Point-record incentives, asymmetric information and dynamic data (revised version)

Jean Pinquet, Georges Dionne, Mathieu Maurice, Charles Vanasse

ECOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Point-record incentives, asymmetric information and dynamic data (revised version)

Jean Pinquet
Georges Dionne
Charles Vanasse
Mathieu Maurice

Juillet 2008

Cahier n° 2008-05

DEPARTEMENT D'ECONOMIE
Route de Saclay
91128 PALAISEAU CEDEX
(33) 1 69333033
http://www.enseignement.polytechnique.fr/economie/
mailto:chantal.poujouly@polytechnique.edu

Point-record incentives, asymmetric information and dynamic data (revised version)[1]

Jean Pinquet[2]
Georges Dionne[3]
Charles Vanasse[4]
Mathieu Maurice[5]

Juillet 2008

Cahier n° 2008-05

Résumé: Les politiques de sécurité routière utilisent souvent des mécanismes incitatifs basés sur les infractions pour améliorer le comportement des conducteurs. Ces mécanismes sont, soit monétaires (amendes, primes d'assurance), soit non monétaires (permis à points). Nous analysons l'efficacité de ces mécanismes dans l'incitation à une conduite prudente. Nous déterminons leurs propriétés théoriques par rapport au nombre de points associés aux infractions et par rapport au temps contrat. Ces propriétés sont ensuite testées empiriquement dans un modèle qui sépare l'aléa moral de l'hétérogénéité inobservée. Nous concluons à la présence d'aléa moral dans les données. Par ailleurs, la prime indicée sur les points introduite en 1992 a réduit de 15% la fréquence d'infractions. Enfin, nous comparons l'efficacité globale de ces différents mécanismes incitatifs et nous calculons des équivalents monétaires pour les infractions et les suspensions de permis.

Abstract: Road safety policies often use incentive mechanisms based on traffic violations to promote safe driving. These mechanisms are both monetary (fines, insurance premiums) and non-monetary (point-record driving licenses). We analyze the effectiveness of these mechanisms in promoting safe driving. We derive their theoretical properties with respect to contract time and accumulated demerit points. These properties are then tested empirically in a model which separates moral hazard from unobserved heterogeneity. We do not reject the presence of moral hazard in the Quebec public insurance regime. Moreover, we verify that the experience rating introduced in 1992 did reduce the frequency of traffic violations by 15%. Lastly, we compare the effectiveness of the different incentive schemes and we derive monetary equivalents for traffic violations and license suspensions.

Mots clés : Mécanismes incitatifs, permis à points, sécurité routière

Key Words : Point-record mechanisms, incentive effects, road safety

Classification JEL: D81, C23

---

[1] Claire Boisvert improved significantly the presentation of the manuscript.
[2] Université Paris 10 et Ecole Polytechnique (courriel: jean.pinquet@polytechnique.edu)
[3] HEC Montréal (courriel: dionne@hec.ca)
[4] TD Asset Managemen
[5] HEC Montréal

1 Introduction

Since the 1970s fatality rates due to road-traffic accidents have decreased steadily in developed countries, although risk exposure increased at the same time (see OECD, 2005). For example, over the last ten years, the road fatality rate decreased by forty percent in France. However, the implied social cost of road accidents is still very high (Doyle, 2005). By 2020, road-traffic accidents should become the third cause of the disability-adjusted life years lost due to disease or injury worldwide (Murray and Lopez, 1997). In 1990, this cause ranked only ninth.

A major reason for the improvement of the situation in the OECD has been the development of incentives for safe driving. Experience rating schemes used by the insurance industry have incentive properties (see Boyer and Dionne, 1989; Abbring et al, 2003). They are supplemented by point-record driving licenses based on traffic violations. In many countries, each convicted traffic offense is filed with a specific number of demerit points. When the accumulated number of points exceeds a given threshold, the driving license is suspended. Point removal clauses are added so that this penalty can be avoided in the long run.[1] A point-record driving license was implemented in Quebec in 1978, together with a no-fault insurance regime for bodily injuries which replaced a tort system.[2] The road fatality rate decreased by fifty percent during the fifteen years that followed.

In Quebec, the *Société de l'Assurance Automobile du Quebec* (referred to as SAAQ in what follows) is a public monopoly which provides coverage for bodily injury. The SAAQ is also in charge of accident prevention and control, including the management of driving licenses. Before 1992, the rating structure for bodily injury insurance was completely flat. The public authorities in Quebec decided to implement an experience rating scheme based on accumulated demerit points, a reform applied from December 1, 1992 onwards. This mechanism was added

to other incentives, i.e., fines, the point-record driving license in force since 1978, and the private sector insurance pricing scheme for property damage.

This paper analyses the incentive properties of both point-record driving license and insurance pricing based on traffic violations. Studies on incentive mechanisms for road safety have been discussed in the economic literature for many years (Peltzman, 1975; Landes, 1982; Boyer and Dionne, 1987). In the presence of asymmetric information, insurers use partial insurance or experience rating to improve resource allocation. Both schemes have been proved to be efficient for handling moral hazard and adverse selection. Different empirical tests have been proposed to measure the effectiveness of such mechanisms for road safety (Sloan et al, 1995; Boyer and Dionne, 1989) or to measure the presence of residual asymmetric information problems in insurers' portfolios (Chiappori and Salanié, 2000; Dionne, Gouriéroux and Vanasse, 2001). More recently, Abbring, Chiappori and Pinquet (2003) designed a new test based on the dynamics of insurance contracts to detect the presence of residual moral hazard. Their model makes it possible to separate the moral hazard effect on accidents from unobserved heterogeneity. They found no evidence of moral hazard in the French car insurance market. The convex structure of the French "bonus-malus" system is used to show that the optimal effort level exerted by a rational policyholder increases after a claim at fault. In our study, insurance pricing is not the major incentive scheme but rather a measure used to complement fines and the point-record driving licence. Moreover, the pricing scheme of the Quebec public automobile insurance is not strictly increasing and convex with respect to past demerit points but is increasing by steps. Finally, time effects are important in Quebec's point-record system, so we cannot apply directly the Abbring et al (2003) econometric methodology.

Insurance pricing may not suffice as a tool for designing an optimal road safety policy since it may not create the appropriate incentives for reckless drivers (Sloan et al, 1995). Bourgeon and Picard (2007) show how point-record driving

licence suspensions provide incentives for road safety among normal drivers (those who respond to the usual incentive schemes) when the judicial system or the insurance market fail to provide optimal incentives. Point-record driving licences also allow the government to incapacitate reckless drivers. Fines for traffic violations may be ineffective for reckless drivers when their amounts are bounded above, either because some drivers would not be able to pay them or for some equity reasons (see also Shavell, 1987). However, fines do reinforce the effectiveness of the point record mechanism by providing more incentives to normal drivers. In the Bourgeon and Picard model which uses only two levels of prevention, the optimal fine must be set at the maximal level and must be neither progressive nor regressive. These authors also discuss the optimality of point removal mechanisms as a screening device. Public intervention can also be justified when there is a significant difference between the private and the social cost of human lives (Viscusi, 1993). Finally, drivers may be unaware of their own accident or infraction probabilities or may misunderstand some features of the incentive environment. With respect to the theoretical contribution of Bourgeon and Picard (2007), we shall test the prediction that the point-record driving licence promotes road safety under moral hazard. We shall also analyse the Quebec public insurance pricing scheme based on past convictions as a progressive fine. It can also be interpreted as a bonus malus scheme.

We present the data base in Section 2 as well as our first empirical results related to the introduction of the new pricing policy implemented in 1992. The point-record mechanisms (driving license suspension and insurance pricing) are described in Section 3 and their incentive properties are investigated in a continuous-time model of optimal behavior that extends the previous literature significantly. These results are then confronted with empirical findings. Identifiability issues created by unobserved heterogeneity are addressed in Section 4. Section 5 presents empirical evidence on the incentive properties of the two point-record mechanisms as well as those of fines. In the spirit of Abbring et

al (2003), we propose a means of disentangling unobserved heterogeneity from incentive effects in a duration model. It involves including an actuarial predictor as an offset variable in the hazard function. The incentives created by the threat of driving license suspension are found to increase with accumulated demerit points. These findings confirm the theoretical analysis. We also find that driving license suspension spells reduce the risks of accidents and traffic violations.

The experience rating system implemented in 1992 has substantially reduced traffic violations among all drivers, whatever their incentive level. We compare the overall effectiveness of the different incentive schemes, and try to link global results to theoretical properties of the relation between safe driving effort and traffic violation risk. Lastly, we derive monetary equivalents for traffic violations and license suspensions. Conclusions are drawn in Section 6 and technicalities are relegated to an appendix available at the URL http://ceco.polytechnique.fr/publications/ (working paper 2008-05).

2 Presentation of the data base and preliminary empirical results

Our data base represents roughly one percent of the SAAQ portfolio. The panel covers the period from January 1, 1983 to December 31, 1996. A first sample of 40,000 license holders was selected at random at the beginning of 1983. Then about 300 young drivers were added each following year.[3] Leaving the motor insurance market is the only cause of attrition in the data base. The attrition rate per year is close to 1.5%, which is very low as compared to the private sector. This attrition result is obviously explained by the monopolistic status of the SAAQ. The endogenous attrition is not very high. It was estimated from a bivariate probit model on traffic offenses and departures from the sample. A

score test for the nullity of the correlation coefficient between the two equations[4] was performed with the regression components set used in Section 5. The null hypothesis was not rejected at a five percent significance level. Hence the attrition risk adds no significant information to the assessment of traffic violation risk.

The personal characteristics of each driver are available on the driving license for the current period. These characteristics are used as regression components in the empirical study. Several types of events are recorded in the data base; they are listed below with related variables in addition to the date. 1) Accidents which have led to a police report. Only those with bodily injury are compensated by the SAAQ. 2) Convicted violations of the Road Safety Code, together with the number of demerit points which are used in the point-record mechanisms. The number of demerit points is based on the severity of the traffic violation. Their distribution is given in Section 3.4. 3) Driving license suspensions, which are spells rather than events, and 4) Premium payments which since the 1992 reform are related to accumulated demerit points. These payments are made every two years at the policyholder's birthday.

