ASSESSING INDIVIDUAL UNEXPLAINED VARIATION IN NON-LIFE INSURANCE

BY

OLA HÖSSJER, BENGT ERIKSSON, KAJSA JÄRNMALM AND ESBJÖRN OHLSSON

ABSTRACT

We consider variation of observed claim frequencies in non-life insurance, modeled by Poisson regression with overdispersion. In order to quantify how much variation between insurance policies that is captured by the rating factors, one may use the coefficient of determination, R^2, the estimated proportion of *total variation* explained by the model. We introduce a novel coefficient of individual determination (CID), which excludes noise variance and is defined as the estimated fraction of *total individual variation* explained by the model. We argue that CID is a more relevant measure of explained variation than R^2 for data with Poisson variation. We also generalize previously used estimates and tests of overdispersion and introduce new coefficients of individual explained and unexplained variance.

Application to a Swedish three year motor TPL data set reveals that only 0.5% of the total variation and 11% of the total individual variation is explained by a model with seven rating factors, including interaction between sex and age. Even though the amount of overdispersion is small (4.4% of the noise variance) it is still highly significant. The coefficient of variation of explained and unexplained individual variation is 29% and 81% respectively.

KEYWORDS

Claim frequency variation, coefficient of determination, coefficient of individual determination, unexplained individual variation, overdispersion, Poisson regression, rating factors.

1. INTRODUCTION

The idea behind modern non-life insurance rating is that each customer should pay a premium as close as possible to the expected value of the cost that he or she causes the company. Consequently, the pure premium (the premium without loading for expenses and cost of capital), should be close to the expected

Astin Bulletin 39(1), 249-273. doi: 10.2143/AST.39.1.2038064

value of the claim cost for each insurance policy. In practice, the actuary tries to fulfill this goal by finding *rating factors* that describe the variation in the expected cost between the policies. These factors are chosen so that the actuarial model will capture as much as possible of the variation in expectation between customers. On the other hand, the risk, i.e. the deviation of the claim cost from its expectation, is of course transferred to the company – in particular, it is not the goal to reduce the variance of the claim cost to zero.

A tariff analysis is most often carried out with the aid of Generalized Linear Models (GLMs), the theory of which is well summarized in McCullagh and Nelder (1989). Application of GLMs to non-life insurance has been considered, among others, by Brockman and Wright (1992). The tariff analysis is usually made separately for claim frequency and average claim severity, using multiplicative models (Jung, 1968). For GLMs, this corresponds to using a log-link function.

In this paper we focus on *claim frequency* under a multiplicative model. Let Y_i be the observed claim frequency for policy i and let γ_j^i denote the *price relativity* for rating factor number j for this policy compared to a reference policy, $j = 1, 2, \ldots, q$. The claim frequency of the multiplicative model can then be written

$$\lambda_i \doteq E(Y_i) = \lambda_0 \gamma_1^i \gamma_2^i \cdots \gamma_q^i, \tag{1}$$

where λ_0 is λ_i for the reference policy. The price relativities are connected to the GLM regression parameters $\beta = (\beta_1, \ldots, \beta_p)$ through the log-link,

$$\lambda_i = \exp(\beta^T x_i),$$

where $x_i = (x_{i1}, \ldots, x_{ip})$ is a vector of 0-1 dummy variables (covariates) indicating which particular parameters that apply to policy i.

In practice, there is always some variation left above the multiplicative model: two policies in the same tariff cell, i.e. with the same values on the rating factors, still have some residual difference in their expectation, unexplained by the multiplicative model. Our aim here is to present measures of explained and unexplained variation. This serves two purposes: (i) it is an aid in choosing rating factors for the model, cf. the use of R^2 in linear regression; (ii) it gives an indication of whether there is a need for experience rating (bonus/malus systems) at the individual level or not.

Several authors have suggested the use of credibility models for so called *optimal* bonus/malus rating, see Lemaire (1995) for an overview. As explained in Ohlsson and Johansson (2006) and Ohlsson (2008), credibility models can be viewed as random effect models, in particular this is convenient in a GLM context. The multiplicative model above then becomes, if U_i denotes the random effect for contract i,

$$E(Y_i | U_i) = \lambda_0 \gamma_1^i \gamma_2^i \cdots \gamma_q^i U_i = \lambda_i U_i. \tag{2}$$

Here $E(U_i) = 1$ and $\text{Var}(U_i)$ can be used as the basis for a measure of the amount of unexplained individual variation, as explained below. Without reference to GLMs, Bühlmann and Gisler (2005, Chapter 4.13) discuss similar models under the name credibility models with "a priori differences". In their Chapter 9, Bühlman and Gisler (2005) also discuss evolutionary credibility models, which allow the U_i for different observational years to have less than the 100% correlation implicitly assumed above.

While the total number of claims is what drives the cost of the insurance company, its variation is not an appropriate starting point for measuring the performance of the chosen rating factors. The rating factors determine how the total premium is distributed among the policy holders, but does not affect the number of claims, the cost or premium income of the company directly in a given portfolio of policies. The goal of a tariff analysis is not to reduce the cost of the company, other things equal, but to get the right price on a competitive market. The latter will, of course, in the end increase the revenue of the company, while the wrong price will result in adverse selection of customers.

To this end we suggest taking an individual perspective when measuring the performance of the chosen tariff, defining the total variation of a portfolio of insurance policies as the average mean square error of prediction, where "average" refers to choosing a policy at random, see Section 2.2.

We use a decomposition of this total variation in the portfolio into three parts: explained individual variation, unexplained individual variation and noise. This is similar to a decomposition defined by Johnson and Hey (1971) and Brockman and Wright (1992, Appendix D), who refer to explained and unexplained individual variation as between cell variance and within cell variance respectively. The *coefficient of determination*, R^2, is defined as the estimated fraction of *total variance* explained by the model. However, the noise part of the total variance, which is the Poisson variance in a model where there is nothing more to explain ($\text{Var}(U_i) = 0$) can never be explained. This suggests that a more relevant index is the *coefficient of individual determination* (CID), defined as the estimated proportion of the *total individual variance* explained by the model. It excludes noise variance and is (close to) one if we manage to explain (almost) all variation between policy means.

In non-life actuarial applications, the likelihood based deviance is often employed for model selection. In the same spirit, coefficients of determination may be defined using likelihood methods and deviance rather than variance decompositions, see for instance Maddala (1983), Cox and Snell (1989, pp. 208-209), Maggee (1990) and Nagelkerke (1991). However, we believe that a variance decomposition of the response variable (in our case claim frequency) is of particular interest to the experimenter, providing an intuitive explanation of the fraction of total variation that can be explained.

It is also of interest to test whether there is more variation left to explain or not. We present tests that generalize those of Venezian (1981, 1990) who only considers the special case with no covariates and constant duration. The tests might be used as an indication of the need for bonus/malus systems and/or a search for additional rating factors. We also present an estimate of the relative amount of overdispersion, ϕ, that differs slightly from the traditional one, based on Pearson's χ^2-statistic in that policies are weighted based on time duration, not estimated claim frequency. We also define coefficients of variation for the exlained and unexplained individual variation, as well as for the noise.

The paper is organized as follows. In Section 2 we define the model and variance decomposition in more detail. Parameter estimation is considered in Subsection 3.1, including definitions of R^2 and CID. Tests of excess variance are discussed in Subsection 3.2 and our findings are applied to Motor TPL (Third Party Liability) insurance in Section 4. We demonstrate, for a tariff with three year durations, that only 0.5% of the total variation (= R^2) and 11% of the total individual variation (= CID) in claim frequencies is explained. The explained and unexplained individual variation have coefficients of variation 29% and 81% respectively. Further discussion of the results is provided in Section 5 and more technical details are gathered in the appendix.

