RESEARCH ARTICLE

Road safety for fleets of vehicles

Georges Dionne[1]*, Denise Desjardins[2], Jean Francois Angers[3]

[1] PhD, HEC Montreal,4.454-3000 Chemin de la Côte-Ste-Catherine, Montreal (Qc) Canada, H3T 2A7.
[2] M.Sc Canada Research Chair in Risk Management, 4.454-3000 Chemin de la Côte-Ste- Catherine, Montreal (Qc) Canada, H3T 2A7
[3] PhD, CIRRELT Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation,AA-5190, 2920, chemin de la Tour Montreal (QC) Canada, H3T 1J4

Abstract

Road safety for fleets of vehicles has been neglected in the insurance literature, mainly because appropriate data and methodology were not available. This article makes a threefold contribution: 1) Produce statistics on current fleets' road safety offences and accidents using a panel of 20 years of data on truck fleets; 2) relate fleets' offences to accidents; and 3) identify and classify the riskiest fleets for insurance ratemaking based on past experience in managing road safety. Our main technical innovation to the insurance literature is in the estimation of fleets' distributions of accidents. For each fleet size (or group of sizes), we estimate the parameters of the negative binomial (NB) distribution of the annual number of accidents according to the characteristics of the fleets, the years, and the number of driver (DRV) and carrier (CAR) road safety violations accumulated in the previous year. When the NB model does not accurately predict the mathematical expectation of the number of accidents of larger fleets, we proceed in two steps. First, we estimate the probability of having zero accidents in a year, and then estimate the negative binomial distribution using the estimated probability of having zero accidents, to weight the zeros of each fleet. To achieve our third objective, we construct risk classes for the vehicle fleets using the predicted accident probabilities obtained from the estimated models. Our results show a substantial heterogeneity between fleets in terms of road safety. This information should be very useful for optimal insurance pricing and better incentives for road safety.

Key Words: *Road safety; truck fleets; professional drivers; road safety infractions; road safety policy; zero-inflated model; negative binomial model.*

JEL codes: D81, G22.

**Corresponding Author*: *Georges Dionne, PhD, Professor of Finance,HEC Montreal, 4.454-3000 Chemin de la Côte-Ste-Catherine, Montreal (Qc) Canada, H3T 2A7. Tel: 514-340-6596. Email: georges.dionne@hec.ca.*

***Received Date**: September 15, 2021, **Accepted Date**: October 15, 2021, **Published Date**: October 29, 2021*

***Citation**: Dionne G, Desjardins D, Angers JF. Road safety for fleets of vehicles. Int J Bank Fin Ins Tech. 2021;1(1):31-59.*

Introduction

There are very few statistical analyses of the road accident risks of owners and operators of heavy vehicles (HVOs) in the insurance literature. Some authors have studied the risks of heavy vehicle drivers, without really assessing the aggregate risk of vehicle fleets [1-7]. These studies of HVOs drivers show that fleet owners can influence driver risk through their road safety management.

Since 1992, the *Société de l'assurance automobile du Québec* (Québec automobile insurance board, SAAQ) has been using violations of the Highway Safety Code to rate bodily injury insurance through driver's license payments. To justify this practice, meticulous analyses of the statistical links between the types of violations and accident rates have been carried out, using robust statistical instruments, to verify whether the number of demerit points accumulated or the number of violations of the Highway Safety Code are closely related to the number of accidents. Estimated coefficients of the effects of violations on crashes can also be used to test whether the demerit points awarded for each violation accurately reflect the relative crash risk of that offense, which is important for public road safety management [8].

The first two objectives of our research are to determine the most common violations committed by HVOs and their drivers, and to establish a statistical link between the types of violations and the numbers of road accidents in which HVOs are involved. The use of traditional counting models is not always appropriate for fleets of a certain size. The annual probability of having zero accidents is strongly affected by fleet size. We use two types of models to predict the mathematical accident expectation using estimated parameters. For larger fleets we use the Zero-inflated Negative Binomial (ZINB) model. This random effect model first estimates the probability of having zero accidents and then estimates the distribution of accidents with the Negative Binomial (NB) model by controlling for the excess zeros. We use the Negative Binomial directly for smaller fleets since we reject the Poisson model.

The third objective of the research is to identify the HVOs most at risk in order to improve insurance pricing based on the relative risks that fleets represent. Insurance pricing based on risk classes will provide an incentive for fleets to be more cautious. The statistical results will help regulators carry out targeted road safety monitoring activities, which can motivate the riskiest businesses to be more prudent. This information will also allow better identification of repeat offenders.

Section 1 proposes a short literature review, and the following section presents the methodology used. Section 3 describes the data obtained from the SAAQ. Section 4 discusses the results of the statistical estimates of accident distributions obtained for HVOs and proposes a calculation of the different risk classes derived from our statistical results in order to increase the incentives for road safety in the presence of information asymmetry among fleet owners, insurers and those responsible for monitoring road safety. It also identifies the riskier fleets. The conclusion discusses the implications of our results for the optimal road safety management of vehicle fleets and proposes an extension to our research.

1. Literature review

The literature review is divided in two sections related to our contribution. The first section reviews the main results related to the methods used to improve incentives for road safety. The second section presents the estimation methods of accidents distributions with panel data and in presence of non-observable heterogeneity. Zero-inflated count models for fleets are also discussed.

1.1. Road safety incentives

One major cause for the improvement of road safety across the world is the development of incentives for safe driving. Regulators have introduced several legal rules to improve road safety, and the insurance industry has improved insurance contracting to reduce asymmetric information with their clients and better monitor their behavior.

Incentive mechanisms for road safety have been investigated in the economic literature for many years [9-17]. Of the many mechanisms proposed, we observe fines for careless driving, point-record driver's licenses, partial insurance (deductible), no-fault accidents, and insurance experience rating. In the latter case, the individual driving history is usually summarized by past accidents or by point-records based on traffic offenses. In this study we will concentrate on past accidents although we control for past driving offences. Experience rating pricing used by the insurance industry has incentive properties [18,19]. By adjusting the premiums individuals pay for their insurance protection according to their past driving behavior, insurers improve the benefits of road safety. In an insurance environment that uses past accidents to set future premiums, those who accumulate accidents have more incentives to drive carefully to reduce their premium. This suggests an empirical test for asymmetric information often referred to as the conditional correlation test.

A number of empirical tests have been proposed to measure the presence of residual asymmetric information problems in insurers' portfolios [18,19] or to measure the efficiency of such mechanisms for road safety [20,21]. These tests were extended by Dionne et al [22] to separate moral hazard from adverse selection, the two main information problems in insurance contracting. Such separation permits to apply better focused incentive schemes for road safety.

1.2. Count data models

Our research is based on trucks accidents. Most of the econometric models applied to count variables that takes nonnegative values start from the Poisson distribution, where the probability of a truck of a given fleet being involved in different accidents (or claims) in a given period is estimated. By definition of the Poisson distribution, the mathematical expectation of the number of accidents is equal to the variance. The exponential form of the distribution introduces a nonlinear relationship between accidents and observed control variables. The regression component can contain continuous variables and such variables can be non-linear. Moreover, it can include categorical variables with a fixed number of possible values such as the size of a fleet or the number of traffic violations obtained by the drivers and the fleet owners. These variables can also introduce non-linear effects [20,21].

The Poisson model is an equidispersion model, meaning that the distribution of accidents can be explained entirely by observable heterogeneity. To take into account of the overdispersion property in the data, we can suppose that the mean parameter has a random term with expected value equal to zero and a positive variance [23-25]. This modelling allows for overdispersion and it considers unobserved heterogeneity that is absent in the Poisson model. The more popular model in this family is the Negative Binomial model (NB).

Unobserved heterogeneity is very important for pricing insurance premiums under asymmetric information [26,27]. Let us now consider panel data that contain observations where the same unit (individual, truck) is observed over several successive periods. There are two possibilities for panel data model estimations in the literature, the fixed effects and the random effects model. In this contribution 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 et al

[23] propose an extension of the Poisson model to panel data. The new model is a hierarchical model that comes directly from the Poisson model. Accidents are distributed according to the NB model with two additional parameters that vary across individuals. One such parameter is the random individual specific effect and the second one is an additional random effect that permit the random individual specific effect to vary over time. Suppose that these parameters follow a beta distribution, we can obtain a closed form solution for the random effects NB model. This model, known as the NB2 model can also be estimated with individual dummies (or other methods) in a fixed effects version.

