WORKSHOP

A GENERALIZATION OF AUTOMOBILE INSURANCE RATING MODELS: THE NEGATIVE BINOMIAL DISTRIBUTION WITH A REGRESSION COMPONENT

By Georges Dionne and Charles Vanasse

Université de Montréal, Canada *

Abstract

The objective of this paper is to provide an extension of well-known models of tarification in automobile insurance. The analysis begins by introducing a regression component in the Poisson model in order to use all available information in the estimation of the distribution. In a second step, a random variable is included in the regression component of the Poisson model and a negative binomial model with a regression component is derived. We then present our main contribution by proposing a bonus-malus system which integrates a priori and a posteriori information on an *individual basis*. We show how net premium tables can be derived from the model. Examples of tables are presented.

Keywords

Multivariate automobile insurance rating; Poisson model; negative binomial model; regression component; net premium tables; Bayes analysis; maximum likelihood method.

Introduction

The objective of this paper is to provide an extension of well known models of tarification in automobile insurance. Two types of tarification are presented in the literature:

1) a priori models that select tariff variables, determine tariff classes and estimate premiums (see Van Eeghen et al. (1983) for a good survey of these models);
2) a posteriori models or bonus-malus systems that adjust individual premiums according to accident history of the insured (see Ferreira (1974), Lemaire (1985, 1988) and Van Eeghen et al. (1983) for detailed discussions of these models).

* Centre de recherche sur les transports and Department of Economics. This research was financed by SSHRC Canada. Comments on a previous version of the paper by Alois Gisler, Christian Gourieroux, Stuart Klugman, Jean Lemaire, David Scott and two anonymous referees were very useful.

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

This study focuses on the selection of tariff variables using multivariate regression models and on the construction of insurance tables that integrates a priori and a posteriori information on an *individual basis*. Our contribution differs from the recent articles in credibility theory where geometric weights were introduced (NEUHAUS (1988), SUNDT (1987, 1988)). In particular, SUNDT (1987) uses an additive regression model in a multiplicative tariff whereas our nonlinear regression model reflects the multiplicative tariff structure.

The analysis begins by introducing a regression component in both the Poisson and the negative binomial models in order to use all available information in the estimation of accident distribution. We first show how the univariate Poisson model can be extended in order to estimate different individual risks (or expected number of accidents) as a function of a vector of individual characteristics. At this stage of the analysis, there is no random variable in the regression component of the model. As for the univariate Poisson model, the randomness of the extended model comes from the distribution of accidents.

In a second step, a random variable is introduced in the regression component of the Poisson model and a negative binomial model with a regression component is derived. We then present our main contribution by proposing a bonus-malus system which integrates explicitly a priori and a posteriori information on an individual basis. Net premium tables are derived and examples of tables are presented. The parameters in the regression component of both the Poisson and the negative binomial models were estimated by the maximum likelihood method.

1. The Basic Model

1.a. *Statistical Analysis*

The Poisson distribution is often used for the description of random and independent events such as automobile accidents. Indeed, under well known assumptions, the distribution of the number of accidents during a given period can be written as

$$\text{(1)} \qquad \operatorname{pr}(Y_i = y) = \frac{e^{-\lambda}\lambda^y}{y!}$$

where y is the realization of the random variable Y_i for agent i in a given period and λ is the Poisson parameter which can be estimated by the maximum likelihood method or the method of moments. Empirical analyses usually reject the univariate Poisson model.

Implicitly, (1) assumes that all the agents have the same claim frequency. A more general model allows parameter λ to vary among individuals. If we assume that this parameter is a random variable and follows a gamma

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

distribution with parameters a and $1/b$ (GREENWOOD and YULE (1920), BICHSEL (1964), SEAL (1969)), the distribution of the number of accidents during a given period becomes

$$\text{pr}(Y_i = y) = \frac{\Gamma(y+a)}{y!\,\Gamma(a)} \frac{(1/b)^a}{(1+1/b)^{y+a}} \tag{2}$$

which corresponds to a negative binomial distribution with $E(Y_i) = \bar{\lambda}$ and $\text{Var}(Y_i) = \bar{\lambda}\left[1 + \frac{\bar{\lambda}}{a}\right]$, where $\bar{\lambda} = ab$.

Again, the parameters a and $(1/b)$ can be estimated by the method of moments or by the maximum likelihood method.

