EFFECTS OF VARIATIONS FROM GAMMA-POISSON ASSUMPTIONS

GARY G. VENTER

Abstract

Two types of variations from negative binomial frequency are considered: mixtures of Poisson other than Gamma, and Poisson parameters that change over time. Any severity distribution can be used instead of the Gamma as a mixing distribution, and Bayesian estimators are easy to calculate from the mixed probabilities. In the case of changing frequencies over time, the Gerber-Jones model is illustrated for calculating credibilities. The Bailey-Simon method is found to be useful for testing model assumptions.

1. INTRODUCTION

A model often used for experience rating assumes that each individual risk has its own Poisson distribution for number of claims, with a Gamma distribution across the population for the Poisson mean. This model has been known since at least 1920 (M. Greenwood and G. Yule [7]), and has been applied to insurance experience rating since at least 1929 (R. Keffer [10]). However, there is meager theoretical support for the Gamma distribution as a mixing function, and the main empirical support given in many studies is that it provides a better fit to the aggregate claim frequency distribution than that given by the assumption that all individuals have the same Poisson distribution; e.g., see Lester B. Dropkin [4], B. Nye and A. Hofflander [12], or R. Ellis, C. Gallup, and T. McGuire [5]. The Poisson assumption for each individual does have theoretical support, but not enough to be regarded as certain. For example, the Poisson parameter could vary over time in random ways, to be discussed further below.

Several alternative models, which, in many cases, fit better than the Poisson, have been presented in the literature; e.g., Gordon Willmot [19, 20], M. Ruohonen [14], W. Hürlimann [8]. Many of these are mixtures of the Poisson by other distributions, such as the inverse Gaussian, reciprocal inverse Gaussian, beta, uniform, noncentral chi-squared, and three-parameter origin shifted Gamma distributions.

The purpose of this paper is to explore the adequacy of the Poisson and Gamma assumptions, the information needed to verify them, and the experience rating consequences of using these assumptions when they do not apply. As will be seen below, there are substantial differences in the experience rating implications of models which have very similar predictions of the aggregate claim frequency distribution. Thus, a model which gives a good fit to this distribution does not necessarily give proper experience rating adjustments. In other words, a model that just fits better than the Poisson is not enough for experience rating use. More detailed records which track individuals over time are needed to determine how much credibility should be given to individual claim experience.

2. PRELIMINARY BACKGROUND

Suppose each risk has its own claim frequency distribution, constant over time, and that the mean of the individual risk annual claim frequency variances is s^2, and the variance of the risk means is t^2. Among linear estimators, the expected squared error in subsequent observations is minimized by the credibility estimator $zx + (1 - z)m$, where m is the overall mean, x is the individual risk annual frequency observed, and for n years of observations, $z = n/(n + K)$, with $K = s^2/t^2$. See, for example, A. Bailey [1], H. Bühlmann [3], W. Jewell [9]. If the restriction to linear estimators is removed, then the Bayesian predictive mean minimizes the expected squared error. Thus, when the Bayes estimator is linear in the observations, it must be the same as the credibility estimator.

This is the case with the Gamma-Poisson model. In fact, if the Gamma has parameters α and β, with mean α/β and variance α/β^2, the Bayesian predictive mean is

$$\frac{\alpha + nx}{\beta + n} = \frac{\alpha}{\beta} \cdot \frac{\beta}{\beta + n} + x \cdot \frac{n}{\beta + n},$$

which is the credibility estimator, as $m = s^2 = \alpha/\beta$ and $t^2 = \alpha/\beta$. The linearity of the Bayes estimate gives a degree of justification to this model. Besides being easy to calculate, it is also less likely than some nonlinear functions to take on exaggerated values in extreme cases. The fact that it is the best linear model implies that even if it is the wrong model, it is the best linear approximation to the Bayes estimator of the actual distribution. The mixed distribution is the negative binomial, with the probability of n claims given by

$$p_n = \frac{(\alpha + n - 1)!\beta^{\alpha}}{(1 + \beta)^{\alpha+n} n!(\alpha - 1)!},$$

or recursively by

$$p_o = \left[\frac{\beta}{1 + \beta}\right]^{\alpha}, \; p_{n+1} = p_n \frac{\alpha + n}{(n + 1)(1 + \beta)}.$$

3. VARIATIONS FROM GAMMA ASSUMPTION

First, the Poisson assumption will be retained, so that each individual is assumed to have a fixed Poisson probability for number of accidents, and variation from the Gamma assumption will be explored.

