ERASMUS

8530/A

ECONOMETRIC INSTITUTE

A STATISTICAL MODEL FOR THE COSTS OF PASSENGER CAR TRAFFIC ACCIDENTS

B.S. VAN DER LAAN AND A.S. LOUTER

REPORT 8530/A

Erasmus

ERASMUS UNIVERSITY ROTTERDAM - P.O. BOX 1738 - 3000 DR ROTTERDAM - THE NETHERLANDS

A STATISTICAL MODEL FOR THE COSTS OF PASSENGER CAR TRAFFIC ACCIDENTS

by

B.S. van der Laan and A.S. Louter

Abstract

The total costs of damage caused by passenger car traffic accidents per distance interval of an individual motorist can be split into two components: the number of accidents which occur to the motorist, and the amount of damage associated with these accidents.

In our study we present a probability model for the number of accidents of an individual motorist. We also discuss models, on empirical results, for the amount of damage. The integration of both models results in a general model for the costs of damage.

From this general model we derive a model for the costs of indemnity to be paid by car insurance companies.

The models are applied to a sample of Dutch car insurance data.

Contents

Rotterdam, November 1985

1. Introduction[1,2]

In this paper we construct a statistical model for the costs of damage of passenger car traffic accidents of Dutch motorists. In the literature concerning risks of accidents, models of this type are described by, for example, Seal (1969), Beard, Pentakäinen and Pesonen (1977), and Panjer and Willmot (1983), and applied by, for example, Weber (1970). We remark that the original development of a model of this type dates back to Lundberg (1909). As distinct from the usual approach, which starts from the number of accidents per unit of time, we consider the number of accidents per distance driven.

Before we can construct a model, we must indicate precisely what we mean by an accident. The National Safety Council (1971, inside back cover), e.g., defines an accident simply as: "accident is that occurance in a sequence of events which usually produces unintended injury, death or property damage". Goeller (1969, p. 170), e.g., defines an accident as follows. "Accident is used as a synonym for traffic collision, which occurs between a moving motor vehicle and a pedestrian, a fixed object, another vehicle, or the roadway (e.g. overturning)." These definitions are not suitable for our purposes, so we modify them in order to distinguish between special types of damage, resulting from accidents. We restrict our study to the analysis of passenger car traffic accidents, hereunder called car accidents. We define, particularly for Dutch circumstances:

"A passenger car traffic accident (or car accident), which occurs to an individual motorist, is defined as the event which results in

a. property-damage to the individual's car, the so-called casco damage, and/or
b. personal injury and/or property damage to one or more third-parties, not

----

1 The authors wish to thank Professor J. Koerts for his valuable contributions. The authors wish also to thank AGO Verzekeringen (after a merger now called AEGON Verzekeringen) at Groningen, a national Dutch insurance company, which provided us the insurance and claim data on which we based the empirical part of the study, and the Netherlands Central Bureau of Statistics at the Hague, which provided us data concerning number of miles driven per year by Dutch motorists.

2 This paper is presented at the International Conference 1985 of the Institute of Statisticians on "Statistical Modelling", Cambridge, Kings College, 26 - 29 June 1985.

being the driver or the passengers of the car, the so-called third-party damage, and/or

c. personal injury to the driver or the passengers of the car,

where the event has been caused by the driver or for which he can be held liable, where the originator of the event cannot be held liable, or when the cause of the event cannot be blamed to a natural person."

The under a. given type of damage concerns in the Netherlands, for example, damage to one's car resulting from a collision with another object (or upset of the car), and parking damage to one's own car. There is also damage caused by particular types of accidents, such as damage resulting from collision with birds or animals, as well as damage resulting from fire, theft, storm, pane crack, etc. The under b. given type of damage concerns the total damage of personal injury to persons other than the driver and his passengers, damage to their properties, and damage to properties of corporate bodies.

The definition given above implies that we regard the case of a collision between two or more vehicles as a single event.

In the Netherlands we find, in connection with car accidents, the following main types of insurances:

- insurance against third-party risks, which covers the risks of causing damage to other persons, not being the driver or the passengers of the car;
- limited casco insurance, which covers the risks of the above mentioned particular types of accidents;
- third-party plus casco insurance, which covers the third-party risks, the limited casco risks, as well as other risks of damage to the insured's car, where the insured can choose different amounts for an additional excess;
- an insurance for particular personal injury protection of the driver and the passengers of the car.

In this study we analyse the following types of insurances and damages:

- third-party insurance in connection with third-party damage; and
- all-risk insurance, which is a third-party plus a casco insurance, with the legaly minimum amount of additional excess, in connection with all-risk damage, which is the sum of third-party and casco damage, excluding damage resulting from particular types of accidents.

We do not analyse the all-risk insurances with higher amounts of excess, because the numbers of insureds with equal amounts of excess are too low to analyse them separately. Moreover, we do not analyse the insurance against damage resulting from particular types of accidents, because the different types of accidents do not frequently occur, except damage resulting from pane crack, their amounts of damage are low in general, and the driver is, in general, not responsible for the damage. Moreover, a claim for damage of a particular type of accident does not influence the no-claim discount, nor is subject to an excess. In other words, the premium remains unchanged, and the company pays the whole amount of the damage. Nor we analyse the insurance for particular injury protection, because we do not have data about this insurance.

We need a model which is suitable for applications in the field of insurance data. As is well known, an insurance company does not always pays the entire amount of damage that is claimed by the insured, due to the excess and to the sum insured. We need, therefore, to be able to derive a model for the costs of indemnity paid by insurance companies from the general model for the costs of damage of road accidents, which occur to an individual motorist.

In Section 2 a general model is constructed. Moreover, we discuss probability models for the number of accidents and for the amount of damage caused by accidents. It is shown that models applicable to insurance data are special cases of the general model. In Section 3 we apply the model to Dutch car insurance data. The paper finish with a discussion of the conclusions.

2. A general model for the costs of car accidents

2.1 A general model

For the construction of a model to explain the total costs of damage from accidents of individual motorists per distance unit we follow the approach of, for example, Seal (1969, p. 12), and Beard et al. (1977, pp. 21 - 22). We make the following assumptions:

(i) the driving of a certain distance interval by an individual motorist is a random experiment with accident as outcomes;

(ii) the amount of damage, associated with an accident, is the outcome of a random experiment;

(iii) the probability distributions of the number of accidents, which occur to an individual motorist, and that of the amount of damage, have specific forms, but the parameter values vary from motorist to motorist, due to factors which influence the risk the motorist is exposed to;

(iv) the number of outcomes, n, of the experiment given under (i) does not depend on the outcomes of the experiment given under (ii); the outcomes of the experiment given under (ii) are mutually independent, they are subject to the same probability law, and they are independent of the distance driven.

On the basis of these assumptions we are able to construct a general model for the costs of damage of road accidents for an individual motorist. We define the following random variables.

N = the number of accidents which occur to an individual motorist driving some distance interval. Let p(n) be the mass function of N, with $n = 0, 1, 2, \ldots.$

Y = the amount of damage of a given accident. Let F(y) be the distribution function of this random variable, with $y > 0$.

X = the total costs of damage of accidents of an individual motorist driving some distance interval. Let G(x) be the distribution function of X, with $x \geq 0$.

We then get

$$X = \begin{cases} 0 & \text{for } N = 0 \\ Y_1 + \ldots + Y_n & \text{for } N = n > 0 \end{cases}$$

Assumption (iv) can now be written more exactly in terms of N and Y_j as follows:

(iv) The sequence $Y_1, Y_2, \ldots, Y_n, \ldots$ is a sequence of independent identically distributed random variables, satisfying $E[|Y_j|] < \infty$, $j = 1, 2, \ldots, n, \ldots$, and N is an integer-valued random variable, satisfying $E[N] < \infty$, whose values n does not depend on the values of the Y_j's.