2. Variance Decomposition and Unexplained Individual Variation

2.1. A Mixed Poisson Model

Consider a portfolio of n insurance policies. For $i = 1, \ldots, n$, let N_i be the observed number of claims during a period of time, t_i, so that $Y_i = N_i / t_i$. It is assumed that conditional on U_i, N_i follows a Poisson distribution with expectation $t_i \lambda_i U_i$, and so the unconditional distribution is a mixed Poisson distribution, i.e.

$$N_i \in \mathrm{Po}(t_i \Lambda_i), \tag{3}$$

where $\Lambda_i = \lambda_i U_i$ and λ_i is given by (2). The Poisson model is frequently used in non-life insurance, see for instance Chapter 2 of Beard et al. (1984).

The unexplained individual variation is captured by the random variable Λ_i – in Motor TPL insurance this variable can be said to capture the *accident proneness* of the driver – with mean

$$E(\Lambda_i) = \lambda_i. \tag{4}$$

We assume a variance function

$$\mathrm{Var}(\Lambda_i) = \xi \lambda_i^a, \tag{5}$$

for the accident proneness for some $a > 0$ and $\xi \geq 0$. When $\xi = 0$, (2)-(4) define a generalized linear model with log link function.

For non-life insurance, $a = 2$ is most well known, since then $\mathrm{Var}(U_i) = \xi$ is independent of the covariate x_i and ξ becomes the squared coefficient of variation of $\Lambda_i|x_i$, a parameter independent of the chosen unit of time. The extension to time varying random effects (see the appendix) is also most natural for $a=2$. When $a=1$, $\xi=\text{Var}(\Lambda_i)/E(\Lambda_i)$ is the relative increase of variance caused by the overdispersion. We regard a as a constant and ξ as an unknown parameter. However, to keep the variance function more flexible, we will not restrict a in advance. See Pocock et al. (1981), Hinde (1982), Breslow (1984) and Lawless (1987) for more details on parameter estimation and choice of variance functions for overdispersed Poisson regression.

2.2. Variance Decomposition

For our purpose of measuring explained and unexplained variation, we first need a measure of the total variability in the portfolio. As explained above, the relevant measure here is not the variance of the total cost for the company, but rather an average of the variance for the individuals. For relevance, the average should be weighted with the time duration t_i, so that a one-year policy has the same impact as two half-year policies. Conceptually, this may be viewed as if we drew a policy at random from the portfolio, giving each policy i a probability proportional to its t_i. The mean claim frequency (of a *randomly* drawn policy) is then

$$\lambda = \sum_i t_i \lambda_i \;/\; \sum_i t_i, \tag{6}$$

where $\sum_i$ is short for $\sum_{i=1}^n$.

The benchmark for measuring the effect of choosing a tariff should be a tariff where all policies are assigned the average λ. (Note that λ differs from λ_0 in (1), which is the claim frequency of a reference policy, chosen to have a price relativity of one for all rating factors.)

The *average mean squared error of prediction* (AMSEP) for all $\{Y_i\}$ is

$$\sigma^2 = \sum_i t_i E\left((Y_i - \lambda)^2\right) / \sum_i t_i. \tag{7}$$

Now, since $E(Y_i|\Lambda_i) = \Lambda_i$ and $E(\Lambda_i) = \lambda_i$ we have the simple decomposition $E((Y_i-\lambda)^2) = (\lambda_i-\lambda)^2 + E((\Lambda_i-\lambda_i)^2) + E((Y_i-\Lambda_i)^2)$. Hence, since $E((Y_i-\Lambda_i)^2) = E(\text{Var}(Y_i|\Lambda_i)) = \lambda_i/t_i$, we can write σ^2 as a sum of three terms

$$\begin{aligned} \sigma^2 &= \sigma_1^2 + \sigma_2^2 + \sigma_3^2 \\ &= \textstyle\sum_i t_i(\lambda_i - \lambda)^2 / \sum_i t_i + \xi \sum_i t_i \lambda_i^a / \sum_i t_i + \sum_i \lambda_i / \sum_i t_i. \end{aligned} \tag{8}$$

The first term of (8), σ_1^2, quantifies explained individual variation, the second term σ_2^2 unexplained individual variation and the third term σ_3^2 represents noise, i.e. the variance in a Poisson model without overdispersion. We will refer to

$$\sigma_{\text{unexp}}^2 = \sigma_2^2 + \sigma_3^2 = \sum_i t_i E\left((Y_i - \lambda_i)^2\right) / \sum_i t_i \tag{9}$$

as the total unexplained variance. Following Venezian (1990), we also use the word excess variance for σ_2^2, since it quantifies the total excess of variance for all Y_i compared to what is expected under a pure Poisson model.

We have argued that the insurance company does not aim at predicting the customers' Y_i, but rather Λ_i. Hence, it would be more relevant to consider the AMSEP for $\{\Lambda_i\}$ rather than $\{Y_i\}$, i.e.

$$\sigma_{\text{ind}}^2 = \sum_i t_i E\left((\Lambda_i - \lambda)^2\right) / \sum_i t_i, \tag{10}$$

which could also be called the *total individual variation*; note that $\sigma_{\text{ind}}^2 = \sigma_1^2 + \sigma_2^2$.

A variance decomposition similar to (8) is defined by Johnson and Hey (1971) and Brockman and Wright (1992) when $a = 2$. The difference is mainly that they sum over tariff cells rather than policies and use a discrete approximation of accident proneness within each cell. With our approach we can handle continuous as well as discrete covariates.

Traditionally, the total variance is decomposed into explained and unexplained variance components, and the explained variance is further divided into various sources of variation. The special feature of (8) is that the *unexplained* variance is split into two terms representing individual variation and noise. It is a special case of a more general variance decomposition introduced by Hössjer (2008) for a large class of mixed regression models, including Poisson, logistic and linear regression.

2.3. Coefficients of Determination

To quantify the proportion of variance explained by the covariates, a traditional R^2-type quantity would be the fraction of the *total* variation,

$$\rho = \frac{\sigma_1^2}{\sigma^2},$$

while we have argued that it would be more relevant to use the fraction of the *total individual* variance,

$$\rho_{\text{ind}} = \frac{\sigma_1^2}{\sigma_{\text{ind}}^2},$$

which excludes the noise variance. Recall here that $\sigma_{\text{ind}}^2 = \sigma_1^2 + \sigma_2^2$.

The upper bound of ρ_{ind} is 1, corresponding to all relevant covariates being used in the model. The upper bound 1 of ρ requires, in addition, that the noise variance has been eliminated, which can only be achieved for very long time durations, t_i. Indeed, it is easy to see that ρ_{ind} is unaffected if all time durations are, for instance, doubled, whereas ρ is increased.

2.4. Measures of Overdispersion and Coefficients of Variation

To assess the amount of unexplained variance several possible quantities could be used, such as ξ or σ_2^2. A more intuitive choice is perhaps $1 - \rho_{\text{ind}} = \sigma_2^2/\sigma_{\text{ind}}^2$, which gives the proportion of total individual variance not explained by the covariates. Alternatively,

$$\phi = \frac{\sigma_{\text{un exp}}^2}{\sigma_3^2} = 1 + \frac{\sigma_2^2}{\sigma_3^2} \tag{11}$$

quantifies, in relative terms, the amount of excess of the total unexplained variance over the noise variance. A value larger than one indicates unexplained individual variation. However, ϕ shares the drawback of ρ in not being invariant with respect to magnified time durations. Alternatively we may use the coefficient of unexplained individual variation

$$\text{CUIV} = \frac{\sigma_2}{\lambda} = \sqrt{\frac{(\phi - 1)\sigma_3^2}{\lambda^2}} \overset{t_i \equiv 1}{=} \sqrt{\frac{\phi - 1}{\lambda}} \tag{12}$$

as a measure of overdispersion. It quantifies the individual unexplained standard deviation in relation to the mean and is invariant with respect to magnified time durations. We argue that CUIV it is a more intuitive and relevant measure of overdispersion than ϕ for actuarial applications.