With the Poisson assumption, each risk has the same mean and variance, so that the average risk variance s^2 is the same as the average risk mean, m. Since the aggregate variance v is $s^2 + t^2$, t^2 is the difference between the aggregate variance and mean (assuming the variance exceeds the mean). Thus, the best linear estimator is the credibility estimator with $K = m/(v - m)$. This is also the Bayes estimator for the Gamma prior distribution having variance t^2.

Other prior distributions may also have variance t^2, but they have different Bayes estimators which are not linear functions of the observations. Two distributions will be shown below for which the aggregate probabilities are much the same as the Gamma provides, but the Bayes estimates are substantially different for some risks. Nonetheless, since the variances are the same, the predictive means from the Gamma prior will be the best linear approximation to the Bayes estimate for either distribution.

The first is the good-risk/bad-risk model. The population has two types of risks; 90% are good risks with a low probability of a claim, and the other 10% are bad risks with a high claim probability. As each risk is still assumed to be Poisson distributed, this model is the Poisson mixed by a two-point prior. If the two Poisson means are a and b, then, over all risks, the probability of n claims is $p_n = .9a^n e^{-a}/n! + .1b^n e^{-b}/n!$, the mean is $m = .9a + .1b$ and the variance is $t^2 = (.9)(.1)(b-a)^2$. Thus the method of moments estimators for a and b are: $a = m - t/3$ and $b = m + 3t$. Given a risk with k claims in n years, the probability (conditional on k) that it is a bad risk is

$$q_k = \frac{1}{1 + 9e^{n(b-a)}\left(\frac{a}{b}\right)^k},$$

by Bayes Theorem. The Bayesian predictive mean for that risk is thus $a(1 - q_k) + bq_k$.

The second model is the inverse Gaussian distribution, discussed in Willmot [19]. This can be parameterized with two parameters b and c with density

$$f(y) = (2\pi c)^{-.5}b^{-1}(b/y)^{1.5}e^{1/2c(2-y/b-b/y)},$$

mean $= b$, and variance $= t^2 = b^2c$. This is a somewhat more skewed distribution than the Gamma. In fact, the skewness is $3c^{1/2}$, and the Gamma with the same first two moments has skewness $2c^{1/2}$. The Poisson mixture has mean $= b$, variance $= b + b^2c$, and the probabilities, p_i, of i claims given by:

$$p_0 = e^{1/c[1-(1+2bc)^{.5}]};$$

$$p_1 = p_0 b(1 + 2bc)^{-.5};$$

$$p_n = \frac{2bc(n - 1)(n - 1.5)p_{n-1} + b^2 p_{n-2}}{(1 + 2bc)n(n - 1)}, \; n > 1.$$

The inverse Gaussian is not obviously related to the Normal distribution. It gets its name from the fact that there is a different, but equivalent, way of parameterizing the distribution that looks like a Normal distribution if you switch the variable and one of the parameters.

The cumulative probabilities can be calculated using the standard Normal cdf $N(x)$ by

$$F(x) = N\left(\frac{x - b}{\sqrt{bcx}}\right) + e^{2/c}\left[1 - N\left(\frac{x + b}{\sqrt{bcx}}\right)\right].$$

As with any Poisson mixture, given a risk with n claims in a period, the Bayesian predictive mean for the number of claims to be observed in a future period is $(n + 1)p_{n+1} \div p_n$. This can be readily verified as follows: let $f(\lambda)$ denote the density for the Poisson parameter λ. Then $p_n = 1/n! \int f(\lambda)e^{-\lambda}\lambda^n d\lambda$. The Bayes predictive mean given n claims is $E(N|n) = E(\lambda/n) = \int \lambda f(\lambda|n)d\lambda = 1/p_n n! \int \lambda f(\lambda)e^{-\lambda}\lambda^n d\lambda$, by Bayes theorem, and the result follows. This implies that any severity distribution can be used as the mixing distribution for a Poisson. The advantage of the inverse Gaussian is that p_n is given by the above recursive formula, while many other distributions would require numerical integration for this.