The estimated parameters can be inconsistent however because of the incidental parameters problem, but the above contributions have shown that the inconsistency may be not important. Estimating the NB2 model can also yield inconsistent random effects estimators because the individual effect term and the vector of observable individual characteristics may be correlated. We can apply the Hausman test statistic to determine if we reject the null hypothesis that the individual effects are not correlated with the variables in the regression component. The NB2 model 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. Angers et al [4,5] show how to introduce the fleet effect in such model.

An additional problem with the NB model lies in applications to data containing large fleets where the annual probability of having zero accident is very low. In such case the zero inflated NB model [28,29] is more appropriate to obtain a better prediction for the zero-accident probability.

2. Methods

To carry out our research, we use several methodologies. We analyze the relative risks between HVOs by fleet size, in order to reduce the effects of unobservable heterogeneity between fleet sizes that we cannot control, such as management of heavy vehicle drivers by business owners or fleets' risk exposure. Further, because we are interested in the distributions of accident numbers, the annual probability of having zero accidents is highly influenced by fleet size. We use two types of models, depending on fleet size, to predict the mathematical expectation of accidents using the estimated parameters. We group fleet sizes together when the number of observations for each size is insufficient to estimate the selected models.

For each fleet size (or group of sizes), we estimate the parameters of the negative binomial (NB) distribution of the annual number of accidents according to the characteristics of the fleets, the years, and the number of driver (DRV) and carrier (CAR) violations accumulated in the previous year. When the NB model does not accurately predict the mathematical expectation of the number of accidents of larger fleets, we proceed in two steps. First, we estimate the probability of having zero accidents in a year, and then estimate the negative binomial distribution using the estimated probability of having zero accidents, to weight the zeros of each fleet. To achieve our third objective, we construct risk classes for the vehicle fleets using the predicted accident probabilities obtained from the estimated models.

Let $Y_{i;t}$ represent the total number of accidents for truck fleet i in year t. To take into account the number of observations equal to 0, we use a zero-inflated model [28,29]. Thus, the probability of observing $y_{i;t}$ accidents is given by:

$$P(Y_{i,t}=y_{i,t})=\begin{cases}\pi_{i,t}+(1-\pi_{i,t})f(0), y_{i,t}=0\\(1-\pi_{i,t})f(y_{i,t}), y_{i,t}=1,2,\ldots\end{cases} \quad (1)$$

where $0\leq\pi_{i,t}\leq 1$. The function $f(y_{i,t})$ represents a probability function on the integers {0; 1;...} [30].

To consider the over-dispersion (variance greater than the mean) of the number of accidents *k,* we opted for the NB2 probability function [23] of parameters λ and *a.*,*i.e.*:

$$f(k)=\begin{cases}\frac{\Gamma(\frac{1}{\alpha}+k)}{\Gamma(\frac{1}{\alpha})\Gamma(k+1)}\left(\frac{1}{1+\alpha\lambda}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\lambda}{1+\alpha\lambda}\right)^{k}, k=0;1;\ldots\\ 0, if\ not.\end{cases} \quad (2)$$

With this model, the first two moments (the mathematical expectation and the variance) of $Y_{i;t}$ are given by:

$$E[Y_{i,t}]=(1-\pi_{i,t})\lambda$$

$$V[Y_{i,t}]=(1-\pi_{i,t})\lambda(1+\alpha\lambda)+\pi_{i,t}(1-\pi_{i,t})\lambda^{2} \quad (3)$$

where α is the over-dispersion parameter. When $\pi_{i,t}$ and *a* are equal to zero, the model corresponds to the Poisson model with a parameter λ.

To link the different explanatory variables, we use a generalized linear model with an exponential function for the parameter λ [24] and with a Logit function for $\pi_{i,t}$. Thus, we have:

$$\lambda(\vec{X})=\exp(\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_pX_p)$$

$$\pi_{i,t}(\vec{Z})=\frac{\exp(\gamma_0+\gamma_1Z_1+\gamma_2Z_2+\ldots+\gamma_pZ_p)}{1+\exp(\gamma_0+\gamma_1Z_1+\gamma_2Z_2+\ldots+\gamma_pZ_p)} \quad (4)$$

The X_j and Z_j represent the different explanatory variables. The parameters (*a*; $\beta_0;\beta_1;\beta_2;\ldots;\beta_p;\ \gamma_0;\gamma_1;\gamma_2;\ldots;\gamma_p$) are estimated by applying the SAS COUNTREG procedure [31], which uses the maximum likelihood method. This model is known as Zero-inflated Negative Binomial (ZINB). It is a random effects model.[1]

If the model is well specified, maximum likelihood theory ensures that the random effects estimators of the NB model obtained with panel data are consistent and asymptotically efficient when they exist [25,33]. The specifications of the two equations in (4) may be different, but this is

1 To our knowledge, very few studies use fixed effects to estimate the ZINB. Majo and van Soest [32] propose a fixed effects model limited to two periods.

not necessary. We can assume that the error terms between the two equations are independent but, again, this is not necessary when the number of observations is very large, as in our study. Staub and Winkelman [33] show, using a Monte Carlo study, that even if the exclusion restrictions are not met, we can obtain unbiased estimators if the correlation is not too high. To reduce the potential correlations between the variables in the two equations, we performed a principal component analysis (PCA) to select the factors or variables to be used in the Logit estimation. For small fleets, the number of observations equal to 0 can be modelled directly very well by using a negative binomial model (see equation 2). To isolate this model, one only has to fix $\pi_{i,t} \equiv 0$ in equations 1 and 3.

3. Data

The data, obtained primarily from the SAAQ, represent the HVO population in Québec over the period 1991-2010, *i.e.* 20 years. In this article, we limit the analysis to heavy trucks and tractors. The database we compiled for this research is the most exhaustive in this area. We were able to link carriers to their drivers. The information received was encrypted, which protected the identity of the vehicle fleets and drivers. This database makes it possible to track the behavior of owners and operators of heavy vehicles over time.

The starting point for building this database was all registered carriers as of December 31, 2010. Vehicle data and vehicle mechanical inspection record data were extracted from the previously selected personal identification numbers (PINs). From these PINs, VINs (Vehicle Identification Numbers) and license plates, we extracted accidents, carrier (CAR) and driver (DRV) violations, and penalties linked to those carriers, vehicles, or drivers. The HVO register was also used. The evaluation process covered several areas, including results of on-site inspections and events (accidents, violations and decommissioning of a heavy vehicle or driver).

4. Results

4.1. Descriptive statistics of variables used in the analysis of HVO accidents

Our data cover the entire population of heavy truck and tractor fleets (SAAQ code BCA) in Québec over 20 years, namely from 1991 to 2010. Detailed data are presented in [34]. (Table 1) shows that the number of HVOs (owners and operators of heavy vehicles) remained fairly stable over the study period: It increased slightly until 1998 and returned to its 1991 level in 2010. The number of small HVOs decreased, while those of other sizes increased. The total number of heavy trucks as of December 31 of each year increased from 91,164 in 1991 to 122,423 in 2010 (see column 1 of Table 2).

There were no major changes in economic activity trends, excluding a decline in agriculture and an increase in construction and other services (Table A1). We note a steady decline in the percentage of new HVOs over time for the years 1991 to 1999, with a slight increase in 1999, followed by a decrease in 2000. For the years 2001 to 2010, the percentage of HVOs that began operations during the year ranged from 6.41 to 7.56. In contrast, mergers have increased, which seems to reflect industry consolidation during the analysis period. Details can be found in [34].

Table 2 gives the numbers and averages of total accidents and accidents with personal injury involving an HVO. The average number of total accidents decreased over the period, with a slight increase in 1999, followed by an almost continuous decrease after that date. The slight increase in 1999 can be explained by a change in the source of the data, because such an increase does not appear in the SAAQ reports for total heavy truck accidents. Our database was compiled in two phases: The first part, covering the period 1991-1998, was prepared in 1999, and the second part, covering the period 1999-2010, was prepared in the years that followed. This difference should not affect the following analyses of accident distributions with individual or non-aggregated data.

The average number of injury accidents involving a heavy truck has fluctuated slightly, although there was a significant increase in 1999 and a sharp decrease in 2010. We see the same stability for violations of the Highway Safety Code (DRV), if we exclude the periods of police strikes in 2005 and 2006. Several carrier violations (CARs) increased in 2000, peaking in 2002. Thereafter, there was a slight decrease in these offences over time, but the level remained higher than in the years prior to 2000**.**

Table 3 provides details on the evolution of driver violations (DRV) over the 20 years of the study. As previously demonstrated for accidents involving passenger vehicle drivers, violations for speeding and failure to obey red lights or stop signs are the most significant in explaining the accident rates of heavy vehicle drivers.