The coefficient of explained individual variation

$$\text{CEIV} = \frac{\sigma_1}{\lambda} \tag{13}$$

and the coefficient of noise variation

$$\text{CNV} = \frac{\sigma_3}{\lambda} \overset{t_i \equiv 1}{=} \frac{1}{\sqrt{\lambda}}, \tag{14}$$

are two other quantities of interest. In fact, the variance decomposition (8) may be restated by decomposing the squared coefficient of variation

$$\mathrm{CV}^2 = \frac{\sigma^2}{\lambda^2} = \mathrm{CEIV}^2 + \mathrm{CUIV}^2 + \mathrm{CNV}^2$$

into three sources of variation.

3. Statistical Inference

3.1. Parameter Estimation

The unknown parameters, a, β and ξ, can be estimated using full maximum likelihood. This requires specification of the distribution of all Λ_i and yields quite complicated parameter estimates.

We will use a simpler approach, where first β is estimated separately by maximum likelihood from a generalized linear model without overdispersion ($\xi = 0$). This facilitates use of standard software and moreover, it can be shown that $\hat{\beta}$ is a consistent and asymptotically normal estimator of β even when $\xi > 0$, see e.g. White (1982).

The next step is to estimate a, as explained in the appendix for the car accidents data set. Given $\hat{\beta}$ and a, we then estimate ξ by

$$\hat{\xi} = \frac{\sum_i \left(t_i \left(Y_i - \hat{\lambda}_i \right)^2 - \hat{\lambda}_i \right)}{\sum_i t_i \hat{\lambda}_i^a}, \tag{15}$$

where $\hat{\lambda}_i = \exp(x_i \hat{\beta}^T)$ and a is regarded as known (for instance the estimated a, but not necessarily so). It is shown in Hössjer (2008) that asymptotically, in the limit of large samples n, $\hat{\xi}$ has a normal distribution with mean ξ when a is regarded as a known constant. An explicit formula for the standard error is also provided there.

The empirical version of the AMSEP for predicting claim frequency with the constant λ in (7) is, using (8),

$$\hat{\sigma}^2 = \hat{\sigma}_1^2 + \hat{\sigma}_2^2 + \hat{\sigma}_3^3 = \frac{\sum_i t_i \left(\hat{\lambda}_i - \hat{\lambda} \right)^2}{\sum_i t_i} + \frac{\sum_i \left(t_i \left(Y_i - \hat{\lambda}_i \right)^2 - \hat{\lambda}_i \right)}{\sum_i t_i} + \frac{\sum_i \hat{\lambda}_i}{\sum_i t_i}, \tag{16}$$

where $\hat{\lambda} = \sum_i t_i \hat{\lambda}_i / \sum_i t_i$. It gives rise to the coefficient of determination

$$R^2 = \hat{\rho} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}^2}, \tag{17}$$

Note that R^2 can be interpreted as the relative decrease in AMSEP we obtain for $\{Y_i\}$ by going from a constant claim frequency $\hat{\lambda}$ to a tariff of $\hat{\lambda}_i$'s, since its denominator is an estimate $\hat{\sigma}^2$ of the AMSEP with λ and the nominator is $\hat{\sigma}^2$ minus an estimate of the AMSEP with the λ_i, i.e. of σ^2_{unexp} in (9).

The empirical version of the more relevant AMSEP for prediction of Λ_i in (10) is

$$\hat{\sigma}^2_{\text{ind}} = \hat{\sigma}^2_1 + \hat{\sigma}^2_2. \tag{18}$$

leading to an alternative to R^2 which we call the *coefficient of individual determination*

$$\text{CID} = \hat{\rho}_{\text{ind}} = \frac{\hat{\sigma}^2_1}{\hat{\sigma}^2_{\text{ind}}}, \tag{19}$$

respectively.

Like R^2, CID can be interpreted as the relative decrease in AMSEP for $\{\Lambda_i\}$ resulting from introducing varying $\hat{\lambda}_i$'s. The denominator is an estimate $\hat{\sigma}^2_{\text{ind}}$ of the AMSEP with λ, see (10), while the numerator the difference of $\hat{\sigma}^2_{\text{ind}}$ and an estimator of the AMSEP with $\hat{\lambda}_i$, i.e. of

$$\sigma^2_2 = \sum_i t_i E\big((\Lambda_i - \lambda_i)^2\big) / \sum_i t_i. \tag{20}$$

When all time durations are equal, the R^2 here is an analogue of the classical R^2 used for (univariate) linear regression models, whereas CID has no such analogue. The reason is that that ρ_{ind} cannot be estimated for linear models, since the two components of the unexplained variance, σ^2_2 and σ^2_3, cannot be identified. On the contrary, CID is computable for both mixed Poisson and logistic regression models, as well as for multivariate linear regression models, see Hössjer (2008), where also standard errors of both R^2 and CID are provided.

In order to estimate ϕ from data, we use

$$\hat{\phi} = \frac{\sum_i t_i \left(Y_i - \hat{\lambda}_i\right)^2}{\sum_i Y_i}. \tag{21}$$

A slightly different version of $\hat{\phi}$ has $\sum_i \hat{\lambda}_i$ in the denominator instead. When all time durations are equal, it follows from the GLM likelihood estimating equations that the two versions are identical (see e.g. McCullagh and Nelder, 1989). A formula for the standard error of $\hat{\phi}$ is provided in the appendix. In the special case of constant time duration and no covariates ($\lambda_i \equiv \lambda$), $\hat{\phi}$ is essentially the index of dispersion, i.e. the ratio of the sample variance and sample mean. In Section 5, we show that (21) is closely related to the Pearson estimate $\hat{\phi}$ used in GLM theory.

The coefficients of variations (12)-(14) are estimated in the natural way, replacing σ_i^2 by $\hat{\sigma}_i^2$ and λ by $\hat{\lambda} = \sum_i t_i \hat{\lambda}_i / \sum_i t_i$.

3.2. Testing Excess Variance

In order to test for excess variance, we formulate the null hypothesis H_0 of no excess variance against the alternative H_1 of a positive excess variance, i.e.

$$\begin{aligned} H_0 &: \xi = 0, \\ H_1 &: \xi > 0, \end{aligned} \tag{22}$$

which is equivalent to testing $\sigma_2^2 = 0$ against $\sigma_2^2 > 0$ or $\phi = 1$ against $\phi > 1$. When the distribution of all Λ_i is specified, one may employ a likelihood ratio test to carry out (22). We will use a simpler approach based on excess variance, which only involves the first two moments of Y_i.

Our starting point is the excess variance statistic $\sum_i t_i (Y_i - \hat{\lambda}_i)^2 - \sum_i Y_i$, which agrees with the numerator of (15), except that $\sum_i \hat{\lambda}_i$ is replaced by $\sum_i Y_i$. (Again, the latter two sums are identical when all time durations are equal.) It is shown in the appendix that the standardized excess variance statistic

$$T = \frac{\sum_i t_i \left(Y_i - \hat{\lambda}_i\right)^2 - \sum_i Y_i}{\sqrt{2 \sum_i \hat{\lambda}_i^2}}, \tag{23}$$

has an approximate standard normal distribution for large samples. Hence, a test with an approximate significance level $1 - \alpha$ rejects H_0 when $T \geq \lambda_\alpha$, where λ_α is the $(1 - \alpha)$-quantile of a standard normal distribution. Since $T = c(\hat{\phi} - 1)$, with $c = \sum_i Y_i / \sqrt{2 \sum_i \hat{\lambda}_i^2}$, we may also regard T as standardized version of $\hat{\phi}$.

For constant time duration and no covariates, (23) amounts to testing overdispersion of stationary count data. Then the denominator of (23) simplifies to $\sqrt{2n}\hat{\lambda}$. This test has been used by Venezian (1981, 1990) for car accident data. An asymptotically equivalent approach, based on a χ^2-approximation of $\sum_i (Y_i - \hat{\lambda})^2$, has been considered by Fisher (1950) and Rao and Chakravarti (1956).