For the sake of comparison, the Gamma, two-point, and inverse Gaussian prior distributions will be fit to a sample of medical malpractice claims by the method of moments, so that the variances will be the same and thus the Gamma-Bayesian estimators will be the best linear approximation to the other two. The sample used is four years of closed claims data from 7,744 internists as reported in Ellis, Gallup, and McGuire [5]. This is for illustration only, as the use of closed claims for pricing insurance has been questioned on various grounds [11]. The number of doctors having various claim counts is shown below:

Number of claims	0	1	2	3	4	5	6
Number of doctors	7,299	386	52	5	1	1	0

This sample has mean .0664 and variance .0834. For the four-year period, by the Poisson assumption, $m = s^2 = .0664$, and thus $t^2 =$ variance $- s^2 = .0170$. Matching the moments m and t^2 will give priors for a four-year Poisson parameter. These prior distributions get the following parameters:

Gamma:	$\alpha = .260$	$\beta = 3.91$
Inverse Gaussian:	$b = .0664$	$c = 3.86$
Two-Point:	$a = .0229$	$b = .458$

Using the formulae above, these parameters result in the following Bayesian predictive means (i.e., the expected number of claims in four years) for risks having the number of claims shown. The percentage errors from using the Gamma when one of the other prior distributions is correct are also given.

BAYESIAN ESTIMATES

Number of Claims:	0	1	2	3	4	5	6
Gamma	.0530	.257	.460	.664	.868	1.070	1.270
Inverse Gaussian	.0540	.223	.521	.852	1.190	1.530	1.860
Two-Point	.0521	.280	.443	.457	.458	.458	.458
Error from using Gamma if true distribution is:							
Inverse Gaussian	−2%	15%	−12%	−22%	−27%	−30%	−32%
Two-Point	2%	−8%	4%	45%	90%	134%	177%

The overall distribution of number of claims predicted by each distribution is given below. These sample and fitted aggregate claim frequencies and the Bayesian means above are shown in Figures 1 and 2.

OVERALL CLAIM PROBABILITIES

Number of Claims:	0	1	2	3	4	5	6
Sample	.943	.0498	.00672	.000646	.000129	.0001290	.0000000
Gamma	.943	.0499	.00640	.000983	.000163	.0000283	.0000051
Inverse Gaussian	.942	.0509	.00568	.000987	.000210	.0000500	.0000127
Two-Point	.943	.0491	.00686	.001010	.000116	.0000106	.0000008
Single Poisson	.936	.0621	.00206	.000046	7.58E-7	1.01E-8	1.11E-10

If the Bayesian estimates are used as experience rating charges, the above two tables together show that small differences in the aggregate probabilities lead to fairly large differences in charges. On a percentage basis, the mixed distribution probabilities differ from each other mostly in the right tails, where the claim data is least reliable. This is also the