1.b. *Optimal Bonus Malus Rule*

An optimal bonus malus rule will give the best estimator of an individual's expected number of accidents at time $(t+1)$ given the available information for the first t periods $(Y_i^1, \ldots, Y_i^t)$. Let us denote this estimator as $\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t)$.

One can show that the value of the Bayes' estimator (i.e. a posteriori mathematical expectation of λ) of the true expected number of accidents for individual i is given by

$$\hat{\lambda}_i^{t+1}(Y_i^1 \ldots Y_i^t) = \int_0^\infty \lambda f(\lambda/Y_i^1 \ldots Y_i^t)\, d\lambda . \tag{3}$$

Applying the negative binomial distribution, the a posteriori distribution of λ is a gamma distribution with probability density function

$$f(\lambda/Y_i^1 \ldots Y_i^t) = \frac{(1/b+t)^{a+\bar{Y}_i} e^{-\lambda(1/b+t)} \lambda^{a+\bar{Y}_i-1}}{\Gamma(a+\bar{Y}_i)}, \tag{4}$$

where $\bar{Y}_i = \sum_{j=1}^{t} Y_i^j$.

Therefore, the Bayes' estimator of an individual's expected number of accidents at time $(t+1)$ is the mean of the a posteriori gamma distribution with parameters $(a+\bar{Y}_i)$ and $((1/b)+t)$:

$$\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t) = \frac{a+\bar{Y}_i}{(1/b)+t} = \bar{\lambda}\left[\frac{a+\bar{Y}_i}{a+t\bar{\lambda}}\right]. \tag{5}$$

Actuarial net premium tables can then be calculated by using (5).

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

2. The Generalized Model

Since past experience cannot, in a short length of time, generate all the statistical information that permits fair insurance tarification, many insurers use both *a priori* and *a posteriori* tarification systems. A priori classification is based on significant variables that are easy to observe, namely, age, sex, type of driver's license, place of residence, type of car, etc. A posteriori information is then used to complete a priori classification. However, when both steps of the analysis are not adequately integrated into a single model, inconsistencies may be produced.

In practice, often linear regression models by applying a standard method out of a statistical package are used for the a priori classification of risks. These standard models often assume a normal distribution. But any model based on a continuous distribution is not a natural approach for count data characterized by many "zero accident" observations and by the absence of negative observations. Moreover, the resulting estimators obtained from these standard models often allow for negative predicted numbers of accidents. Regression results from count data models are more appropriate for a priori classification of risks.

A second criticism is linked to the fact that *univariate* (without regression component) statistical models are used in the Bayesian determination of the individual insurance premiums. Consequently, insurance premiums are function merely of time and of the past number of accidents. The premiums do not vary *simultaneously* with other variables that affect accident distribution. The most interesting example is the age variable. Let us suppose, for a moment, that age has a significant negative effect on the expected number of accidents. This implies that insurance premiums should decrease with age. Premium tables derived from univariate models do not allow for a variation of age, even if they are a function of time. However, a general model with a regression component would be able to determine the specific effect of age when the variable is statistically significant.

Finally, the third criticism concerns the coherency of the two-stage procedure using different models in order to estimate the same distribution of accidents.

In the following section we will introduce a methodology which responds adequately to the three criticisms. First, count data models will be proposed to estimate the individual's accident distribution. The main advantage of the count data models over the standard linear regression models lies in the fact that the dependent variable is a count variable restricted to non-negative values. Both the Poisson and the negative binomial models with a regression component will be discussed. Although the univariate Poisson model is usually rejected in empirical studies, it is still a good candidate when a regression component is introduced. Indeed, because the regression component contains

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

many individual variables, the estimation of the individual expected number of accidents by the Poisson regression model can be statistically acceptable since it allows for heterogeneity among individuals. However, when the available information is not sufficient, using a Poisson model introduces an error of specification and a more general model should be considered. Second, we will generalize the optimal bonus-malus system by introducing all information from the regression into the calculation of premium tables. These tables will take account of time, accident record and the individual characteristics.

2.a. *Statistical Analysis*

Let us begin with the Poisson model. As in the preceding section, the random variables Y_i are independent. In the extended model, however, λ may vary between individuals. Let us denote by λ_i the expected number of accidents corresponding to individuals of type i. This expected number is determined by k exogenous variables or characteristics $x_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$ which represent different a priori classification variables. We can write

$$\lambda_i = \exp(x_i\beta) \tag{6}$$

where β is a vector of coefficients $(k \times 1)$. (6) implies the non-negativity of λ_i.