Between 1985 and 1996, the average yearly frequencies of accidents with bodily injuries, accidents of all types (not including jointly-agreed reports to private insurers) and traffic violations are equal to, respectively, 1.4%, 6.7% and 16.9%. Figure 1 represents the relative frequencies derived from a one year centered moving average.[5] There is an overall decline in the frequency of accidents, whereas the frequency of traffic violations remains more stationary. This may seem surprising, but it is explained by the evolution of the traffic control environment. For instance, the number of traffic control devices such as radars increased during the 1980s and 1990s. An increase in the rate of traffic offenses recorded by devices or police officers among those committed explains this relationship. Figure 1 shows evidence of several periods where the frequency of traffic violations increased along with opposite variations in the frequencies of

accidents. A step-up in traffic control during these periods may well explain such observations. A traffic violation committed by a driver must be selected twice in order to be filed with demerit points. It must first be recorded by a control device or a police officer. We already mentioned that the related selection rate increased in the past. Second, the recorded traffic violation must be convicted. The filing of a traffic violation is somewhat discretionary. After the 1992 reform for instance, people are being forced to pay more in premiums given demerit points, and we might expect policemen to be more hesitant to hand them out, and to give warnings instead.[6] The conviction rate is less likely to vary with time than the recording rate.

Insert Figure 1 about here

In Figure 1, a downturn is also observed for the frequency of traffic violations just before the date (December 1, 1992) of the reform which introduced the experience rating structure based on demerit points. Notice that the reform was announced four months before its enforcement, which may explain this lag.[7] On average, the annual frequency of traffic violations was equal to 17.6% before the reform and 15.4% afterwards, which corresponds to a 12.5% decrease. The 1992 reform can be interpreted as a laboratory experiment to test whether an exogenous change in the use of memory reduces traffic violations. But the lower rate of traffic violations following the 1992 Quebec reform may be due to the change of other factors that may influence the driver behavior. Identifying the influence of these factors necessitates a control group that is not affected by the policy change. Unfortunately, we do not have access to such a control group since the insurer is a monopoly and bodily insurance is compulsory.[8] In Section 5.2, we shall link the average decrease of the frequency of traffic violations before and after the 1992 reform to the overall effectiveness of the different incentive schemes.

Monetary and non monetary incentives for safe driving are based on traffic violations as well as the optimal behavior models designed in Section 3. However

the actual social cost of road traffic is caused by accidents. To reconcile these two approaches, let us mention two results. First, demerit points are good predictors of accidents. This is well documented in the literature and is confirmed on our data in Section 5.1. Second, the global stationarity of convicted traffic violations frequency observed in Figure 1 concurs with a probable decrease in the frequency of committed traffic violations (see the aforementioned developments on selection rates). Lowering traffic violation risk through point-record mechanisms should also lower accident risk and the related social cost.

Finally, Figure 1 shows that accidents with bodily injuries evolve in much the same way as all those recorded in the SAAQ file. We include accidents of all types in the empirical analysis in order to obtain more stable results.[9]

3 Incentive effects of point-record mechanisms

3.1 Point-record mechanisms in Quebec

In this section, we describe Quebec's point-record mechanisms which are based on traffic violations, both monetary (insurance premiums) and non-monetary (point-record driving license). Comparisons are given with respect to the mechanisms used by other countries. We investigate the incentive properties of point-record mechanisms in Sections 3.2 to 3.4.

In many countries nowadays, driving license suspensions are based on demerit points. In Quebec, demerit points are assigned to convictions for traffic offenses and their number depends on the traffic violation severity. When the accumulated number of demerit points reaches or exceeds a given threshold, the driving license is suspended. Before January 1990 this threshold was set at twelve in Quebec and has been increased to fifteen since then.

In order to mitigate the social cost of driving license suspension, point removal systems exist for most real-world point-record driving licenses. In Que-

bec, the demerit points related to a given driving offense are removed after two years. Hence, driving license suspensions will depend on the demerit points recorded during the last two years. The French system is similar, with a three year seniority for the redemption of offenses and a twelve point threshold. New York State follows the same logic as Quebec and France (with an eighteen month seniority and an eleven point threshold). The average number of demerit points per convicted offense is equal to 2.4 in Quebec. It takes about six traffic violations within two years to trigger a license suspension, an unlikely outcome when the annual traffic violation frequency is equal to 16.9%. But heterogeneity of risks is high and a point-record driving licence is also an incapacitating device of risky and reckless drivers through the licence suspensions. Another point removal system consists in cancelling all the demerit points after a given period of violation-free driving. This mechanism was recently implemented in Spain, with a two year period. Utah has a point-record driving license similar to the Spanish one.

The experience rating structure introduced by the SAAQ in December 1, 1992 links each premium paid every two years to the demerit points accumulated over the previous two years. The rating structure is given in Section 3.4. Once the premium is paid, the driver is reinstated with a fresh zero point record. Thus the length of the record relevant to the derivation of optimal behavior never exceeds two years.

3.2 Basic model for a point-record driving license without point removal

Bourgeon and Picard (2007) analyze the incentive effects of point-record driving licenses. Their model uses a binary effort variable. We extend their approach with a continuous effort level. Hence the effectiveness of effort may also be a continuous function of contract time, a desirable property for empirical valida-

tion. We show that under fairly general conditions, a rational policyholder's safe driving effort will increase with the number of accumulated demerit points.

We suppose that the driving license is revoked when the driver reaches a total of N demerit points. For the sake of simplicity, each convicted traffic violation is linked to one supplementary demerit point in this section.[10] A driver with a suspended driving license is reinstated after a period D with a fresh zero-point record like that of a beginner.[11] The duration D may be fixed or random in the model. In Quebec, a licence suspension is of random length because drivers must pass a new exam after a given period before recovering their driving license. A rational driver maximizes his expected lifetime utility expressed in \$ and derived from:

- An instantaneous driving utility, d_u.
- A time-dependent disutility of effort, denoted as $e(t)$.[12] This effort level is linked with an instantaneous traffic violation frequency risk, denoted as $\lambda(e(t))$.[13] The hazard function λ is assumed to be a positive, decreasing and strictly convex function of the effort level.

In this section, we suppose that there is no point removal mechanism. In that case, the lifetime expected utility (we assume an infinite horizon) will depend only on the number n of accumulated demerit points. The Bellman equation on the expected utility leads to

$$u_n = \frac{d_u}{r} - \frac{\lambda_*(u_n - u_{n+1})}{r}, \ (0 \leq n < N), \tag{1}$$

where r is a discount rate, and where λ_* is defined as follows

$$\lambda_*(\Delta u) \underset{def}{=} \min_{e \geq 0} e + [\lambda(e) \times \Delta u]. \tag{2}$$

Technical details can be found in Appendix A.1. In equation (2), Δu is the lifetime utility loss between the current state and the one reached after an additional traffic offense. Once quantified, Δu is the monetary equivalent of this

traffic violation. Values of this type are derived in Section 5.2. The objective function minimized in (2) is the disutility flow of both effort (short term component) and the expected lifetime utility loss (long term). The function λ_* is the convex dual of the hazard function λ. All the u_n are lower than $u_{\max} = d_u/r$, the private lifetime driving utility without the point-record driving license. Equation (1) means that $\lambda_*(u_n - u_{n+1})/r$ is the minimal private utility cost of the point-record mechanism for a driver with n demerit points. In our models, the optimal effort level and its related effectiveness depend on the argument of λ_*. The cycle of lifetime utilities is closed with a link between u_0 and u_N, the lifetime expected utility just after the suspension of the driving license. For instance, if the private disutility of driving license suspension is only the loss of driving utility during a period D, we have that

$$u_N = \beta u_0,\ \beta = E[\exp(-rD)]. \tag{3}$$

The utilities are then derived from the recurrence equations (1) and (3). Optimal effort depends on the variation of lifetime utility as it minimizes the function defined in equation (2). Hence, in this setting, optimal effort will depend on the number n of accumulated demerit points but not on time, and we denote it as e_n. It is shown in Appendix A.1 that e_n increases with n for any given value of N. Then the related frequency of violations $\lambda_n = \lambda(e_n)$ decreases with n.

Fines represent another monetary incentive scheme applied in Quebec during the whole period investigated in this study. Let us denote $\overline{fa}$ as the average fine for a traffic violation conviction. Since fines and premiums are low in comparison to average wealth, we leave out risk aversion. With fines, incentives are effective if

$$e_n > 0 \Leftrightarrow u_n - u_{n+1} + \overline{fa} > \frac{-1}{\lambda'(0)}. \tag{4}$$

This means that the average fine is added to the utility loss in the argument of λ_*, which determines optimal effort. If fines are combined with the preceding

point-record driving license, the optimal effort still increases with n for a given value of the average fine.

3.3 Point-record driving licenses with point removal

In Quebec, each traffic violation is redeemed at the end of a two-year period. Integrating this feature in the optimal behavior model is difficult, as all the seniorities of non-redeemed driving offenses must be included as state variables in the dynamic programming equations. Lifetime utility is expected to increase with time for a given number of demerit points accumulated. Optimal effort depends on the difference between the present utility and a substitute utility (i.e., that reached after an additional traffic violation). With the point removal system in force in Quebec, the substitute utility increases with time as does the present utility. Time should have more value for worse situations, hence the substitute utility should increase faster than the present utility. Thus optimal effort should decrease with time. Besides, we prove in Appendix A.2 that optimal effort is continuous at the time of a point removal. This property will be tested empirically in Section 5.1. Optimal effort is then expected to increase with each traffic violation in order to compensate for the decreasing link between time and effort. On the whole, the incentive properties of the point removal system in force in Quebec are close to those of a mechanism without point removal when it comes to the number of demerit points accumulated.

3.4 Incentive effects of premiums indexed on demerit points: The example of Quebec

Table 1 presents the rating structure enforced for each driving license on the first contract birthday following December 1, 1992. The premium paid every two years after this date depends on the number of demerit points accumulated in the last two years. It does not represent the total premium for bodily injury

insurance but the additional premium related to demerit points. This average premium is equal to \$54.60, and complements a yearly driving license fee for insurance coverage equal to \$107.