The distribution function of X can be written as

$$G(x) = Pr[X \leq x] = Pr[Y_1 + \dots + Y_N \leq x]$$

$$= \sum_{n=0}^{\infty} Pr[Y_1 + \dots + Y_n \leq x \mid N = n] \; Pr[N = n]$$

Because of assumption (iv) we get

$$(2.1) \quad G(x) = \sum_{n=0}^{\infty} Pr[Y_1 + \dots + Y_n \leq x] \; Pr[N = n]$$

According to Feller (1971, pp. 143 - 148), we define $F^{n*}(x)$ as the n-fold convolution of F(x), and $F^{0*}(x)$ as the atomic distribution, concentrated at the origin, hence

$$F^{n*}(x) = Pr[Y_1 + \dots + Y_n \leq x]$$

where

$$F^{n*} = F^{(n-1)*} \times F$$

and

$$F^{0*}(x) = 0 \quad \text{if} \quad x < 0$$
$$= 1 \quad \text{if} \quad x \geq 0$$

Then, substituting p(n; k) for Pr[N = n], (2.1) becomes (cf., for e.g., Feller (1971, p. 159), and Seal (1977, p.12))

$$G(x) = \sum_{n=0}^{\infty} F^{n*}(x) \; p(n)$$

The expected costs of damage for an individual motorist driving a unit distance equals (cf. Feller (1971, p. 167)

$$E[X] = E[N] \; E[Y]$$

and the variance is given by (cf. Feller (1971, p. 167)

$$\text{var}(X) = E[N]\ \text{var}(Y) + \{E[Y]\}^2\ \text{var}(N)$$

Now we have obtained a general framework for the total costs of damage of accidents for an individual motorist. When the distribution function F(y) and the mass function p(n) are known, or estimated, we are able, in principle, to derive the distribution function G(x). This implies that we are able, in principle, to estimate the total costs of damage of accidents of an individual motorist.

An insurance company, however, does not always pay indemnity for the entire amount of damage resulting from an accident. It pays indemnity for the amount of the claim, i.e. the amount of the damage reduced by the amount of the excess, and, moreover, it pays at most the amount of the sum insured, reduced by the amount of the excess. Therefore, we want to adjust the general model to a model for the costs of claims, suitable for applications with insurance claim data.

It is of importance to remark that an insurance company does not have data about all accidents. It has information only about accidents where indemnity has been paid, i.e., about claims. The number of claims is always smaller than the total number of accidents. For this, three reasons can be given.

1. Not all risks, related to accidents, can be insured; the policy always contain some restrictions. Moreover, not all motorists have a casco insurance.
2. Claims for casco damage, having an amount of damage lower than the excess, are not considered for indemnity.
3. The no-claim discount can rise to some hundred guilders. It may be profitable for an insured not to claim a particular damage when its amount is lower than the sum of the excess and the no-claim discount.

A consequence of this is that accidents having low amounts of damage are underrepresented in the set of data an insurance company has available.

There is known few about the behaviour of insureds with respect to the no-claim rule. Therefore, it is difficult to take into account this behaviour. Because of the lack of information on this point, we start from the (far-reaching) assumption that all accidents with an amount of damage exceeding the excess, are claimed.

Given some arbitrary interval $\mu = [a, b]$, where a represents the amount of the excess and b the sum insured, we define the following random variables.

K = the number of claims, i.e., the number of accidents with amounts of damage equal to or exceeding a. Let $p(k; \mu)$ be the mass function of K with $k = 0, 1, 2, \ldots$.

Z = the size of the claim of a given accident, where

$$Z = \begin{cases} 0 & \text{if } Y < a \\ Y - a & \text{if } Y \in [a, b] \\ b - a & \text{if } Y > b \end{cases}$$

Let $H(z; \mu)$ be the distribution function of Z, with $z > 0$.

U = the total costs of claims of accidents of an individual insured during the driving of some distance interval. Let $G^*(u; \mu)$ be the distribution function of U, with $u \geq 0$.

The probability distributions of K, Z, and U can be related to the probability distributions of, respectively, Y, N and X. In this case K can be understood as a random sum of N stochastic Bernoulli-variables $R_1, \ldots, R_N$, with parameter $p = Pr[Y \geq a] = 1 - F(a)$. The mass function of K is given by

$$(2.2) \quad p(k; \mu) = Pr[K = n] = \sum_{i=k}^{\infty} Pr[K = n \mid N = i] \, Pr[N = i]$$

$$= \sum_{i=k}^{\infty} \binom{i}{k} p^k (1 - p)^{i-k} \, Pr[N = i]$$

It can easily be derived that the mathematical expectation of K equals

$$(2.3) \quad E[K] = p \, E[N]$$

and that its variance is equal to

$$(2.4) \quad var(K) = p^2 \, var(N) + p \, (1 - p) \, E[N]$$

It is easy to verify that, if N is, for example, Poisson distributed with parameter λ, K is also Poisson distributed, with parameter $\xi = p\lambda$. Similarly, if N is negative binomial distributed with parameters r $(r > 0)$ and q $(0 < q <$

1), then K is also negative binomial distributed with parameters $s = r$ and $Q = q/(p + q - pq)$. Similar relationships can be shown for other distributions for N. Thus from the estimates of the values of the parameters of the probability distribution of the number of claims per mile travelled, we can derive estimates of the values of the parameters of the probability distribution of the number of accidents per mile travelled, given that the value of p is known, or can be estimated.

The probability distributions of Z can be expressed as a function of the distribution function of Y as follows

$$(2.5)\quad H(z, \mu) = \begin{cases} 0 & \text{for } z < 0 \\ F(z+a) & \text{for } 0 \leq z < b - a \\ 1 & \text{for } z \geq b - a \end{cases}$$

The distribution function of

$$U = \begin{cases} 0 & \text{for } K = 0 \\ Z_1 + \dots + Z_k & \text{for } K = k > 0 \end{cases}$$

equals

$$(2.6)\quad G^*(u; \mu) = \Pr[U \leq u] = \Pr[Z_1 + \dots + Y_k \leq u]$$

$$= \sum_{k=0}^{\infty} \Pr[Z_1 + \dots + Z_k \leq u \mid K = k] \Pr[K = k]$$

$$= \sum_{k=0}^{\infty} \Pr[Z_1 + \dots + Z_k \leq u] \Pr[K = n]$$

The last equality sign follows from the result that, if the sequence $Y_1, \dots, Y_N$ is a sequence of N independent random variables, the sequence $Z_1, \dots, Z_K$ is a sequence of K independent random variables. Defining $H^{n*}(u; \mu)$ as the n-fold convolution of $H(u; \mu)$ and $H^{0*}(u; \mu)$ as the atomic distribution, concentrated at the origin, (2.6) can be written as

$$(2.7)\quad G^*(u;\mu) = \sum_{k=0}^{\infty} H^{n*}(u;\mu)\,\Pr[K = k]$$

It is clear that, if we take as particular interval for μ the interval $[0, \infty)$, (2.7) becomes (2.1).

The moments of U can be expressed in moments of K and Z, and therefore (with the help of (2.3) and (2.4)) in moments of N and Z. They cannot be expressed in moments of Y and N, because the distribution of Z is truncated.

The estimation of the values of the parameters of the distributions of K and Z is straight on. The unknown parameters of the distribution of Z are the parameters of the distribution of Y. These parameters can be estimated on the basis of the conditional distribution of Y, given $Y \in \mu$, with the help of a proper computer program. We can estimate the value(s) of the parameter(s) of the distribution of N on the basis of the estimate(s) of the value(s) of the parameter(s) of the distribution of K and the estimate of the values of $p = \Pr[Y \geq a]$. Thus we conclude that it is possible to estimate the values of the parameters of the model for the costs of accidents on the basis of claim data.

2.2 Probability models for the number of accidents

In the extensive literature on accident statistics we mostly find the number of accidents in which an individual motorist was involved during a time interval of length $t > 0$. We can then assume that the number of accidents is Poisson distributed. For reviews of probability models for the number of accidents used in insurance and road accident studies, we refer to Kupper (1960), Seal (1969), pp. 13 - 19 and Erlander (1971). The parameters of the distributions are not equal for all motorists, but vary under different circumstances, different qualities of the motorist, and other risk factors which influence the risk the motorist is exposed to.