The probability specification becomes

$$\Pr(Y_i = y) = \frac{e^{-\exp(x_i\beta)}(\exp(x_i\beta))^y}{y!}. \tag{7}$$

It is important to note that λ_i is not a random variable. The model assumes implicitly that the k exogenous variables provide enough information to obtain the appropriate values of the individual's probabilities. The β parameters can be estimated by the maximum likelihood method (see HAUSMAN, HALL and GRILICHES (1984) for an application to the patents — R & D relationship). Since the model is assumed to contain all the necessary information required to estimate the values of the λ_i, there is no room for a posteriori tarification in the extended Poisson model. Finally, it is easy to verify that (1) is a particular case of (7).

However, when the vector of explanatory variables does not contain all the significant information, a random variable has to be introduced into the regression component. Following GOURIEROUX MONFORT and TROGNON (1984), we can write

$$\lambda_i = \exp(x_i\beta + \varepsilon_i) \tag{8}$$

yielding a random λ_i. Equivalently, (8) can be rewritten as

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

$$\lambda_i = \exp(x_i\beta)\, u_i \tag{9}$$

where $u_i \equiv \exp(\varepsilon_i)$.

As for the univariate negative binomial model presented above, if we assume that u_i follows a gamma distribution with $E(u_i) = 1$ and $\operatorname{Var}(u_i) = 1/a$, the probability specification becomes

$$\operatorname{pr}(Y_i = y) = \frac{\Gamma(y+a)}{y!\,\Gamma(a)} \left[\frac{\exp(x_i\beta)}{a}\right]^{y} \left[1 + \frac{\exp(x_i\beta)}{a}\right]^{-(y+a)} \tag{10}$$

which is also a negative binomial distribution with parameters a and $\exp(x_i\beta)$. We will show later that the above parameterization does not affect the results if there is a constant term in the regression component.

Then $E(Y_i) = \exp(x_i\beta)$ and $\operatorname{Var}(Y_i) = \exp(x_i\beta)\left[1 + \dfrac{\exp(x_i\beta)}{a}\right]$.

We observe that $\operatorname{Var}(Y_i)$ is a nonlinear increasing function of $E(Y_i)$. When the regression component is a constant c, $E(Y_i) = \exp(c) = \bar{\lambda}$ and

$$\operatorname{Var}(Y_i) = \bar{\lambda}\left[1 + \frac{\bar{\lambda}}{a}\right]$$

which correspond, respectively, to the mean and variance of the univariate negative binomial distribution.

DIONNE and VANASSE (1988) estimated the parameters of both the Poisson and negative binomial distributions with a regression component. A priori information was measured by variables such as age, sex, number of years with a driver's license, place of residence, driving restrictions, class of driver's license and number of days the driver's license was valid. The Poisson distribution with a regression component was rejected and the negative binomial distribution with a regression component yielded better results than the univariate negative binomial distribution (see Section 3 for more details).

An extension of the Bayesian analysis was then undertaken in order to integrate a priori and a posteriori tarifications on an individual basis.

2.b. *A Generalization of the Optimal Bonus Malus Rule*

Consider again an insured driver i with an experience over t periods; let Y_i^j represent the number of accidents in period j and x_i^j, the vector of the k characteristics observed at period j, that is $x_i^j = (x_{i1}^j, \ldots, x_{ik}^j)$. Let us further suppose that the true expected number of accidents of individual i at period j, $\lambda_i^j(u_i, x_i^j)$, is a function of both individual characteristics x_i^j and a random

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

variable u_i. The insurer needs to calculate the best estimator of the true expected number of accidents at period $t+1$. Let $\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^{t+1})$ designate this estimator which is a function of past experience over the t periods and of known characteristics over the $t+1$ periods.