4. Car-Accident Data

We will analyze Swedish car accident data from If P&C Insurance Company. A detailed description of the data set can be found in Järnmalm (2006). Car accidents are registered for customers having a uninterrupted 3 year period in between January 1, 2002 and December 31, 2005. Shorter durations, in total approximately 30% of the total portfolio are thus excluded. The rating factors

TABLE 1

THE RATING FACTORS USED FOR THE CAR ACCIDENTS DATA SET.

Rating factor *j*	**Class**	**Variable**	k_j	**Class Description**
1: Customer years	1 2 3 4	0-2 3-5 6-10 11-	4	No. of years a customer has been insured in the company.
2: Geographic zone	0-18		19	A division of Sweden into 19 geographical zones.
3: Age of car	1 2 3 4 5 6	0-2 3-5 6-8 9-12 13-16 17-	6	
4: Premium class	0-9		10	Premium class is determined by type of car.
5: Driving distance	1-5		5	Five intervals of reported driving distances. A larger class index corresponds to a longer distance.
6: Sex			2	The sex of the customer.
7: Age	1 2 3 ⋮ 13	0-24 25-26 27-29 ⋮ 75-	13	The age of the customer. The classes 4-12 have five year intervals, 30-34, ..., 70-74.

are defined at the beginning of the risk period. The age depending factors, e.g. Age of car, are for this reason not as accurate as possible, the advantage on the other hand is that each individual's characteristics are kept in one data record. Hence, although our methodology in principle handles varying duration, the present data set has $t_i \equiv 1$, measuring time in three year intervals. The size of the data set is $n = 439\,283$, and customers report a total of 29405 accidents during the three year period.

The seven rating factors of the model are presented in Table 1. Although our variance decomposition handles continuous covariates, we have followed the current practice at If P&C and discretized the continuous covariates. This implies that rating factor j is divided into k_j classes. For $j = 1, \ldots, 5$, each class within the given rating factor has a distinct regression coefficient β_r, except for the class of the reference policy, which is chosen to have a fixed regression coefficient, 0, not included in β. We model interaction between sex and age

TABLE 2

ESTIMATED RELATIVE INCREASE OF THE ACCIDENT RATE, $\exp(\hat{\beta}_r)$ FOR SELECTED RATING FACTOR CLASSES AND CORRESPONDING WALD CONFIDENCE INTERVALS (CIS) WITH (APPROXIMATE) COVERAGE PROBABILITY 95%. FOR EACH (COMBINED) RATING FACTOR, WE HAVE ONLY INCLUDED THE TWO CLASSES WITH MINIMAL AND MAXIMAL $\exp(\hat{\beta}_r)$. THE CIS ARE CALCULATED WITH THE HELP OF THE STANDARD SOFTWARE (PROC GENMOD IN SAS) FOR GLM LOGLINK POISSON REGRESSION ML-ESTIMATION. HENCE THE OVERDISPERSION IS MODELED SLIGHTLY DIFFERENTLY THAN FOR THE MIXED POISSON DISTRIBUTION (3). THIS DOES NOT CHANGE THE PARAMETER ESTIMATES $\hat{\beta}_r$, BUT THE CIS ARE SLIGHTLY AFFECTED. THE DIFFERENCE IS HOWEVER NEGLIGIBLE, SINCE THE AMOUNT OF OVERDISPERSION $\hat{\xi}$ IS SMALL.

Rating factor	**Class**	$\exp(\hat{\beta}_r)$	$\exp(I_{\beta_r})$
Intercept		0.0590	(0.0470,0.0742)
Customer years	1	1.2364	(1.1922,1.2822)
	4	1.0000	(1.0000,1.0000)
Geographic zone	2	0.5475	(0.5056,0.5925)
	16	1.0021	(0.9468,1.0607)
Age of car	2	1.0641	(1.0242,1.1055)
	6	0.7279	(0.6811,0.7779)
Premium class	1	0.3988	(0.2484,0.6371)
	6	1.5873	(1.2815,1.9662)
Driving distance	1	0.8203	(0.7898,0.8520)
	5	1.2545	(1.1739,1.3407)
Sex/age	Female/13	1.4593	(1.3475,1.5471)
	Female/10	0.8740	(0.8125,0.9400)

(rating factors 6 and 7), giving k_6k_7 combined classes, of which one is chosen as reference. The covariates are chosen as $x_{i1} = 1$ (the intercept) and, for $r > 1$, $x_{ir} = 1$ if individual i belongs to the given (combined) class and 0 otherwise. The total number of regression coefficients is

$$p = 1 + \sum_{j=1}^{5}\left(k_j - 1\right) + (k_6k_7 - 1) = 65.$$

Table 2 shows results of a standard GLM analysis, including parameter estimates $\hat{\beta}_r$ and the associated confidence intervals of some selected regression coefficients. The average estimated claim frequency is

$$\hat{\lambda} = n^{-1}\sum_i \hat{\lambda}_i = n^{-1}\sum_{i=1}^{n} Y_i = 0.0669$$

per three year intervals.

The next step is to test for overdispersion. For the data set analyzed by Venezian (1990), the overdispersion is highly significant. Our conclusion is the same, since the test statistic for excess variance is

$$T = 19.89, \tag{24}$$

so that the null hypothesis of no excess variance is rejected at level 0.001. As a comparison, $T = 23.30$ for a model with no covariates. Hence the rating factors only decrease the significance of excess variance marginally.

To assess more explicitly the impact of the rating factors, we estimated the three components of the empirical variance decomposition (16) as

$$\begin{aligned} \hat{\sigma}_1^2 &= 0.0003732, \\ \hat{\sigma}_2^2 &= 0.0030, \\ \hat{\sigma}_3^2 &= \hat{\lambda} = 0.0669. \end{aligned}$$

Inserting these values into (17) and (19), we get surprisingly low coefficients of determination

$$\begin{aligned} R^2 &= 0.0053, \\ \text{CID} &= 0.1120. \end{aligned} \tag{25}$$

Only about 0.5% of the total variation and 11% of the total individual variation is thus explained by the rating factors. In Figure 1 both R^2 and CID are plotted as functions of time, assuming all policies in the portfolio have the same time duration τ years. The two individual variances are constant, whereas the noise variance σ_3^2 is inversely proportional to τ. Hence R^2 increases with τ whereas CID is constant. We notice that $R^2 = 0.18\%$ if $\tau = 1$ and that $\tau = 60.2$ is required in order for R^2 to reach 0.5CID. Of course, in practice, the time duration τ cannot be varied in this way. An insurance contract usually lasts for one year. On the other hand, it is still of interest to consider claims over several years, and then several policies may remain unchanged for at least, say, five years. In any case, Figure 1 illustrates that noise is by far the dominating source of variation for time durations used in practice and that unrealistically long durations would be required in order to reduce noise variance significantly.