Insert Table 1 about here

In this section, the incentive properties of this rating structure are analysed separately from the point-record driving license. An important input is the distribution of demerit points for a given driving offense, which we left out in Section 3.2. Denoting f_j as the proportion of traffic violations with j demerit points, we have the following values

$$f_1 = 4.71\%;\ f_2 = 52.32\%;\ f_3 = 38.34\%;\ f_4 = 2.83\%;\ f_5 = 1.80\%. \qquad (5)$$

Note that f_5 actually refers to offenses with five points and more. From Table 1, we see that the premium is a step function of the accumulated demerit points. Because of the local non-convexity of the premium, the incentives may not always increase with the number of demerit points accumulated. Let us consider for instance a policyholder just before her contract birthday. The incentive level will be stronger with two accumulated demerit points than with four. With four points, it is indeed less than likely that the next traffic offense will trigger an increase in premium. The corresponding probability is $2.83 + 1.80 = 4.63\%$, if we assume that the distribution of the f_j is independent of the accumulated demerit points. The incentives for safe driving are stronger at a two-point level because the probability of climbing a step in the rating structure after a traffic offense is close to one. The aforementioned result stands in contrast to the one obtained by Abbring et al (2003) for the French "bonus-malus" scheme and its exponential structure.

Let us design an optimal behavior model based on this rating structure. Once the premium is paid, the driver is reinstated with a fresh zero point record. Hence the optimal control model can be designed with the next contract birthday

as the horizon. Let π_n be the premium paid for a n point record. As we here discard the point-record driving license and its possible deprivation, the driving utility is no longer a parameter. On the other hand, we retain fines in the incentives for safe driving. We denote $v_n(t)$ as the optimal expected disutility of premiums and fines until the next contract birthday, where t is the seniority of the last birthday and n is the number of demerit points accumulated since that date. We have the terminal conditions

$$v_n(T) = \pi_n, \ \forall n = 0, \dots, N. \tag{6}$$

With the notations of Section 3.2, the optimal effort level depends on the argument of λ_*. If incentives are related to fines and insurance premiums, this argument is the sum of the average fine $\overline{fa}$ and of the expected variation of $v_n(t)$ after a traffic offense (see Appendix A.3). Hence optimal effort is determined by $\overline{fa} + \Delta v_n(t)$, with $\Delta v_n(t) = \left(\sum_{j \ / \ f_j>0} \ f_j \, v_{\min(n+j,N)}(t)\right) - v_n(t)$. Derivations show that the average of $\Delta v_n(t)$ with respect to n does not vary much with time. The terminal values of Δv_n are derived from equations (5), (6) and Table 1. For $n = 0, 2, 4$, we obtain

$$\Delta v_0(T) = \ \$2.32; \ \Delta v_2(T) = \ \$47.65; \ \Delta v_4(T) = \ \$3.43. \tag{7}$$

The incentives with two points accumulated are much stronger than with four points, which confirms the analysis following equation (5). The overall average of $\Delta v_n(t)$ with respect to t and n is close to \$12, a nine percent increase with respect to an \$130 average fine. In Section 5.2, this increase will be compared with the variation in traffic violation frequency before and after the reform.

4 Description and indentification issues on count data in insurance

Frequency risk models in insurance are addressed at length by the actuarial literature. Actuarial models use mixtures of Poisson models to describe the dynamics of the data. Their main limitation is that identification issues are not taken into account, since the observed dynamics are supposed to be created only by the revelation of unobserved heterogeneity. If random effects are time-independent, the predictor summarizing the individual history (the "bonus-malus" coefficient) decreases with risk exposure (bonus) and increases with the number of events (malus). Consider for instance a mixture of Poisson processes with a hazard function $\lambda_i(t)\varepsilon_i$ for policyholder i. The parameter $\lambda_i(t)$ depends on the observable individual information. The multiplicative random effect ε_i verifies $E(\varepsilon_i) = 1$; $V(\varepsilon_i) = \sigma^2$. Actuarial predictors are based on expectations of the type $E(\varepsilon_i \mid N_{i,t})$, where $N_{i,t}$ is the number of insurance claims made by policyholder i between 0 and t. These conditional expectations can be derived in a parametric setting, for instance if ε_i follows a Gamma distribution (Dionne, Vanasse, 1989). Semiparametric derivations with a linearity constraint on the shape of the predictor can also be retained. This is known as the "linear credibility" approach (Bühlmann, 1967). The two approaches lead to the same "bonus-malus" coefficient

$$\widehat{E}(\varepsilon_i \mid N_{i,t}) = \frac{1 + (\widehat{\sigma^2} \times N_{i,t})}{1 + (\widehat{\sigma^2} \times \widehat{\Lambda}_{i,t})},\ \Lambda_{i,t} = E(N_{i,t}) = \int_0^t \lambda_i(s)ds. \tag{8}$$

This formula reflects the continuous time-effect of the revelation of unobserved heterogeneity on the one hand. On the other hand, there is a jump of the predictor at each event occurrence. The estimated hazard function which integrates experience rating is then equal to $\widehat{\lambda}_i(t) \times \widehat{E}(\varepsilon_i \mid N_{i,t})$.

Disentangling incentive effects from unobserved heterogeneity is an identification issue. The basic strategy is to obtain statistics which are invariant with

respect to the mixing distribution related to hidden features in the risk distribution. Abbring et al (2003) provide an inference strategy when the hazard function is multiplied by a constant β after each event (accident for instance) and does not vary with time. Assessing the existence of moral hazard amounts to estimating β and testing for $\beta < 1$ if the marginal benefit of effort increases with the number of claims. Time effects do however exist in the point-record mechanism in force in Quebec, so we cannot apply this approach here. In Section 3.3, we showed that safe driving effort induced by the point-record driving license increases with the number of demerit points and decreases with time if this number is greater than zero. The induced duration-event effects on traffic violation risk are opposed to those created by the revelation of unobserved heterogeneity. In non-life insurance, empirical hazard functions related to frequency risks usually increase with claims and decrease with time. This justifies the "bonus-malus" systems and means that incentive effects do not outweigh the revelation effect on this type of data.

5 Empirical results on the incentive effects of point-record mechanisms

5.1 Point-record driving license

In this section, we analyze the data before the 1992 reform which introduced the experience rating scheme based on demerit points. Thus the point-record driving license interacts only with fines. Regressions are performed from January 1985 (we need a two-year history to derive the accumulated demerit points) to December 1992, date of the reform enforcement. We try to obtain a confirmation of the theoretical findings of Sections 3.2 and 3.3 (i.e., the effort level increases globally with the number of demerit points accumulated and decreases with the seniority of non redeemed traffic violations, if any), and to confirm the presence

of moral hazard in the data.

The whole history of traffic violations is useful in assessing the revelation of unobserved heterogeneity, whereas the last two years are enough to determine the incentive level. We use a two-step estimation strategy. First, we derive an actuarial predictor which is updated every month, and include it as a constant (an "offset" variable) in the hazard functions of convicted traffic offenses and accidents. Second, a proportional hazards model (Cox, 1972) is used to estimate these hazard functions. We retained the following specification

$$\lambda_i^j(t) = \exp(x_i(t)\beta_j) \times g_j(cdp_i(t)) \times k_j(nsps_i(t)) \times BM_i^j(m(t)) \times h_j^{S_i(t)}(c_i(t)). \quad (9)$$

In equation (9), $\lambda_i^j(t)$ is the hazard function of type j ($j = 1$: traffic violation or $j = 2$: accident) for driver i at calendar time t. Regression components which do not refer to the individual driver record are denoted by the line-vector $x_i(t)$. We retained the gender, driving license class, place of residence, age of the driver and calendar effects related to years and months.[14] The number of demerit points accumulated in the last two years is denoted as $cdp_i(t)$, and a decreasing shape is expected for g_1 from the theoretical model of Sections 3.2 and 3.3. The variable $nsps_i(t)$ is the number of past driving license suspension spells. The link with traffic violation risk should be decreasing if such a spell increases the perceived driving utility or the perceived risk level. The actuarial predictor is denoted as $BM_i^j(m(t))$ and is updated each month, with $m(t)$ the month related to t. Its estimation is discussed later on.

Effort is expected to decrease with time only if the number of demerit points accumulated is greater than zero. Hence we specified a stratified proportional hazards model.[15] The baseline hazard functions $h_j^{S_i(t)}$ depend on the risk type j and on the stratum $S_i(t)$. There are two strata, depending on whether the variable $cdp_i(t)$ is equal to zero or not. Lastly, contract time $c_i(t)$ is integrated into the baseline hazard function h_j. The function c_i is set equal to zero at the beginning of the whole period. Then it is reset to zero at each event which trig-

gers a variation of the accumulated demerit points (i.e., traffic violation or point removal). This event-driven operation should eliminate interactions between calendar and contract-time effects for the stratum associated to $cdp > 0$.

In equation (9), the actuarial predictor BM_i^j is assumed to reflect the revelation of unobserved heterogeneity rather than incentives, whereas the functions $g_j(cdp)$ and h_j^S are first related to the event and time effects of incentives. We do not, however, pretend to disentangle exactly the revelation of unobserved heterogeneity from the incentive effects with this specification because the actuarial predictor is calibrated on the observed dynamics, which result from both effects. From the theoretical model in Section 3.3, we expect effort to increase with the number of demerit points accumulated, under moral hazard. This is globally true in Table 2, where the function g_1 decreases beyond seven points. It is worth mentioning that the SAAQ warns the policyholders when their accumulated demerit points reach a seven point threshold.[16] For robustness, another empirical result (available upon request) is consistent with an effective effort beyond seven demerit points. Indeed, when equation (9) is estimated on traffic violations without the actuarial predictor included as an offset variable, the estimated function g_1 increases from one to seven points and then decreases. As the revelation effect of a traffic violation is always positive, this result can be explained only by an opposite incentive effect of the point-record driving license beyond seven points.[17] As a consequence, the actual revelation effect of the traffic violation record should be stronger than that given by the actuarial predictor, calibrated on the observed dynamics. To compensate for this bias, the actual incentive effect of accumulated demerit points should be stronger than its estimation in Table 2.