If we assume that the u_i are independent and identically distributed over time and that the insurer minimizes a quadratic loss function, one can show that the optimal estimator is equal to:

$$\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^{t+1})$$

$$(11) \qquad = \int_0^\infty \lambda_i^{t+1}(u_i, x_i^{t+1}) f(\lambda_i^{t+1}/Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^t)\, d\lambda_i^{t+1}.$$

Applying the negative binomial distribution to the model, the Bayes' optimal estimator of the true expected number of accidents for individual i is:

$$(12) \qquad \hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t; x_i^1, \ldots, x_i^{t+1}) = \dot{\lambda}_i^{t+1}\left[\frac{a+\bar{Y}_i}{a+\bar{\lambda}_i}\right]$$

where $\lambda_i^j = \exp(x_i^j\beta)\,u_i \equiv (\dot{\lambda}_i^j)\,u_i$, $\bar{\lambda}_i = \sum_{j=1}^{t} \dot{\lambda}_i^j$ and $\bar{Y}_i = \sum_{i=1}^{t} Y_i^j$.

When $t = 0$, $\hat{\lambda}_i^1 = \dot{\lambda}_i^1 \equiv \exp(x_i^1\beta)$ which implies that only a priori tarification is used in th first period. Moreover, when the regression component is limited to a constant c, one obtains:

$$(13) \qquad \hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t) = \bar{\lambda}\left[\frac{a+\bar{Y}_i}{a+t\bar{\lambda}}\right]$$

which is (5). This result is not affected by the parametrization of the gamma distribution.

It is important to emphasize here some characteristics of the model. In (13) only individual past accidents ($Y_i^1, \ldots, Y_i^t$) are taken into account in order to calculate the individual expected numbers of accidents over time. All the other parameters are population parameters. In (12), individual past accidents and characteristics are used simultaneously in the calculation of individual expected numbers of accidents over time. As we will show in the next section, premium tables that take into account the variations of both individual characteristics and accidents can now be obtained.

Two criteria define an optimal bonus-malus system which has to be fair for the policyholders and be financially balanced for the insurer. It is clear that the estimator proposed in (12) is fair since it allows the estimation of the individual

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

risk as a function of both his characteristics and past experience. From the fact, that $E(E(A/B)) = E(A)$, it follows that the extended model is financially balanced:

$$E(\hat{\lambda}_i^{t+1}(Y_i^1, \ldots, Y_i^t;\ x_i^1, \ldots, x_i^{t+1})) = \lambda_i^{t+1} \text{ since } E(u_i) = 1\,.$$

3. Examples of Premium Tables

As mentioned above, Dionne and Vanasse (1988) estimated the parameters of the Poisson regression model (β vector) and of the negative binomial regression model (β vector and the dispersion parameter a) by the maximum likelihood method. They used a sample of 19 013 individuals from the province of Québec. Many *a priori* variables were found significant. For example, the age and sex interaction variables were significant as well as classes of driver's licences for bus, truck, and taxi drivers. Even if the Poisson model gave similar results to those of the negative binomial model, it was shown (standard likelihood ratio test) that there was a gain in efficiency by using a model allowing for overdispersion of the data (where the variance is greater than the mean): the estimate of the dispersion parameter of the negative binomial regression $\hat{a}$ was statistically significant (asymptotic t-ratio of 3.91). The usual χ^2 test generated a similar conclusion. The latter results are summarized in Table 1:

TABLE 1

ESTIMATES OF POISSON AND NEGATIVE BINOMIAL DISTRIBUTIONS WITH A REGRESSION COMPONENT

Individual number of accidents in a given period	Observed numbers of individuals during 1982-1983	Predicted numbers of individuals for 1982-1983	
		Poisson *	Negative binomial *
0	17,784	17,747.81	17,786.39
1	1,139	1,201.59	1,131.05
2	79	60.56	86.21
3	9	2.88	8.18
4	2	.15	.98
5+	0	0	0
	19,013	$\chi^2 = 29.91$ $\chi^2_{2,95} = 5.99$	$\chi^2 = 1.028$ $\chi^2_{1,95} = 3.84$
		Log Likelihood = −4,661.57	Log Likelihood = −4,648.58

* The estimated β parameters are published in DIONNE-VANASSE (1988) and are available upon request. $\hat{a} = 1.47$ in the negative binomial model.

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

The univariate models were also estimated for the purpose of comparison. Table 2 presents the results. The estimated parameters of the univariate negative binomial model are $\hat{a} = .696080$ and $(1/\hat{b}) = 9.93580$ yielding $\hat{\lambda} = .0701$. One observes that $\hat{a} = 1.47$ in the multivariate model is larger than $\hat{a} = .6961$ in the univariate model. This result indicates that part of the variance is explained by the a priori variables in the multivariate model.