The relative excess variance is estimated as

$$\hat{\phi} = 1.0442. \tag{26}$$

and the coefficients of variation as

$$\begin{aligned} \widehat{\text{CEIV}} &= 0.289 \\ \widehat{\text{CUIV}} &= 0.813 \\ \widehat{\text{CNV}} &= 3.866 \\ \widehat{\text{CV}} &= 3.962. \end{aligned}$$

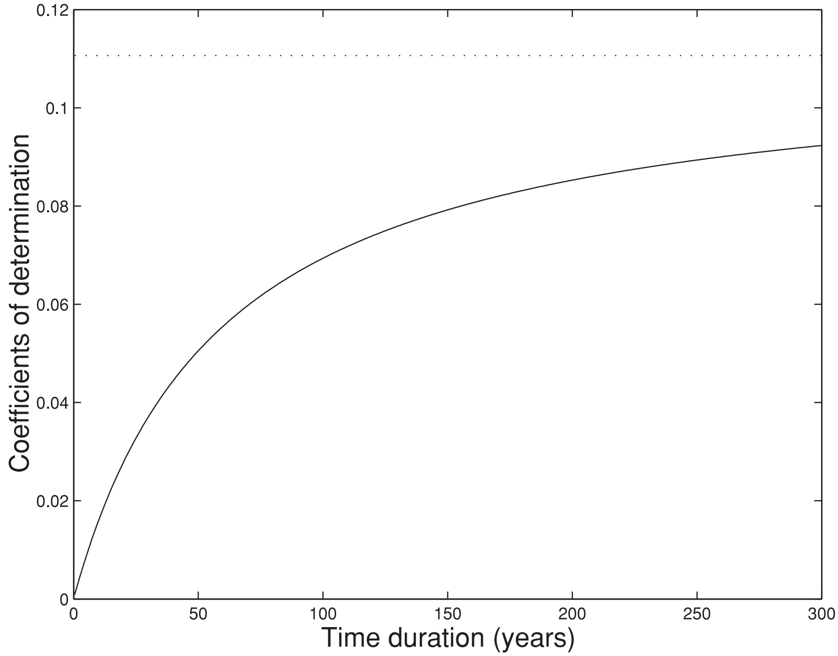

FIGURE 1: Plot of R^2 (solid line) and CID (dotted line) versus τ for the car accident data set, assuming all policies remain in the portfolio for τ years, with $\hat{\sigma}_1^2(\tau) = \hat{\sigma}_1^2$, $\hat{\sigma}_2^2(\tau) = \hat{\sigma}_2^2$ and $\hat{\sigma}_3^2(\tau) = 3\hat{\sigma}_3^2/\tau$.

Hence, the explained standard deviation is about 29% and the unexplained individual standard deviation about 81% of the average claim frequency. The noise coefficient of variation is much larger, 387%. As explained above, this is due to the short time durations.

In order to give confidence intervals for (selected) parameter estimates, we need to estimate the variance function, which requires estimation of a and ξ. A regression method, explained in the appendix, yields

$$(\hat{a}, \hat{\xi}_0) = (1.3051, 0.0946). \tag{27}$$

For the final estimate (15) of the dispersion parameter ξ we use two different values of a, which is regarded as known, and obtain

$$\hat{\xi} = \begin{cases} 0.0442, & \text{if } a = 1, \\ 0.0979, & \text{if } a = 1.3. \end{cases}$$

Here $a = 1.3$ is taken from the initial regression analysis (27) and $a = 1$ is chosen to yield a simple overdispersion model.

We report 95% Wald confidence intervals in Table 2 for selected regression parameters and in Table 4 for ρ, ρ_{ind}, ϕ and ξ. We notice that I_ρ, $I_{\rho_{\text{ind}}}$ and I_ϕ are very insensitive to the choice of a (1 or 1.3). This is due to the small amount of excess variance in the data, making the exact model of overdispersion less crucial. Hence we recommend using the simpler model with $a = 1$. Two versions of I_ϕ are reported based on standard errors defined in the appendix. The parametric model assumes a gamma distributed accident proneness, wheres the nonparametric model only includes the first four moments of $\Lambda_i | x_i$. They give essentially the same confidence intervals.

5. Discussion

In this paper, we have defined a general framework for quantifying explained and unexplained variation of claim frequencies in non-life insurance, including a new coefficient of determination, CID, generalizations of previously used estimates of relative overdispersion ($\widehat{\phi}$) and test statistics for overdispersion (T), as well as new coefficients of explained and unexplained individual variation, CEIV and CUIV respectively.

Our purpose is not to solve new actuarial problems, but rather to provide new, intuitive quantities that hopefully give new insights as regards to the quality of the chosen model. Our hope is that CID, CEIV and CUIV all become valuable and much used tools when the actuary selects rating factors and classes within each rating factor.

An application to a Swedish car accident data set reveals that the amount of overdispersion is highly significant, but yet small in relation to noise variance. This manifests itself by CID being much larger than R^2 and is explained by the fact that time durations in Motor TPL are very short in relation to claim frequencies. Similar analyses for other countries (see Järnmalm, 2006) show that although the amount of overdispersion varies, it is persistently significant but yet small in relation to noise variance.

Surprisingly, the proportion of explained variance is still very small after removing noise variance. We obtained CID = 11.2%, whereas higher, but still low values CID = 35.8% and CID = 31.2% can be deduced from the variance decompositions of Johnson and Hey (1971) and Brockman and Wright (1992) respectively. The low value of CID obtained for our data set and model may have several reasons:

1) The multiplicative risk assumption is only approximately correct. In particular, our model only includes interaction between two of the seven rating factors in Table 1.
2) The number of classes within some rating factor could be increased or replaced by continuous covariates.
3) The true claim frequencies λ_i may be time varying, not constant. See the appendix for more discussion on this topic.

TABLE 3

MEAN AND EXCESS VARIANCE FOR PREMIUM GROUPS (SEE APPENDIX).

I_j	$\tilde{\lambda}_j$	$\tilde{\sigma}^2_{\text{excess},j}$	$\|I_j\|$	Accidents
(0.000, 0.015)	0.0113	−0.00142	1513	13
(0.015, 0.025)	0.0200	0.00029	627	13
(0.025, 0.035)	0.0317	0.00091	7462	226
(0.035, 0.045)	0.0410	0.00144	35494	1448
(0.045, 0.055)	0.0505	0.00132	78455	3933
(0.055, 0.065)	0.0600	0.00323	100717	6092
(0.065, 0.075)	0.0697	0.00345	84785	5945
(0.075, 0.085)	0.0796	0.00350	57082	4518
(0.085, 0.095)	0.0896	0.00541	35490	3239
(0.095, 0.105)	0.0993	0.00196	20130	1962
(0.105, 0.115)	0.1093	0.00327	9942	1057
(0.115, 0.125)	0.1193	0.00483	4345	517
(0.125, 0.135)	0.1294	−0.00430	1818	220
(0.135, 0.145)	0.1394	0.00921	836	122
(0.145, 0.155)	0.1493	0.03867	391	61
(0.155, 0.165)	0.1590	0.14955	126	23
(0.165, 0.175)	0.1696	0.14359	42	9
(0.175, 0.185)	0.1806	0.06495	15	5
(0.185, 0.195)	0.1888	0.00247	8	2
(0.195, 0.205)	0.1988	−0.15927	3	0
(0.205, 0.215)	0.2097	−0.16571	2	0

TABLE 4

WALD CONFIDENCE INTERVALS $I_\theta = (\hat{\theta} - \lambda_{\alpha/2} d_{\hat{\theta}}, \hat{\theta} + \lambda_{\alpha/2} d_{\hat{\theta}})$ OF VARIOUS PARAMETERS θ. THE ASYMPTOTIC COVERAGE PROBABILITY IS 95% ($\alpha = 0.05$) AND $d_{\hat{\theta}}$ IS THE STANDARD ERROR OF $\hat{\theta}$. THE ASSUMED a IS EITHER 1 OR 1.3 AND I_ϕ IS EITHER NONPARAMETRIC (NP) OR PARAMETRIC (P).

θ	a	I_θ
ρ	1	(0.0049, 0.0058)
ρ	1.3	(0.0049, 0.0058)
ρ_{ind}	1	(0.0967, 0.1274)
ρ_{ind}	1.3	(0.0967, 0.1274)
ϕ	1	$(1.0383, 1.0500)_{\text{NP}}$
ϕ	1.3	$(1.0383, 1.0500)_{\text{NP}}$
ϕ	1	$(1.0384, 1.0500)_{\text{P}}$
ξ	1	(0.0383, 0.0500)
ξ	1.3	(0.0849, 0.1108)

4) A number of unknown individual characteristics are not included in the model. For instance, the annual driving distance is self-reported and may differ from the true one. Car drivers use different roads with varying risks, and this variation is only to some extent captured by geographical zone. The individual ability to drive safely is only to some extent explained by sex/age. Other factors, such as psychological make-up and drinking habits, cannot be included in the model.
5) Inclusion of customers with time duration less than three years in the portfolio may increase CID. These drivers typically have higher claim frequencies than average.

