant for the other groups of vehicles/maximum-axle support indicate that these groups of vehicles run lower risks that those in the reference group.

- *Vehicle's type of use:* 3 dichotomous variables for vehicles transporting goods have been created:

compr =	1	if the vehicle is meant for commercial use, including transportation of goods without a CTQ permit;
tbrgn =	1	if the vehicle is meant for transportation of goods but other than in bulk, which requires a CTQ permit;
tbrvr =	1	if the vehicle is meant for transportation of bulk goods.

The group of vehicles transporting bulk goods is used as the reference category. Coefficients estimated as negative and significant for other groups of vehicles/types-of-use indicate that these groups of vehicles run lower risks than the reference group.

- Six (6) variables have been created to measure the number of convictions per vehicle accumulated in the year preceding year *t* by one or more drivers:
    - *Number of violations for speeding per vehicle, committed in the year preceding year t.* A positive sign is predicted because more speeding violations should, on average, generate more accidents.
    - *Number of violations for driving with a suspended license per vehicle, committed in the year preceding year t.* A positive sign is predicted because more driving with a suspended license should, on average, generate more accidents.
    - *Number of violations for running a red light per vehicle, committed in the year preceding year t.* A positive sign is predicted because more incidences of running a red light should, on average, generate more accidents.
    - *Number of violations for failure to obey a stop sign or a signal from a traffic officer per vehicle, committed in the year preceding year t.* A positive sign is predicted because more incidents of failure to respect a stop sign or a signal from a traffic cop should, on average, generate more accidents.
    - *Number of violations for failure to wear a seat belt per vehicle, committed in the year preceding year t.* A positive sign is predicted because more incidents of failure to wear a seat belt should, on average, generate more accidents.

- *Number of violations other than those mentioned per vehicle, committed in the year preceding year t.* A positive sign is predicted because a greater number of violations other than those mentioned should, on average, generate more accidents.

APPENDIX B: ESTIMATION RESULTS FOR FLEETS OF MORE THAN FOUR TRUCKS

Table B20: Estimation of the parameters of the distribution of the number of annual truck accidents for the 1991-1998 period (fleet of more than four trucks and trucks with two periods or more): Poisson and NB2 models

Explanatory variables	Poisson model		NB2model	
	Coefficient	**Standard deviation**	**Coefficient**	**Standard deviation**
Constant	-3.5145*	0.0495	-3.5211*	0.0524
Number of years as carrier at 31 December	-0.0372*	0.0032	-0.0377*	0.0034
Sector of activity in 1998				
Other sector	-0.3248*	0.0923	-0.3180*	0.0964
General public trucking	0.0913*	0.0242	0.0964*	0.0262
Bulk public trucking	Reference group		Reference group	
Private trucking	0.1714*	0.0214	0.1776*	0.0232
Short-term rental firm	0.4264*	0.0301	0.4416*	0.0329
Size of fleet				
5	Reference group		Reference group	
6 to 9	0.0654*	0.0140	0.0661*	0.0150
10 to 20	0.1622*	0.0132	0.1649*	0.0142
21 to 50	0.1596*	0.0142	0.1644*	0.0153
More than 50	0.1705*	0.0133	0.1720*	0.0142
Number of days authorized to circulate year before	1.7167*	0.0328	1.7231*	0.0339
Number of violations of trucking standards year before				
For overload	0.1375*	0.0119	0.1413*	0.0135
For excessive size	0.1725	0.0964	0.1786	0.1071
For poorly secured cargo	0.2669*	0.0374	0.2720*	0.0433
For failure to respect service hours	0.2507*	0.0668	0.2557*	0.0785
For failure to pass mechanical inspection	0.2330*	0.0327	0.2449*	0.0374
For other reasons	0.3083*	0.0758	0.2846*	0.0885
Type of vehicle use				
Commercial use including transport of goods without C.T.Q. permit	-0.0748*	0.0210	-0.0813*	0.0229
Transport of other than "bulk" goods	-0.0065	0.0232	-0.0118	0.0253
Transport of "bulk" goods	Reference group		Reference group	
Type of fuel				
Diesel	Reference group		Reference group	
Gas	-0.3387*	0.0140	-0.3400*	0.0148
Others	-0.2869*	0.0735	-0.2859*	0.0769
Number of cylinders				
1 to 5 cylinders	0.3369*	0.0424	0.3352*	0.0454
6 to 7 cylinders	0.3725*	0.0130	0.3732*	0.0137
8 or more than 10 cylinders	Reference group		Reference group	
Number of axles				
2 axles (3,000 to 4,000 kg)	-0.1840*	0.0202	-0.1859*	0.0215
2 axles (more than 4,000 kg)	-0.1308*	0.0134	-0.1344*	0.0145
3 axles	-0.0678*	0.0137	-0.0723*	0.0148
4 axles	-0.1951*	0.0178	-0.1951*	0.0191
5 axles	-0.1850*	0.0159	-0.1864*	0.0171