For all offences, with the exception of failure to obey stop signs, there were significant decreases in 2005 and 2006, explained by the police strike. Speeding violations continued to be the most prevalent. New violations added to the regulations in 2001 did not reach significant volumes, nor did violations related to cell phone use, added in 2008.

Regarding carrier violations (CAR), presented in (Table 4), many of the infractions (equipment, road signs, traffic rules, driving hours, hazardous materials) increased significantly after 1999, the year a reform on road safety management came into effect. More traditional offences such as overloading, oversizing, improper stowage, and mechanical inspections did not change significantly after the 1999 reform. We used variables of heavy truck characteristics by year as control variables in the different analyses of accident distributions.

Table 1: *Fleet size as of December 31 of the current year.*

This table presents the number of HVOs (owners and operators of heavy vehicles) over the period 1991 to 2010. They remained fairly stable over the study period.

Year	1	2	3	4-5	6-9	10-20	21-50	More than 50	Number of HVO's
1991	28,466	5,371	2,238	1,815	1,123	665	258	104	40,040
1992	28,602	5,445	2,319	1,836	1,102	665	246	99	40,314
1993	28,607	5,648	2,308	1,888	1,154	679	240	102	40,626
1994	29,453	5,699	2,348	2,035	1,175	711	253	110	41,784
1995	29,523	5,722	2,354	1,955	1,244	710	262	115	41,885
1996	29,555	5,735	2,352	2,019	1,215	738	262	111	41,987
1997	29,675	5,820	2,402	2,031	1,325	750	282	116	42,401
1998	29,504	5,819	2,461	2,118	1,376	790	295	130	42,493
1999	27,691	5,490	2,431	2,195	1,366	830	320	136	40,459
2000	26,727	5,471	2,424	2,130	1,482	849	326	148	39,557
2001	25,936	5,414	2,407	2,209	1,430	881	332	148	38,757
2002	25,581	5,362	2,415	2,135	1,505	857	330	152	38,337
2003	25,657	5,350	2,409	2,216	1,560	903	366	151	38,612

2004	25,870	5,432	2,404	2,301	1,576	945	379	166	39,073
2005	25,811	5,578	2,438	2,314	1,647	981	382	173	39,324
2006	26,008	5,583	2,527	2,303	1,630	998	391	175	39,615
2007	26,255	5,620	2,528	2,337	1,668	1,003	398	183	39,992
2008	25,586	5,557	2,580	2,367	1,649	1,046	399	187	39,371
2009	25,514	5,597	2,569	2,385	1,717	1,051	394	181	39,408
2010	25,716	5,834	2,622	2,482	1,770	1,124	436	186	40,170

Table 2: *Accidents, casualties, HVO driver traffic violations (DRV) and carrier violations (CAR).*
This table gives the numbers and averages of total accidents and accidents with personal injury involving an HVO.
* means that the violations reported in year *t* are those of year *t*−1 to explain the accidents as of date *t*.

Year	No.of Trucks	No.of accidents	Casualty rates	No.of DRV* violations	No.of CAR* violations	Average accidents	Average casualty rate	Av.DRV* violations	Av.CAR* violations
1991	91,164	12,958	1,465	7,956	6,281	0.142	0.016	0.087	0.07
1992	91,303	12,325	1,437	6,903	4,518	0.135	0.016	0.076	0.05
1993	92,229	13,166	1,589	7,715	6,453	0.143	0.017	0.084	0.07
1994	96,618	13,861	1,621	8,620	6,111	0.143	0.017	0.089	0.06
1995	97,108	13,506	1,458	10,819	8,563	0.139	0.015	0.111	0.09
1996	97,568	12,042	1,397	11,540	8,550	0.123	0.014	0.118	0.09
1997	102,532	13,451	1,709	12,587	8,992	0.131	0.017	0.123	0.09
1998	105,475	12,599	1,586	11,117	6,551	0.119	0.015	0.105	0.06
1999	104,346	13,707	1,866	11,213	5,303	0.131	0.018	0.107	0.05
2000	105,575	14,635	1,996	10,926	8,252	0.139	0.019	0.103	0.08
2001	105,403	13,474	1,863	8,673	14,086	0.128	0.018	0.082	0.13
2002	107,355	14,079	1,998	14,020	16,309	0.131	0.019	0.131	0.15
2003	110,525	14,398	2,045	12,445	15,710	0.13	0.019	0.113	0.14
2004	113,763	14,366	2,147	12,125	12,401	0.126	0.019	0.107	0.11
2005	116,465	14,466	2,227	12,400	12,983	0.124	0.019	0.106	0.11
2006	116,974	13,085	1,771	7,360	11,603	0.112	0.015	0.063	0.1
2007	118,773	14,030	1,838	8,401	11,319	0.118	0.015	0.071	0.1
2008	118,811	14,079	1,746	10,836	12,745	0.118	0.015	0.091	0.11
2009	118,436	11,646	1,487	10,896	14,169	0.098	0.013	0.092	0.12
2010	122,423	8,838	1,148	9,962	12,723	0.072	0.009	0.081	0.1

Table 3: *Driver offenses (DRV).*
This table provides details on the evolution of driver violations (DRV) over 1990-2009.

Year	Speed	Red light	Stop sign	Seat belt	Cell phone	Additions in 2001	Other DRV
1990	3,961	1,390	1,190	1,064	0	0	351
1991	3,878	1,059	931	739	0	0	296
1992	4,325	1,052	1,013	879	0	0	446
1993	5,006	1,193	1,049	882	0	0	490
1994	6,523	1,308	1,134	1,245	0	0	609
1995	7,083	1,346	1,204	1,282	0	0	625
1996	8,519	1,070	1,051	1,406	0	0	541
1997	7,567	1,213	1,090	639	0	0	608
1998	7,636	1,318	1,079	842	0	0	338
1999	7,210	1,367	1,154	639	0	0	556
2000	5,365	1,194	1,002	495	0	0	617
2001	9,386	1,255	1,132	1,204	0	284	759
2002	7,861	1,242	1,148	1,055	0	416	723
2003	8,334	983	1,062	762	0	498	486
2004	8,285	1,044	1,082	878	0	585	526
2005	4,212	884	1,005	598	0	212	449
2006	4,866	841	1,068	930	0	242	454
2007	6,727	951	1,093	1,174	0	365	526
2008	6,035	967	1,078	1,560	355	423	478
2009	5,058	828	975	1,492	822	364	423

Table 4: *Carrier offences (CAR).*
This table presents the carrier violations (CAR) over the period 1990-2009.

Year	Axle overload	Total overload	Dimension	Stowage	Hazardous materials	Driving hours	Mechanical inspection	Equipment	Road signs	Traffic rules	Pre-departure inspection	Other
1990	2,044	1,544	1,467	503	50	45	239	31	0	76	215	67
1991	1,831	1,301	496	342	42	12	252	23	0	7	131	81
1992	1,543	1,831	583	449	93	53	1,543	44	0	1	249	64
1993	1,793	1,821	625	459	129	76	810	34	0	2	315	47
1994	3,224	2,063	779	759	145	180	725	22	0	12	581	73
1995	3,283	2,610	805	532	227	195	396	14	0	18	431	39
1996	3,922	2,442	780	487	173	167	474	21	0	12	474	40
1997	3,306	1,475	471	318	58	153	345	6	0	6	359	54
1998	1,756	1,431	505	342	82	134	285	12	11	82	480	183
1999	1,144	1,927	328	509	47	485	483	437	989	669	980	1,254
2000	2,773	1,904	620	592	0	1,068	357	607	1,800	1,027	1,246	2,092
2001	2,583	1,745	778	980	0	1,175	376	651	2,552	1,492	1,289	2,688

2002	3,330	1,388	687	906	14	1,059	335	523	2,596	1,372	871	2,629
2003	2,563	1,075	499	738	246	837	290	395	1,864	1,223	588	2,083
2004	3,208	1,254	494	686	262	713	270	391	2,142	1,174	536	1,853
2005	2,828	1,382	464	560	307	653	276	399	1,549	933	500	1,752
2006	2,589	1,401	527	892	222	628	305	402	1,302	956	450	1,645
2007	2,885	1,519	515	1,073	306	672	396	390	2,042	1,120	398	1,429
2008	2,748	2,032	552	1,205	274	1,159	740	493	1,960	1,238	433	1,335
2009	1,764	2,404	473	886	179	931	602	425	1,632	1,094	995	1,338

4.2. Regression results on total annual HVO accidents

In order to estimate the relative risks of different HVOs, we first estimated their accident distributions over the period 1991-2010. To better focus on the comparable relative risks of HVOs, we perform analyses by fleet size. We group the larger fleets together when the numbers by size are not sufficient to accurately estimate the model parameters. We use all available information to control for the determinants that could affect road accidents. For example, our regressions contain variables on carriers' economic sector and on the characteristics of the vehicles they own. We also used years to consider the evolution of accidents over the course of the analysis period.