Since individual variation of claim frequencies is very complex, we don't state that 1-5) are enough to guarantee a CID of 100%, simply that they to some extent explain the low CID found in our data set. For more discussion on this theme we refer to Haight (2001), Lemaire (1995) and Brockman and Wright (1992).

Our work can be extended in several ways. A first extension is to consider time varying covariates (see item 4 above) and random effects, as described in more detail in the appendix.

A second extension is to use overdispersed Poisson distributions (ODP) rather than mixed Poisson distributions. For ODPs, the parameter ϕ is defined directly in terms of the variance function;

$$v_i = \mathrm{Var}(Y_i) = \phi\lambda_i / t_i, \tag{28}$$

for all policies $i = 1, \ldots, n$, see McCullagh and Nelder (1989). In general, for mixed Poisson distributions, (28) does not hold, and the definition of ϕ in (11) cannot be reformulated in terms of the variance function of individual policies. An exception is $a = 1$ and $t_i \equiv t$, in which case (28) is satisfied for mixed Poisson distributions as well, with $\phi = 1 + \xi t$.

Formally, the variance decompositions (8) and (16) can be defined for ODPs, provided we change the interpretation of v_i to that of (28). This in turn provides us with R^2 and CID for ODPs. The interpretion of unexplained individual variance and CID is less clear though, since ODP is not a mixed model, having no random effects.

A third extension, when p/n is non-negligible, is to account for reduced degrees of freedoms when defining R^2 and CID (see Hössjer, 2008), as well as $\widehat{\phi}$ and T. For our data set, this adjustment has a minor effect, since $p/n = 1.48 \cdot 10^{-4}$.

A fourth extension is to replace t_i by other weights w_i when defining λ, σ^2, the variance decomposition (8), ρ, ρ_{ind} and ϕ. Various weighting schemes are discussed in Hössjer (2008). One possibility is inverse variance weighting $w_i = t_i / \lambda_i$. This choice of weights results in all policies having approximately the same contribution to the unexplained part of σ^2, since

$$w_i \mathrm{Var}(Y_i) \approx 1,$$

where the approximation is exact in absence of overdispersion. Since these weights involve unknown parameters, we use estimated weights

$$\hat{w}_i = t_i / \hat{\lambda}_i \tag{29}$$

to compute $\hat{\lambda}$, $\hat{\sigma}^2$, the empirical variance decomposition (16), R^2, CID and $\hat{\phi}$. We may also generalize the version of T with $\sum_i \hat{\lambda}_i$ instead of $\sum_i Y_i$ in the numerator to

$$T = \frac{\sum_i \hat{w}_i \left(Y_i - \hat{\lambda}_i\right)^2 - \sum_i \hat{w}_i \left(\hat{\lambda}_i / t_i\right)}{\sqrt{2 \sum_i \hat{w}_i^2 \left(\hat{\lambda}_i / t_i\right)^2}} \overset{(29)}{=} \frac{\chi^2 - n}{\sqrt{2n}} = 19.864, \tag{30}$$

where

$$\chi^2 = \sum_i \frac{t_i \left(Y_i - \hat{\lambda}_i\right)^2}{\hat{\lambda}_i} = 457\,901$$

is the unscaled Pearson statistic (Pearson, 1900) for Poisson regression. The version of $\hat{\phi}$ with $\sum_i \hat{\lambda}_i$ in the denominator is generalized to

$$\hat{\phi} = \frac{\sum_i \hat{w}_i \left(Y_i - \hat{\lambda}_i\right)^2}{\sum_i \hat{w}_i \left(\hat{\lambda}_i / t_i\right)} \overset{(29)}{=} \frac{\chi^2}{n} = 1.0424, \tag{31}$$

which agrees with the Pearson definition of $\hat{\phi}$, except for using n instead of $n-p$ in the denominator. We notice that (30) and (31) only differ marginally from (24) and (26). Hence, for our data set, it seems that the choice of weights is not crucial. This is probably due to the fact that all time durations are equal and the estimated claim frequencies $\hat{\lambda}_i$ vary quite little for the majority of policies. For other tariffs, this may not be the case and then it is of interest to compare how various weighting schemes affect the coefficients of determination, test of excess variance, estimated overdispersion and coefficients of variation in terms of efficiency and power.

A fifth extension would be to include claim severity. Assuming X_{ij} is the j^{th} claim severity of the i^{th} policy we may variance decompose the observed cost rates

$$Z_i = \sum_{j=1}^{N_i} X_{ij} / t_i$$

with weights $w_i = t_i$. An alternative approach is to treat claim severity separately and condition on the observed $N_i = n_i$. This leads to variance decomposition of the average claim costs

$$Z_i = \sum_{j=1}^{n_i} X_{ij} / n_i,$$

for all policies with $n_i > 0$, using weights $w_i = n_i$.

APPENDIX

Estimating the variance parameter *a*. We divide the estimated individual claim frequencies $\hat{\lambda}_i$ into 21 intervals (see Table 3), henceforth denoted as premium groups. Let I_j be the j^{th} premium group ($j = 1, \ldots, 21$) and

$$\tilde{\lambda}_j = \sum_{i \in I_j} \hat{\lambda}_i / |I_j|,$$

$$\tilde{\sigma}^2_{\text{excess},j} = \sum_{i \in I_j} \left(Y_i - \hat{\lambda}_i\right)^2 / |I_j| - \tilde{\lambda}_j$$

the estimated average premium and excess variance within I_j. Assuming a power relation $E(\tilde{\sigma}^2_{\text{excess},j}) = \xi \tilde{\lambda}_j^a$, a weighted linear regression of $\log(\tilde{\sigma}^2_{\text{excess},j})$ against $\log(\tilde{\lambda}_j)$ is employed, with weights proportional to $|I_j|$. Since $\tilde{\sigma}^2_{\text{excess},j}$ is unreliable (and sometimes negative) for small premium groups, we only include $I_3, \ldots, I_{12}$ in the regression analysis, resulting in (27), where $\hat{\xi}_0$ is different from (15), which assumes a to be known. The estimate (27) is quite stable. Further exclusion of i) I_{12} gives $(\hat{a}, \hat{\xi}_0) = (1.3234, 0.0997)$ and ii) I_3 and I_{12} gives $(\hat{a}, \hat{\xi}_0) = (1.2971, 0.0930)$. In Figure 2, the pairs $(\tilde{\lambda}_j, \tilde{\sigma}^2_{\text{excess},j})$, $j = 3, \ldots, 12$ are plotted together with fitted variance curves based on (27) and a second curve with $a = 1$ and only ξ being estimated.

Asymptotic normality of $\hat{\phi}$ and the numerator of (23). Define

$$S = (S_1, S_2) = \left(\sum_i \lambda_i, \sum_i t_i v_i\right),$$

$$\hat{S} = (\hat{S}_1, \hat{S}_2) = \left(\sum_i Y_i, \sum_i t_i (Y_i - \hat{\lambda}_i)^2\right),$$

where $v_i = \text{Var}(Y_i)$. We will prove that asymptotically, in the limit of large samples, $\hat{S}$ has a bivariate normal distribution with mean S and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