Explanatory variables	Poisson model		NB2model	
	Coefficient	**Standard deviation**	**Coefficient**	**Standard deviation**
6 axles or more	Reference group		Reference group	
Number of violations with demerit points year before				
For speeding	0.2819*	0.0105	0.2930*	0.0122
For driving under suspension	0.5355*	0.0461	0.5713*	0.0558
For running a red light	0.4070*	0.0262	0.4200*	0.0299
For ignoring stop sign or traffic agent	0.4735*	0.0280	0.4843*	0.0321
For not wearing a seat belt	0.1910*	0.0331	0.1969*	0.0367
Observation period				
1991	0.0109	0.0268	0.0146	0.0290
1992	-0.0221	0.0242	-0.0188	0.0262
1993	-0.0817*	0.0224	-0.0811*	0.0241
1994	-0.0147	0.0204	-0.0129	0.0220
1995	0.0044	0.0188	0.0050	0.0202
1996	-0.0373**	0.0177	-0.0355	0.0191
1997	-0.1443*	0.0176	-0.1438*	0.0189
1998	Reference group		Reference group	
$\hat{\delta}$			0.8032*	0.0203
Number of observations:	336,772		336,772	

* Significant at 1%. ** Significant at 5%.

Table B21: Estimation of the parameters of the distribution of the number of annual truck accidents for the 1991-1998 period (fleet of more than four trucks and trucks with two periods or more): Hausman's model and Gamma-Dirichlet model

Explanatory variables	Hausman's model		Gamma-Dirichlet model	
	Coefficient	**Standard deviation**	**Coefficient**	**Standard deviation**
Constant	-0,0155	0,0977	-3.8291*	0.0825
Number of years as carrier at 31 December	-0,0385*	0,0038	-0.0402*	0.0067
Sector of activity in 1998				
Other sector	-0,3136*	0,1072	-0.1683	0.1556
General public trucking	0,0984*	0,0293	0.1446*	0.0402
Bulk public trucking	Reference group		Reference group	
Private trucking	0,1756*	0,0261	0.2460*	0.0625
Short-term rental firm	0,4713*	0,0369	0.5943*	0.0625
Size of fleet				
5	Reference group		Reference group	
6 to 9	0,0646*	0,0161	0.0008	0.0918
10 to 20	0,1465*	0,0158	0.0527**	0.0219
21 to 50	0,1396*	0,0170	0.0485**	0.0243
More than 50	0,1372*	0,0160	0.0517**	0.0245
Number of days authorized to circulate in year before	1,7278*	0,0338	2.1333*	0.0338
Number of violations of trucking standards in year before				
For overload	0,1099*	0,0133	0.0846*	0.0219
For excessive size	0,1533	0,1032	0.1637	0.0986
For poorly secured cargo	0,2285*	0,0410	0.1772*	0.0396
For failure to respect service hours	0,2265*	0,0735	0.1666**	0.0720
For failure to pass mechanical inspection	0,1813*	0,0358	0.1107*	0.0343
For other reasons	0,2944*	0,0855	0.1744**	0.0803
Type of vehicle use				
Commercial use including transport of goods without C.T.Q. permit	-0,1008*	0,0254	-0.1661*	0.0252
Transport of other than "bulk" goods	-0,0307	0,0280	-0.1011*	0.0279
Transport of "bulk" goods	Reference group		Reference group	
Type of fuel				
Diesel	Reference group		Reference group	
Gas	-0,3520*	0,0167	-0.3529*	0.0152
Others	-0,2912*	0,0843	-0.2664*	0.0787
Number of cylinders				
1 to 5 cylinders	0,3360*	0,0526	0.1431*	0.0460
6 to 7 cylinders	0,3725*	0,0156	0.3658*	0.0140
8 or more than 10 cylinders	Reference group		Reference group	
Number of axles				
2 axles (3,000 to 4,000 kg)	-0,1853*	0,0242	-0.3534*	0.0237
2 axles (more than 4,000 kg)	-0,1503*	0,0166	-0.3076*	0.0167
3 axles	-0,1087*	0,0170	-0.0937*	0.0167
4 axles	-0,201*2	0,0218	-0.1180*	0.0208
5 axles	-0,1967*	0,0190	-0.1738*	0.0193
6 axles or more	Reference group		Reference group	
Number of violations with demerit points year before				
For speeding	0,2433*	0,0118	0.1681*	0.0114
For driving under suspension	0,4711*	0,0518	0.3874*	0.0490