Many authors emphasize that the number of miles driven per year is an important risk factor. Mehring (1962) shows that there is a significant association between the number of accidents per year and the annual mileage. Jorgensen (1969) considers the notion "accident density", which is the number of accidents per kilometre per year. He uses this notion, however, in combination with a given road classification. Weber (1970) and Ferreira (1971) also consider the importance of the annual mileage, but for lack of relevant data on this variable, they did not incorporate it into their studies. Ferreira (1971), p. 115 states that "territorial location and number of miles driven are regarded as direct indications of vehicle exposure to potential hazards and would be considered preferable to biographical factors such as age of driver even if the statistical correlation between each of these characteristics and drivers' accidents experience were the same". One of the groups of factors which the characteristics are generally grouped into, is the group of "exposure-related factors related to where, when and how often a vehicle is driven" (Ferreira (1971, p. 115)). White (1976) concludes that "several measures of exposure (...) have been proposed and succesfully used in specific situations. However, the most common measure and the one used herein is driving distance expressed in vehicle miles of travel." Foldvary (1975, 1976, 1977, 1978 and 1979) studies road accident per miles travelled "to produce and analyse accident involvement rates relative to exposure expressed in miles travelled for a large selection of driver-vehicle variables and their combination" (Foldvary (1975, p. 191)). Dutt, Reinfurt and Stutts (1977) also consider accidents per (million) miles, instead of accidents per year. Van der Laan (1979) shows the importance of the number of miles driven per year on the basis of Dutch statistical data.

To incorporate the relation between the number of accidents and the annual mileage, we can include this factor as an explanatory variable for the Poisson parameter λ, as suggested by Weber (1970). In our opinion it is more satisfactory to consider the number of accidents as a linear function of the distance covered rather then as a linear function of the time period, as already proposed by Van der Laan (1971).

We, therefore, make the following assumptions concerning an individual motorist:

(1) The number of accidents in non-overlapping distance intervals are stochastically independent.

(2) The probability of a certain number of accidents in a given small distance interval is dependent only on the length of that interval and is independent of the place on the distance axis. The probability of one accident in each small distance interval of length h equals

$$p(1;\, h) = \Pr[N = 1] = \lambda h + o(h)$$

where λ is a positive constant. The term o(h) denotes a quantity which is of smaller order of magnitude than h, i.e., $\lim_{h \to 0} \frac{o(h)}{h} = 0$.

(3) The probability of two or more accidents in an interval of length h equals

$$\sum_{n=2}^{\infty} p(n; h) = o(h)$$

These assumptions are the regular Poisson postulates. Given these assumptions it is easy to derive that the number of accidents an individual motorist has during the driving of some distance interval is Poisson-distributed.[3]

We can ask whether these assumptions are valid in practical situations. In the literature about accident analysis several authors consider the three problems associated with these assumptions. For example, Beard et al. (1977) discuss the assumptions given above in Section 2.3 (pp. 8 - 10) and in Chapter 10 (pp. 110 - 131).

Assumption 1 does not lead to many problems in car insurance. They can "often be met by defining (...) a risk unit as a combination of all those risks lying near to each other, between which contamination is possible" (Beard et al. (1979, p. 10)). Assumption 3 does not give problems either. "Two motor vehicles may collide giving rise to a double event. (...) This difficulty can, however, be circumvented by a suitable choice of definition, for example by regarding the case of collisions between two cars as a single claim" (Beard et al. (1979), p. 10).

The problems with respect to assumption 2 are much more difficult to deal with, since it can be expected that the parameter λ is a function of several characteristics, e.g.

a. The driving of some distance interval by a particular motorist may lead to a higher probability of having one or more accidents than driving another distance interval of the same length, due to environmental factors such as the conditions of the road surface, the type of classification of parts of the road (road crossings, silent country-side roads, etc.), the amount of traffic,

---

3 For a derivation of the Poisson model, see, e.g., Cramér (1954, pp. 104 - 106), Feller (1968, pp. 446 - 448), Hogg and Craig (1978, pp. 99 - 100), or Beard, Pentikaïnen and Pesonen (1979, Appendix A).

and so on.

b. The probability of a particular motorist having one or more accidents while driving a particular distance interval is subject to change, due to several risk factors. For example, the time of the day or the time of the year this distance interval is covered. It makes a great difference whether some distance interval is covered during a rush hour or during a quiet hour of the day, during the day or during the night, during working-days or during holidays. The weather conditions also influence the probability of having one or more accidents while driving a particular distance interval (for example, a dry road or icy patches on the road). Personal factors about the driver at the time he is driving, such as health and traffic mentality, influence the probability of having one or more accidents. We can also expect that the state of the car will influence the probability of having one or more accidents. Other examples of risk factors of these type can be given. Arbous and Kerrich (1951) list as the more important of the psycho-physiological qualities of individuals (pp. 344 - 345): health, age and experience, alcohol, and fatigue. They arrive at a list of just these four on the basis of the results of prior studies of, among others, Newbold (1926), Farmer and Chambers (1933), and Vernon (1936). Mehring (1962) considers a number of possible risk factors. He distinguishes four groups of risk factors: risk factors related to the car, to the area, to the driving intensity and to the driver. He discusses the influence of impersonal risk factors such as the state of the road, of objective risk factors related to the driver such as sex, age, civil status, duration of the possession of the driving licence and occupation, as well as of subjective risk factors related to the driver such as intelligence, character, natural ability, health conditions and traffic mentality.

c. We doubt whether the probability of a certain number of accidents in a given interval is independent of the place on the distance axis. This probability will depend on the driver's driving experience, and therefore on the length of the distance he has covered since he began driving.

d. Besides the problems mentioned above, we run into an estimation problem. Assuming λ is a constant for an individual motorist, we have to consider the problem that the value of λ can vary from the one individual motorist to another, due to factors the motorist is exposed to.

With respect to these problems the following discussion may be of relevance.

Firstly, the distance axis may be divided into intervals with equal value of the parameter λ. Then the Poisson distribution holds for each interval. Because the sum of a number of Poisson variables is again Poisson distributed, we can use the Poisson distribution for the whole distance covered.

Therefore, let some individual motorist drive some given distance within some given time period. Assume

(i) that the network of the roads, where the motorist covers the given distance, can be classified into a number of parts (or sets of parts) of the network of roads, where for each part the roads, the traffic density, etc. are comparable, in such a way that the three basic assumptions are satisfied;

(ii) that the time period can be divided into a number of s time intervals (or sets of intervals), in such a way, that in time intervals with equal length the probability distribution of the number of accidents is the same, and that the numbers of accidents in non-overlapping time periods are stochastically independent;

(iii) that during the driving of the k distance units the motorist's health conditions, the weather conditions and the state of the car does not change.

We define the interval δ_{ij} as the length of a distance interval which is covered driving at the part of the network of roads i (i = 1, ..., r), during a time interval j (j = 1, ..., s). Further we define, for each interval, λ_{ij} as the average number of accidents in a distance unit of length 1. Then, according to the basic assumptions, the number of accidents, N_{ij}, which occur to an individual motorist while driving the distance interval δ_{ij}, is Poisson distributed with parameter equal to $\delta_{ij}\lambda_{ij}$. The distribution of the number of accidents N, occurring to an individual motorist who drives a distance with a length equal to the sum of the lengths of rs intervals, is given by

$$p(n;\ \lambda,\ \delta) = \frac{(\lambda\delta)^n}{n!}\, e^{-\lambda\delta} \qquad n = 0, 1, 2, \ldots$$

where

$$N = \sum_{i=1}^{r}\sum_{j=1}^{s} N_{ij}\ , \quad \lambda = \frac{1}{\delta}\sum_{i=1}^{r}\sum_{j=1}^{s}\lambda_{ij}\delta_{ij} \quad \text{and} \quad \delta = \sum_{i=1}^{r}\sum_{j=1}^{s}\delta_{ij}$$

Secondly, assuming λ_{ij} is not a constant, but varies owing to changes in the weather condition, the driver's health, etc., we can consider the different values of λ_{ij} as the outcomes of a random variable Λ_{ij} which follows some probability law, as is proposed by many authors, such as Greenwood and Yule (1920). If, for example Λ_{ij} is chosen to be gamma distributed, the resulting unconditional probability distribution of the number of accidents is negative binomial. The choice of the gamma distribution as the prior distribution is obvious, because this distribution is the natural conjugate of the Poisson distribution. Kupper (1960) remarks: "Die Einfachheit der entstehenden Waarscheinlichkeitsverteilung rechtfertigt diese Wahl" (p. 457). Kupper suggests other prior distributions, such as the beta and the truncated normal distribution. We think that the beta distribution is not a proper distribution for λ_{ij}, because its values are, in principle, unbounded from above. Instead of the beta distribution, we introduce the lognormal distribution as a prior distribution for λ_{ij}.