Insert Table 2 about here

The number of past driving license suspensions generates interesting results in the traffic violation equation. One suspension spell entails a 5.6% reduction in traffic violation frequency, and two bring about a 13.1% reduction. One

possible explanation is that the perceived driving utility of drivers increases after a driving licence suspension spell. Another interpretation points to an availability effect (Tversky, Kahneman, 1973), where the subjective estimation of the frequency of an event is based on how easily a related outcome can be brought to mind.[18] Table 2 indicates that the number of demerit points accumulated has less influence on accident risk than on traffic violation risk. A possible interpretation is that we cannot separate at fault from no-fault accidents. In the literature, the incentive effect is usually higher with at fault accidents. Besides, drivers nearing the license suspension threshold might also apply opportunistic strategies regarding traffic violations (e.g. paying more attention to radars) without otherwise modifying their attitude towards road traffic risk. As with traffic violations, past driving license suspension spells reduce accident risk.

Traffic violations with a seniority greater than two years are redeemed and do not have an incentive effect on the drivers. In order to use this result when disentangling unobserved heterogeneity from moral hazard in the duration dimension, we estimated a dynamic random effects specification. Hence the seniority of past traffic violations will be taken into account by the actuarial predictor. Let us denote $N_{i,y}^j$ the number of type j events observed for driver i and year y. The parameter of the related Poisson distribution is $\lambda_{i,y}^j \, \varepsilon_{i,y}^j$, where the first component depends on the observable information and the second one is the dynamic random effect reflecting the residual effect of hidden information. Random effects are supposed stationary, with an expectation equal to one and i.i.d. between the individuals. If $\widehat{\lambda}_{i,y}^j$ is the estimation in the Poisson model without random effects[19], a consistent estimation of the covariances between the random effects is the following

$$\widehat{Cov}(\varepsilon_{i_0,y_0}^j, \, \varepsilon_{i_0,y_0-h}^1) = \frac{\sum_{i,y}(N_{i,y}^j - \widehat{\lambda}_{i,y}^j)(N_{i,y-h}^1 - \widehat{\lambda}_{i,y-h}^1)}{\sum_{i,y} \widehat{\lambda}_{i,y}^j \, \widehat{\lambda}_{i,y-h}^1} \text{ (for } h > 0 \text{ or } j \neq 1). \tag{10}$$

The sums are derived on all the possible couples i, y from the sample and the

integer lag h is assumed to be lower than the maximum length of individual histories. The variance (case $h = 0,\ j = 1$) is estimated from an overdispersion of residuals (see Appendix A.4). The covariance estimated in (10) clearly reflects the predictive ability of a traffic violation on the frequency risk of type j assessed h years after.

Insert Table 3 about here

Table 3 exhibits a decreasing shape for both covariances series. This means that the predictive ability of past traffic violations on both risk types decreases with the seniority. If the seniority is larger than two years, this result stems only from a revelation effect since the traffic violation has been redeemed. The bonus-malus coefficient $BM_i^j(m)$ given in equation (9) is obtained from an affine probabilistic regression of a multiplicative random effect $\varepsilon^j_{i,y(m)}$ (related to driver i, month m (related to the year $y(m)$) and type j event) on the number of traffic violations recorded for the driver for each past month. More details are given in Appendix A.4. As the second goal of this section is to estimate the time effects of incentives from the observed dynamics, we need results on the behavior of actuarial coefficients. Their time decay is usually lower for a claimless history than with the basic actuarial approach of (8), but stronger since the last claim, if any. In this case indeed, the continuous aging of past events supplements the increase in risk exposure (see Pinquet, Guillén, Bolancé (2001)).

Insert Figure 2 about here

Two baseline hazard functions on traffic violation risk are presented in Figure 2. They are estimated on the stratum with $cdp > 0$ (resp. with $cdp = 0$), from the ratio between the number of traffic violations and risk exposure, expressed in equation (9). Contract-time is less than two years as it represents the time elapsed since the last variation in accumulated demerit points. The frequency of traffic violations decreases by 57% (resp. 43%) during the two years

on the stratum with $cdp > 0$ (resp. with $cdp = 0$). This decrease is explained first by the actuarial coefficient. For the stratum with $cdp > 0$, the average value of the actuarial predictor decreases from 1.93 to 1.15, whereas the corresponding average varies from 1.10 to 0.85 for the other stratum[20]. As explained earlier, the sharper decrease of the predictor in the presence of recent past traffic violations ($cdp > 0$) is due to their continuous aging, which supplements the increase in risk exposure. The hazard functions are globally stationary. We do not obtain the increasing property expected from the theory for the stratum $cdp > 0$. A possible explanation is that the revelation effect of past traffic violations regarding frequency risk is underestimated in Table 3 when the lag is less than two years, as this effect is counteracted by the incentives. If the covariances in Table 3 were derived solely from unobserved heterogeneity, they would be greater for lags less than two years, and this would increase the ageing effect of past claims in the actuarial predictor. As a result, the decrease of the actuarial predictor would be sharper on the stratum with $cdp > 0$, and the residual baseline hazard would be globally increasing.

Let us test another prediction of the theoretical model, which is the continuity of effort at the time of a point removal. If a traffic violation is followed by a two year violation free record, the baseline hazard function increases from 0.967 (the terminal value of the baseline hazard function related to $cdp > 0$) to 0.977, which corresponds to the initial value of the other hazard function. On the other hand, the actuarial predictor is continuous at the time of a point removal. Hence the continuity property of effort is almost fulfilled.

5.2 Incentive effects of the 1992 reform and monetary equivalents for traffic violations and license suspensions

In Section 2, we mentioned a 12.5% decrease in the average frequency of traffic violations before and after the reform which introduced the experience rating structure based on demerit points. This result is slightly modified if we control with the regression components used in Table 2. A regression estimated from 1985 to 1996 with the covariates of Section 5.1 and a dummy related to the period following December 1, 1992 associates the reform with a 15% decrease.[21]

The results of Section 3.4 (for instance equation (7)) suggest that the number of demerit points accumulated since the last birthday should influence the effectiveness of the reform. However, we did not obtain significant results in this direction. The drivers' limited knowledge of the environment could explain this poor results, a point developed later.

Let us make an overall comparison of the three incentive schemes. We use the model without point removal of Section 3.2 to analyze the incentives for drivers in Quebec. Before the 1992 reform, fines were supplemented by a point-record driving license. Optimal effort after n non redeemed traffic violations depends on the argument of λ_*, which is the dual of the hazard function λ (see equation (2)). This argument is equal to $\overline{fa} + u_n - u_{n+1}$ from equation (4). We recall that the average fine $\overline{fa}$ is equal to \$130. Besides, the 1992 reform entails an average increase in the argument of λ_* equal to \$12 from Section 3.4.[22] At this point, it seems interesting to relate the optimal risk level and the argument of λ_*, which determines the incentive level. This relation can be assessed from the elasticity between optimal frequency risk and the argument of λ_*. When the incentives are effective, it can be shown that this elasticity is less than -1 if and only if $\log(\lambda)$ is a concave function of effort (elasticity and concavity are considered locally: see Appendix A.5 for a proof). A global elasticity equal to

-1 is linked to an exponential decay of λ. With $\lambda(e) = \lambda(0) \times \exp(-\alpha e)$, the optimal risk level as a function of Δu (the argument of λ_*) is equal to $1/(\alpha \Delta u)$ if the incentives are effective.

Let us assess the global effectiveness of the 1992 reform. As the reform entailed a significant reduction in traffic violation risk no matter the number of demerit points accumulated, we can assume that incentives are effective for a representative driver.[23] Effective incentives lead us to analyse the elasticity between traffic violation risk and the argument of λ_*. Suppose that we leave out the modifications of lifetime utility variations due to the aggregation of incentive schemes. Then we can relate:

- On the one hand, a 15% reduction in the frequency of traffic violations after the 1992 reform.

- On the other hand, a relative increase in the argument of λ_* lying between 9% and 10%. Indeed, the 1992 reform entails a \$12 average increase in the argument of λ_*. This increase supplements the other arguments of λ_*, i.e., the \$130 average fine and the utility variation for the point-record driving license. In Table 2, the point-record driving license offers significant incentives to careful driving beyond a seven point threshold, a result corresponding to only a minority of drivers (1.4%). The contribution of the point-record driving license to the argument of λ_* is low as compared with fines.

This suggests that the elasticity between the optimal frequency risk and the argument of λ_* is less than -1 in this case. This result is linked to a locally concave shape of $\log(\lambda)$ for the representative driver. However external effects could also explain the reduction in the frequency of traffic violations. We cannot eliminate these effects because there is no control group that is not affected by the reform. Besides, the elasticity would be modified if the distribution of demerit points for

a given driving offense (see equation (5)) was wrongly perceived by the drivers. The \$12 average of the increase in the argument of λ_* is low, due to the high frequency of drivers without demerit points since the last birthday (87%), and to the low incentive level of the reform for these drivers (see equation (7)). This incentive level mostly depends on the probability of moving up a step in the premium schedule after an additional traffic violation, which must be associated to four demerit points or more. If the perceived frequency of corresponding traffic violations was greater than the actual one (i.e., $f_4 + f_5 = 2.83 + 1.80 = 4.63\%$), the incentive level would increase as well as the variation of the argument of λ_* induced by the reform. In that case the elasticity would be closer to zero.

Lastly, let us assess monetary equivalents for a traffic violation and a license suspension. The monetary equivalent of a traffic violation for a driver is the loss of lifetime utility, which depends on the number of traffic violations accumulated. A value can be derived from the effectiveness of effort estimated in Table 2 and from the aforementioned link between efficiency of effort and the incentive level. An additional traffic violation beyond seven accumulated demerit points entails a reduction of traffic violation frequency close to twenty percent. Although these drivers cannot be seen as representative, we will apply the elasticity derived from the preceding developments. If a 9% increase in the argument of λ_* entails a 15% reduction in the frequency of traffic violations, a 20% decrease of traffic violation frequency is associated with a 12% increase in the argument of λ_*. The implied loss of lifetime utility depends on the traffic violation frequency risk λ but mostly on the discount rate r (see Appendix A.6). With $\lambda = 0.15$, the monetary equivalent of an additional traffic violation for these drivers belongs to the interval [\$120, \$195] if $r = 3\%$, and to [\$41.1, \$55.7] if $r = 6\%$. Besides, the growth rate of this monetary equivalent with respect to the number n of non redeemed traffic violations falls between r/λ_n and r/λ_{n+1}, where λ_n is the optimal traffic violation frequency related to n. Monetary costs for license suspensions are then obtained by adding the costs of traffic violations until the crossing of

the demerit point threshold. Starting from a zero-point record and assuming that six traffic violations are needed to entail a license suspension, the monetary cost of a license suspension is bounded by \$700 and \$1178 if $r = 3\%$, and if $\lambda_0 = 0.17$; $\lambda_1 = 0.17$; $\lambda_2 = 0.16$; $\lambda_3 = 0.15$; $\lambda_4 = 0.12$; $\lambda_5 = 0.09$.[24] Besides, a misperception of the environment could modify the monetary equivalents of traffic violations and license suspensions. We argued that an overestimation of the frequency of severe traffic violations would increase the perceived incentive effect of the 1992 reform. In that case, the results obtained for monetary equivalents should be upgraded.