Explanatory variables	Hausman's model		Gamma-Dirichlet model	
	Coefficient	**Standard deviation**	**Coefficient**	**Standard deviation**
For running a red light	0,3388*	0,0286	0.2688*	0.0271
For ignoring stop sign or traffic agent	0,4040*	0,0306	0.3266*	0.0290
For not wearing a seat belt	0,1662*	0,0356	0.1159*	0.0340
Observation period				
1991	0,0204	0,0309	0.0972**	0.0492
1992	-0,0176	0,0276	0.0800	0.0429
1993	-0,0783*	0,0251	0.1227*	0.0371
1994	-0,0129	0,0225	0.2290*	0.0313
1995	0,0077	0,0203	0.2475*	0.0260
1996	-0,0330	0,0188	0.1598*	0.0215
1997	-0,1416*	0,0182	-0.0473**	0.0187
1998	Reference group		Reference group	
$\hat{a}$	58,5410*	4,2211		
$\hat{b}$	1,8359*	0,0420		
$\hat{\nu}$			1.9656*	0.0433
$\hat{\kappa}$			21.9790*	0.5530
$\hat{\delta}$			4.7239*	0.3283
Number of observations:	336,772		336,772	

* Significant at 1%. ** Significant at 5%.

REFERENCES

Abowd J-M., Kramarz, F., Margolis, D.N., 1999. High wage workers and high wage firms. Econometrica 67, 251-333.

Allison, P.D., Waterman, R.P., 2002. Fixed-effects negative binomial regression models. Sociological Methodology 32, 247-265.

Angers J-F., Desjardins D., Dionne G., Guertin F., 2006. Vehicle and fleet random effects in a model of insurance rating or fleets of vehicles. Astin Bulletin 36, 25-77.

Baltagi B.H., 1995. Econometric Analysis of Panel Data Wiley, Chichester.

Boucher J-P, Denuit, M., Montserrat, G., 2008. Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. Casualty Actuary Society 2, 135-162.

Boyer, M., Dionne, G., Vanasse, C., 1992. Econometric Models of Accident Distributions. In Dionne, G. (Ed.), Contributions to Insurance Economics, Springer, 169-213.

Cameron, A.C., Trivedi, P.K., 2013. Count panel data. Mimeo, University of California.

Cameron, A.C., Trivedi, P.K., 1986. Econometric models based on count data: Comparisons and applications of some estimators and tests. Journal of Applied Econometrics 1, 29–54.

Dionne, G., Gagné, R., Gagnon, F., Vanasse, C., 1997. Debt, moral hazard and airline safety: An empirical evidence. Journal of Econometrics 79, 379-402.

Dionne G., Gagné, R., Vanasse, C., 1998. Inferring technological parameters from incomplete panel data. Journal of Econometrics 87, 303-327.

Dionne G., Vanasse, C., 1992. Automobile insurance ratemaking in the presence of asymmetrical information. Journal of Applied Econometrics 7, 149-165.

Dionne G., Vanasse, C., 1989. A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. Astin Bulletin 19, 199-212.

Frangos N., Vrontos, S.D., 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 vehicle plans? In Dionne, G., Laberge-Nadeau, C. (Eds.), Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation. Kluwer Academic Press, pp. 101-117.

Gouriéroux C., Monfort A., Trognon A., 1984. Pseudo maximum likelihood methods: Applications to Poisson models. Econometrica 52, 701-720.

Gouriéroux C., 1999. Statistiques de l’assurance, Economica.

Gradshteyn I.S., Ryzhik, I.M., 1980. Table of integrals, series and products. Academic Press Inc., New York.

Greene, W.H., 2005. Functional form and heterogeneity in models for count data. In W. Greene (Ed.), Foundations and Trends in Econometrics, vol. 1, no 2, pp. 115-218.

Greene, W.H., 2004. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. Econometrics Journal 7, 98-119.

Hausman J.A., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the patents – R&D relationship. Econometrica 52, 909-938.

Hausman J.A., Wise, D.A., 1979. Attrition bias in experimental and panel data: The Gary income maintenance experiment. Econometrica 47, 455-473.

Holmstrom, B., 1982. Moral hazard in teams. The Bell Journal of Economics 13, 324-340.

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.

Pinquet J., 2013. Experience rating in non-life insurance. In Dionne, G. (Ed.), Handbook of Insurance. Springer, New York, pp. 471-586.

Purcaru, O., Denuit, M., 2003. Dependence in dynamic claim frequency credibility models. *Astin Bulletin* 33, 1, 23-40.

Scheffé H., 1999. The Analysis of Variance. Wiley, New York.

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

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.