A disadvantage of this solution is that it may be sometimes difficult to obtain a proper prior distribution for λ_{ij}. Moreover, this additional assumption may result in unmanageable formulae, if we do not put restrictions on the values of the paramaters of the distribution of the random variable Λ_{ij}, as will be shown below.

The density function of Λ_{ij} is given by

$$f(\lambda_{ij};\ \alpha_{ij},\ \beta_{ij}) = \frac{\beta_{ij}^{\alpha_{ij}}}{\Gamma(\alpha_{ij})}\ \lambda_{ij}^{\alpha_{ij}-1}\ e^{-\beta_{ij}\lambda_{ij}} \qquad \lambda_{ij} \geq 0$$

The unconditional density function of N_{ij}, given that the conditional distribution is Poisson, is given by

$$g(n_{ij};\ \alpha_{ij},\ \beta_{ij}) = \frac{\Gamma(n_{ij} + \alpha_{ij})}{n_{ij}!\ \Gamma(\alpha_{ij})} \left(\frac{\beta_{ij}}{\beta_{ij} + \delta_{ij}}\right)^{\alpha_{ij}} \left(\frac{\delta_{ij}}{\beta_{ij} + \delta_{ij}}\right)^{n_{ij}}$$

(where $n_{ij} = 0, 1, 2, \ldots$). Thus for each combination of some part of the network of roads i and some time interval j, the number of accidents N_{ij} is negative binomial distributed.

The distribution of $N = \Sigma\,\Sigma\,N_{ij}$, the number of accidents driving $\delta = \Sigma\,\Sigma\,\delta_{ij}$ distance units, however, is only negative binomial distributed, if δ_{ij}/β_{ij} is some constant for each i and j, implying that there would be a functional relationship between δ_{ij} and β_{ij}. Therefore, one has to choose between a solution where the resulting distribution is manageable, but where a more or less unrealistic assumption is made, and a solution where the resulting distribution is mathematically difficult to manage.

Thirdly, the difficulty in estimating the probability distribution of an individual motorist from a heterogeneous population of motorists can be met in different ways. 1) We can try to construct homogeneous groups of motorists, and fit Poisson distributions for each group. Problems may arise with respect to the construction of such homogeneous groups of motorists: the sample may be too small, or there may be insufficient information about discriminating factors. 2) We can consider λ as a function of a set of (explanatory) variables, and try to estimate λ on the basis of the values of these variables. 3) We can consider the different values of λ as outcomes of a random variable Λ, which follows some probability law.

Fourthly, if we want to predict future number of accidents, we meet the difficulty that we are in principle completely ignorant of the kind of distance intervals the motorist will drive, the points of time in which these distance intervals will be covered, the future weather conditions and the future driver's conditions, and so on. Moreover, we are ignorant of the probability distribution of Λ.

The final answer to the problems mentioned above can only be given by empirical investigations, with samples of sufficiently large size. As Cramér (1955) points out: "There can be no doubt that (...) assumptions (...) do involve some idealization of observed facts. Nevertheless, it is believed that they are sufficiently realistic to serve as a first approximation that may be practically useful in many cases" (p. 20). We assume, therefore, as a first approximation, the three basic assumptions to be fullfilled. Moreover, we fit three compound Poisson distributions, where the Poisson parameter is, subsequently, gamma, truncated normally and lognormally distributed.

2.3 Probability models for the amount of damage

There is a lot of literature about models for the number of accidents, but no extensive literature on probability models for the amount of damage of accidents. There is no theoretical derivation of a distribution for the amount of damage. Beard et al. (1977) state that "the existence of the distribution function S(z) [of the size of the claims, and therefore of the amount of the damage (added by the authors)] is an axiom in the theory of risk" (p. 18). An extensive review of distributions, which can be considered as distributions for the amount of damage of accidents, is given by Kupper (1962). However, he does not give empirical evidence for the applicability of these distributions. Kupper mentions the following distributions: the gamma distribution (already suggested by Bühlmann (1958)), the lognormal distribution (already suggested by Henry (1938), Latsche (1956) and Bühlmann (1958)), the exponential distribution (already suggested by Cramér (1955, pp. 40 - 43), Hofmann (1955), Amneter (1957) and Segerdahl (1957)), the Pearson type V and type VI distributions, the one-parameter and two-parameter Pareto distributions (also suggested by Depoid (1961)), and the Weibull and the truncated normal distributions. Weber (1970) suggests a mixture of two exponential distributions, and a compound exponential distribution. Beard et al. (1979, pp. 36 - 37), mention the exponential, the Pareto, the lognormal, and the Pearson type IV distribution.

Bühlmann and Hartman (1956) found that a truncated lognormal distribution fitted very well to data of industrial accidents. Leimkuhler (1963) concludes with a lognormal probability distribution for the amount of damage of accidents on the basis of a sample of 200 accidents involving tractor-semi-trailers. Weber (1970) found that a mixture of two exponential probability distributions provided a close fit to his sample of claim sizes. Van der Laan and Boermans (1970) found that the lognormal probability distribution agrees very well with a sample of claim sizes of about 1000 accidents reported, in contrary with the Pareto and the exponential distribution. We must keep in mind that these four studies have been based on samples which are probably not drawn from a homogeneous population. Newhouse et al. (1980) suggest the Box-Cox family of distributions. They fit that family of distributions with success to a set of data involving medical expenses. Hogg and Klugman (1983) state that "size-of-loss distributions in casualty insurance have mainly approximate Pareto, generalized Pareto, Burr and log-t distributions"[4]. They fit these probability models to two data sets, one involving hurricane losses and one involving malpractice claims.

The estimation of the parameters of stated probability distributions will

---

4 Cf. Hogg and Klugman (1983, p. 91).

give rise to similar problems to those met when we discussed the estimation of the parameter(s) of the probability distributions of the number of accidents. The parameters of the probability distribution of the amount of damage will vary in principle from the one motorist to another. We must, therefore, try to construct groups of motorists which are homogeneous with respect to the amount of damage.

As there are no theoretical considerations which lead to a probability model for the amount of damage, we shall let our choice depend on empirical considerations. Five theoretical distributions are considered: the lognormal distribution, the gamma distribution, the sum of two exponential distributions, a compound exponential distribution (where the parameter of the exponential distribution is assumed to be gamma distributed), and the normal distribution, after the observations are transformed by the Box-Cox transformation. This transformation reads as follows. Let Y be some continuous random variable. Then Box and Cox (1964) state that for some distributions of Y, there is a value of λ, such that the random variable

$$Z = \frac{Y^{\lambda} - 1}{\lambda} \qquad \text{for } \lambda \neq 0$$

$$= \log Y \qquad \text{for } \lambda = 0$$

is normally distributed.

3. An application of the model to Dutch car insurance data

3.1 The data

We have available insurance and claim data concerning 11981 passenger car insurances with respect to the years 1971 and 1972, which we obtained from the insurance company mentioned in footnote 1. This company could not provide data on the number of miles the insureds drive per year. They could only give information regarding the insured's statement if he drives more or less than 20000 kilometres (is 12427.5 miles) per year. The data, therefore, had to be supplemented with data on the number of miles the insureds drive per year.

These data were provided by the Netherlands Central Bureau of Statistics from an investigation on the possession and the use of passenger cars in the Netherlands. How both sets of data were combined is described by Van der Laan and Louter (1985). They also give a detailed description of the data used in the present paper.

Concerning each insured we have information about variables relating to the insured, such as his age and the degree of urbanization of his residence; relating to the car, such as its list price, its age, its maximum speed and its weight; and relating to the insurance conditions, such as the type of insurance, the (maximum) casco sum insured, the additional excess casco (if present), the area of coverage (home, or home and abroad), and the number of claim-free years.

In the case of some insureds, changes have occurred in the values of some of these variables, during the time of observation. When a change took place, we have information about the type of change, as well as about the month in which the change occurred.

With respect to each (reported) accident we have information about the month in which the accident occurred, the amount of indemnity paid for third-party damage as well as for casco damage, and information whether the accident was a particular one.