6 Conclusion

In this article, we analyse the properties of policies designed to promote safe driving. Three important incentive mechanisms for road safety are used in Quebec. The incentive effects of the point-record driving license increase with the number of demerit points accumulated. This confirms the presence of moral hazard in the data. The point-record driving license acts as an incapacitating device for reckless drivers. Also, past suspension spells entail a significant reduction in the frequency of traffic violations and accidents.

Fines are on average the most efficient device, but the absence of memory entails a uniform incentive effect for given characteristics of the policyholder. We designed our incentive models with a representative driver, but there is of course heterogeneity in the individual parameters, such as the threshold beyond which the incentives are effective. We did not have wealth variables at hand, and an interesting empirical issue would have been to cross such variables with a reform dummy in risk assessment.

The experience rated premium based on accumulated demerit points is a monetary point-record mechanism. The empirical results exhibit a rather uniform effectiveness after its enforcement in 1992, i.e., a 15% decrease in the

frequency of traffic violations. Its incentive effects do not strictly increase with the accumulated demerit points, however, because of the steps in the rating structure. The actual incentive effect of the reform looks more like that induced by an increase in the average fine. The SAAQ modified its rating policy in 2008, with a premium increase from the first demerit point. This should enhance the effectiveness of the premium schedule for the majority of drivers with a violation-free record.

In this study we have not examined in detail the long-term evolution of accidents, for we did not have access to the control variables of interest. Over recent years, many road safety initiatives had an impact on accidents but did not necessarily have any effect on violations. These initiatives include such measures as: occasional campaigns on fatal accidents; increased police patrols to reduce speeding; and designated driver campaigns to prevent drinkers from driving. It is also worth noting that the decline in deaths and serious injuries can be explained by vehicular improvements and the wearing of seat belts. Such measures are complementary to those studied in this article.

Notes

[1]These clauses and their incentive properties are detailed in Section 3.

[2]The North American continent preceded Europe in the design of such systems. Point-record driving licenses were introduced in 1947 in the USA. Germany, France, and Spain implemented these mechanisms in 1974, 1992 and 2005, respectively.

[3]Selecting at random one percent of the new licence holders every year would of course have been a preferable sampling procedure. One thousand new licence holders would then have been selected every year, as the entry rate in the SAAQ portfolio is close to 2.5%.

[4]Binary variables related to traffic offenses and attrition were created on a monthly basis, and explained with the covariates used in Section 5.1. The score test statistic is equal to 0.34. Hence we do not reject the nullity of the correlation coefficient at the usual significance levels.

[5]We begin in 1985 in order to match the regressions which follow, as a two-year history is needed to derive the accumulated demerit points. Data are first averaged over one year, to

account for strong seasonal effects. A centered moving average derived on five fortnights is then performed twice in order to reduce the volatility of the series.

[6]We thank a referee for suggesting this interpretation.

[7]Drivers with a contract birthday falling between the announcement of the reform and its enforcement are not incited by the experience rated premium before this birthday. Incentives exist otherwise (for these drivers after the birthday, and for all the other drivers). A referee suggested using this natural experiment in order to disentangle the incentive effects of the reform from calendar effects. We did not obtain significant results. Four months is however a short period, and on average only one driver out of twelve was not incited by the rating scheme during the period.

[8]See Manning et al (1987) for the use of a control group in the assessment of a cost sharing modification in the health insurance market.

[9]Important variables in the regressions such as the number of accumulated demerit points have low frequencies for the highest values. It is hard to make an accurate estimation if the frequency of events is low, as it is the case for accidents with bodily injury.

[10]All the traffic offenses recorded in the data base are linked to convictions, which is the condition for the addition of demerit points.

[11]This reinstatement can be seen as a removal of demerit points. In the paper, we consider a point removal mechanism to be a cancellation of demerit points applied before the suspension of the driving license.

[12]Safe driving effort can also reduce the expected disutility of accidents. If $e \rightarrow \delta(e)$ is the implied decrease in the disutility flow, replacing e by $e - \delta(e)$ in the model includes the influence of safe driving effort on accident disutility.

[13]The hazard function $\lambda(e)$ corresponds to a probability $p(e)$ in discrete time incentive models.

[14]Comprehensive regressions based on two-year periods can be found in Dionne, Maurice, Pinquet, and Vanasse (2001).

[15]Stratification in a proportional hazards model means that Cox likelihoods (of a multinomial logit type) are derived for each stratum and then multiplied together. In other words, an individual with an observed event is assumed to have competed only with other individuals in the same stratum and at risk at the same date. However, the same coefficients for the covariates are used across all strata.

[16]On the other hand, the drivers are not informed when offenses are redeemed.

[17]The license suspension threshold increased in January 1990 from twelve to fifteen points. We tested the effect of this adjustment and did not obtain significant differences in the results.

[18]We thank a referee for suggesting this interpretation.

[19]Data were observed from 1983 to 1992, and we retained the covariates denoted as x in equation (9). The regression is performed on a monthly basis because risk exposure is updated monthly in the derivation of actuarial coefficients.

[20]The two regression components which determine the greatest variations in the additive score ($x\beta$ in equation (9)) during the two years are age (-16% and -8% on the two strata) and gender (-12% and -7%).

[21]We retained the covariates used in Table 3, except for dummies related to years and the number of past license suspension spells. The estimated additive parameter for the reform dummy is equal to -0.163, and the related standard deviation is equal to 0.008. Hence the reform effect is conclusive at the usual tests significance levels.

[22]In Section 3.4, we derived expected disutilities $v_n(t)$ until the next contract birthday. They can be associated with a negative lifetime utility $u_n(t)$. We have

$$u_n(t) = -v_n(t) + \exp(-r(T-t))\, u_0(0); \quad u_0(0) = \frac{-v_0(0)}{1-\exp(-rT)}.$$

From the preceding equation, we have $u_n(t) - u_{n+1}(t) = v_{n+1}(t) - v_n(t)$. The average increase in disutility after a traffic offense is equal to the corresponding decrease in lifetime utility.

[23]From equation (4), a sufficient condition to have this result is that the average fine is higher than the threshold $-1/\lambda^{'}(0)$ beyond which the incentives are effective.

[24]These values comply with the relative risks estimated in Table 2.

References

[1] Abbring, Jaap, Pierre-André Chiappori, and Jean Pinquet, "Moral Hazard and Dynamic Insurance Data," *Journal of the European Economic Association* 1 (2003), 767-820.

[2] Boyer, Marcel and Georges Dionne, "The Economics of Road Safety," *Transportation Research*, 21B (5) (1987), 413-431.

[3] Boyer, Marcel and Georges Dionne, "An Empirical Analysis of Moral Hazard and Experience Rating," *The Review of Economics and Statistics* 71 (1989), 128-134.

[4] Bourgeon, Jean-Marc and Pierre Picard, "Point-Record Driving License and Road Safety: An Economic Approach," *Journal of Public Economics* 91 (2007), 235-258.

[5] Bühlmann, Hans, "Experience Rating and Credibility," *ASTIN Bulletin* 4, (1967), 199-207.

[6] Chiappori, Pierre-André and Bernard Salanié, "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy* 108 (2000), 56-78.

[7] Cox, David R., "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, Series B, 34 (1972), 187-220.

[8] Dionne, Georges and Charles Vanasse, "A Generalization of Automobile Insurance Rating Models: The Negative Binomial Distribution with a Regression Component," *ASTIN Bulletin* 19 (1989), 199-212.

[9] Dionne, Georges, Christian Gouriéroux, and Charles Vanasse, "Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment," *Journal of Political Economy* 109 (2001), 444-453.

[10] Dionne, Georges, Mathieu Maurice, Jean Pinquet, and Charles Vanasse, "The Role of Memory in Long-Term Contracting with Moral Hazard: Empirical Evidence in Automobile Insurance," Working paper 01-05, HEC Montréal (2001). URL: http://neumann.hec.ca/gestiondesrisques/01-05.pdf

[11] Doyle, Joseph J., "Health Insurance, Treatment and Outcomes: Using Auto Accidents as Health Shocks," *The Review of Economics and Statistics* 87, 2 (2005), 256-270.

[12] Landes, Elisabeth M. "Insurance, Liability, and Accidents: A Theoretical and Empirical Investigation of the Effect of No-fault Accidents," *Journal of Law and Economics* 25 (1982), 49-65.

[13] Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, Arleen Leibowitz, and Susan Marquis, "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review* 77 (1987), 251-277.

[14] Murray, Christopher J.L., and Alan D. Lopez, "Alternative Projections of Mortality and Disability by Cause 1990-2020: Global Burden of Disease Study," *The Lancet* 349 (1997), 1498-1504.

[15] OECD International Road Traffic and Accident Data Base (2005). URL: http://cemt.org/IRTAD/.

[16] Peltzman, Sam, "The Effects of Automobile Regulation," *Journal of Political Economy* 83 (1975), 677-725.

[17] Pinquet, Jean, Montserrat Guillén, and Catalina Bolancé, "Allowance for the Age of Claims in Bonus-Malus Systems," *ASTIN Bulletin* 2, 31 (2001), 337-348. URL: http://www.casact.org/library/astin/.

[18] Shavell, Steven, "The Optimal Use of Nonmonetary Sanctions as a Deterrent," *American Economic Review* 77 (1987), 584- 592.

[19] Sloan, Frank A., Bridget A. Reilly, and Christoph Schenzler, "Effects of Tort Liability and Insurance on Heavy Drinking and Driving," *Journal of Law and Economics*, 38 (1995), 49-77.