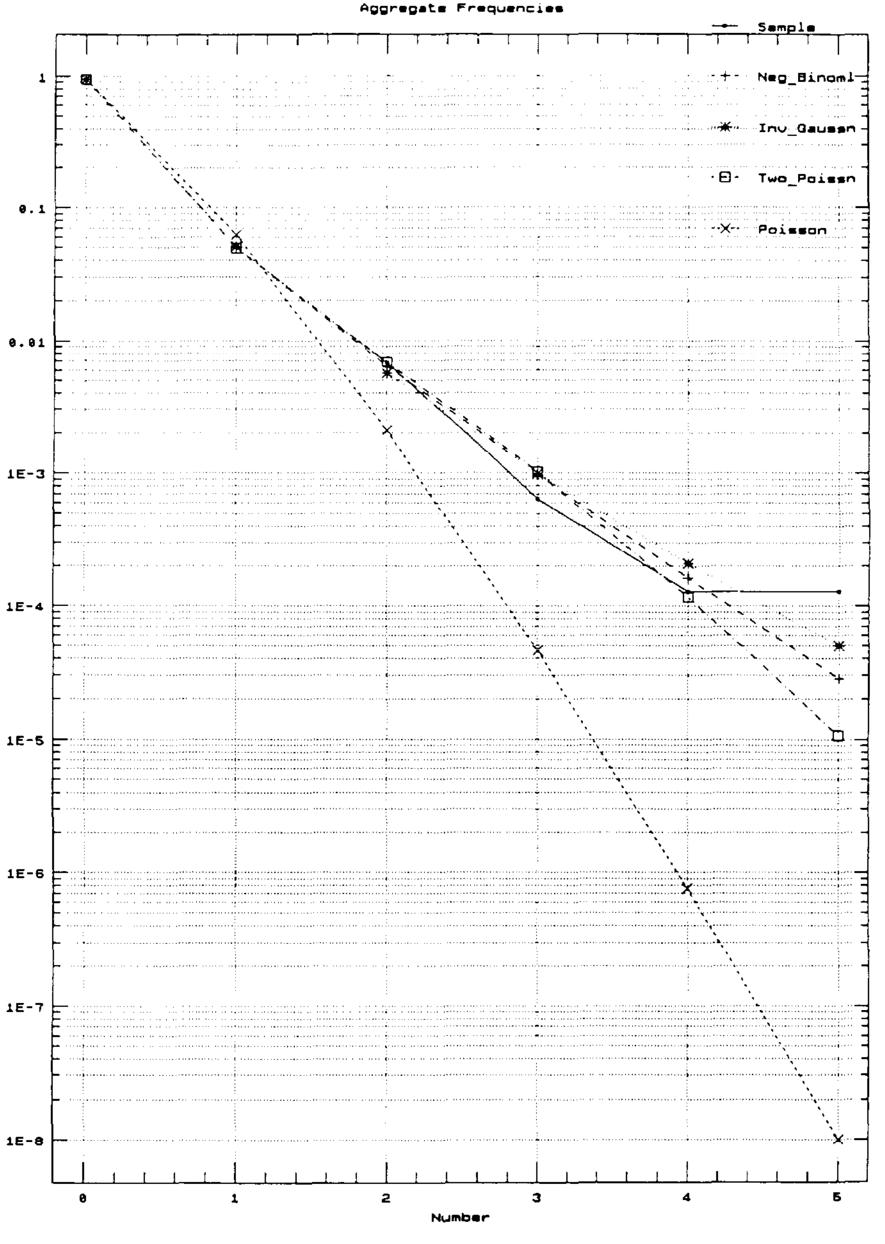
Aggregate Frequencies
Sample
Neg_Binoml
Inv_Gaussn
Two_Poissn
Poisson
1
0.1
0.01
1E-3
1E-4
1E-5
1E-6
1E-7
1E-8
0
1
2
3
4
5
Number

FIGURE 1

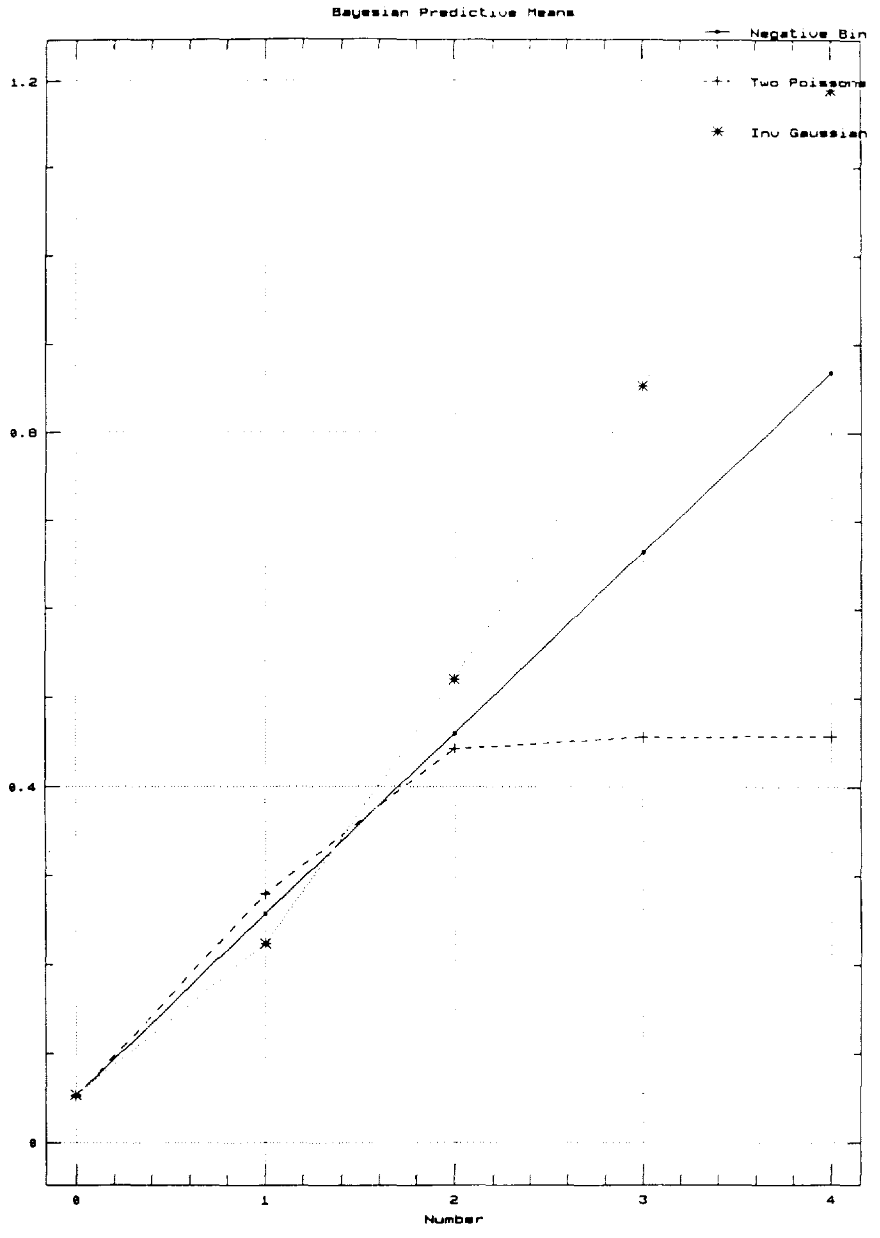
Bayesian Predictive Means
Negative Bin
Two Poissons
Inv Gaussian
1.2
0.8
0.4
0
0
1
2
3
4
Number