The information available does not entirely answer the purpose for which it is used in the present paper. For example, we cannot investigate whether female drivers show an accident behaviour which is different from that of male drivers, because we lack information about the sex of the insured. Further, we would like to have available information about the total number of accidents. The given information concerns information about the number of reported accidents, i.e., for which indemnification has been paid. We lack information about an unknown number of unreported accidents. Finally, the information about the annual mileage is rather inaccurate. The data provided by the Netherlands Central Bureau of Statistics are based on estimates from the inquired vehicle owners. White (1976) concludes that vehicle "owners tend to overestimate VMT (vehicle miles of travel) on low usage vehicles and underestimate the annual mileage on high usage vehicles. As a result, the use of owner estimates of annual VMT will invalidate accident involvement rate comparisons among those vehicle groups which differ with regard to annual usage." However, he also points out that "individual estimates may be combined to form a very reliable estimate of the mean annual VMT." Unfortunately, more reliable information on annual mileage of Dutch vehicle owners is not available, so we must make the best of what we have.

Van der Laan and Louter (1985) explore which of the variables, about which information is available, make a significant contribution to the "explanation" of the number of claims per mile travelled, and of the amount of damage. It appears that the most important risk factors for the number of claims are the number of claim-free years, in particular, and the insured's age and the degree of urbanization, to a less extent. There appears to be a weak association between the claim size and factors such as the price of the car, the area of coverage of the insurance and the degree of urbanization. On the basis of their results we conclude to a number of more or less homogeneous groups of insureds. For a selection from these groups we fit probability distributions to the number of claims per mile travelled and to the amount of damage.

3.2 Fitting distributions to the data

In this subsection we give, for both categories of insurances, the result of the process of fitting of four probability distributions to the number of claims per mile travelled, and of five probability distributions to the size of claim.

It is a matter of course that there are big differences between the number of miles the insureds drove per year. The average number of miles of all insureds considered is about 10500 miles and the standard deviation is about 4500 miles. We consider, subsequently, the number of claims of those, more or less, homogeneous groups of insureds, who drove at least m miles during the period considered, in their first m miles, where m equals 5000, 10000 and 15000, respectively.

The parameters of the distributions are estimated on the basis of the maximum likelihood estimation procedure. The goodness of fit is compared by means of the chi-square test statistic, X^2, although it is known that when this test statistic is based on maximum likelihood estimators, it is not χ^2-distributed (cf. Chernoff and Lehmann (1954)).

We remark that the variance of the negative binomial distribution, as well as the variance of the Poisson lognormal distribution, always exceeds the corresponding mathematical expectation. In the cases that the sample mean exceeds the sample variance, the fit of this distributions tends to an unconditional Poisson distribution. The Poisson truncated normal and the Poisson lognormal distribution show similar results. In such cases we will not fit the three compound Poisson distributions. This implies that we do not have an alternative for the Poisson distribution in these cases.

Because of the low average number of claims per m miles, on the one hand, and the division of the sample of insureds in more or less homogeneous groups, on the other hand, the number of insureds having two or more claims per m miles is relatively small. Therefore, it is sometimes impossible to test the goodness of fit with the help of the chi-square test, because the number of degrees of freedom is sometimes less than one.

Table 1 shows the results of the fits. Group 1 is the group of third-party insureds, group 2 that of all-risk insureds. The results of both groups are of interest for the model of the costs of claims. The results of group 2 can also be used to estimate the values of the parameters of the model for the costs of accidents. It appears that the Poisson distribution fits sometimes rather well and sometimes very good in all cases of third-party claims, and also in the cases of all-risk claims for insureds having 0 or 1 claim free year. The compound Poisson distributions are not applied in six of the twelve cases, where the sample mean exceeds the sample variance. The fit of the negative binomial distribution is good in the other six cases, although it must be remarked that there are relatively few classes with over one observation. The results of the fits of the other two compound Poisson distributions indicate that they are comparable with or inferior to the fit of the negative binomial distribution. Moreover, we point out that these fits take much computer time.

Next, we consider the five distribution of the claim size. The parameters of the distributions are estimated with the method of maximum likelihood, and the goodness of fit is tested by means of the chi-square test statistic, X^2. We compute for each fit the P-value related to a value obtained by the value of X^2, as if it is χ^2-distributed with k-r-1 degrees of freedom, where k is the number of classes and r is the number of unknown parameters. Chernoff and Lehmann (1954) proved that the real value of the P-value is higher than the one based on the $\chi^2(k-r-1)$ distribution, but lower than the one based on the $\chi^2(k-1)$ distribution. Hence, we must take into account that the real P-values exceed the computed values.

The third-party claim sizes are, in general, equal to the amounts of third-party damage associated with the accidents. Therefore, we can fit unconditional distributions to the third-party claim size. With respect to the all-risk insureds, we restrict the analysis of the claim size to insureds, who have the minimal excess of Dfl. 100 or Dfl. 150. We fit conditional distributions to the all-risk claim size. Firstly, we take as point of truncation Dfl. 150, secondly, we take Dfl. 500. The reason for this higher point of truncation is, that we expect that claims with low sizes are underrepresented in the sample. For, it can be profitable for an insured not to claim a low amount of damage, so that he does not lose his no-claim discount. Moreover, claims for casco damage of accidents caused by drivers of 23 years old or younger are underrepresented, because these drivers had an additional excess of Dfl. 150 for each claim.

Table 2 gives the approximating P-values for a selection of groups of insureds. From these results it is clear that the lognormal distribution and the Box-Cox family fit very well to the set of third-party claim data. For all-risk claim sizes, and point of truncation Dfl. 150, only the Box-Cox family shows a good fit. To get acceptible fits of the other distributions it is clear that we must not truncate the distribution at a point close to the amount of the excess. Moreover, we see that the gamma distributions gives good fits only in the case of all-risk claims with a high point of truncation.

Figure 1 gives a picture of the goodness of the fit of the conditional lognormal distribution and the conditional Box-Cox family in the case of all-risk claims of all all-risk insureds with point of truncation equal to Dfl. 500. Although the P-values corresponding to these fits are rather high, the histogram of the empirical distribution suggest that the set of data is not homogeneous.

It is known that the maximum likelihood estimation method does not produce the highest values of the chi-square test statistic. The highest values are obtained if we apply the minimum chi-square estimation method. However, this method leads to computation problems if we base the statistic on equiprobable classes, for, then the statistic is not a continuous function of the parameters. We can approximate the optimum estimates of the parameters by means of a trial and error procedure. This is done in the case of the set of all third-party claims, for the lognormal distribution and the Box-Cox family. The approximating value of the P-values are 0.1919 and 0.3097, respectively. These values are much higher than those we obtained by applying the maximum likelihood method, where the corresponding values are 0.0056 and 0.0405. The result of all four fits are pictured in Figure 2, where the number 1 denotes

Table 1 Frequency distributions of the number of claims per m miles

number of claims	observed frequency	Poisson	Poisson gamma	Poisson truncated normal	Poisson lognormal
Group 1: third-party claims, all insureds, m = 5000 miles, 7276 insureds					
0	6965	6961.0	6965.3	6965.4	6965.5
1	301	308.1	299.8	299.6	299.5
2	9	6.9	10.9	11.0	11.0
> 2	1				
X^2		1.54	0.08	.10	.11
Group 1: third-party claims, all insureds, m = 10000 miles, 1449 insureds					
0	1332	1331.1	1331.8	1331.9	1331.8
1	111	113.0	111.5	111.4	111.6
> 1	6	4.9	5.6	5.7	5.6
X^2		0.27	0.03	0.02	0.04
Group 1: third-party claims, all insureds, m = 15000 miles, 725 insureds					
0	660	657.4	659.8	659.7	659.5
1	59	64.4	59.8	60.0	60.4
> 1	6	3.3	5.4	5.3	5.1
X^2		2.77	0.09	0.11	0.20
Group 1: third-party claims, insureds with 0 or 1 claim free year, m = 5000 miles, 1523 insureds					
0	1319	1326.0	*)	*)	*)
1	197	183.7			
2	7	12.7			
> 2	0	0.6			
X^2		4.18			
Group 1: third-party claims, insureds with 0 or 1 claim free year, m = 10000 miles, 205 insureds					
0	147	150.0	*)	*)	*)
1	52	46.8			
2	6	7.3			
> 2	0	0.8			
X^2		1.69			