The tables of all regression results from the study include 21 regressions, corresponding to different fleet sizes and vehicle types. As indicated above, the choices of statistical models estimated depend on the fleet size. Given that we are interested in the annual distribution of fleet accidents, our natural starting point is the family of count distributions containing the Poisson and negative binomial distributions where there is over-dispersion (variance greater than the mean). For small fleets, we rejected the Poisson distribution and retained the negative binomial distribution because the alpha over-dispersion parameter was always positive. For these sizes, we also estimated the negative binomial with random effects to consider the panel aspect of the data. Because the main results are essentially the same between the two models, we interpret them by using the results of the negative binomial. The results of both models are presented in the tables.

Fleet size can influence the annual probability of having zero accidents, so we had to adapt our modeling to consider the fact that for larger fleets, the probability of having zero accidents in a year is very low. We also observed that the NB model did not estimate this probability adequately. As mentioned above, for these fleets we estimated the accident distributions by grouping different fleet sizes. In addition, we considered the fact that fleets of different sizes have different probabilities of having zero accidents. Therefore, we first estimated the probability of having zero accidents and used the estimated probability to weight the zeros in the regressions of the accident distributions for fleet sizes of 20 trucks and more. For these fleet sizes, it was necessary to use the negative binomial distribution with an overweight for the zeros of the largest fleets in the group to obtain predicted numbers of accidents that matched the observed frequencies.

An important part of our problem is to ascertain whether there is a statistical relationship between accidents and cumulative violations. We used the most frequent violations of the Highway Safety Code for drivers (DRV) and carriers (CAR). To reduce the simultaneity problem, we used the cumulative violation rates of the previous year *t-1* in the estimation of the accident distributions of the current year *t*. The variables are defined in (Table A2).

The results of the 21 total accident regressions are presented in [34]. We summarize the main results of our analyses below and present the estimations for two fleet sizes in more detail in tables 5 and 6.

The most interesting and stable results pertain to the numbers of violations that drivers accumulated in the previous year to explain the current year's accident numbers. For all fleet sizes below 50 trucks, the variables speeding, and failure to obey a red light or stop sign are significant at 1% with a positive coefficient to explain the following year's fleet accident numbers. As expected, DRV violations for failure to stop at a red light or a stop sign have the highest significant positive coefficients. Not wearing a seatbelt has a positive and significant coefficient at 1% for all fleet sizes below 50 trucks, except for the 8-truck group, where it is not significant at 10%. For fleets between 51 and 150 vehicles, non-wearing of seatbelts is non-significant, whereas the other violations have positive coefficients significant at 7% and better. Finally, for fleet sizes of more than 150 vehicles, only the speeding violation is significant with the expected sign, at 5%.

Two factors may explain the non-significance of some DRV variables for larger fleets. First, we have very few observations in this fleet category: Only 721 over the entire analysis period for the category of 150 trucks and more. This reduces the degrees of freedom, an important dimension when estimating non-linear models of this nature with panel data. Second, managers of larger fleets may exercise more stringent control over their drivers.

The statistical relationships between carrier violations (CAR) in one year and accidents in the following year are less significant than those for DRV violations. The violations most frequently significant at 5%, with a positive coefficient, are those for axle overload, total overload, improper stowage, failure to perform mechanical inspection and absence of pre-trip vehicle inspection. Violations for driving hours, hazardous materials and oversize are significant with a positive sign for very small fleets only. The results are less stable than those for DRV violations across sizes, but the most consistent violations with the highest significant positive coefficients are axle overloading, failure to check the vehicle's condition before departure, poor stowage, and failure to perform a mechanical inspection of the vehicle before departure. The years should be interpreted in relation to the year 1999. The years 2009 and 2010 were found to have negative signs with very high orders of magnitude.

In general, we are satisfied with the results obtained, except for very large fleets of more than 150 trucks, for which we have very few significant variables due to the fact that we have very few observations.

What most distinguishes the regressions in tables 5 and 6 is the frequency of having zero accidents for a fleet. It is 83% in (Table 5), but it is only 12.4% in (Table 6). This clearly justifies the use of an estimation model that incorporates this frequency into the estimation of the 21-50 vehicle group, given that the NB model does not integrate this frequency adequately. Without consideration of this frequency, the estimation of the distribution of these accidents would not meet the properties of a counting model. Note in (Table 6) that the percentage of zero crashes is much higher in the sample (12.4%) than the estimated probability (10.4%). This is why we use, in (Table 7), the negative binomial regression with an overweight for zeros, obtained from the Logit model presented in the second part of the table, to estimate the model.

In order to decrease potential correlations between the two models, we used control variables obtained from a PCA to estimate the probability of having zero accidents using the Logit model. The results of the Logit model are presented at the bottom of (Table 7), and the factors used are defined in (Table A3). We note, in (Table 7), that the percentage of zero accidents for a fleet in the sample is now very close to the estimated probability.

Table 5: *Estimated number of accidents involving HVO of size 2.*
This table presents the estimation results of the Negative binomial model with and without random effects.
Pr indicates the significance of the estimated parameter.

Variable name	Negative binomial (NB)		NB with random effects	
	Coefficient	**Pr > |t|**	**Coefficient**	**Pr > |t|**
Constant	-1.4214	<0.0001	1.1307	<0.0001
Main economic activity of HVO				
Activity not specified	-0.0674	0.0033	-0.0878	0.0005
Trucking (reference)	-----	-----	-----	-----
Passenger transportation	0.1759	<0.0001	0.1610	<0.0001
Agriculture and agriculture related services	-0.6983	<0.0001	-0.7340	<0.0001
Food and tobacco	0.0605	0.0962	0.0650	0.1320
Associations of leisure or Finance	-0.0950	0.0558	-0.0976	0.0767
Furniture	-0.0838	0.1572	-0.1012	0.1565
Timber harvesting and paper	-0.3846	<0.0001	-0.4064	<0.0001
Construction	-0.3380	<0.0001	-0.3462	<0.0001
Other	-0.1608	<0.0001	-0.1653	<0.0001
Average age of HVO vehicles	-0.0482	<0.0001	-0.0460	<0.0001
Average maximum number of axles of HVO	0.1171	<0.0001	0.1203	<0.0001
HVO started during the year	-0.2768	<0.0001	-0.3231	<0.0001
Number of axle overload violations	0.2098	<0.0001	0.1346	<0.0001
Number of total overload violations	0.1457	<0.0001	0.1209	<0.0001
Number of oversize violations	0.0637	0.1442	0.0413	0.3054
Number of stowage violations	0.2204	<0.0001	0.2080	<0.0001
Number of hazardous material violations	0.2620	0.0134	0.0397	0.6993
Number of driving violations	0.1932	0.0001	0.1180	0.0019
Number of mechanical inspection violations	0.3297	<0.0001	0.2760	<0.0001
Number of pre-departure inspection violations	0.2895	<0.0001	0.2356	<0.0001
Number of speeding violations	0.2395	<0.0001	0.1758	<0.0001
Number of red light violations	0.3899	<0.0001	0.2427	<0.0001
Number of stop sign violations	0.4148	<0.0001	0.3042	<0.0001
Number of seat belt violations	0.3160	<0.0001	0.2482	<0.0001

Year of accident accounted for by dichotomous variables					
Dispersion parameter	0.9956	<0.0001	-----	-----	
a	-----	-----	25.9681	<0.0001	
b	-----	-----	2.0112	<0.0001	
Number of HVOs		32,886		32,886	
Number of observations		110,570		110,570	
Likelihood log		-58,597		-57,841	
AIC		117,283		115,774	
BIC		117,716		116,216	
Statistics	**N**	**Average**	**Standard deviation**	**Min**	**Max**
Probability of zero accidents	110,570	0.8352	0.0743	0.0342	0.9792
Percentage of zero observations	110,570	0.8341	0.3720	0.0000	1.0000
Mathematical expectation of accidents	110,570	0.2104	0.2095	0.0213	27.8920
Average number of accidents	110,570	0.2085	0.5498	0.0000	23.0000

Table 6: *Estimation of the number of accidents of heavy trucks of HVO sizes 21 to 50 with NB.*

This table presents the estimation results of the Negative binomial model with and without random effects for larger fleets.

Pr indicates the significance of the estimated parameter.