FIGURE 2

area in which the Bayesian estimates diverge the most. It is not clear that goodness-of-fit measures could be of much help in selecting among these distributions either, because without theoretical reason to support one prior distribution or another, a good fit in the left tails has questionable relevance to the right tails.

Nonetheless, as great as the differences are, they are small compared to those from using the overall mean of .0664 for each risk; i.e., not experience rating at all. If the population were known to consist of some mixture of risks, each with fixed Poisson distributed claims exposure, the two-point model might be the most justifiable in this case because the penalty for the right tail risks, whose exposures are not clearly understood, would be less. On the other hand, if it were known from other investigations that the high-frequency doctors were actually quite bad, a more heavily-tailed model might be justified.

4. VARIATIONS FROM POISSON ASSUMPTION

Two types of variation from the Poisson assumption are considered below. First, each risk may have some distribution of claim counts other than Poisson, with that distribution invariant over time. Second, each risk may have a fixed distribution for each time period, but the mean changes each period, with the degree of change coming from a distribution that is invariant over time.

The first case could arise if the risk has a Poisson distribution for each period, but the Poisson parameter is drawn at random each period from a prior distribution. The variance for a year would be the sum of the expected Poisson variance and the variance of the Poisson means from the prior. This is different from the second case because the Poisson parameters are drawn from a fixed distribution each year, while in the second case the incremental change in the parameter is so drawn. Both cases allow variation among risks in addition to the greater variation each risk can display due to the relaxation of the Poisson assumption. For instance, the good-risk/bad-risk model could end up being a mixture of two Negative Binomials instead of two Poissons. Since there are many possibilities for distributional assumptions, the analysis of these cases will be carried out only for the linear approximations to the Bayesian estimates, that is for the credibility estimators.

In the first case, the credibility estimator is the same as discussed above, $zx + (1 - z)m$, with $z = n/(n + s^2/t^2)$, again where s^2 is the expected individual risk variance and t^2 is the variance of the risk means.

This result did not depend on any Poisson assumption. However, without the Poisson assumption, it is not as easy to determine what s^2 and t^2 should be. They still add to the total variance v, but $v - m$ might not be a good estimator of t^2, as s^2 may be greater than m. In fact, in the extreme case, $s^2 = v$ and $t^2 = 0$, when all risks have the same (non-Poisson) distribution of claims. In this case, $z = 0$ and the best estimator for any risk is the overall mean m.

As Nye and Hofflander [13] point out, this is not an appealing model, because most people believe there is some inherent difference among risks. However, it is also plausible that risks would have some degree of instability over time as well, and the question becomes how much of the difference $v - m$ can be attributed to each effect. Data such as the distribution of claims in period 2 for the risks with no claims in period 1 would be useful for making such determinations. If the no-claim risks from period 1 had the same distribution in period 2 as did all the other risks, for example, it would lend support to the conclusion that all risks are fundamentally the same. On the other hand, if they had better-than-average experience, the credibility that should be attributed to those risks could then be estimated.