Table 1 Continued

number of claims	observed frequency	Poisson	Poisson gamma	Poisson truncated normal	Poisson lognormal
Group 1: third-party claims, insureds with 0 or 1 claim free year, m = 15000 miles, 97 insureds					
0	67	66.9	*)	*)	*)
1	24	24.8			
> 1	6	5.2			
X^2		0.14			
Group 2: all-risk claims, all insureds, m = 5000 miles, 3061 insureds					
0	2888	2881.5	2887.8	2886.3	2887.2
1	161	174.2	162.2	164.9	163.4
2	12	5.4	10.3	9.3	9.5
> 2	0		0.7	0.5	0.8
X^2		9.19	1.02	1.44	1.49
Group 2: all-risk claims, all insureds, m = 10000 miles, 1879 insureds					
0	1689	1668.7	1688.5	1677.2	1686.9
1	160	198.0	163.2	180.4	166.7
2	27	12.2	23.0	19.2	20.6
> 2	3		4.3	2.2	4.7
X^2		33.37	1.14	5.86	2.88
Group 2: all-risk claims, all insureds, m = 15000 miles, 647 insureds					
0	560	551.8	559.8	557.7	559.2
1	73	87.8	73.9	77.2	75.2
2	12	7.4	11.2	10.6	10.3
> 2	2		2.1	1.6	2.2
X^2		8.57	0.08	0.51	0.36
Group 2: all-risk claims, insureds with 0 or 1 claim free year, m = 5000 miles, 777 insureds					
0	625	629.2	*)	*)	*)
1	140	132.8			
2	12	14.0			
> 2	0	1.0			
X^2		1.75			

Table 1 Continued

number of claims	observed frequency	Poisson	Poisson gamma	Poisson truncated normal	Poisson lognormal
Group 2: all-risk claims, insureds with 0 or 1 claim free year, m = 10000 miles, 368 insureds					
0	249	251.6	*)	*)	*)
1	101	95.7			
2	15	18.2			
> 2	3	2.5			
X^2		0.96			
Group 2: all-risk claims, insureds with 0 or 1 claim free year, m = 15000 miles, 145 insureds					
0	87	90.7	*)	*)	*)
1	50	42.5			
2	6	10.0			
> 2	2	1.8			
X^2		3.08			

*) The sample mean exceeds the sample variance in this case.

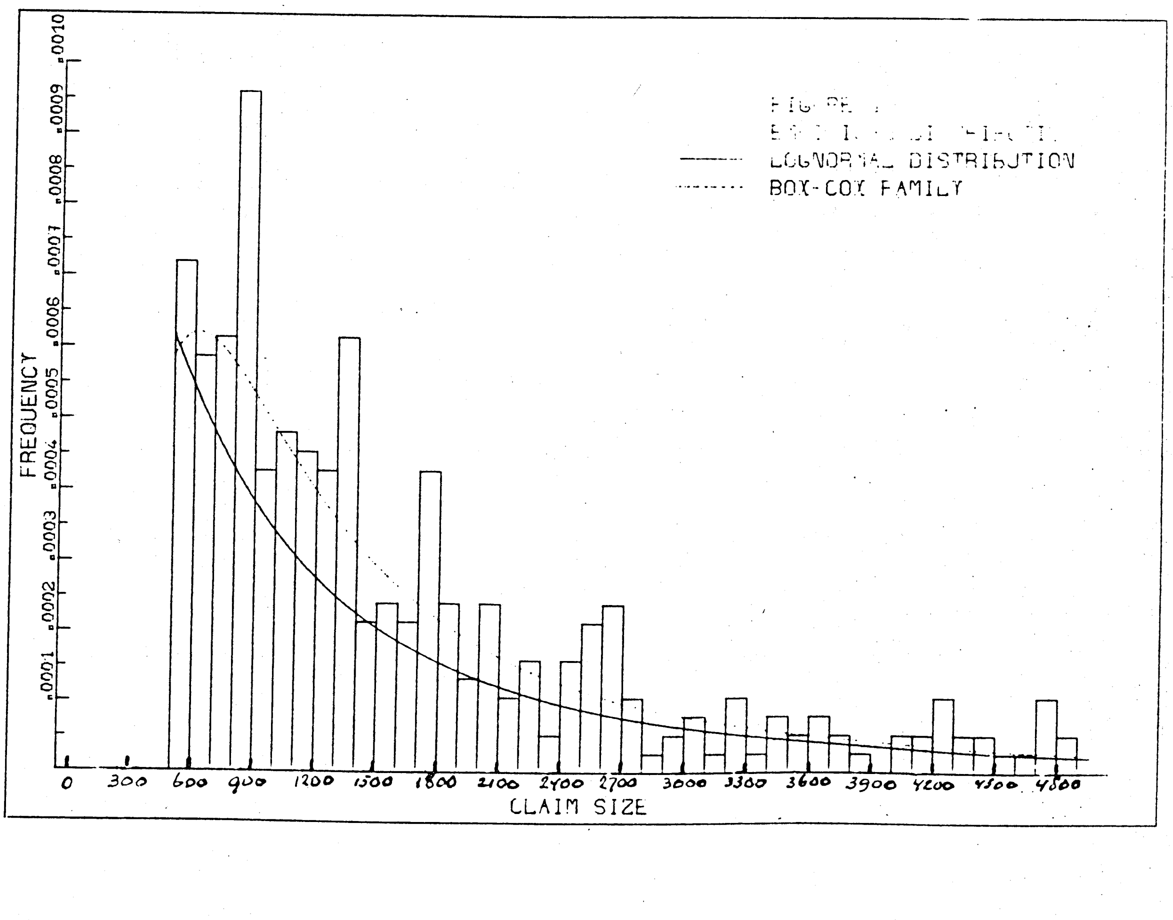
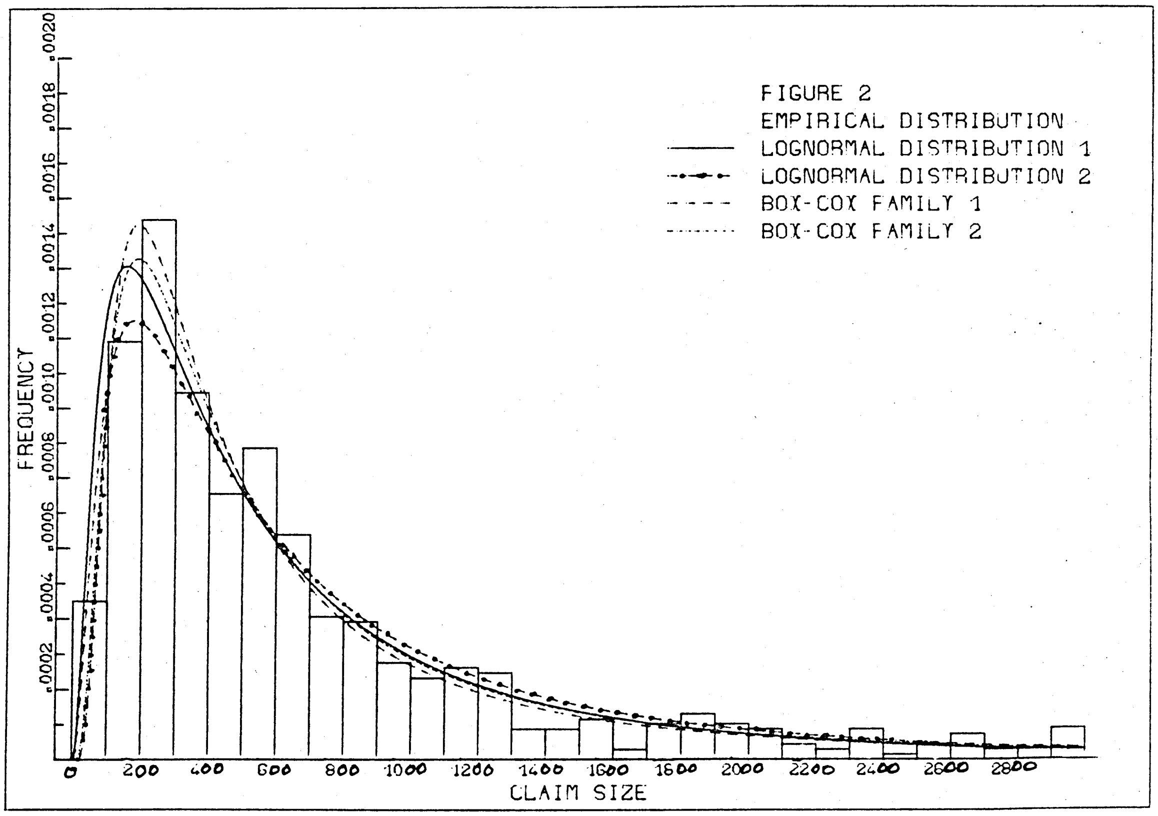