Using the estimated parameters of the univariate negative binomial distribution presented above, table 3 was formed by applying (14) where \$ 100 is the first period premium ($t = 0$):

$$\hat{P}_i^{t+1}(Y_i^1, \ldots, Y_i^t) = 100 \frac{(\hat{a} + \bar{Y}_i)}{(\hat{a} + t\hat{\lambda})} \tag{14}$$

In Table 3, we observe that only two variables may change the level of insurance premiums, i.e. time and the number of accumulated accidents. For example, an insured who had three accidents in the first period will pay a premium of \$ 462.43 in the next period, but if he had no accidents, he would have paid only \$ 90.86.

From (14) it is clear that no additional information can be obtained in order to differentiate an individual's risk. However, from (12), a more general pricing formula can be derived:

$$\hat{P}_i^{t+1}(Y_i^1 \ldots Y_i^t;\ x_i^1 \ldots x_i^{t+1}) = M\hat{\lambda}_i^{t+1} \left[\frac{\hat{a} + \bar{Y}_i}{\hat{a} + \hat{\lambda}_i} \right] \tag{15}$$

TABLE 2

ESTIMATES OF UNIVARIATE POISSON AND NEGATIVE BINOMIAL DISTRIBUTIONS

Individual number of accidents in a given period	Observed numbers of individuals during 1982-1983	Predicted numbers of individuals for 1982-1983	
		Poisson (exp $\hat{c}$ = 0.0701)	Negative binomial ($\hat{a}$ = 0.6960; $1/\hat{b}$ = 9.9359)
0	17,784	17,726.60	17,785.28
1	1,139	1,241.86	1,132.05
2	79	43.50	88.79
3	9	1.02	7.21
4	2	0.02	.61
5+	0	0	0
	19,013	$\chi^2 = 133.06$; $\chi^2_{2,95} = 5.99$	$\chi^2 = 2.21$; $\chi^2_{1,95} = 3.84$
		Log Likelihood = −4950.28	Log Likelihood = −4916.78

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

TABLE 3

UNIVARIATE NEGATIVE BINOMIAL MODEL

$\hat{a} = .696080 \quad \hat{\lambda} = .0701$

t \ $\bar{Y}_i$	0	1	2	3	4
0	100.00				
1	90.86	221.38	351.91	462.43	612.96
2	83.24	202.83	322.42	442.01	561.60
3	76.81	187.15	297.50	407.84	518.19
4	71.30	173.72	276.15	378.58	481.00
5	66.52	162.09	257.66	353.23	448.80
6	62.35	151.92	241.49	331.06	420.63
7	58.67	142.95	227.23	311.52	395.80
8	55.40	134.98	214.56	294.15	373.73
9	52.47	127.85	203.23	278.61	353.99

where $\hat{\lambda}_i^{t+1} \equiv \exp(x_i^{t+1}\hat{\beta})$, $\hat{\bar{\lambda}}_i \equiv \sum_{j=1}^{t} \exp(x_i^j\hat{\beta})$,

and M is such that

$$1/I \sum_{i=1}^{I} \hat{\lambda}_i^{t+1} M = \$ \, 100$$

when the total number of insureds is I.

This general pricing formula is function of time, the number of accumulated accidents and the individual's significant characteristics in the regression component. In consequence, tables can now be constructed more generally by using (15). First, it is easy to verify that each agent does not start with a premium of \$ 100. In Table 4, for example, a young driver begins with

TABLE 4

NEGATIVE BINOMIAL MODEL WITH A REGRESSION COMPONENT

Male, 18 years old in period 0, region 9, class 42

t \ $\bar{Y}_i$	0	1	2	3	4
0	280.89				
1	247.67	416.47	585.27	754.07	922.87
2	217.46	365.66	513.86	662.07	810.27
3	197.00	331.26	465.53	599.79	734.06
4	180.06	302.78	425.50	548.23	670.95
5	165.81	278.81	391.82	504.82	617.83
6	153.64	258.36	363.07	467.79	572.50
7	79.85	134.28	188.70	243.12	297.55
8	76.92	129.35	181.77	234.19	286.62
9	74.20	124.76	175.33	225.90	276.46

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

$ 280.89. Second, since the age variable is negatively significant in the estimated model, two factors, rather than one, have a negative effect on the individual's premiums (i.e. time and age). In Table 4, the premium is largely reduced when the driver reaches period seven at 25 years old (a very significant result in the empirical model).