[20] Tversky, Amos and Daniel Kahneman, "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology*, 5 (1973), 207-232.

[21] Viscusi, Kip W., "The Value of Risks to Life and Health," *Journal of Economic Literature*, 31 (1993), 1912-1946.

TABLE 1:—SAAQ INSURANCE PREMIUMS FOR BODILY INJURY AS A FUNCTION OF ACCUMULATED DEMERIT POINTS SINCE THE LAST CONTRACT BIRTHDAY

Accumulated demerit points (last two years)	Premium for the next two years (Canadian $)	Frequency (%)
0,1,2,3	50	93.7
4,5,6,7	100	4.9
8,9,10,11	174	1.1
12,13,14	286	0.2
15 and more	398	0.1

TABLE 2.—ESTIMATION OF THE HAZARD FUNCTION FOR TRAFFIC VIOLATION AND ACCIDENT FREQUENCY RISKS

Variable	Level	Frequency (%)	Traffic violation risk	Accident risk
nsps: Number of past driving license suspension spells	0 (*)	98.96	0	0
	1	0.94	-0.058	-0.064
			(0.022)	(0.046)
	2	0.09	-0.140	-0.519
			(0.062)	(0.168)
	3 and more	0.01	-0.091	-0.147
			(0.156)	(0.410)
cdp: Number of demerit points accumulated (last two years)	0 point	76.60	stratum	stratum
	1 point (*)	0.39	0	0
	2 points	9.36	0.100	0.073
			(0.060)	(0.107)
	3 points	6.23	0.119	0.192
			(0.061)	(0.107)
	4 points	1.92	0.124	0.065
			(0.062)	(0.111)
	5 points	2.09	0.155	0.121
			(0.062)	(0.110)
	6 points	1.25	0.104	0.120
			(0.063)	(0.113)
	7 points	0.72	0.102	0.005
			(0.065)	(0.118)
	8 points	0.55	-0.032	0.101
			(0.067)	(0.120)
	9 points	0.43	-0.133	0.152
			(0.071)	(0.125)
	10 points	0.32	-0.184	0.084
			(0.072)	(0.127)
	11 points	0.06	-0.051	-0.192
			(0.104)	(0.223)
	12 points	0.04	-0.625	0.087
			(0.147)	(0.230)
	13-14 points	0.04	-0.283	-0.347
			(0.120)	(0.265)

(*): Reference level. Additive coefficients and level frequencies are weighted by duration. Standard errors are in parentheses. Additional regression variables are: gender, driving license class (9 levels), place of residence (16 levels), age of the driver (5 slopes) as well as calendar effects related to years (8 levels) and months (12 levels).

Number of observations: 3,587,654 duration-events of at most one month, derived from 41,290 driving licenses.
Global test for the nullity of coefficients (traffic violations): likelihood ratio statistic = 19416.71.; degrees of freedom = 62; limit significance level < 0.0001.
Global test for the nullity of coefficients (accidents): likelihood ratio statistic = 4464.91; degrees of freedom = 62; limit significance level < 0.0001.

TABLE 3.—ESTIMATED MOMENTS OF RANDOM EFFECTS USED IN THE PREDICTION

	$\widehat{Cov}\left(\varepsilon_{i,y}^{1}, \varepsilon_{i,y-h}^{1}\right)$	$\widehat{Cov}\left(\varepsilon_{i,y}^{2}, \varepsilon_{i,y-h}^{1}\right)$
$h = 0$	0.981	0.636
$h = 1$	0.800	0.482
$h = 2$	0.745	0.368
$h = 3$	0.731	0.336
$h = 4$	0.704	0.344
$h = 5$	0.705	0.293
$h = 6$	0.648	0.289
$h = 7$	0.673	0.288
$h = 8$	0.636	0.342
$h = 9$	0.608	0.296

$\varepsilon_{i,y}^{j}$: Multiplicative random effect for driver *i*, in period *y*, and risk of type *j*.

Figure 1: Relative frequencies (in percentage) for traffic violations and accidents

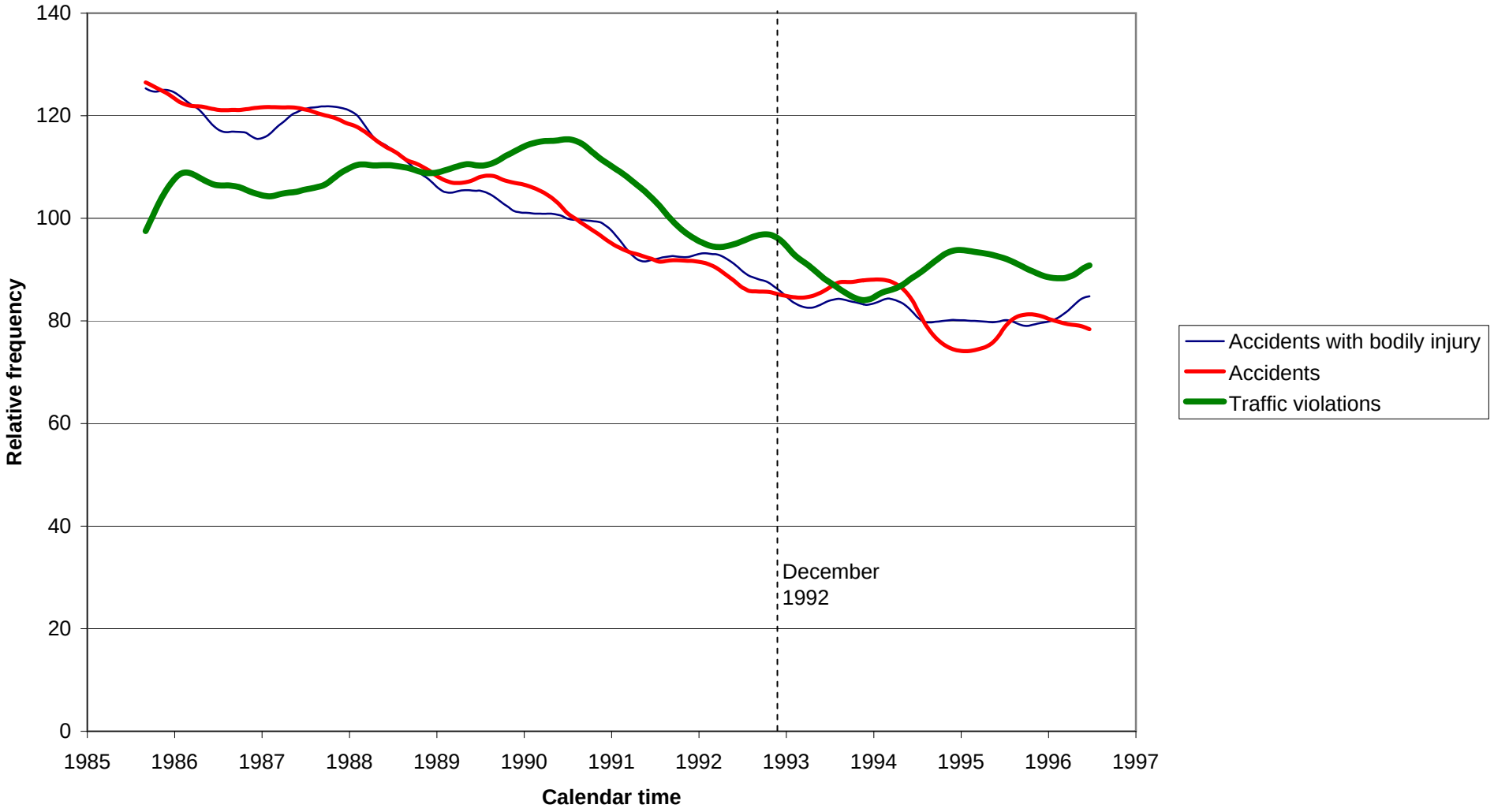
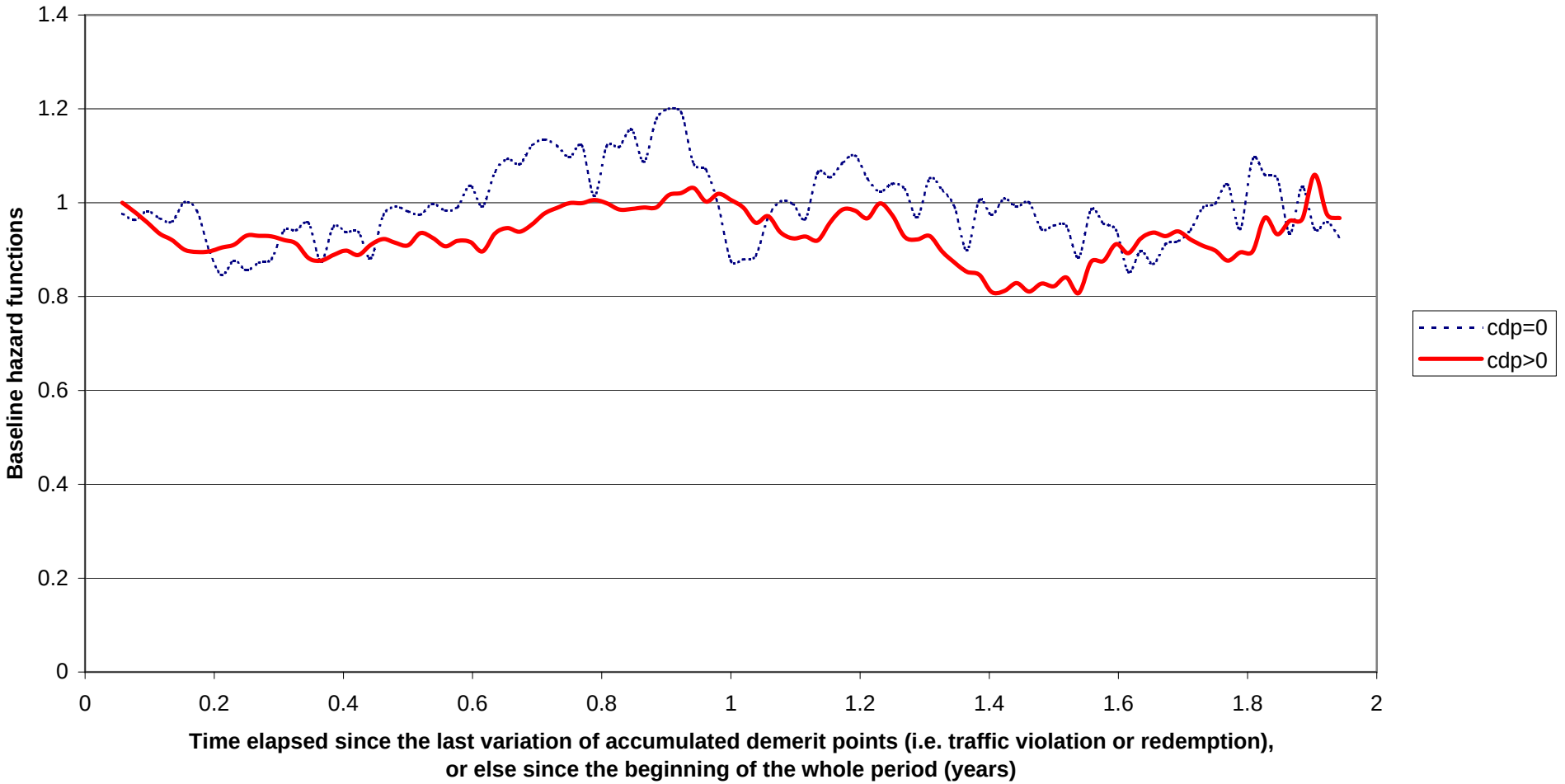