One model of the latter case is provided by H. Gerber and D. Jones [6]. The individual risk mean changes each year by a random amount taken from a distribution with mean 0 and variance d^2. For year 1, the risk means are distributed around an overall mean m with variance t^2. (Thus, for year 2, assuming independence, the mean is still m, but the variance is $t^2 + d^2$, etc.) The distribution of actual results around a risk's mean for a given year has variance s^2. Then, given a risk with losses X_i in year i, the linear least square estimator C_{i+1} for the next year's losses is given iteratively by:

$$C_1 = m; \qquad C_{i+1} = z_i X_i + (1 - z_i) C_i$$

where

$$z_1 = \frac{L}{L+1}; \qquad z_{i+1} = \frac{z_i + J}{z_i + J + 1};$$

$$L = t^2/s^2; \qquad J = d^2/s^2.$$

This follows from [22: p. 428] by taking $v = s^2$ and $w = t^2$.

Gerber and Jones had a somewhat more general framework, in that m and t^2 were any prior mean and variance for the (conditional) mean of X_1, not necessarily arising from a distribution of risks around a grand mean. Working through the iterative definition gives

$$C_{i+1} = m \prod_{j=1}^{i} (1 - z_j) + \sum_{j=1}^{i} X_j z_j \prod_{h=j+1}^{i} (1 - z_h),$$

which shows how the credibility for an observation decreases in estimating ever later future observations. When all the past observations are 0, the estimate reduces to the first term above.

One study that provided data that could be used to evaluate the above cases was that of Robert Bailey and LeRoy Simon [2]. They estimated the credibility of one, two, and three years of driver experience as 1 minus the relative claim frequency of drivers with one or more, two or more, and three or more years without a claim prior to the experience period. The results for five driver classifications are shown below.

CREDIBILITY FOR CLAIM FREE EXPERIENCE

Class	1 year	2 years	3 years
1	.046	.068	.080
2	.045	.060	.068
3	.051	.068	.080
4	.071	.085	.099
5	.038	.050	.059

Bailey and Simon note that the additional credibility for years past the first is less than would be anticipated. Fitting $n/(n + K)$ to this data by row by least squares for K gives the percentage errors:

PERCENTAGE ERRORS IN CREDIBILITY WITH $n/(n + K)$

Class	1 year	2 years	3 years	K
1	−30	−8	+14	30.4
2	−38	−10	+16	35.1
3	−37	−8	+14	30.0
4	−41	−6	+17	23.2
5	−39	−9	+13	41.6

The large errors and systematic signs on the errors are indicative that the standard credibility model is inappropriate. Any other method of fitting the K's would have the same result. Also, the credibility from this model being too high in the third year and too low in the first indicates that the relevance of a year's data declines as it ages, which suggests that the changing mean model may apply. Fitting this by using least squares to find J and z_1 gives the following percentage errors:

PERCENTAGE CREDIBILITY ERRORS USING GERBER-JONES MODEL

Class	1 year	2 years	3 years	z_1	J
1	−4	+3	−1	.0098	.018
2	−9	+7	−1	.0035	.020
3	−8	+7	−3	.0065	.022
4	−13	+11	−3	.0024	.033
5	−11	+8	−2	.0044	.016

This fits much better, but it does use more parameters and still has systematic sign changes. The extra parameter helps because it does the right thing—it decreases the relative credibility for the older years, reflecting changing risk conditions over time. Nonetheless, there still seem to be aspects of the data that are not captured by the model.

The Bailey-Simon results support the use of changing parameter models over fixed parameter mixed models, Poisson or not. They also support the use of experience rating over not experience rating, and their work points to the kind of data that is needed to compute experience credits and debits.

Other studies using similarly detailed data have also rejected the stable Poisson assumption for automobile insurance. These include Venezian [16], who found that a two-point Poisson model with shifting driver probabilities between the two parameters fit well to California data from 1961–1963, and Richard Woll [21], who found problems with both the Gamma and the stable Poisson hypotheses using four years of North Carolina data published in 1970. These findings do not challenge the value of experience rating, but they do tend to reduce the credibilities that would apply.