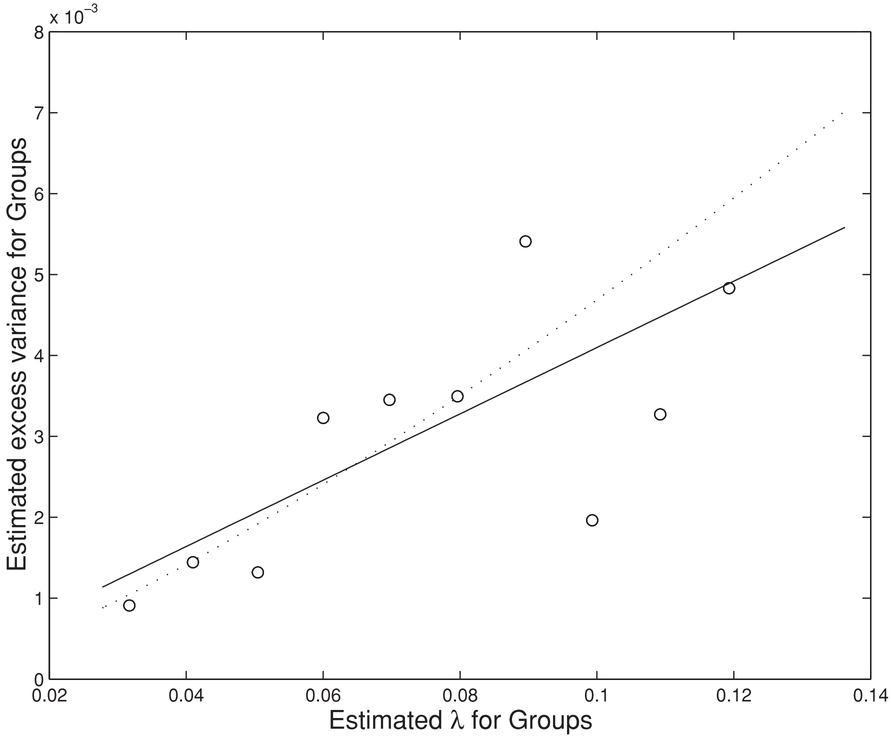

FIGURE 2: Plot of $\tilde{\sigma}^2_{\text{excess},j}$ against $\tilde{\lambda}_j$ for premium groups $j = 3, \ldots, 12$, together with fitted variance curves based on estimates (27) (dotted line), $a = 1$ and estimated $\hat{\xi}_0$ (solid line).

From this asymptotic normality of $\widehat{\phi}$ follows. Indeed,

$$\widehat{\phi} = g(\widehat{S}),$$
$$\phi = g(S)$$

where $g(S_1, S_2) = S_2/S_1$. Let $G = g'(S) = (-\phi,1)/S_1$. Then, by Taylor expanding g around S, it follows that $\widehat{\phi}$ is asymptotically normal with mean ϕ and variance

$$\sigma^2_{\widehat{\phi}} = G\Sigma G^T = (\sigma_{22} - 2\phi\sigma_{12} + \phi^2\sigma_{11}) / S_1^2. \tag{A.1}$$

Similarly, let $\widehat{C}$ be the numerator of (23) and write

$$\widehat{C} = \textstyle\sum_i \left(t_i(Y_i - \hat{\lambda}_i)^2 - Y_i\right) = k(\widehat{S}),$$
$$C = \textstyle\sum_i (t_i v_i - \lambda_i) = k(S),$$

where $k(S) = S_2 - S_1$. Putting $K = k'(S) = (-1, 1)$ we find that $\widehat{C}$ is an asymptotically normal estimator of C with asymptotic variance

$$\sigma^2_{\widehat{C}} = K\Sigma K^T = \sigma_{22} - 2\sigma_{12} + \sigma_{11}. \tag{A.2}$$

Following the lines of proof in Hössjer (2008), one verifies that

$$\widehat{S} = S + \sum_i (Y_i - \lambda_i, t_i(Y_i - \lambda_i)^2 - t_i v_i) + o_p(n^{1/2}), \tag{A.3}$$

where the the last term is small in probability compared to $n^{1/2}$ and hence asymptotically negligible. A consequence of (A.3) is that the impact of replacing λ_i by $\widehat{\lambda}_i$ in the definition of $\widehat{S}$ has no effect on the asymptotic distribution. It follows from (A.3) that

$$\begin{aligned} \sigma_{11} &= \textstyle\sum_i v_i, \\ \sigma_{12} &= \textstyle\sum_i \tau_i, \\ \sigma_{22} &= \textstyle\sum_i \kappa_i, \end{aligned} \tag{A.4}$$

where $\tau_i = t_i E\big((Y_i - \lambda_i)^3\big)$ and $\kappa_i = t_i^2 E\big(((Y_i - \lambda_i)^2 - v_i)^2\big)$. Inserting (A.4) into (A.1) and (A.2) we obtain

$$\begin{aligned} \sigma^2_{\widehat{\phi}} &= S_1^{-2} \textstyle\sum_i \big(\kappa_i - 2\phi\tau_i + \phi^2 v_i\big), \\ \sigma^2_{\widehat{C}} &= \textstyle\sum_i \big(\kappa_i - 2\tau_i + v_i\big). \end{aligned} \tag{A.5}$$

To compute standard errors, we replace S_1, ϕ, v_i, τ_i and κ_i by estimates and obtain

$$\begin{aligned} d^2_{\widehat{\phi}} &= \big(\textstyle\sum_i Y_i\big)^{-2} \textstyle\sum_i \big(\hat{\kappa}_i - 2\widehat{\phi}\hat{\tau}_i + \widehat{\phi}^2 (Y_i - \hat{\lambda}_i)^2\big), \\ d^2_{\widehat{C}} &= \textstyle\sum_i \big(\hat{\kappa}_i - 2\hat{\tau}_i + (Y_i - \hat{\lambda}_i)^2\big), \end{aligned} \tag{A.6}$$

One option is to proceed nonparametrically and put

$$\begin{aligned} \hat{\tau}_i &= t_i\big((Y_i - \hat{\lambda}_i)^3 - \hat{v}_i(Y_i - \hat{\lambda}_i)\big), \\ \hat{\kappa}_i &= t_i^2\big((Y_i - \hat{\lambda}_i)^2 - \hat{v}_i\big)^2, \\ \hat{v}_i &= \hat{\lambda}_i / t_i + \hat{\xi}\hat{\lambda}_i^a. \end{aligned}$$

We added the second term $-\hat{v}_i(Y_i - \hat{\lambda}_i)$ in the definition of $\hat{\tau}_i$ in order to guarantee that $\hat{\Sigma}$ is positive (semi)definite and thus $d^2_{\widehat{\phi}}$ and $d^2_{\widehat{C}}$ are non-negative.

Alternatively, a parametric approach is to assume a gamma distribution for all Λ_i. For instance, if $a = 1$, $\Lambda_i \in \Gamma(\lambda_i/\xi, \xi)$, where $\Gamma(\alpha,\beta)$ has density

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0.$$

Hence $t_i\Lambda_i \in \Gamma(\lambda_i/\xi, t_i\xi)$, and $N_i = t_i Y_i$ has a negative binomial distribution $\mathrm{Nbin}(\lambda_i/\xi, 1/(1+t_i\xi))$. From moments of negative binomial distributions we obtain

$$\begin{aligned} v_i &= t_i^{-1}\lambda_i(1 + \xi t_i), \\ \tau_i &= t_i^{-1}\lambda_i(1 + 3\xi t_i + 2\xi^2 t_i^2), \\ \kappa_i &= 2\lambda_i^2(1 + \xi t_i)^2 + t_i^{-1}\lambda_i(1 + 7\xi t_i + 12\xi^2 t_i^2 + 6\xi^3 t_i^3), \end{aligned} \tag{A.7}$$

and their estimated analogues by plugging in $\hat{\lambda}_i$ and $\hat{\xi}$. Since ξ is often very small for non-life insurance data the higher order powers of ξ make little contribution to the standard errors. When $\xi = 0$, we obtain the denominator of (23) from (A.5) and (A.7). □

Extending the variance decomposition to time-varying covariates and random effects. For simplicity, assume that time is counted in units of years and that all t_i are integer valued and extend (2) and (3) to

$$\begin{aligned} \lambda_{ij} &= \exp(\beta^T x_{ij}), \\ N_i &= \textstyle\sum_{j=1}^{t_i} Y_{ij}, \\ Y_{ij} &\in \mathrm{Po}(\lambda_{ij} U_{ij}), \end{aligned} \tag{A.8}$$

where λ_{ij}, $x_{ij} = (x_{ij1}, \ldots, x_{ijp})$ and U_{ij} is the time varying price relativity, covariate vector and random effect of policy i during year j respectively. We assume that Y_{ij} and Y_{lk} are conditionally independent given U_{ij} and U_{lk}, that U_{ij} and U_{lk} are independent when $i \neq l$ and that

$$\begin{aligned} E(U_{ij}) &= 1, \\ \mathrm{Cov}(U_{ij}, U_{ik}) &= \xi \lambda_{ij}^{a/2-1} \lambda_{ik}^{a/2-1} r_{k-j} \end{aligned} \tag{A.9}$$

for some autocorrelation function r.