Variable name	**NB**		**NB with random effects**	
	Coefficient	**Pr > |t|**	**Coefficient**	**Pr > |t|**
Constant	-1.7903	<.0001	-1.5258	<.0001
Main economic activity of HVO				
Activity not specified	-0.6327	<.0001	-0.5225	<.0001
Trucking (reference)	-----	-----	-----	-----
Passenger transportation	-0.0337	0.4034	0.0894	0.1258
Agriculture and agriculture related services	-0.5340	<.0001	-0.2814	0.0835
Food and tobacco	-0.0040	0.9352	0.2645	0.0015
Associations of leisure or Finance	-0.1817	0.0131	0.1406	0.1682
Furniture	-0.7267	0.1001	-0.2450	0.6659
Timber harvesting and paper	-0.5586	<.0001	-0.3444	0.0160
Construction	-0.1741	<.0001	-0.0581	0.3203
Other	-0.1237	0.0010	-0.0410	0.4909
Average age of HVO vehicles	-0.0142	<.0001	-0.0177	0.0002
Average maximum number of axles of HVO	-0.0482	<.0001	0.0395	0.0042
HVO started during the year	-0.2537	<.0001	-0.2311	<.0001
Number of axle overload violations	0.0260	<.0001	0.0108	0.0262

Number of total overload violations	0.0263	0.0161	0.0193	0.0176
Number of oversize violations	-0.0098	0.5567	-0.0056	0.6403
Number of stowage violations	0.0318	0.1064	0.0113	0.4428
Number of hazardous material violations	-0.0185	0.6387	0.0445	0.1363
Number of driving violations	-0.0014	0.9096	0.0084	0.3449
Number of mechanical inspection violations	0.0287	0.0565	0.0071	0.4559
Number of pre-departure inspection violations	0.0361	0.1550	0.0315	0.0926
Number of speeding violations	0.0341	<.0001	0.0207	0.0000
Number of red light violations	0.0761	<.0001	0.0271	0.0159
Number of stop sign violations	0.1056	<.0001	0.0301	0.0199
Number of seat belt violations	0.0873	<.0001	0.0393	0.0020
Year of accident accounted for by dichotomous variables				
Dispersion parameter	0.4618	<.0001	-----	-----
a	-----	-----	7.6711	<.0001
b	-----	-----	3.9547	<.0001
Number of HVOs	1,229		1,229	
Number of observations	6,440		6,440	
Likelihood log	-16,282		-15,248	
AIC	32,655		30,588	
BIC	32,960		30,899	

Statistics	**N**	**Average**	**Standard deviation**	**Minimum**	**Maximum**
Probability of zero accidents	6,440	0.1044088	0.0598838	0.0001	0.452
Percentage of zero observations	6,440	0.1237578	0.3293306	0	1.0000000
Mathematical expectation of accidents	6,440	4.8472516	2.8994031	0.9590	134.908
Average number of accidents	6,440	4.7569876	4.5540819	0	76.000000

Table 7: *Estimated number of accidents of HVO heavy trucks sizes 21 to 50 with the NB and considering excess zeros.*

This table presents the estimation results of the Negative binomial model with and without random effects for larger fleets when considering excess zeros with the Logit model. Pr indicates the significance of the estimated parameter.

Variable name	**Negative binomial logit (overweight for zeros)**	
	Coefficient	**Pr > |t|**
Constant	-1.6082	<0.0001
Main economic activity of HVO		
Activity not specified	-0.5212	<0.0001
Trucking (reference)	-----	-----
Passenger transportation	-0.0687	0.0722
Agriculture and agriculture related services	-0.5687	<0.0001
Food and tobacco	-0.0564	0.2221
Associations of leisure or Finance	-0.2406	0.0005
Furniture	-0.7985	0.0569
Timber harvesting and paper	-0.5845	<0.0001
Construction	-0.2054	<0.0001
Other	-0.1244	0.0006
Average age of HVO vehicles	-0.0123	0.0003
Average maximum number of axles of HVO	-0.0611	<0.0001
HVO started during the year	-0.2194	0.0002
Number of axle overload violations	0.0228	<0.0001
Number of total overload violations	0.0192	0.0542
Number of oversize infractions	-0.0114	0.4481
Number of stowage violations	0.0218	0.2281
Number of hazardous material violations	-0.0293	0.4273
Number of driving hour violations	-0.0035	0.7595
Number of mechanical verification violations	0.0264	0.0490
Number of infractions pre-departure check	0.0245	0.2986
Number of speeding violations	0.0278	<0.0001
Number of red light violations	0.0643	<0.0001
Number of stop sign violations	0.0885	<0.0001
Number of seat belt violations	0.0687	<0.0001
Year of accident accounted for by dichotomous variables		
Dispersion parameter	0.3672	<0.0001
Probability of having zero accidents	Logit	
Constant	-9.6510	<0.0001
Trucking	0.7181	0.0074
Number of years	0.1082	0.3027
Number of axles	-0.2510	0.0162

Start HVO	-0.0302	0.6738
Offences CAR 1	-4.1327	0.0024
Offences CAR 2	-0.6606	0.1491
DRV Offences	-2.2247	<0.0001
Number of HVO	1,229	
Number of observations	6,440	
Likelihood log	- 16,151	
AIC	32,408	
BIC	32,767	

Statistics	N	Average	Minimum	Maximum
Probability of zero accidents	6,440	0.1246566	0.0001	0.4698
Percentage of zero observations	6,440	0.1237578	0	1.0000000
Mathematical expectation of accidents	6,440	4.8047781	0.9670	87.9479
Average number of accidents	6,440	4.7569876	0	76.0000000

Another distinction is the size of the violation coefficients between tables 5 and 7, although the coefficients for DRV offences are all highly significant in both tables. The coefficients in (Table 5) indicate a greater sensitivity of crash violations with respect to violations. This can be explained by greater heterogeneity in road safety management between the larger fleets.

4.3. Identification of HVO risk classes

Many theoretical contributions were published since the 1970s to account for stylized facts related to information problems observed in insurance markets.[2] Partial insurance, such as deductible and coinsurance contracts, can be justified by asymmetric information. However, these static contracts have been shown to be dominated by dynamics contracts based on the information from past experience when long-term relationships exist between the principal (less informed) and the agent (more informed). This is because past experience in road safety convictions and accidents add information on the long run behavior of more informed participants [21,22,23]. In particular, it was demonstrated that multi-period contracts with memory outperform those without memory under full commitment by the insurance industry. Here, under asymmetric information, the relevant trade-off is between insurance availability at a low price and incentives that consider the benefits and costs of road safety in a hierarchical framework comprising insurers, regulators, fleet owners and drivers. In this section, we use fleets' past experience in road safety activities to predict road accidents and to set different risk classes for given fleet sizes.[3]

To achieve our third objective, we construct risk classes using model estimation results for vehicle fleets. We want to identify vehicle fleets at risk for road safety. The results are presented in (Table 8) for selected fleet sizes. To perform this task, we used the results obtained from the regressions on total heavy truck crashes such as those in tables 5 and 7. After determining that we would have five risk classes per size of HVO, we first predicted the number of annual accidents for each HVO in each year it was present in our database by calculating its annual mathematical accident expectation, namely the sum of the products of the estimated coefficients of the variables in the

2 Usually, insured are assumed to have more information about their risk or prevention activities than the insurer or the regulator (adverse selection: [35]; and moral hazard: [36]).
3 For an analysis of the deterrent effects of traffic enforcement see [6].

regressions and the values of the different variables in the regressions. We then ordered the mathematical expectation of accidents and constructed 100 ordered groups of 1% of the observations (percentiles).

To determine the percentile of the least risky class for a given HVO size, we used the percentage of HVOs with zero accidents. For example, in the group of one-truck HVOs, 93% of the fleets had zero accidents. We then placed the HVOs with the lowest mathematical expectation up to the 93^{rd} percentile, *i.e.* those with an average mathematical expectation of 8.22%, in the first risk class. Confidence intervals of calculated mathematical expectations are not reported but are available.

We still had 7% of HVOs to classify. We therefore used the seven remaining 1% groups, which we classified into four risk classes. To determine the sizes of the four remaining classes, we analyze the accident averages of the seven remaining groups and classify them according to the ranges of observed accident numbers. Once the percentages of the number of observations in each remaining class have been set, we rank the mathematical accident expectations in ascending order within each class. For example, the second risk class of 1 truck HVOs contains 3% of the HVOs with an average mathematical expectation of 18.89%, and so on. The size 1 HVOs most at risk for road safety are those with an average mathematical expectation of 49.97%. They numbered 5,310 over the 20 years of the study and represented 1% of the observations. The size 3 fleets most at risk of accidents have an average mathematical expectation (per truck) of 51.46% and represent 1.2% of the population of this fleet size.