For the purpose of comparison, Table 4 was normalized such that the agent starts with a premium of $ 100. The results are presented in table 5a. The effect of using a regression component is directly observed. Again the difference between the corresponding premiums in Table 3 and Table 5a come from two

TABLE 5a

TABLE 4 DIVIDED BY 2.8089

t \ $\bar{Y}_i$	0	1	2	3	4
0	100.00				
1	88.17	148.27	208.36	268.46	328.55
2	77.42	130.18	182.94	235.70	288.46
3	70.13	117.93	165.73	213.53	261.33
4	64.10	107.79	151.48	195.18	238.87
5	59.03	99.26	139.49	179.72	219.95
6	54.70	91.98	129.26	166.54	203.82
7	28.43	47.81	67.18	86.55	105.93
8	27.38	46.05	64.71	83.37	102.04
9	26.42	44.42	62.42	80.42	98.42

TABLE 5b

COMPARISON OF BASE PREMIUM AND BONUS-MALUS FACTOR COMPONENTS

	Univariate Model			Individual of Table 4		
	Base Premium	Bonus Malus Factor		Base Premium *	Bonus Malus Factor	
t \ $\bar{Y}_i$		0	1		0	1
0	100.00	1.0000		280.89	1.0000	
1	100.00	0.9086	2.2138	280.89	0.8817	1.4827
2	100.00	0.8324	2.0283	280.89	0.7742	1.3018
3	100.00	0.7681	1.8715	280.89	0.7013	1.1793
4	100.00	0.7130	1.7372	280.89	0.6410	1.0779
5	100.00	0.6652	1.6209	280.89	0.5903	0.9926
6	100.00	0.6235	1.5192	280.89	0.5470	0.9198
7	100.00	0.5867	1.4295	154.67	0.5163	0.8682
8	100.00	0.5540	1.3498	154.67	0.4973	0.8363
9	100.00	0.5247	1.2785	154.67	0.4797	0.8066

* To be compared with Table 5a, this column should be divided by 2.8089.

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

sources: the individual in Table 5a has particular a priori characteristics while all individuals are implicitly assumed identical in Table 3 and age is significant when the individual reaches period seven (25 years old). Finally, the above comparison shows that the Bonus-Malus factor is now a function of the individual's characteristics as suggested by (12). Table 5b separates the corresponding base premium and Bonus-Malus factor components of the total premiums in the first two columns of Table 3 and Table 4.

Moreover, when the insured modifies significant variables, new tables may be formed. In Table 4 the driver was in region # 9 (a risky region in Quebec) and had a standard driving license.

TABLE 6

NEGATIVE BINOMIAL MODEL WITH A REGRESSION COMPONENT
SAME INDIVUDUAL AS IN TABLE 4, MOVED TO MONTREAL IN PERIOD 4

t \ $\bar{Y}_i$	0	1	2	3	4
0	280.89				
1	247.67	416.47	585.27	754.07	922.87
2	217.46	365.66	513.86	662.07	810.27
3	197.00	331.26	465.53	599.79	734.06
4	119.65	201.19	282.73	364.28	445.82
5	113.18	190.32	267.45	344.59	421.73
6	107.38	180.56	253.74	326.92	400.11
7	56.98	95.81	134.65	173.48	212.32
8	55.47	93.28	131.08	168.89	206.69
9	54.04	90.87	127.70	164.53	201.36

TABLE 7

NEGATIVE BINOMIAL MODEL WITH A REGRESSION COMPONENT
SAME INVIDUAL AS IN TABLE 4, MOVED TO MONTREAL IN PERIOD 4,
CHANGED FOR CLASS 31 (TAXI) IN PERIOD 5

t \ $\bar{Y}_i$	0	1	2	3	4
0	280.89				
1	247.67	416.47	585.27	754.07	922.87
2	217.46	365.66	513.86	662.07	810.27
3	197.00	331.26	465.53	599.79	734.06
4	119.65	201.19	282.73	364.28	445.82
5	291.65	490.42	689.19	887.96	1086.73
6	256.00	430.48	604.95	779.42	953.90
7	127.26	213.99	300.72	387.45	474.18
8	119.97	201.73	283.49	365.25	447.02
9	113.47	190.80	268.13	345.47	422.80

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

Now if the individual moves from region # 9 to a less risky region (Montreal, for example) in period 4, the premiums then change (see Table 6).