Let $\lambda_i = \sum_{j=1}^{t_i} \lambda_{ij}/t_i$ and $\Lambda_i = \sum_{j=1}^{t_i} \lambda_{ij} U_{ij}/t_i$ be the average price relativity and accident proneness of contract i. Then (A.9) gives an extension

$$\mathrm{Var}(\Lambda_i) = \xi \tilde{\lambda}_i^a, \tag{A.10}$$

of the variance function (5), where

$$\tilde{\lambda}_i^a = t_i^{-2} \sum_{j,k=1}^{t_i} \lambda_{ij}^{a/2} \lambda_{ik}^{a/2} r_{k-j}.$$

We notice that $\tilde{\lambda}_i \leq \lambda_i$ when $a \leq 2$, with equality if and only if $r_k \equiv 1$ and either $a = 2$ or $\lambda_{ij} \equiv \lambda_i$. Hence, the variance of the averaged accident proneness is reduced when the random effects or price relativities are time-varying.

There are at least two ways to extend the variance decomposition of Section 2. The first option is to retain (7) with $Y_i = N_i/t_i$, as before. This yields a variance decomposition for which the explained variance σ_1^2 and the noise variance σ_3^2 agree with (8), whereas the unexplained individual variance

$$\sigma_2^2 = \xi \sum_i t_i \tilde{\lambda}_i^a / \sum_i t_i.$$

is decreased as soon as $\tilde{\lambda}_i < \lambda_i$ for at least one policy i. The conclusion is that both R^2 and CID are increased when random effects within each policy are time varying.

The second option is to consider the variance of the annual claim frequencies,

$$\sigma^2 = \sum_{i=1}^{n} \sum_{j=1}^{t_i} E\left(\left(Y_{ij} - \lambda\right)^2\right) / \sum_i t_i.$$

By similar calculations as in Section 2, this yields quite a different variance decomposition with

$$\begin{aligned}
\sigma_1^2 &= \left(\textstyle\sum_i t_i(\lambda_i - \lambda)^2 + \sum_{i,j}\left(\lambda_{ij} - \lambda_i\right)^2\right) / \textstyle\sum_i t_i. \\
\sigma_2^2 &= \xi \textstyle\sum_{i,j} \lambda_{ij}^a / \sum_i t_i, \\
\sigma_3^2 &= \textstyle\sum_i t_i \lambda_i / \sum_i t_i.
\end{aligned}$$

The noise variance σ_3^2 is increased for a variance decomposition based on annual claim frequencies (by a factor t if $t_i \equiv t$) in agreement with the discussion in Section 4. The explained variance, σ_1^2, is enlarged as well, since we are able to explain not only λ_i but also the variation of λ_{ij} around λ_i. The unexplained individual variance, σ_2^2, also increases when $a \geq 1$, since then $\sum_{i,j} \lambda_{ij}^a \geq \sum_i t_i \lambda_i^a$.

The conclusion is that in general, R^2 decreases if annual claim frequencies Y_{ij} are considered rather than the averaged ones Y_i, due to the increased noise variance. On the other hand, CID may increase or decrease, depending on how variable the price relativities and random effects are within each policy. If the random effects are constant ($r_k \equiv 1$), then CID typically increases when annual claim frequencies are considered.

Acknowledgement

Ola Hössjer's work was supported by the Swedish Research Council, contract number 621-2005-2810. The authors wish to thank two anonymous referees, whose comments significantly increased the quality of the manuscript.

References

Beard, R.E., Pentikainen, T. and Pesonen, E. (1984) *Risk Theory, The Stochastic Basis of Insurance (3rd edition).* Chapman and Hall.

Breslow, N.E. (1984) Extra-Poisson variation in log-linear models. *Appl. Statist.* **33(1)**, 38-44.

Brockman, M.J. and Wright, T.S. (1992) Statistical motor rating: Making effective use of your data. *Journal of the Institute of Actuaries* **119(111)**, 457-543.

Bühlmann, H. and Gisler, A. (2005) *A Course in Credibility Theory and its Applications.* Springer Universitext.

Cox, D.R. and Snell, E.J. (1989) *The analysis of binary data*, 2nd ed., Chapman and Hall, London.

Fisher, R.A. (1950) The significance of deviations from expectation in a Poisson series. *Biometrics* **6**, 17-24.

Haight, F.A. (2001) Accident proneness: The history of an idea. Institute of Transportation Studies, University of California, Irvine, USA.

Hinde, J. (1982) Compound Poisson regression models. In *GLIM 82: Proceedings of the International Conference in Generalized Linear Models* (R. Gilchrist, ed.), pp. 199-121, Springer, Berlin.

Hössjer, O. (2008) On the coefficient of determination for mixed regression models. *Journal of Statistical Planning and Inference* **138**, 3022-3038.

Johnson, P.D. and Hey, G.B. (1971) Statistical studies in motor insurance. *Journal of the Institute of Actuaries* **97**, 199.

Jung, J. (1968) On automobile insurance ratemaking. *ASTIN Bulletin* **5**, 41.

Järnmalm, K. (2006) *Measures of the remaining systematic variance between individuals when divided into individual premium groups in non-life insurance*. Master Thesis, Mathematical Statistics, Stockholm University, Report 2006:15. (In Swedish.)

Lawless, J.F. (1987) Negative binomial and mixed Poisson regression. *Canadian J. Statist.* **15(3)**, 209-225.

Lemaire, J. (1995) *Bonus-Malus Systems in Automobile Insurance*. Springer.

Maddala, G.S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

Magee, L. (1990) R^2 measures based on Wald and likelihood ratio joint significance tests. *Am. Statistician* **44**, 250-253.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, second edition, Chapman and Hall.

Nagelkerke, N.J.D. (1991) A note on a general definition of the coefficient of determination. *Biometrika* **78(3)**, 691-692.

Ohlsson, E. (2008) Combining generalized linear models and credibility models in practice. *Scandinavian Actuarial Journal* **2008(4)**, 301-314.

Ohlsson, E. and Johansson, B. (2006) Exact credibility and Tweedie models. *ASTIN Bulletin* **36(1)**, 121-133.

Pearson, K. (1900) On a criterion that a given system of deviations from the probable in case of a correlated system of variables in such that it can be reasonably supposed to have arisen from a random sampling. *Phil. Mag.* **50(5)**, 157-75.

POCOCK, S.J., COOK, D.G. and BERESFORD, S.A.A. (1981) Regression of area mortality rates on explanatory variables: what weighting is appropriate? *Appl. Statist.* **30**, 286-295.
RAO, C.R. and CHAKRAVARTI, I.M. (1956) Some small sample tests for significance for a Poisson distribution. *Biometrics* **12**, 264-282.
VENEZIAN, E.C. (1981) Good drivers and bad drivers – a Markov model of accident proneness. *Proceedings of the Casualty Actuarial Society*, **LXVII**, 65-85.
VENEZIAN, E.D. (1990) The distribution of automobile accidents – are relativities stable over time? *Proceedings of the Casualty Actuarial Society*, **LXXVII**, 309-336.
WHITE, H. (1982) Maximum likelihood under misspecified models. *Econometrica* **50**, 1-25.

OLA HÖSSJER (corresponding author)
Department of Mathematics
Stockholm University, S-106 91 Stockholm,
Sweden
E-Mail: ola@math.su.se
Fax: +46-8-612 67 17,
URL: www2.math.su.se/~ola/