In (Table 9), we redo the exercise and modify two aspects. On the one hand, we represent the risk classes for the year 2010 and not an average for all years. The advantage of this method is that it provides a shorter-term and more operational view of the pricing. In addition, violations of both types are now aggregated in the regressions. Their parameters are presented in (Table 10). As seen above, (Table 10) shows a similar discrepancy between the coefficients of the DRV and CAR variables. They are now all statistically significant for both types of infractions by fleet size. For each fleet size, we verify that the parameters of the aggregate DRV violations are larger than those of CAR, which is consistent with the results in tables 5 and 7. These parameters were estimated with the 20-year data. The other regression parameters are not reported but are very similar to those in the previous regressions. However, for the calculation of mathematical expectations of accidents, only 2010 data are used for fleet characteristics, and 2009 data are used for both types of infractions. An insurer could therefore use stable parameters estimated over several years (to be updated from time to time) to rank fleets from one year to another using current and previous year's information.

Note also that we use the same percentiles as in (Table 8) to divide the five risk classes for comparison purposes, but also to ensure the stability of results. Because we only use the fleets present in 2010, we have fewer observations in each risk class. We note that the mathematical expectations are lower in the risk classes because, on average, the accident frequencies have decreased over time. For size 1, we have 254 bad risks with an average accident frequency of 19.25%, compared with an average of 4.63% for the lowest risk class. The worst risks accumulated 2.73 DRV points and 4.0 CAR points, while those in the least risky class accumulated, on average, 0.14 DRV points and 0.16 CAR points. These figures clearly affirm great heterogeneity between fleets of the same size.

Table 8: *Risk classes of different sizes of HVOs calculated with total accidents over all years*.

This table identifies the average fleet risk classes over all years in using econometric estimation results.

Five classes are identified for documented fleet sizes.

	Risk class				
HVO with 1 heavy truck	**1**	**2**	**3**	**4**	**5**
Number of observations	491,872	15,872	10,544	5,310	5,310
% of 528,908	93%	3%	2%	1%	1%
Average number of accidents	0.0836	0.1801	0.2047	0.2480	0.3910
Mathematical expectation of accident	0.0822	0.1889	0.2217	0.2688	0.4997
Mathematical expectation of accident per truck	0.0822	0.1889	0.2217	0.2688	0.4997
HVO with 2 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number of observations	92,295	12,633	2,343	2,223	1,076
% of 110,570	83%	11%	2%	2%	1%
Average number of accidents	0.1685	0.3428	0.4285	0.5551	0.8634
Mathematical expectation of accident	0.1660	0.3410	0.4512	0.5697	1.2262
Mathematical expectation of accident per truck	0.0830	0.1705	0.2256	0.2849	0.6131
HVO with 3 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number of observations	36,079	4,642	5,141	1,792	597
% of 48,251	74.8%	9.6%	10.7%	3.7%	1.2%
Average number of accidents	0.2601	0.4946	0.5684	0.7991	1.1223
Mathematical expectation of accident	0.2587	0.4572	0.5789	0.8126	1.5439
Mathematical expectation of accident per truck	0.0862	0.1524	0.1930	0.2709	0.5146
HVO from 10 to 20 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number of observations	4,307	3,344	4,218	2,611	2,555
% of 17,035	25%	20%	25%	15%	15%
Average number of accidents	1.0697	1.5472	2.1280	2.6032	3.8204
Mathematical expectation of accident	1.1274	1.5664	1.9974	2.5648	4.1352
Mathematical expectation of accident per truck	0.1006	0.1299	0.1524	0.1768	0.2586
HVO from 21 to 50 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number of observations	754	1 378	2 003	1 574	781
% of 6,440	12%	21%	31%	24%	12%

Average number of accidents	2.1366	3.0965	4.5926	5.8727	8.4609
Mathematical expectation of accident	2.2397	3.1706	4.2496	5.9051	9.4412
Mathematical expectation of accident per truck	0.0937	0.1275	0.1549	0.1746	0.2335
HVO from 51 to 150 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number of observations	148	571	723	482	144
% of 2,069	7%	28%	35%	23%	7%
Average number of accidents	5.1824	7.5639	12.062	17.886	23.833
Mathematical expectation of accident	5.2862	7.8729	11.298	17.603	30.798
Mathematical expectation of accident per truck	0.0887	0.1246	0.1607	0.1815	0.2581
HVO of more than 150 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number of observations	108	125	236	144	108
% of 721	15%	17%	33%	20%	15%
Average number of accidents	10.019	20.504	29.013	40.424	94.722
Mathematical expectation of accident	10.153	17.918	27.546	45.455	96.354
Mathematical expectation of accident per truck	0.0505	0.0884	0.1238	0.1541	0.2007

Table 9: *Risk classes of different sizes of HVO calculated with total accidents for 2010.*

This table identifies the fleet risk classes for the year 2010 in using econometric estimation results. Five classes are identified for documented fleet sizes.

	Risk class in 2010				
HVO with 1 heavy truck	**1**	**2**	**3**	**4**	**5**
Number	23,149	746	486	260	254
% of 24,895	93%	3%	2%	1%	1%
Mathematical expectation of accident	0.0463	0.0942	0.1054	0.1231	0.1925
Average DRV demerit points in 2009	0.1434	0.4853	1.2058	1.5885	2.7283
Average accumulated carrier points in 2009	0.1634	0.6367	1.0206	1.7769	4.0079
HVO with 2 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number	4,805	636	117	116	58
% of 5,789	84%	11%	2%	2%	1%
Mathematical expectation of accident	0.0953	0.1785	0.2201	0.2614	0.4622

Average DRV demerit points in 2009	0.2435	1.0330	1.4701	2.2500	3.9655
Average accumulated carrier points in 2009	0.3589	1.1431	2.1111	3.1897	5.6207
HVO with 3 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number	1,957	258	265	104	27
% of 2,611	75%	10%	10%	4%	1%
Mathematical expectation of accident	0.1565	0.2630	0.3212	0.4279	0.6853
Average DRV demerit points in 2009	0.3091	0.9496	1.6075	2.8269	5.4074
Average accumulated carrier points in 2009	0.5422	1.0736	1.8755	3.4423	8.3704
HVO from 10 to 20 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number	280	224	281	168	168
% of 1,121	25%	20%	25%	15%	15%
Mathematical expectation of accident	0.6937	0.8840	1.0804	1.3517	1.9251
Average DRV demerit points in 2009	0.7429	1.7902	2.5801	4.0536	7.1964
Average accumulated carrier points in 2009	1.4036	2.0893	3.0534	5.0238	8.8155
HVO from 21 to 50 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number	53	90	135	24	51
% of 434	12%	21%	31%	24%	12%
Mathematical expectation of accident	1.5754	1.9248	2.3874	3.3150	5.1936
Average DRV demerit points in 2009	1.5472	2.3333	5.2370	7.5143	14.7255
Average accumulated carrier points in 2009	2.2453	4.1667	6.6444	8.3714	17.3725
HVO with 51 to 150 heavy trucks	**1**	**2**	**3**	**4**	**5**
Number	9	39	49	32	9
% of 138	6.5%	28.3%	35.5%	23.2%	6.6%
Mathematical expectation of accident	3.4742	5.2194	7.0988	10.7321	16.8773
Average DRV demerit points in 2009	7.3333	5.9487	12.5102	19.3125	35.8889
Average accumulated carrier points in 2009	5.6667	9.7179	14.6939	23.5313	23.6667
HVO of more than 150 heavy trucks	**1**	**2**	**3**	**4**	**5**

Number	6	6	15	11	8
% of 46	13%	13%	32%	24%	17%
Mathematical expectation of accident	6.7379	9.6246	15.8590	25.1876	60.4193
Average DRV demerit points in 2009	13.3333	13.667	30.8000	37.5455	82.6250
Average accumulated carrier points in 2009	15.5000	15.000	22.8000	29.7273	41.3750

Table 10: *Parameters of aggregate DRV and CAR violations.*

This table documents the estimated parameters of aggregate driver violations (DRV) and carrier violations (CAR) obtained from the Negative binomial model. Pr indicates the significance of the estimated parameter.