Table 2 P-values of the fits of five distributions to the claim size

	lognormal	gamma	mixed exponential	compound exponential	Box-Cox family
Group 1: third-party claims					
car insurances with car price < Dfl. 10000					
inland coverage only	0.3666	0.0006	0.0006	0.0062	0.2269
inland and abroad coverage	0.1056	0.0000	0.0000	0.0000	0.7652
total	0.0478	0.0000	0.0000	0.0000	0.0589
car insurances with car price ≥ Dfl. 10000					
inland coverage only	0.4073	0.1511	0.2346	0.1718	0.2346
inland and abroad coverage	0.0703	0.0000	0.0022	0.0492	0.2500
total	0.4064	0.0001	0.0292	0.0432	0.5967
all third-party car insurances					
inland coverage only	0.7202	0.0001	0.0001	0.0008	0.2906
inland and abroad coverage	0.1379	0.0000	0.0000	0.0000	0.3227
total	0.0056	0.0000	0.0000	0.0000	0.0405
Group 2: all-risk claims, point of truncation: Dfl. 150					
insureds living in					
- villages and small towns	0.3662	0.0000	0.2613	0.0140	0.6449
- medium sized and big towns	0.0216	0.0005	0.0203	0.1695	0.5371
total	0.0201	0.0000	0.0060	0.0013	0.0373
Group 2: all-risk claims, point of truncation: Dfl. 500					
insureds living in					
- villages and small towns	0.6245	0.2210	0.1102	0.4101	0.4574
- medium sized and big towns	0.2963	0.2505	0.2535	0.4747	0.3257
total	0.0831	0.0056	0.0333	0.0801	0.1896



the application of the minimum chi-square method, and the number 2 that of the maximum likelihood method. The picture shows clearly that the Box-Cox family 1 gives the best fit.

4. Conclusions

In this paper we derived a statistical model for the costs of claims of insureds, deduced from a model for the costs of passenger car traffic accidents. We showed that, given some assumptions, it is possible to estimate the parameters of the distributions of the number of accidents and of the amount of damage on the basis of claim data of insureds.

The application of the model was hampered by the fact that we had only available a limited set of data relating to 1 or 2 years. The number of insureds with 2 claims or more appeared to be, relatively, small. Therefore, we could not expect satisfactory results when we fit distributions to the number of claims. Nevertheless, the results obtained indicate that the Poisson distribution fits very well to the number of accidents occurred to an individual motorist diving some distance interval. The value of the Poisson parameter can be estimated more accurate from insurance and claim data, when we consider it as a function of a set of characteristics relating to the insured, his car, and the type of insurance, in spite of the no-claim rule.

We cannot derive, on theoretical grounds, a distribution for the amount of damage; however, two distributions show a very good fit to our set of claim data: the lognormal distribution and the Box-Cox family. The estimation of the parameter λ of the Box-Cox family is hampered by the fact that its maximum likelihood estimator cannot be determined analytically.

References

Amneter, H. (1957), "Die Ermittlung der Risikogewinne im Versicherungswesen auf risicotheoretischer Grundlage", Mitteilungen der Vereinigung schweizerischen Versicherungsmathematiker, Band 57, pp. 124 - 200.

Arbous, A.G. and J.E. Kerrich (1951), "Accident statistics and the concept of accident pronenes", Biometrics, Vol. 7, pp. 340 - 432.

Beard, R.E., T. Pentikäinen and E. Pesonen (1977), Risk theory, The stochastic basis of insurance, Second edition, Chapman and Hall, London.

Box, G.E.P. and D.R. Cox (1964), "An analysis of transformations" (with discussion), Journal of the Royal Statistical Society, Series B, Vol. 26, pp. 211 - 252.

Bühlmann, H. (1958), "Ein theoretischer Beitrag zur stochastischen Erfassung der Gesamtbetriebsunfallkosten", Mitteilungen der Vereinigung schweizerischen Versicherungsmathematiker, Band 58, pp. 53 - 65.

Bühlmann, H. and W. Hartmann (1956), "Änderungen in der Grundgesamtlichkeit der Betriebsunfallkosten", Mitteilungen der Vereinigung schweizerischen Versicherungsmathematiker, Band 56, pp. 303 - 320.

Chernoff, H. and E.L. Lehmann (1954), "The use of maximum likelihood estimates in χ^2 tests for goodness of fit", The Annals of Mathematical Statistics, Vol. 25, pp. 579 - 586.

Cramér, H. (1954), The elements of probability theory and some of its applications, Almqvist & Wiksell, Stockholm, and John Wiley & Sons, New York.

Cramér, H. (1955), "Collective risk theory, A survey of the theory from the point of view of the theory of stochastic processes", in: Jubilee Volume of the Försäkringsaktiebolaget Skandia, 1855 - 1955, Stockholm.

Depoid, P. (1961), "Problèmes statistique de l'assurance 'automobile'", Bulletin de l'Institut International de Statistique, Tome XXXVIII, 3e Livraison, pp. 3 - 21.

Dutt, A.D., D.W. Reinfurt and J.C. Stutts (1977), "Accident involvement and crash injury rates; an investigation by make, model and year of car", Accident Analysis and Prevention, Vol. 9, pp. 275 - 283.

Erlander, S. (1971), "A review of some statistical models used in automobile insurance and road accident studies", Accident Analysis and Prevention, Vol. 3, pp. 45 - 75.

Farmer, E. and E.G. Chambers (1933), "Tests for accident proneness", Industrial Health Research Board, Report No. 68.

Feller, W. (1968), An introduction to probability theory and its applications, Vol. I, 3rd edition, John Wiley & Sons, Inc., New York.

Feller, W. (1971), An introduction to probability theory and its applications, Vol. II, 2nd edition, John Wiley & Sons, Inc., New York.

Ferreira, J. (1971), Some analytical aspects of driver licencing and insurance Regulation, Technical Report No. 58, Operation Research Center, Massachusetts Institute of Technology, Cambridge, U.S.A.

Foldvary, L.A. (1975), "Road accident involvement per miles travelled - I", Accident Analysis and Prevention, Vol. 7, pp. 191 - 205.

Foldvary, L.A. (1976), "Road accident involvement per miles travelled - II", Accident Analysis and Prevention, Vol. 8, pp. 97 - 127.

Foldvary, L.A. (1977), "Road accident involvement per miles travelled - III" Accident Analysis and Prevention, Vol. 9, pp. 22 - 54.

Foldvary, L.A. (1978), "Road accident involvement per miles travelled - IV", Accident Analysis and Prevention, Vol. 10, pp. 143 - 176.

Foldvary, L.A. (1979), "Road accident involvement per miles travelled - V", Accident Analysis and Prevention, Vol. 11, pp. 75 - 99.

Goeller, B.F. (1969), "Modelling the traffic-safety system", Accident Analysis and Prevention, Vol. 1, pp. 167 - 204.

Greenwood, M. and G.D. Yule (1920), "An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurance of multiple attacks of decease or of repeated accidents", Journal of the Royal Statistical Society, Vol. 83, pp. 255 - 279.

Henry, M. (1938), "Etude sur le coût moyen des sinistres en assurance Responsabilité Civile Auto", Bulletin de l'Institut des Actuaires Francais.

Hofmann, M. (1955), "Über zusammengesetzte Poisson-Prozesse und ihre Anwendungen in der Unfallversicherung", Mitteilungen der Vereinigung schweizerischen Versicherungsmathematiker, Band 55, pp. 499 - 575.

Hogg, R.V. and A.T. Craig (1978), Introduction to mathematical statistics, 4th edition, The Macmillan Company, New York.

Hogg, R.V. and S.A. Klugman (1983), "On the estimation of long tailed skewed distributions with actuarial applications", Journal of Econometrics, Vol. 23, pp. 91 - 102.