Figure 2: Baseline hazard functions for traffic violation risk, population stratified by accumulated demerit points (cdp) in the last two years (strata cdp=0 and cdp>0)
One month moving average, vertical unit=initial value for the stratum defined by cdp>0



A Appendix

A.1 Incentive effects of point-record driving licenses: Model without point removal

The Bellman equation on the expected utility is

$$u_n = \max_{e\geq 0} (d_u - e)dt + (\exp(-rdt) \times [((1-\lambda(e)dt) \times u_n) + (\lambda(e)dt \times u_{n+1}) + o(dt)]) .$$

Then we obtain

$$0 = \max_{e\geq 0} (d_u - e) - (r + \lambda(e))u_n - r\, u_{n+1},$$

and equation (1).

Let us give the main properties of the function

$$\lambda_* : \Delta u \rightarrow \min_{e\geq 0} e + [\lambda(e) \times \Delta u] = \min_{e\geq 0} h(\Delta u, e) ,$$

with λ a positive, decreasing and strictly convex hazard function. The related optimal effort level is equal to

$$e_{opt}(\Delta u) = \arg\min_{e\geq 0} h(\Delta u, e) \Rightarrow$$

$$e_{opt}(\Delta u) = 0 \text{ if } \Delta u \leq \frac{-1}{\lambda^{'}(0)};\ e_{opt}(\Delta u) = \left(\lambda^{'}\right)^{-1}\left(\frac{-1}{\Delta u}\right) \text{ if } \Delta u \geq \frac{-1}{\lambda^{'}(0)}. \quad (11)$$

Hence the dual function λ_* is defined on the real line as the optimal effort. From the last equation, we obtain

$$\Delta u \leq \frac{-1}{\lambda^{'}(0)} \Rightarrow \lambda_*(\Delta u) = \lambda(0) \times \Delta u, \quad (12)$$

and the dual function is linear in the neighborhood of 0, which corresponds to no effort. The dual function λ_* is strictly increasing since λ is strictly positive. If $\Delta u \geq 0$, we have that:

$$\lambda_*(\Delta u) = h(\Delta u, e_{opt}(\Delta u)) \geq e_{opt}(\Delta u) \Rightarrow \lim_{\Delta u \rightarrow +\infty} \lambda_*(\Delta u) \geq \lim_{\Delta u \rightarrow +\infty} e_{opt}(\Delta u) = +\infty.$$

Hence λ_* is an increasing homeomorphism on the real line.

The dual of a convex function is concave. This can be proved with geometrical arguments (see Rockafellar (1996)), or by the envelope theorem. We have

$$h'_{\Delta u}(\Delta u, e) = \lambda(e) \Rightarrow \lambda'_*(\Delta u) = h'_{\Delta u}(\Delta u, e_{opt}(\Delta u)) = \lambda(e_{opt}(\Delta u)). \tag{13}$$

Hence λ_* is concave from the assumptions on λ and from the properties of e_{opt}.

We give a proof of the increasing property of the optimal effort level as a function of accumulated demerit points. From equation (1), we obtain

$$u_n - u_{n+1} = \lambda_*^{-1}\left(r(u_{\max} - u_n)\right),\ u_{\max} = \frac{d_u}{r}\ (0 \leq n < N). \tag{14}$$

The sequence $(u_n)_{0\leq n\leq N}$ is decreasing since we have $u_{\max} \geq u_n$. Plugging this result into equation (14) implies that the sequence $(u_n - u_{n+1})_{0\leq n<N}$ is increasing. The optimal effort level is denoted as e_n, and expressed as

$$e_n = \arg\ \min_{e\geq 0}\ e + [\lambda(e) \times (u_n - u_{n+1})] = e_{opt}(u_n - u_{n+1}),$$

for $0 \leq n < N$, where e_{opt} is defined from (11). As e_{opt} is an increasing function, the optimal effort is an increasing function of the number of demerit points for any given value of the license suspension threshold.

Let us specify the condition under which incentives are effective. From (14) and (12), we obtain

$$e_n > 0 \Leftrightarrow u_n - u_{n+1} = \frac{d_u - ru_n}{\lambda(0)} > \frac{-1}{\lambda'(0)} = \underline{\Delta u}. \tag{15}$$

If fines are included in the incentives, u_{n+1} is replaced by $u_{n+1} - \overline{fa}$ in equation (1), which leads to the recurrence equation (see Figure 3)

$$d_u - ru_n = \lambda_*(u_n - u_{n+1} + \overline{fa})$$

$$\Leftrightarrow u_{n+1} = u_n + \overline{fa} - \lambda_*^{-1}(d_u - ru_n) = g(u_n). \tag{16}$$

The fixed point of g is the lifetime driving utility if fines were the only incentive scheme, i.e.

$$\widetilde{u}_{\max} = \frac{d_u - \lambda_*(\overline{fa})}{r}.$$

We of course assume that $d_u > \lambda_*(\overline{fa})$, i.e. $\widetilde{u}_{\max} > 0$. If the two incentives are mixed, we have $u_n \leq \widetilde{u}_{\max}$ and we deduce from (16) the properties of utilities and of optimal effort levels as functions of n that we obtained in the first place. Besides, we have

$$e_n > 0 \ , \forall n, \ \Leftrightarrow \overline{fa} + u_n - u_{n+1} > \underline{\Delta u} \ , \forall n.$$

This condition is fulfilled if

$$\overline{fa} > \underline{\Delta u} = -1/\lambda^{'}(0),$$

in which case the incentives are effective at every level.

Notice that in this setting the optimal effort depends on the lifetime utility but not on the fines. Indeed, optimal effort depends on the argument of λ_*. From equation (16), this argument is equal to:

$$u_n - u_{n+1} + \overline{fa} = \lambda_*^{-1}(d_u - ru_n).$$

A.2 Incentive effects of point-record driving licenses: Model with point removal

The Bellman equation on a holistic incentive model can be written as follows

$$d_u - ru(S) + \left(\frac{d}{dt}\left[u(S_t)\right]\right)_{t=0^+} = \lambda_*(\overline{fa} + u(S) - E\left[u(TR(S))\right]). \qquad (17)$$

The state variables S are the seniorities of each non redeemed traffic offense (if any), the related demerit points and the seniority of the last contract birthday if the premium is included in the incentives. The related lifetime utility is $u(S)$. The state S_t is reached from S with an eventless history (no traffic offense, point

removal or contract birthday) of duration t. The parameters d_u and $\overline{fa}$ are the driving utility flow and the average fine, and $E\left[u(TR(S))\right]$ is the lifetime utility averaged with transition probabilities on the state(s) reached from S after a traffic offense. Continuity equations on utility at the time of a point removal or of a contract birthday (in the latter case, the increase in lifetime utility is equal to the disutility of the premium) and the equation linking the utility of a beginner and the utility just after a license suspension define the solution together with equation (17).

Let us prove the continuity of optimal effort after a point removal in a system where each traffic violation is redeemed beyond a given seniority threshold, equal to T. We suppose that each traffic violation is associated to one demerit point, and that incentives are related to fines and to the point-record driving license. The state variables are then the seniorities of each non redeemed traffic offense, if any. Let us denote these variables as

$$S = (t_1, \ldots, t_n),\ 0 \leq t_1 < \ldots < t_n < T.$$

The corresponding optimal effort is denoted as $e(S)$. Then the states reached without traffic offense before the next point removal are

$$S_t = (t_1 + t, \ldots, t_n + t),\ 0 \leq t < T - t_n.$$

We denote the state reached from S after an additional traffic offense (if $n < N$) as

$$(0, t_1, \ldots, t_n) = TR(S).$$

Since the lifetime utility is continuous after a point removal, we have the following result:

$$n \geq 1 : \lim_{t \to (T-t_n)^-} u(S_t) = u(S^R),\ S^R = (t_1 + T - t_n, \ldots, t_{n-1} + T - t_n).$$

The state S^R is reached from S if there is no traffic offense before the first point

removal. Then it is easily seen that

$$\lim_{t\to(T-t_n)^-} u\left[TR(S_t)\right] = u\left[TR(S^R)\right] = u(0, t_1 + T - t_n, \ldots, t_{n-1} + T - t_n).$$

This means that the left continuity at $T - t_n$ of the map $t \to u(S_t)$ also holds for the map $t \to u\left[TR(S_t)\right]$, which is associated with the states reached after an additional traffic offense. The reason is that redemption of past offenses occurs regardless of the future individual history.

From the three last equations, we obtain

$$\lim_{t\to(T-t_n)^-} e(S_t) = e(S^R)$$

and the continuity property of the optimal effort level. Since we expect a global increasing link between optimal effort and the accumulated demerit points, the time-effect should globally be decreasing in order to fulfill this continuity property.