HVO	**Negative binomial**		
	Parameter	**Standard deviation**	**Pr > |t|**
1 heavy truck for years 1991-2010			
Number of CAR violations	0.2671	0.0092	<0.0001
Number of DRV violations	0.3476	0.0099	<0.0001
2 heavy trucks for years 1991-2010			
Number of CAR violations	0.1869	0.0092	<0.0001
Number of DRV violations	0.2597	0.0103	<0.0001
3 heavy trucks for years 1991-2010			
Number of CAR violations	0.1221	0.0088	<0.0001
Number of DRV violations	0.2006	0.0103	<0.0001
10 to 20 heavy trucks for years 1991-2010			
Number of CAR violations	0.0375	0.0029	<0.0001
Number of DRV violations	0.0769	0.0039	<0.0001
21 to 50 heavy trucks for years 1991-2010			
Number of CAR violations	0.0161	0.0025	<0.0001
Number of DRV violations	0.0376	0.0031	<0.0001
51 to 150 heavy trucks for years 1991-2010			
Number of CAR violations	0.0060	0.0024	<0.0001
Number of DRV violations	0.0166	0.0028	<0.0001
More than 150 heavy trucks for years 1991-2010			
Number of CAR violations	0.0096	0.0026	<0.0001
Number of DRV violations	0.0042	0.0021	<0.0001

5. Discussion and conclusion

Our research has implications for road safety managers and, more specifically, those responsible for the road safety of owners and operators of heavy vehicles. It is also aimed at managers of transport companies, drivers of heavy vehicles, insurers, and regulators of road transport companies.

One important contribution of our research is that we compiled an original 20-year database for HVOs. This article concentrated on heavy trucks, although other types of vehicles were analyzed.

The first immediate outcome of our research is that we have developed a methodology to identify the individual risks of HVOs. In line with the recent literature on road safety incentives [2,3,7] our method consists in calculating the annual mathematical expectation of crashes for each HVO for the coming year. We show that these mathematical expectations are a function of the characteristics of HVOs in the current period, and of driver and carrier safety code violations (to a lesser extent) in the previous year. The statistical results show that past offenses are significant in explaining the relative risks of HVOs, which are stable across fleet sizes in general.

Several violations of the Highway Safety Code by drivers (DRV) and carriers (CAR) are determining in explaining fleet accidents. The main DRV offenses are speeding, failure to stop at a red light and failure to stop at a stop sign, while the main CAR violations are axle overload, total overload, and failure to perform a mechanical inspection. This information could be used to better target the oversight of road safety regulations, as discussed in [12,13].

In a second step, we constructed risk classes by fleet size. We found considerable heterogeneity between vehicle fleets in terms of road safety. The different risk classes constructed in this research categorize the riskiest vehicle fleets by size. The use of these risk classes to price insurance should encourage the riskiest fleets to increase their prevention activities and motivate the least risky to continue to be cautious. In the medium term, we should see a decline in road accidents for vehicle fleets if this type of pricing is applied [18,20,22].

An extension of our research would be to test how the implementation of this type of insurance pricing affects accident rates. It is not necessary to apply the change to all fleets. Rather, it should be implemented on a voluntary basis. This will allow a control group and an experimental group to be identified. The use of difference-in-difference and firm propensity score-matching methodologies will make it possible to detect the differences between the two groups and to verify whether the use of risk classes has a causal effect on road safety [37, 38].

Acknowledgments

Research funded by the Société de l’assurance automobile du Québec, in collaboration with the Fonds de recherche sur la société et la culture du Québec and the Fonds de recherche du Québec – Santé, and by SSHRC Canada. We would like to thank Claire Boisvert for her remarkable contribution to the preparation of this document. We thank Benoit Dostie for his comments.

References

1. Savage I. A structural model of safety and safety regulation in the truckload trucking industry. Transportation Research Part E: The Logistics and Transportation Review. 2011;47:249-62.
2. Savage I. Reflections on the economics of transportation safety. Res Transp Econ. 2012; 43:1-8.
3. https://truckingresearch.org/wp-content/uploads/2015/09/ATRI-Predicting-Truck-Crash-Involvement-2011-Update.pdf
4. Angers JF, Desjardins D, Dionne G, et. al. Vehicle and fleet random effects in a model of insurance rating for fleets of vehicles. ASTIN Bulletin. 2006;36:25-77.
5. Angers JF, Desjardins D, Dionne G, et. al. Modelling and estimating individual and firm effects with count panel data. ASTIN Bulletin. 2018; 48:1049-78.
6. Tay R. General and specific deterrent effects of traffic enforcement: Do we have to catch offenders to reduce crashes? J Transp Econ Policy. 2005;39:209-23.
7. Desjardins D, Dionne G, Lu Y. Hierarchical random effect model for insurance pricing of vehicles belonging to a fleet. 2021.

   https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3788515.
8. https://saaq.gouv.qc.ca/fileadmin/documents/publications/espace-recherche/infractions-sanctions-routier-2009-2018.pdf
9. Peltzman S. The effects of automobile regulation. J Polit Econ. 1975;83:677-726.
10. Landes EM. Insurance, liability, and accidents: A theoretical and empirical investigation of the effect of no-fault accidents.J Law Econ. 1982;25:49-65.
11. Graham JD, Garber S. Evaluating the effects of automobile safety regulation. J Policy Anal Manage. 1984;3:206-24.
12. Boyer M, Dionne G. The economics of road safety. Transport Res B-Meth. 1987;21:413-31.
13. Blomquist GC. The Regulation of Motor Vehicle and Traffic Safety. (1stedn). Springer Netherlands. 1988:pp.1-130.
14. Devlin RA. Liability versus no-fault automobile insurance regimes: An analysis of the experience in Québec. In: Dionne G. (Eds). Contributions to Insurance Economics. (1stedn). Springer, Dordrecht. 1992:pp.499-520.
15. Cummins JD, Weiss MA. Incentive effect of no-fault on automobile insurance: Evidence from insurance claim data. In: Dionne, G. (Eds), Contributions to Insurance Economics. (1stedn). Springer, Dordrecht. 1992;pp.445-70.
16. Cummins JD, Phillips R, Weiss MA. The incentive effects of no-fault automobile insurance. J Law Econ. 2001;44:427-64.
17. Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation. Dionne G, Labergen Nadeau C. (Eds). Springer US. 1999:pp.1-351.
18. Abbring J, Chiappori PA, Pinquet J. Moral hazard and dynamic insurance data. J Eur Econ. 2003;1:767-820.

19. Dionne G, Michaud PC, Pinquet J. A review of recent theoretical and empirical analyses of asymmetric information in road safety and automobile insurance. Res Transp Econ. 2013;43:85-97.

20. Pinquet J. Experience rating in nonlife insurance. In: Dionne G. (Ed.). Handbook of Insurance. (2ndedn). Springer New York. 2013:pp.471-85.

21. Cohen A, Siegelman P. Testing for adverse selection in insurance markets. J Risk Insur. 2010;77:39-84.

22. Dionne G, Michaud PC, Dahchour M. Separating moral hazard from adverse selection and learning in automobile insurance: Longitudinal evidence from France. J Eur Econ. 2013;11:897-91.

23. Hausman JA, Hall BH, Griliches Z. Econometric models for count data with an application to the patents- R&D Relationship. Econometrica. 1984 52:909-38.

24. Boyer M, Dionne G, Vanasse C. Econometric models of accident distributions. In: Dionne G. (Ed.), Contributions to Insurance Economics. (1stedn). Dordrecht Springer. 1992;pp.169-213.

25. Cameron AC, Trivedi PK. Regression Analysis of Count Data. (1stedn). Econometric society monograph Cambridge University Press. 1998:pp.1-432.

26. Dionne G, Vanasse C. Automobile insurance ratemaking in the presence of asymmetrical information.J Appl Econom. 1992;7:149-165.

27. Dionne G, Vanasse C. A generalization of automobile insurance rating models: The negative binomial distribution with a regression component. ASTIN Bull. 1989;19: 199-212.

28. Mullahy J. Specification and testing of some modified count data models. J Econom. 1986;33:341-365.

29. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992;34:1-14.

30. Creel M, Loomis JB. Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California.Am J Agric Econ. 1990;72:434-41

31. https://support.sas.com/resources/papers/proceedings/pdfs/sgf2008/322-2008.pdf

32. Majo MC, Van Soest AHO. The fixed-effects zero-inflated Poisson model with an application to health care utilization. (CentER Discussion Paper no. 2011-083). Econometrics. 2011.

33. Staub KE, Winkelmann R. Consistent estimation of zero-inflated count models. Health Econ. 2012;22:673-86.

34. https://chairegestiondesrisques.hec.ca/wp-content/uploads/2015/04/15.01.pdf

35. Rothschild M, Stiglitz J. Equilibrium in insurance markets: An essay on the economics of imperfect information. Q J Econ. 1976;90:629-49.

36. Shavell S. On moral hazard and insurance.Q J Econ. 1979;93:541-62.

37. Elvik R. Paradoxes of rationality in road safety policy.Res Transp Econ. 2013;43:62-70.

38. Dionne G, Liu Y. Effects of insurance incentives on road safety: Evidence from a natural experiment in China. Scand J Econ. 2021;123:453-77.