Having two accidents, he now pays $ 282.73 in period 4 instead of $ 425.50. Finally, if the driver decides to become a Montreal taxi driver in period 5, the following results can be seen in Table 7.

Again, having two accidents, he now pays $ 689.19 in period 5 instead of $ 267.45.

CONCLUDING REMARKS

In this paper, we have proposed an extension of well-known models of tarification in automobile insurance. We have shown how a bonus-malus system, based only on a posteriori information, can be modified in order to take into account simultaneously a priori and a posteriori information on an individual basis. Consequently, we have integrated two well-known systems of tarification into a unified model and reduced some problems of consistencies. We have limited our analysis to the optimality of the model.

One line of research is the integration of accident severity into the general model even if the statistical results may be difficult to use for tarification (particularly in a fault system). Recent contributions have analyzed different types of distribution functions to be applied to the severity of losses (LEMAIRE (1985) for automobile accidents, CUMMINS et al. (1988) for fire losses, and HOGG and KLUGMAN (1984) for many other applications). Others have estimated the parameters of the total loss amount distribution (see SUNDT (1987) for example) or have included individuals' past experience in the regression component (see BOYER and DIONNE (1986) for example). However, to our knowledge, no study has ever considered the possibility of introducing the individual's characteristics and actions in a model that isolates the relationship between the occurence and the severity of accidents *on an individual basis*.

REFERENCES

BICHSEL, F. (1964) Erfahrungs-Tarifierung in der Motorfahrzeughaftpflichtversicherung. *Mitt. Verein. Schweiz. Versicherungs-Mathematiker* **64**, 119–130.

BOYER, M. and DIONNE, G. (1986) La tarification de l'assurance automobile et les incitations à la sécurité routière: une étude empirique. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* **122**, 293–322.

CUMMINS, J. D., DIONNE, G., MCDONALD, J. B. and PRITCHETT, B. M. (1988) Application of the GB2 Distribution in Modeling Insurance Loss Processes. Mimeo, Department of Insurance, University of Pennsylvania.

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press

DIONNE, G and VANASSE, C. (1988) Automobile Insurance Ratemaking in the Presence of Asymmetrical Information. Mimeo, CRT and Economics Department, Université de Montréal.
FERREIRA, J. (1974) The Long-Term Effects of Merit Rating Plans on Individual Motorists. *Operations Research* **22**, 954–978.
GOURIEROUX, C., MONFORT, A. and TROGNON, A. (1984) Pseudo Maximum Likelihood Methods: Application to Poisson Models. *Econometrica* **52**, 701–720.
GREENWOOD, M. and TYLE, G. U. (1920) An Inquiry into the Nature of Frequency Distribution of Multiple Happenings. *Journal of Royal Statistical Society* **A 83**, 255–279.
HAUSMAN, J., HALL, B.H. and GRILICHES Z. (1984) Econometric Models for Count Data with an Application to the Patents — R & D Relationship. *Econometrica* **52**, 910–938.
HOGG, R.V. and KLUGMAN S.A. (1984) *Loss Distributions*. Wiley, New-York.
LEMAIRE, J. (1985) *Automobile Insurance: Actuarial Models*. Kluwer-Nighoff, Boston, 248 pages.
LEMAIRE, J. (1988) A Comparative Analysis of Most European and Japanese Bonus-Malus Systems. *Journal of Risk and Insurance* LV, 660–681.
NEUHAUS, W. (1988) A Bonus-malus System in Automobile Insurance. *Insurance: Mathematics & Economics* **7**, 103–112.
SEAL, H. L. (1969) *Stochastic Theory of a Risk Business*. Wiley, New-York.
SUNDT, B. (1987) Two Credibility Regression Approaches for the Classification of Passenger Cars in a Multiplicative Tariff". *ASTIN Bulletin* **17**, 41-70.
SUNDT, B. (1988) Credibility Estimators With Geometric Weights. *Insurance: Mathematics & Economics* **7**, 113–122.
VAN EEGHEN, J., GREUP, E.K. and NIJSSEN, J.A. (1983) *Surveys of Acturial Studies: Rate Making*. Research Department, Nationale — Nederlanden N.V., 138 pages.

GEORGES DIONNE
CRT, Université de Montréal, C.P. 6128, Succ. A, Montreal, H3C 3J7, Canada.

https://doi.org/10.2143/AST.19.2.2014909 Published online by Cambridge University Press