A.3 Incentive effects of the experience rating system

Let us derive the Bellman equation on the expected disutility function given in (18), including an average fine of fa_j for a j demerit point traffic violation. The optimal disutility function is obtained from the program

$$v_n(t) = \min_{e\geq 0} edt + (\exp(-rdt) \times (1 - \lambda(e)dt) \times v_n(t + dt))$$

$$+ \left(\exp(-rdt) \times \left[\sum_{j\ /\ f_j>0} f_j\, \lambda(e)dt \times \left[v_{\min(n+j,N)}(t + dt) + fa_j\right]\right]\right) + o(dt),$$

which leads to

$$0 = v_n^{'}(t) + \lambda_* \left(\overline{fa} + \left(\sum_{j\ /\ f_j>0} f_j\, v_{\min(n+j,N)}(t)\right) - v_n(t)\right) - rv_n(t),$$

with $\overline{fa} = \sum_{j\ /\ f_j>0}\ f_j \times fa_j$ the average fine. Then we obtain the Bellman equation

$$v_n^{'}(t) = rv_n(t) - \lambda_* \left(\overline{fa} + \Delta v_n(t)\right),\ (0 \leq n \leq N). \tag{18}$$

A.4 Actuarial predictors with dynamic random effects

A consistent estimation of the variance of the random effect related to traffic violations is

$$\widehat{V}(\varepsilon^1_{i_0,y_0}) = \frac{\sum_{i,y}(N^1_{i,y} - \widehat{\lambda}^1_{i,y})^2 - N^1_{i,y}}{\sum_{i,y}\left(\widehat{\lambda}^1_{i,y}\right)^2}. \tag{19}$$

The moment-based estimators given in equations (10) and (19) can be improved by a link between the expectation and the variance of dependent variables (Liang, Zeger (1986)).

The bonus-malus coefficient $BM^j_i(m)$ given in equation (9) is obtained from an affine probabilistic regression of a multiplicative random effect $\varepsilon^j_{i,y(m)}$ related to driver i, month m and type j event with respect to the number of traffic violations recorded for the driver for each past month, and denoted as N^1_{i,m_1} $(m_1 < m)$. With the assumption $E(\varepsilon^j_{i,y}) = 1\ \forall i, j, y$, the predictor is given by

$$BM^j_i(m) = \widehat{E}(\varepsilon^j_{i,y(m)} \mid N^1_{i,m_1\ (m_1<m)})$$

$$= 1 + {}^t\widehat{Cov}(SN^1_{i,m}, \varepsilon^j_{i,y(m)}) \left[\widehat{V}(SN^1_{i,m})\right]^{-1} (SN^1_{i,m} - \widehat{E}\left(SN^1_{i,m}\right)), \tag{20}$$

where $SN^1_{i,m} = \underset{m_1<m}{vec}(N^1_{i,m_1})$ is the stacked vector of numbers of past traffic violations and where conditional expectation is restricted to affine regression. All the moments in equation (20) are estimated from the $\widehat{\lambda}^1_{i,m} = \widehat{E}(N^1_{i,m})$ and from the estimated moments of random effects.

A.5 Overall comparisons of incentive schemes

As a conclusion, let us derive the link given in Section 5.2 between the elasticity of the optimal frequency of traffic violations and the argument of λ_*, which determines the optimal effort level. We perform a local expansion around a value Δu^0 of the argument of λ_*, in a situation where the incentives are effective

(i.e. $\Delta u^0 > \underline{\Delta u} = -1/\lambda^{'}(0)$). If we write

$$e^0 = e_{opt}(\Delta u^0),\ e^0 + de = e_{opt}(\Delta u^0 + d\Delta u),$$

the equations

$$1 + \lambda^{'}(e^0)\Delta u^0 = 0;\ 1 + \left[\lambda^{'}(e^0 + de)\left(\Delta u^0 + d\Delta u\right)\right] = 0$$

lead to

$$de = \frac{-\lambda^{'}(e^0)}{\lambda^{''}(e^0)\Delta u^0} d\Delta u + o\left(d\Delta u\right),$$

and to

$$\frac{d\lambda}{\lambda(e^0)} = \frac{\lambda^{'}(e^0)}{\lambda(e^0)} de = \frac{\left[-\lambda^{'}(e^0)\right]^2}{\lambda(e^0) \times \lambda^{''}(e^0)} \times \frac{d\Delta u}{\Delta u^0}.$$

Hence the aforementioned elasticity is equal to $\left(\lambda^{'}\right)^2 / \lambda\lambda^{''}$. Now we have that

$$\left(\log \lambda\right)^{''} = \frac{\lambda^{''}}{\lambda} - \left(\frac{\lambda^{'}}{\lambda}\right)^2 = \frac{\lambda^{''}}{\lambda}\left(1 + \frac{\left(\lambda^{'}\right)^2}{\lambda\lambda^{''}}\right).$$

Then the conclusions given in Section 3.2 are easily obtained.

A.6 Monetary equivalents of traffic violations

Let us suppose that the increase in the argument of λ_* is close to 12% after a traffic violation. This is the value retained in Section 5.2 for a driver with seven demerit points accumulated, which corresponds to $n = 3$ traffic violations on average. As the argument of λ_* in the model without point removal and with fines is equal to $\overline{fa} + u_n - u_{n+1}$ (see equation 16) we have that

$$\overline{fa} + u_{n+1} - u_{n+2} = 1.12 \times \left(\overline{fa} + u_n - u_{n+1}\right). \tag{21}$$

We shall compare the utility losses $u_{n+1} - u_{n+2}$ and $u_n - u_{n+1}$ from the recurrence equation on lifetime utility, and obtain a monetary equivalent of an additional traffic violation from a derivation of the utility loss $u_n - u_{n+1}$. We have

$$u_{n+1} = g(u_n),\ g(u) = \overline{fa} + u - \lambda_*^{-1}(d_u - ru)$$

(see Figure 3). From the equality $\lambda_*^{'}(\Delta u) = \lambda(e_{opt}(\Delta u))$ (see equation (13)), we obtain

$$g^{'}(u_n) = 1 + \frac{r}{\lambda_n},\ \lambda_n = \lambda(e_{opt}(\overline{fa} + u_n - u_{n+1})).$$

The parameter λ_n is the frequency risk corresponding to the optimal effort exerted with n traffic violations accumulated. As λ_n decrease with n, we have that

$$1 + \frac{r}{\lambda_n} \leq \frac{u_{n+1} - u_{n+2}}{u_n - u_{n+1}} = \frac{g(u_n) - g(u_{n+1})}{u_n - u_{n+1}} \leq 1 + \frac{r}{\lambda_{n+1}}. \tag{22}$$

From equations (21) and (22), we obtain

$$\left(\frac{r}{\lambda_n^{'}} - 0.12\right) \times (u_n - u_{n+1}) = 0.12 \times \overline{fa} = 15.6\,\$,\ \lambda_{n+1} \leq \lambda_n^{'} \leq \lambda_n.$$

The monetary equivalent of an additional traffic violation is then bounded as follows:

$$\Leftrightarrow \frac{15.6\,\$}{\frac{r}{\lambda_{n+1}} - 0.12} \leq u_n - u_{n+1} \leq \frac{15.6\,\$}{\frac{r}{\lambda_n} - 0.12}.$$

Section 5.2 provides numerical examples with $\lambda_n = 0.15,\ \lambda_{n+1} = 0.12$. The monetary cost of a license suspension follows from a sum of the items related to traffic violations and from the inequalities given in (22).

References

[1] Liang, Kung-Yee, and Scott L. Zeger, "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika* 73 (1986), 13-22.

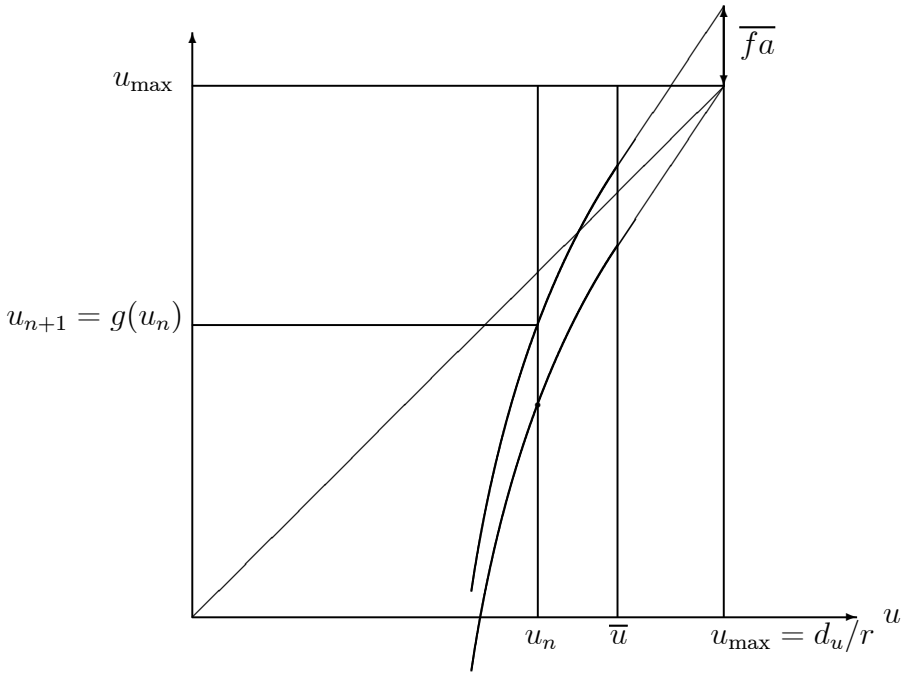
[2] Rockafellar, Ralph T., *Convex Analysis* (Princeton University Press, 1996).

Figure 3

Recurrence equation on the lifetime utility function

Point-record driving license without fines: $u_{n+1}^0 = f(u_n^0)$

Point-record driving license with fines: $u_{n+1} = g(u_n)$



$$f(u) = u - \lambda_*^{-1}(r(u_{\max} - u)),\ g(u) = f(u) + \overline{fa}.$$

Effective incentives condition with and without fines

$$e_n > 0 \Leftrightarrow u_n < \overline{u} = u_{\max}\left(1 + \frac{\lambda(0)}{\lambda'(0) \times d_u}\right).$$