Appendix

Table A1: *HVO main economic activity over the study period.*

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Activity not specified	10,306	10,548	10,780	13,371	15,095	16,346	17,462	18,557	10,214	9,558
Trucking	8,736	8,405	8,572	8,497	8,445	8,483	8,586	8,527	8,062	8,134
Passenger transportation	43	47	49	46	39	40	45	41	52	52
Other transportation	107	105	105	95	85	86	82	78	139	141
Other services related to transportation	256	273	290	266	243	223	208	206	480	500
Wholesale automobile, parts and accessories trade	96	97	96	90	79	69	61	63	97	101
Retail automobile, parts and accessories trade	864	888	892	820	734	681	642	613	943	909
Transportation equipment industries	76	73	68	54	52	51	49	46	65	70
Pipeline transportation	7	9	7	6	6	6	5	4	23	23
Air transportation	15	16	14	15	10	9	9	9	24	25
Rail transportation and services	14	14	13	13	10	10	9	9	11	11
Water transportation	30	30	31	30	28	27	27	25	43	49
Agriculture	3,750	3,800	3,791	3,777	3,667	3,669	3,618	3,449	2,837	2,712
Services related to agriculture	88	95	96	89	87	81	77	74	86	93
Wholesale trade in agricultural products	59	59	60	53	50	44	41	42	58	57
Wholesale trade in agricultural machinery, equipment and supplies	475	475	485	440	394	343	333	320	496	491
Food	1,388	1,385	1,379	1,252	1,129	1,034	958	882	1,479	1,465
Association and leisure activities	268	290	304	278	253	225	207	179	335	336
Furnishings	741	741	732	672	593	538	508	468	671	669
Timber harvesting and paper	826	842	864	802	741	694	658	628	974	945
Clothing and accessories	336	322	310	288	263	220	197	184	272	258
Construction	5,501	5,646	5,578	5,131	4,624	4,204	3,957	3,690	5,203	5,181
Fishing and trapping	13	10	9	10	9	9	9	11	18	13
Materials processing	552	546	517	489	445	417	390	362	470	466
Oil and gas	322	336	346	330	305	276	264	251	276	272
Communications	185	199	233	204	190	180	154	145	294	287

Human services	754	776	781	704	635	575	543	488	899	889
Administrative and public services	1,313	1,398	1,390	1,379	1,358	1,341	1,313	1,288	1,442	1,412
Business services	193	196	184	160	137	123	116	111	384	404
Tobacco	28	30	26	20	18	19	18	17	14	17
Textiles	83	82	80	74	62	53	52	49	65	64
Finance, real estate and insurance	404	377	355	310	271	238	227	199	983	932
Mining industry	345	355	358	337	302	274	271	252	324	309
Other commercial ventures	1,376	1,373	1,352	1,246	1,129	1,027	958	898	1,548	1,599
Other industries	406	396	405	365	333	312	292	267	422	428
None									481	442
Health care	78	74	68	67	60	57	51	56	101	106
Unknown	6	6	6	4	4	3	4	5	174	137
	2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Activity not specified	9,113	8,971	8,961	8,992	8,916	8,974	9,035	8,710	8,520	8,583
Trucking	8,039	8,116	8,364	8,780	9,044	9,149	9,110	8,703	8,475	8,560
Passenger transportation	53	52	59	68	78	66	67	66	66	71
Other transportation	145	152	163	182	189	197	215	237	264	313
Other services related to transportation	487	507	503	509	539	546	554	566	592	636
Wholesale automobile, parts and accessories trade	104	101	96	100	105	115	116	127	120	127
Retail automobile, parts and accessories trade	907	925	938	964	975	1,010	1,037	1,051	1,056	1,090
Transportation equipment industries	64	66	75	87	89	94	98	98	101	113
Pipeline transportation	18	17	18	17	17	17	17	16	17	18
Air transportation	25	28	24	26	30	29	33	33	34	38
Rail transportation and services	12	10	11	11	12	14	15	16	14	13
Water transportation	44	45	43	43	44	39	39	40	38	38
Agriculture	2,587	2,503	2,479	2,388	2,277	2,250	2,217	2,207	2,198	2,208
Services related to agriculture	88	94	94	101	99	106	119	122	131	140
Wholesale trade in agricultural products	58	57	66	67	64	69	68	78	81	77
Wholesale trade in agricultural machinery, equipment and supplies	494	486	470	461	458	452	454	447	452	450
Food	1,467	1,481	1,467	1,474	1,450	1,442	1,464	1,439	1,464	1,455
Association and leisure activities	324	337	348	358	346	364	376	375	392	399
Furnishings	690	670	652	653	638	632	644	636	637	624

Timber harvesting and paper	912	857	840	854	838	833	810	778	798	830
Clothing and accessories	242	222	189	175	157	138	134	125	123	111
Construction	5,254	5,248	5,372	5,450	5,641	5,789	5,985	6,138	6,444	6,706
Fishing and trapping	17	17	16	18	22	20	20	25	27	31
Materials processing	465	445	437	431	431	423	435	432	434	429
Oil and gas	262	250	260	251	245	237	237	231	225	215
Communications	288	291	289	290	316	323	330	341	325	327
Human services	887	882	895	906	903	909	959	990	1,021	1,050
Administrative and public services	1,379	1,262	1,256	1,262	1,268	1,291	1,295	1,286	1,297	1,328
Business services	390	379	406	431	439	442	469	509	549	567
Tobacco	16	15	14	13	10	7	6	9	9	8
Textiles	60	50	57	53	51	48	48	40	37	36
Finance, real estate and insurance	874	849	819	780	782	738	741	677	670	714
Mining industry	310	314	312	313	319	311	305	316	306	320
Other commercial ventures	1,616	1,605	1,629	1,645	1,660	1,687	1,726	1,709	1,702	1,740
Other industries	423	424	429	433	428	446	453	464	462	474
None	418	390	350	290	256	194	138	107	78	56
Health care	114	120	119	113	113	112	111	102	94	98
Unknown	111	99	92	84	75	102	112	125	155	177

Table A2: *Variable descriptions*.

Variable name	Description
Main economic activity of HVO	
Activity not specified	Main economic activity of the fleet not specified
Trucking (reference)	Trucking is the main economic activity of the fleet
Passenger transportation	Passenger transportation is the main economic activity of the fleet
Agriculture and agriculture related services	Agriculture and agriculture related services is the main economic activity of the fleet
Food and tobacco	Food and tobacco is the main economic activity of the fleet
Associations of leisure activities or Finance	Associations and leisure activities, or Finance, real estate and insurance is the main economic activity of the fleet
Furniture	Furniture is the main economic activity of the fleet
Timber harvesting and paper	Timber harvesting and paper is the main economic activity of the fleet
Construction	Construction is the main economic activity of the fleet

Other	The main economic activity is other than those mentioned above
Average age of HVO vehicles	Average age of the fleet vehicles during the year
Average maximum number of axles of HVO	Average maximum number of vehicles axles to obtain registration
HVO started during the year	Dummy variable equal to one if the fleet started its activities during that year Equal to zero otherwise
Number of axle overload violations	Number of axle overload violations by the fleet during the previous year
Number of total overload violations	Number of total overload violations by the fleet during the previous year
Number of oversize violations	Number of oversize violations by the fleet during the previous year
Number of stowage violations	Number of stowage violations by the fleet during the previous year
Number of hazardous material violations	Number of hazardous material violations by the fleet during the previous year
Number of driving violations	Number of driving violations by the fleet during the previous year
Number of mechanical inspection violations	Number of mechanical inspection violations by the fleet during the previous year
Number of pre-departure inspection violations	Number of pre-departure inspection violations by the fleet during the previous year
Number of speeding violations	Number of speeding violations by the drivers of the fleet during the previous year
Number of red light violations	Number of red light violations by the drivers of the fleet during the previous year
Number of stop sign violations	Number of stop sign violations by the drivers of the fleet during the previous year
Number of seat belt violations	Number of seat belt violations by the drivers of the fleet during the previous year

Table A3: *Definition of factors used in the principal component analysis.*

Factor	Description
Number of years	Score of the mean and standard deviation of the age of HVO vehicles in years and standard deviation of the mass of these vehicles
Number of axles	Score of the mean and standard deviation of the maximum number of axles of HVO vehicles and mean mass of these vehicles
Start HVO	Score if the HVO started or merged during the year
CAR1 offences	Score of mean number of violations per HVO truck for axle overloading,

	total overloading, oversize, and improper stowage
CAR2 offences	Score of average number of violations per HVO truck for driving hours, failure to perform mechanical or pre-trip inspection.
DRV offences	Average number of speeding, red light, stop sign and seatbelt violations per HVO truck