Jorgensen, N.O. (1969), "A model for forecasting traffic accidents", Accident Analysis and Prevention, Vol. 1, pp. 261 - 266.

Kupper, J. (1960), "Wahrscheinlichkeitstheoretische Modelle in der Schadenversicherung. Teil I: Die Schadenzahl", Blätter der deutschen Gesellschaft für Versicherungsmatematik, Vol. 5, pp. 451 - 503.

Kupper, J. (1962), "Wahrscheinlichkeitstheoretische Modelle in der Schadenversicherung. Teil II: Schadenhöhe und Totalschaden", Blätter der deutschen Gesellschaft für Versicherungsmatematik, Vol. 6, pp. 95 - 130.

Latscha, R. (1956), "Zur Anwendung der kollectiven Risikotheorie in der schweizerischen obligatorischen Unfallversicherung", Mitteilungen der Vereinigung schweizerischen Versicherungsmathematiker, Band 56, pp. 275 - 302.

Leimkuhler, F. (1963), Trucking of radioactive materials: safety vs economy in highway transport, The John Hopkins Press, Baltimore.

Lundberg, F. (1909), "Über die Theorie der Rückversicherung", Transactions of the VIth International Congress of Actuaries, Vol. 1, pp. 877 - 955.

Mehring, J. (1962), "Die Schadenstruktur in der KraftfahrtHaftpflichtversicherung von Personenwagen", Blätter der Deutschen Gesellschaft für Versicherungsmathematik, Band VI, pp. 23 - 41.

National Safety Council (1971), Accident facts, 1971 edition, National Safety Council, Chicago.

Newbold, E.M. (1926), "Contribution to the study of the human factors in the causation of accidents", Industrial Health Research Board, Report No. 68.

Newhouse, J.P., J.E. Rolph, B. Mori and M. Murphy (1980), "The effect of deductibles on the demand for medical care services", Journal of the American Statistical Society, Vol. 75, pp. 525 - 533.

Panjer, H.H. and G.E. Willmot (1983), "Compound Poisson models in actuarial risk theory", Journal of Econometrics, Vol. 23, pp. 63 - 76.

Seal, H.L. (1969), Stochastic theory of a risk business, John Wiley & Sons, Inc., New York.

Segerdahl, C.-O. (1957), "If a risk business goes bankrupt, when does it occur? A basis for fixing net retention", Transactions of the XVth International Congress of Actuaries, New York, pp. 280 - 287.

Van der Laan, B.S. (1971) "A note on the probability distributions of the number and the amount of the claims of motor car accidents", Report 7103 of the Econometric Institute, Erasmus University Rotterdam, Rotterdam.

Van der Laan, B.S. (1979), "Motorists and accidents (an empirical study)", Report 7901 of the Econometric Institute, Erasmus University Rotterdam, Rotterdam.

Van der Laan, B.S. and F.J.J. Boermans (1970), "Some probability distributions in the field of motor car damages (an empirical study)", Report 7001 of the Econometric Institute, Erasmus University Rotterdam, Rotterdam.

Van der Laan, B.S. and A.S. Louter (1985), "On the number of claims and the amount of damage of passenger car traffic accidents in The Netherlands", Report 8529/A of the Econometric Institute, Erasmus University Rotterdam, Rotterdam.

Vernon, H.M. (1936), Accidents and their prevention, Cambridge University Press, Cambridge.

Weber, D.C. (1970), A stochastic model for automobile accident experience, Institute of Statistics, Mimeograph Series No. 651, North Carolina State University, Raleigh, North Carolina.

White, S.B. (1976), "On the use of annual vehicle miles of travel estimates from vehicle owners", Accident Analysis and Prevention, Vol. 8, pp. 257 - 261.

LIST OF REPORTS 1985

8500 "Publications of the Econometric Institute Second Half 1984: List of Reprints 378-400, Abstracts of Reports".

8501/O **R.M. Karp, J.K. Lenstra, C.J.H. McDiarmid and A.H.G. Rinnooy Kan,** "Probabilistic analysis of combinatorial algorithms: an annotated bibliography", 26 pages.

8502/E **M.E. Homan,** "Verschillen in consumptie tussen één en tweekostwinnerhuishoudens: een eerste analyse", 13 pages.

8503/O **A.W.J. Kolen,** "The round-trip p-center and covering problem on a tree", 17 pages.

8504/O **H.C.P. Berbee, C.G.E. Boender, A.H.G. Rinnooy Kan, C.L. Scheffer, R.L. Smith and J. Telgen,** "Hit-and-run algorithms for the identification of nonredundant linear inequalities", 32 pages.

8505/S **H. Brozius and L. de Haan,** "On limiting laws for the convex hull of a sample", 10 pages.

8506/O **J.B.G. Frenk and A.H.G. Rinnooy Kan,** "On the rate of convergence, to optimality of the LPT rule - postscript", 5 pages.

8507/E **P. Kooiman, H.K. van Dijk and A.R. Thurik,** "Likelihood diagnostics and Bayesian analysis of a micro-economic disequilibrium model for retail services", 35 pages.

8508/O **C.G.E. Boender, A.H.G. Rinnooy Kan and J.R. de Wit,** "A Bayesian procedure for the (s,Q) inventory problem", 26 pages.

8509/E **B.M.S. van Praag, S. Dubnoff and N.L. van der Sar,** "From judgments to norms: measuring the social meaning of income, age and educaton", 35 pages.

8510/S **B. Bode, J. Koerts and A.R. Thurik,** "On shopkeepers' pricing behaviour", 19 pages.

8511/M **H. Bart, I. Gohberg and M.A. Kaashoek,** "Exponentially dichotomous operators and inverse Fourier transforms", 70 pages.

8512/M **J. Brinkhuis,** "Normal integral bases and embedding of fields", 13 pages.

8513/E **P.M.C. de Boer,** "On the relationship between Revankar's and Lu-Fletcher's production function", 5 pages.

8514/S **L. de Haan,** "Extremes in higher dimensions: the model and some statistics", 15 pages.

8515 Publications of the Econometric Intitute First Half 1985: List of Reprints 401-414, Abstracts of Reports.

8516/A **A.R. Thurik and A. Kleijweg**, "Cyclical effects in retail labour productivity", 20 pages.

8517/C **B.M.S. van Praag and J. van Weeren**, "The impact of past experiences and anticipated future on individual income judgements", 24 pages.

8518/A **A.R. Thurik and J. Koerts**, "Behaviour of retail entrepreneurs". 16 pages.

8519/B **M. Hazewinkel, J.F. Kaashoek and B. Leynse**, "Pattern formation for a one dimensional evolution equation based on Thom's River Basin Model". 24 pages.

8520/A **H.K. van Dijk, T. Kloek and C.G.E. Boender**, "Posterior moments computed by mixed integration", 25 pages.

8521/B **J. Brinkhuis**, "Testing concavity and quasi-concavity is easy", 12 pages.

8522/A **R. Harkema**, "Minimum sample size requirements for maximum likelihood estimation of some demand models", 25 pages.

8523/B **A.C.F. Vorst**, "The general linear group of discrete Hodge algebras", 10 pages.

8524/A **A.M.H. Gerards and A.W.J. Kolen**, "Polyhedral combinatorics in combinatorial optimization", 25 pages.

8525/A **B.J. Lageweg, J.K. Lenstra, A.H.G. Rinnooy Kan and L. Stougie**, "Stochastic integer programming by dynamic programming", 19 pages.

8526/A **J.B. Orlin**, "A dual version of Tardo's algorithm for linear programming", 9 pages.

8527/A **B.M.S. van Praag**, "Household cost functions and equivalence scales", 24 pages (report 8424/E revised)

8528/A **S. Schim van der Loeff**, "Limited information maximum likelihood estimation of a subsystem of nonlinear equations", .. pages.

8529/A **B.S. van der Laan and A.S. Louter**, "On the number and the amount of damage of a passenger car traffic accidents in the Netherlands", ..pages.

8530/A **B.S. van der Laan and A.S Louter**, "A statistical model for the costs of passenger car traffic accidents", 30 pages.

Until report 8515 this series was devided in E(conometrics), S(tatistics), M(athematics) and O(perations Research). From report 8516 on it will be

A. Economics, Econometrics and Operations Research

B. Mathematics

C. Miscellaneous