The situation is not necessarily the same for medical malpractice, in that greater training is required prior to licensing, so learning by doing should have a smaller effect. However, there is anecdotal evidence to the contrary. In one study that followed individuals across time periods, Venezian, Nye, and Hofflander [18] were not able to reject the stable Poisson hypothesis using a chi-square test. However, the results of Venezian [17] suggest that there is not enough data in their sample to detect moderate deviations from Poisson by this test. Other tests, such as computing the Bailey-Simon credits and debits deserved by class, would be possible from their data. Another caveat is that since the sample contains only large claims, it may be better approximated by the Poisson than would data using all claims, due to the effect of the severity probability of a loss being large (see Joseph Schumi [15]).

In conclusion, aggregate frequencies are not adequate to verify either the Poisson or the Gamma hypothesis. Variations from the fixed Poisson assumption are likely, and would tend to lower the credibility which should be given to risk experience; variations from the Gamma assumption could lower or raise it. The Bailey-Simon method provides a good way to test proper credibilities, and the Gerber-Jones model gives a method to model changing frequencies over time.

REFERENCES

[1] Arthur L. Bailey, "A Generalized Theory of Credibility," *PCAS* XXXII, 1945.

[2] Robert A. Bailey, and LeRoy J. Simon, "An Actuarial Note on the Credibility of Experience of a Single Private Passenger Car," *PCAS* XLVI, 1959.

[3] H. Bühlmann, "Experience Rating and Credibility," *ASTIN Bulletin* 4, 1967.

[4] Lester B. Dropkin, "Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records," *PCAS* XLVI, 1959.

[5] R. Ellis, C. Gallup, and T. McGuire, "Should Medical Professional Liability Insurance Be Experience Rated?" *The Journal of Risk and Insurance*, 57.1, 1990.

[6] H. Gerber and D. Jones, "Credibility Formulas of the Updating Type," *Transactions of the Society of Actuaries* 27, 1975.

[7] M. Greenwood and G. Yule, "An Inquiry into the Nature of Frequency Distribution of Multiple Happenings," *Journal of the Royal Statistical Society*, A 83, 1920.

[8] W. Hürlimann, "On Maximum Likelihood Estimation for Count Data Models," *Insurance: Mathematics and Economics*, 9.1, 1990.

[9] W. Jewell, "Credible Means Are Exact Bayesian for Simple Exponential Families," *ASTIN Bulletin* 8.1, 1974.

[10] R. Keffer, "An Experience Rating Formula," *Transactions of the Actuarial Society of America* 30, 1929.

[11] R. Michael Lamb, "Uses of Closed Claim Data for Pricing," *Pricing Property and Casualty Insurance Products*, New York, Casualty Actuarial Society, 1980.

[12] B. Nye and A. Hofflander, "Experience Rating in Medical Professional Liability Insurance," *The Journal of Risk and Insurance* 60.1, 1988.

[13] B. Nye and A. Hofflander, "Experience Rating in Medical Professional Liability Insurance: Author's Reply," *The Journal of Risk and Insurance* 61.2, 1989.

[14] M. Ruohonen, "A Model for the Claim Number Process," *ASTIN Bulletin* 18.1, 1988.

[15] Joseph R. Schumi, "A Method to Calculate Aggregate Excess Loss Distributions," *CAS Forum*, Spring 1989.

[16] Emilio C. Venezian, "Good and Bad Drivers—A Markov Model of Accident Proneness," *PCAS* LXVIII, 1981.

[17] Emilio C. Venezian, "The Distribution of Automobile Accidents—Are Relativities Stable over Time?" *PCAS* LXXVII, 1990.

[18] E. Venezian, B. Nye, and A. Hofflander, "The Distribution of Claims for Professional Malpractice: Some Statistical and Public Policy Aspects," *The Journal of Risk and Insurance* 61.4, 1989.

[19] Gordon E. Willmot, "Mixed Compound Poisson Distributions," *ASTIN Bulletin* 16S, 1986.

[20] Gordon E. Willmot, "Sundt and Jewell's Family of Discrete Distributions," *ASTIN Bulletin* 18.1, 1988.

[21] Richard G. Woll, "A Study of Risk Assessment," *PCAS* LXVI, 1979.

[22] Gary G. Venter, in *Foundations of Casualty Actuarial Science*, Casualty Actuarial Society, New York, 1990, Chapter 7, "Credibility," p. 375.